The Epilarynx in Speech

by

Scott Reid Moisik
B.A., University of Calgary, 2006
M.A., University of Victoria, 2008

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Linguistics

© Scott Reid Moisik, 2013
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy
or other means, without the permission of the author.
Supervisory Committee

The Epilarynx in Speech

by

Scott Reid Moisik
B.A., University of Calgary, 2006
M.A., University of Victoria, 2008

Supervisory Committee

Dr. John H. Esling (Department of Linguistics)
Supervisor

Dr. Ewa Czaykowska-Higgins (Department of Linguistics)
Co-Supervisor

Dr. Sonya Bird (Department of Linguistics)
Departmental Member

Dr. Bryan Gick (Department of Linguistics, University of British Columbia)
Outside Member
Abstract

Supervisory Committee
Dr. John H. Esling (Department of Linguistics)
Supervisor
Dr. Ewa Czaykowska-Higgins (Department of Linguistics)
Co-Supervisor
Dr. Sonya Bird (Department of Linguistics)
Departmental Member
Dr. Bryan Gick (Department of Linguistics, University of British Columbia)
Outside Member

This dissertation examines the phonetic and phonological functioning of the supraglottal part of the larynx, the epilarynx, from an articulatory-physiological perspective. The central thesis is that, through constriction, the epilarynx physically couples the vocal folds to the supralaryngeal vocal tract. This basic principle is important in explaining a wide range of speech phenomena, such as the mechanism of glottal stop, creaky and harsh (“constricted”) phonation, interaction between vocal fold state and lingual state, and the coordination of phonatory and vowel quality as voice quality, which underlies many register-like patterns. Furthermore, oscillation of the epilarynx and (typically) the vocal folds below is the basis for “growl”, which is demonstrated to have numerous expressions in speech, both phonetically and phonologically.

The thesis is explored by detailed examination of three functions of the epilarynx: (1) epilaryngeal vibration, (2) epilaryngeal interaction with the vocal folds, and (3) epilaryngeal interaction with the supralaryngeal vocal tract. Phonetic evaluations of these functions include physiological, theoretical, and taxonomic considerations, imaging data (obtained with laryngeal and lingual ultrasound, simultaneous laryngoscopy and laryngeal ultrasound, and videofluoroscopy), and computational modeling.
These phonetic evaluations are then taken as the basis for a model of lower vocal tract phonology. Traditional models of such sounds do not accommodate the epilarynx. Rather than positing new distinctive features, an alternative approach is taken. A theoretical model is proposed which is framed in terms of “phonological potentials”, which are the biases associated with physical principles that underlie the formation of phonological systems and patterns. In the context of epilaryngeal function, the phonological potentials are expressed in terms of synergistic relations amongst gross physiological states that either support or hinder epilaryngeal constriction. These biases are argued to exert an articulation-based typological skewing on phonemic systems and patterning, and numerous cases are examined in support of this claim.
Table of Contents

Supervisory Committee .......................................................................................................................... ii
Abstract.................................................................................................................................................. iii
Table of Contents ....................................................................................................................................... vii
List of Tables ............................................................................................................................................ viii
List of Figures .......................................................................................................................................... ix
Acknowledgments.................................................................................................................................... xii
Dedication................................................................................................................................................ xvi

Chapter 1: INTRODUCTION................................................................................................................. 1
  1.1 The epilarynx as a speech production system .................................................................................. 4
  1.2 Dissertation hypotheses and outline ...............................................................................................11

Chapter 2: EPILARYNGEAL PRELIMINARIES .................................................................................... 15
  2.1 Anatomo-physiological aspects of the epilarynx ............................................................................. 15
  2.1.1 Epilarynx anatomy: Defining the epilarynx ............................................................................... 15
  2.1.2 Epilarynx physiology .................................................................................................................. 27
  2.2 Observations of the epilarynx in data and theory .......................................................................... 47
  2.2.1 Epilaryngeal observation difficulties: Focus on x-ray and laryngoscopy................................. 48
  2.2.2 Previous evidence and theory for the speech activity of the epilarynx ..................................... 55
  2.2.3 Esling and colleagues ................................................................................................................. 65
  2.2.4 Beyond Esling and colleagues ..................................................................................................... 78
  2.3 Nomenclature and taxonomy of lower vocal tract sounds categories ............................................ 84
  2.4 Chapter summary: Epilaryngeal preliminaries .............................................................................. 89

Chapter 3: EPILARYNGEAL VIBRATION ............................................................................................. 91
  3.1 The epilarynx as a source mechanism ..............................................................................................91
  3.1.1 Phonetic classification of epilarynx vibration ............................................................................. 93
  3.1.2 A survey of paralinguistic and linguistic uses of epilarynx vibration ....................................... 99
  3.1.3 A tour of the different types of epilaryngeal vibration .............................................................. 105
  3.1.4 Epilaryngeal vibration acoustics: Agul and N|u ....................................................................... 119
  3.2 High-speed laryngoscopy of aryepiglottic trilling in Iraqi Arabic .................................................. 132
  3.2.1 Methodology .............................................................................................................................. 134
  3.2.2 Results and discussion .............................................................................................................. 136
  3.2.3 Summary of high-speed laryngoscopy of AE trilling in Iraqi Arabic ......................................... 145
  3.3 A two-trap-door model of aryepiglottic vibration .......................................................................... 146
  3.3.1 Previous models of epilaryngeal vibration .................................................................................. 147
  3.3.2 Principles of design for the aryepiglottic vibration model ......................................................... 152
  3.3.3 Simulation results ...................................................................................................................... 165
  3.3.4 Discussion .................................................................................................................................. 170
  3.3.5 Summary of the aryepiglottic vibration model .......................................................................... 178
  3.4 Chapter summary: Epilaryngeal vibration ...................................................................................... 179

Chapter 4: THE EPILARYNX AND THE VOCAL FOLDS ..................................................................... 181
  4.1 Visualizing vocal-ventricular contact with laryngeal ultrasound .................................................. 182
  4.1.1 Laryngeal ultrasound methodology ........................................................................................... 185
  4.1.2 Results ....................................................................................................................................... 187
  4.1.3 Discussion ................................................................................................................................... 190
  4.2 SLLUS observation of lower epilarynx function and larynx height ............................................... 192
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1 SLLUS methodology</td>
<td>194</td>
</tr>
<tr>
<td>4.2.2 Optical flow analysis</td>
<td>196</td>
</tr>
<tr>
<td>4.2.3 Qualitative analysis of laryngoscopy</td>
<td>199</td>
</tr>
<tr>
<td>4.2.4 SLLUS results</td>
<td>200</td>
</tr>
<tr>
<td>4.2.5 Discussion</td>
<td>206</td>
</tr>
<tr>
<td>4.2.6 Summary: Lower epilarynx function and larynx height</td>
<td>212</td>
</tr>
<tr>
<td>4.3 Modeling vocal-ventricular fold coupling</td>
<td>213</td>
</tr>
<tr>
<td>4.3.1 Introduction: The larynx – more than just the vocal folds</td>
<td>214</td>
</tr>
<tr>
<td>4.3.2 Method: The Vocal-Ventricular Fold Coupling (VVFC) Model</td>
<td>222</td>
</tr>
<tr>
<td>4.3.3 Results: General overview of model performance:</td>
<td>239</td>
</tr>
<tr>
<td>4.3.4 Simulation case studies: Overview</td>
<td>243</td>
</tr>
<tr>
<td>4.3.5 Discussion</td>
<td>257</td>
</tr>
<tr>
<td>4.3.6 Summary of the vocal-ventricular fold coupling model</td>
<td>267</td>
</tr>
<tr>
<td>4.4 Chapter summary: The epilarynx and the vocal folds</td>
<td>269</td>
</tr>
<tr>
<td>Chapter 5: THE EPILARYNX AND THE SUPRALARYNGEAL VOCAL TRACT...</td>
<td>270</td>
</tr>
<tr>
<td>5.1 Visualizing epilaryngeal constriction continuum with videofluoroscopy</td>
<td>271</td>
</tr>
<tr>
<td>5.1.1 Videofluoroscopy methodology</td>
<td>272</td>
</tr>
<tr>
<td>5.1.2 Videofluoroscopy results and discussion</td>
<td>275</td>
</tr>
<tr>
<td>5.1.3 Videofluoroscopy summary</td>
<td>283</td>
</tr>
<tr>
<td>5.2 SLLUS observations of upper epilaryngeal stricture</td>
<td>284</td>
</tr>
<tr>
<td>5.2.1 SLLUS methodology</td>
<td>286</td>
</tr>
<tr>
<td>5.2.2 SLLUS results and discussion</td>
<td>286</td>
</tr>
<tr>
<td>5.2.3 SLLUS summary</td>
<td>292</td>
</tr>
<tr>
<td>5.3 Vowel quality change as a function of larynx height</td>
<td>293</td>
</tr>
<tr>
<td>5.3.1 Methods</td>
<td>295</td>
</tr>
<tr>
<td>5.3.2 Results</td>
<td>296</td>
</tr>
<tr>
<td>5.3.3 Discussion</td>
<td>302</td>
</tr>
<tr>
<td>5.3.4 Summary</td>
<td>307</td>
</tr>
<tr>
<td>5.4 The epilarynx and vowels: A model of subtle and strong effects</td>
<td>308</td>
</tr>
<tr>
<td>5.4.1 Subtle Effects: Open vowel susceptibility to epilaryngeal stricture</td>
<td>309</td>
</tr>
<tr>
<td>5.4.2 More dramatic effects: The “neo-tongue” &amp; the “neo-pharynx”</td>
<td>317</td>
</tr>
<tr>
<td>5.5 Chapter summary: The epilarynx and the supralaryngeal vocal tract</td>
<td>326</td>
</tr>
<tr>
<td>Chapter 6: LOWER VOCAL TRACT PHONOLOGY: THEORY</td>
<td>328</td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>329</td>
</tr>
<tr>
<td>6.1.1 Scope of work: Guttural / post-velar phonology and beyond</td>
<td>331</td>
</tr>
<tr>
<td>6.1.2 Broader theoretical concerns: The phonetics-phonology interface</td>
<td>335</td>
</tr>
<tr>
<td>6.1.3 Broader theoretical concerns: Phonological features or potentials?</td>
<td>340</td>
</tr>
<tr>
<td>6.2 Previous models of lower vocal tract phonology</td>
<td>350</td>
</tr>
<tr>
<td>6.2.1 Glottocentrism-linguocentrism in guttural/post-velar phonology</td>
<td>351</td>
</tr>
<tr>
<td>6.2.2 Early approaches to features of lower vocal tract sounds</td>
<td>352</td>
</tr>
<tr>
<td>6.2.3 Feature geometric representations of gutturals/post-velars</td>
<td>357</td>
</tr>
<tr>
<td>6.2.4 A short excursus on OT and lower vocal tract sounds</td>
<td>363</td>
</tr>
<tr>
<td>6.2.5 Lingual-laryngeal encoding in Articulatory Phonology</td>
<td>365</td>
</tr>
<tr>
<td>6.2.6 Summary: General problems with previous approaches</td>
<td>368</td>
</tr>
<tr>
<td>6.3 LVT Phonology: A model of articulation-based phonological potentials</td>
<td>369</td>
</tr>
<tr>
<td>6.3.1 Physiological components of epilaryngeal operation</td>
<td>370</td>
</tr>
</tbody>
</table>
6.3.2 Phonological potentials of the epilarynx and lower vocal tract.......................... 374
6.3.3 Phonological potentials: Formalism ................................................................. 388
6.3.4 Predictions about the pathways of lower vocal tract phonology ...................... 397
Chapter 7: LOWER VOCAL TRACT PHONOLOGY: ARGUMENTS.............................. 401
  7.1 Expressions of epilaryngeal vibration in phonology......................................... 401
    7.1.1 Distinctive epilaryngeal vibration in Khoisan ........................................ 403
    7.1.2 Epilaryngeal vibration and pharyngeal consonants .................................. 408
    7.1.3 The dual nature of epilaryngeal vibration ............................................... 416
    7.1.4 Epilaryngeal vibration and tone systems: Zhenhai and Bai ......................... 420
  7.2 The epilarynx and the vocal folds .................................................................... 427
    7.2.1 Strong and weak glottal stops ................................................................. 428
    7.2.2 Laryngeal-pharyngeal interaction in Tigre ............................................... 435
    7.2.3 Hypopharyngeal allophony in Amis ......................................................... 443
    7.2.4 Pharyngeal genesis in Southern Wakashan and Ejectives ......................... 453
    7.2.5 Relatively open vowels and glottal stops .................................................. 469
  7.3 The epilarynx and the supralaryngeal vocal tract ............................................ 481
    7.3.1 Pharyngeals, pharyngealization, and vowels ............................................ 482
    7.3.2 Register contrasts and vowel quality ....................................................... 500
    7.3.3 “ATR languages” and their relationship to register .................................... 511
  7.4 Discussion: The phonological epilarynx and beyond ....................................... 531
    7.4.1 Summary of the LPP account ................................................................. 531
    7.4.2 Summary of what is outside the scope of the model ................................. 538
    7.4.3 Comparison of the LPP to mainstream models of phonology ..................... 540
    7.4.4 Features revisited: [±constricted epilaryngeal tube] .................................. 542
  7.5 Summary and directions for phonology .......................................................... 545
Chapter 8: CONCLUSION ......................................................................................... 551
  8.1 The theory of the epilarynx in speech: A summary .......................................... 552
    8.1.1 Epilaryngeal vibration ............................................................................... 554
    8.1.2 The epilarynx and lingual-laryngeal linkage ............................................. 555
  8.2 Directions for future epilarynx-related research .............................................. 559
    8.2.1 Future empirical studies of the epilarynx .................................................. 559
    8.2.2 Future computational studies of the epilarynx .......................................... 561
    8.2.3 Future phonological research related to the epilarynx ................................. 563
    8.2.4 Extended domains of epilaryngeal research ............................................. 565
  8.3 Epilaryngeal coda ............................................................................................ 567
Bibliography ........................................................................................................... 568
Appendix A: List of languages ................................................................................ 613
Appendix B: Definition of terms .............................................................................. 615
Appendix C: Transcription of Burkikhan and Tpig Agul ......................................... 618
Appendix D: List of mathematical symbols ............................................................. 622
Appendix E: List of muscles and muscle abbreviations .......................................... 623
Appendix F: Phonological extensions .................................................................... 624
List of Tables

Table 2.1: Laryngoscopic natural language data collected by Esling ............................................. 71
Table 2.2: Unconstricted and constricted laryngeal states and vocal fold parameters .... 73
Table 2.3: Major hypotheses arising from the research of Esling ........................................... 77
Table 3.1: Aryepiglottic trill duration and average frequency .................................................... 137
Table 3.2: Aryepiglottic fold mucosa parameters in comparison those to ST95 ........... 157
Table 3.3: Aryepiglottic fold trapdoor parameters in comparison to M92 ................. 159
Table 3.4: Circuit parameter formulae for airflow in the aryepiglottic model .......... 162
Table 4.1: Biomechanical parameter ranges .............................................................................. 240
Table 4.2: Specific biomechanical parameters used in the case studies ......................... 244
Table 5.1: Larynx raising in Arabic pharyngeal consonants .................................................. 285
Table 5.2: Formant change cross-participant averages relative to neutral height ....... 300
Table 6.1: Cross-linguistic variation in guttural/post-velar class membership .......... 332
Table 6.2: Lower vocal tract issues that are addressed in this dissertation ..................... 334
Table 6.3: Lower vocal tract issues not addressed ............................................................... 334
Table 6.4: Czakowksa-Higgins’ feature matrix for uvulars and pharyngeals ................ 355
Table 6.5: Definitions of the components of epilaryngeal control ......................................... 371
Table 6.6: Typical concomitant phonetic properties of epilaryngeal states .................... 373
Table 6.7: Gross (physiological) states of the lower vocal tract ............................................ 375
Table 6.8: Hypopharyngeal stop potentials .............................................................................. 386
Table 7.1: Trigo’s representation !Xóõ phonation types .......................................................... 406
Table 7.2: Aryepiglottic trilling in Iraqi Arabic ................................................................. 410
Table 7.3: Narrow phonetic transcriptions of Burkikhan and Tpig Agul ....................... 412
Table 7.4: Distribution of whisper, whispery voice, and growl in Zhenhai .................. 422
Table 7.5: Tonal-register contrasts in Bai ............................................................................. 424
Table 7.6: Minimal pairs in Quianviní Zapotec: creaky vs. interrupted vowels ......... 430
Table 7.7: Hypopharyngeal allophony in Amis ................................................................. 447
Table 7.8: Caucasian languages with pharyngealization .................................................... 450
Table 7.9: Pharyngeal genesis in the history of Southern Wakashan ......................... 454
Table 7.10: Synchronic activity of Pharyngeal Genesis ....................................................... 455
Table 7.11: Gradient of uvular proneness to induce epilaryngeal stricture ............... 463
Table 7.12: Vowels of Udi ................................................................................................. 492
Table 7.13: Vowels of Even ............................................................................................... 496
Table 7.14: Vowel subsets in !Xóõ, Ju’hoansi, and N|uu ................................................. 501
Table 7.15: Phonetics of register in Bor Dinka ................................................................. 504
Table 7.16: Vowel alternations in Bor Dinka ..................................................................... 504
Table 7.17: “ATR” Systems ............................................................................................. 513
Table 8.1: Future empirical (imaging) research ................................................................. 560
Table 8.2: Future computational modeling research I ....................................................... 562
Table 8.3: Future computational modeling research II .................................................... 563
Table 8.4: Predicted quantal biomechanical-articulatory effects of the epilarynx .... 564
Table 8.5: Epilaryngeal vibration in speech sound systems .............................................. 627
Table 8.6: Epilarynx interaction with the vocal folds ......................................................... 628
Table 8.7: Epilarynx interaction with the supralaryngeal vocal tract ............................ 629
List of Figures

Figure 1.1: Basic epilarynx anatomy ................................................................. 1
Figure 2.1: The epilarynx as a tube-in-a-tube .................................................. 18
Figure 2.2: Anatomical sketch of the larynx .................................................... 20
Figure 2.3: Laryngoscopic views of the epilaryngeal tube ............................... 24
Figure 2.4: Functional planes of the larynx ...................................................... 28
Figure 2.5: Fink’s continuum of laryngeal folding .......................................... 31
Figure 2.6: Laminagrams showing laryngeal plication and its effect on the epilarynx 32
Figure 2.7: Upturned/inlet (a) vs. downturned/outlet (b) valves ....................... 37
Figure 2.8: Intralaryngeal musculature associated with the epilarynx ............... 42
Figure 2.9: Different styles of x-ray tracing .................................................... 50
Figure 2.10: Ladefoged’s one-dimensional, abduction-adduction continuum ...... 57
Figure 2.11: Gauffin’s two-dimensional model of laryngeal states in speech ...... 57
Figure 2.12: Place of articulation redrawn from Sounds of the World’s Languages ... 67
Figure 2.13: Esling’s (2005) Laryngeal Articulator Model (LAM) .................. 74
Figure 3.1: Variation in epilaryngeal aperture topology and vibration patterns ......................................................................................................................... 109
Figure 3.2: Variations on epilaryngeal vibration patterns ................................ 110
Figure 3.3: Voiced, “billowing” vibration of the inner aryepiglottic mucosa ....... 111
Figure 3.4: Voiced vibration of the inner aryepiglottic mucosal “ring” ............... 112
Figure 3.5: Lower, inner aryepiglottic mucosa vibration ................................ 113
Figure 3.6: Voiced aryepiglottic trilling ........................................................... 114
Figure 3.7: Voiceless aryepiglottic trilling with left-side cuneiform oscillation ... 115
Figure 3.8: A voiceless epiglottal trill in Somali (/hanii/ ‘testicle’) ....................... 117
Figure 3.9: High-speed laryngoscopy data (500 fps) of voiced aryepiglottic vibration . 121
Figure 3.10: Voiced epilaryngeal vibration in Burkikhan Agul ‘bridges’ .......... 124
Figure 3.11: Voiceless epilaryngeal vibration in Burkikhan Agul ‘wheys’ ........... 125
Figure 3.12: Voiced epilaryngeal vibration following [ʔ] in Burkikhan Agul ‘to cry’ .. 127
Figure 3.13: Voiced epilaryngeal vibration following Tpig Agul ‘crest’ ............... 128
Figure 3.14: Voiced epilaryngeal vibration in Nuu ‘to fly’ ................................. 130
Figure 3.15: Voiced epilaryngeal vibration in Nuu ‘chameleon’ ....................... 131
Figure 3.16: Singleton voiceless aryepiglottic trill [h] in /raːhiːl/ ‘to travel’ ............ 139
Figure 3.17: Geminate voiceless aryepiglottic trill [hrː] in /raːhiːl/ ‘to travel a lot’ .... 140
Figure 3.18: Singleton voiced aryepiglottic trill [ʕ] in /saːʃiːd/ ‘happy’ ............... 141
Figure 3.19: Geminate voiced aryepiglottic trill [ʕʕ] in /saːʃiːd/ ‘make people happy’ . 142
Figure 3.20: Trace of simulated and empirical results for vocal-ventricular phonation. 151
Figure 3.21: Estimated epilarynx mid-sagittal area during voiced aryepiglottic trilling 156
Figure 3.22: Mechanical model of the aryepiglottic folds ................................. 158
Figure 3.23: Geometry of the aryepiglottic fold ............................................. 160
Figure 3.24: Equivalence circuit the aryepiglottic model ................................. 162
Figure 3.25: Areas enclosed during simulated phases of aryepiglottic oscillation 163
Figure 3.26: Pressure changes in the epilarynx .............................................. 164
Figure 3.27: Simulation of vocal fold vibration .............................................. 165
Figure 3.28: Simulation of voiceless aryepiglottic “trilling” .......................... 166
Figure 3.29: Simulation of voiced aryepiglottic “trilling” .............................. 167
Figure 3.30: A comparison of glottal source spectra ............................................. 168
Figure 3.31: Simulation of voiced aryepiglottic trilling at a G-F0 of 65 Hz .................. 169
Figure 3.32: Simulation of voiced aryepiglottic trilling at a G-F0 of 124 Hz ............... 170
Figure 3.33: Vocal fold closure alternation pattern during voiced aryepiglottic trilling 172
Figure 4.1: Illustration of ultrasound probe placement ........................................ 186
Figure 4.2: Structural registration in coronal laryngeal ultrasound ............................. 188
Figure 4.3: Coronal laryngeal ultrasound of glottal stop ....................................... 189
Figure 4.4: Validation of the optical flow algorithm ............................................... 198
Figure 4.5: Velocity and position of the metal bar ............................................... 199
Figure 4.6: Indicators of laryngeal constriction ..................................................... 200
Figure 4.8: SLLUS data for glottal stop without height suppression ....................... 201
Figure 4.9: SLLUS of creaky voice with deliberate height suppression .................. 204
Figure 4.10: Longitudinal nature of the effect of the epilarynx on the vocal folds ....... 207
Figure 4.11: The relationship amongst larynx height, constriction, and pitch control .. 210
Figure 4.12: The effect of larynx height ............................................................... 211
Figure 4.13: Abstraction of VVFC into low-dimensional model .............................. 225
Figure 4.14: Spatial abstraction and mass movement during the glottal cycle .......... 229
Figure 4.15: Overview data of model performance ............................................. 241
Figure 4.16: Model performance as a function of percentage-wise increasing VVFC ... 243
Figure 4.17: Plots for case study #1 ................................................................... 247
Figure 4.18: Plots for case study #2 ................................................................... 249
Figure 4.19: Model state plots for Simulation #4 .................................................. 251
Figure 4.20: Model state plots for Simulation #5 .................................................. 253
Figure 4.21: Simulation of glottal stop ................................................................. 256
Figure 5.1: Parameter extraction from videofluoroscopy data ................................. 273
Figure 5.2: Videofluoroscopic images of the epilarynx .......................................... 276
Figure 5.3: VFS data for [a½a] ........................................................................ 279
Figure 5.4: VFS data for [a½a] ........................................................................ 280
Figure 5.5: VFS data for [a½a] ........................................................................ 282
Figure 5.6: SLLUS data for [i½i] .............................................................. 288
Figure 5.7: SLLUS data for [i½i] .............................................................. 289
Figure 5.8: SLLUS data for [i½i] .............................................................. 290
Figure 5.9: First three resonances of a hard-walled, uniform tube ....................... 294
Figure 5.10: Larynx height and formant frequency ............................................ 297
Figure 5.11: Average larynx height change by vowel ......................................... 298
Figure 5.12: Average formant change by larynx height condition and vowel .......... 299
Figure 5.13: Actual larynx height data regressed on formant frequency ............. 300
Figure 5.14: Vowel quality and passive epilarynx narrowing ............................... 311
Figure 5.15: Midsagittal tracings of the vocal tract based on MRI data .................. 313
Figure 5.16: Midsagittal schematic of passive epilaryngeal stricture in [a] .............. 314
Figure 5.17: Coronal section of the larynx illustrating stricture bias of [a] ............. 316
Figure 5.18: Palato-pharyngo-epiglottal configuration in !Xôô ................................ 318
Figure 5.19: “Neo-tongue” formation during “double-bunching” epilaryngealization 320
Figure 5.20: Role of strong middle genioglossus contraction in “double bunching” .... 322
Figure 5.21: Tube-abstraction explanation of the “neo-tongue” and “neo-pharynx” 324
| Figure 6.1: Phonological potentials and phonological reality of /h/ | 346 |
| Figure 6.2: Five variations on guttural/post-velar feature geometry | 359 |
| Figure 6.3: Representations of post-velars following Rose (1996, p. 80) | 362 |
| Figure 6.4: Key components of epilarynx operation | 371 |
| Figure 6.5: Schematic of synergistic relations in the LPP | 376 |
| Figure 6.6: Vocalic correspondence of several gross states | 378 |
| Figure 6.7: Broad pharyngeal regions of constriction and gross states | 381 |
| Figure 6.8: Palatal/palato-pharyngeal gross states (tube abstraction) | 382 |
| Figure 6.9: Epilaryngeal gross states (tube abstraction) | 384 |
| Figure 6.10: Schematic showing hypopharyngeal stop potentials along a continuum | 386 |
| Figure 6.11: Schematization of phonological potentials | 390 |
| Figure 6.12: The “big picture”: phonological potentials and realities | 394 |
| Figure 6.13: Simplified illustration of the potential glottal stop subsystem | 396 |
| Figure 7.1: Potential hypopharyngeal stop systems in the LPP | 433 |
| Figure 7.2: Failed GCLC analysis of Tigre laryngeal-pharyngeal neutralization | 439 |
| Figure 7.3: Potential Tigre sub-systems | 441 |
| Figure 7.4: Amis pharyngeal dorsalization in a GCLC analysis | 448 |
| Figure 7.5: Potential hypopharyngeal subsystems relevant to Amis | 452 |
| Figure 7.6: Pharyngeal genesis in GCLC models | 458 |
| Figure 7.7: Potential pharyngeal genesis, part I | 464 |
| Figure 7.8: Potential pharyngeal genesis, part II | 465 |
| Figure 7.9: Potential pharyngeal genesis, part III | 467 |
| Figure 7.10: Relationship between glottal stop and open vs. close vowels | 479 |
| Figure 7.11: Relationship between glottal fricative and open vs. close vowels | 480 |
| Figure 7.12: Forming the neo-tongue and neo-pharynx | 490 |
| Figure 7.13: Two species of pharyngeal in the LPP | 498 |
| Figure 7.14: Potential unconstricted-constricted registers | 508 |
| Figure 7.15: X-ray tracings of Akan and Ateso [±ATR] vowel sets | 515 |
| Figure 7.16: Cross-height harmony (“ATR”) in the LPP | 521 |
| Figure 7.17: From consonantal contrast to register to cross-height contrast in the LPP | 529 |
| Figure 7.18: Synergistic relations in the LPP | 533 |
| Figure 8.1: Potential unconstricted-constricted registers | 624 |
| Figure 8.2: Cross-height harmony (“ATR”) in the LPP | 625 |
| Figure 8.3: Linkage of voicing, register, and cross-height harmony | 626 |
Acknowledgments

The dust has settled. At the end of the day (dissertation), I am overwhelmed with gratitude that I am loved by somebody like Carly, my best friend and wife, whose support was instrumental in seeing me through to the completion of this document. I believe that the meaning of life emerges from sharing it with those you love, and I am one of the lucky ones to be in the company of someone so dedicated and patient, sharp and intelligent, mischievous and tenacious, worldly and diplomatic, and, above all, deeply loving and caring. Thanks for putting up with me.

I’m also blessed to have my family: Ryan, Mom, Dad, Grandma, Lorne, Sandy, you’ve all been so supportive over the years. Many thanks to Ronn and Colleen, Ben and John, Eunice and Bill, and Sharon, too. When you integrate all of this family support, it amounts to a lot of love – enough to have seen me through to this stage. I love you guys.

At moments like these, I also think of the spirits in my life\(^1\): Jordan, your words still echo within me (I miss you); Mike Dobrovolsky, I’ll strive to live up to the example you set: \(\text{teşekkür ederim}\); Grandad, I can still hear your laughter, and it makes me smile.

Then there are my committee members, whose mettle and resolve to tackle this growling, snarling beast of a dissertation leaves me feeling deeply indebted. In my eyes, you’re all A+++ professors:

John (Esling), to me you are the larynx guru and a legend of many voices. Thanks for the laughs, for letting me indulge my inner artist, for showing me what you can learn about speech by listening to the Muppets, and for your encouragement to attend all of

\(^1\) Like the spirits of Yoda, Obi Wan, and Anakin at the end of Return of the Jedi – the original not the redux.
those conferences – it has made all the difference. I hope to pay it forward one day. A sincere thank you.

Ewa, you were always there to catch me at the right moments and prevent me from falling to pieces. Thanks for your belief in me, for your guidance and caring for me (and Carly), and for helping me see value in my work. Vive la Phonologie²!

Bryan, how will I ever repay your support? You’ve taught this man to fish³. Thanks for your friendship, encouragement, and generosity, for the countless hours of stirring discussion on Skype, and for opening my eyes to the world beyond the larynx⁴. And thanks for showing me how to pay my “brass tax”, too.

Sonya, you’re an excellent role model for what makes a great linguist/scientist and teacher: I’ve been taking notes! Thanks for all of your help over the years by taking my work seriously and by asking those questions that go deeper and deeper.

John (Ohala), funny how our paths keep crossing. I hope this continues to be the case. Thank you for your sober view of phonological theory and for your intoxicating jokes. I’ll do my best to follow the example you’ve set. (And thanks for Russell, too – it’s so interesting how the knowledge has been “out there” for a good while.)

For those department members spared the eye strain of my dissertation but still instrumental in helping me along the path to a PhD, I also extend a deeply felt thank you. Leslie Saxon, Alex D’Arcy, Hua Lin, and Su Urbanczyk have all contributed so much to my growth as an academic and as a person. There are many UVic grad students who have helped me survive in the field of battle that is grad student life. Special mention goes to

---

² What else but [±voice] explains why I sometimes substitute “b” for “p” (and vice versa) when typing on a keyboard?
³ Something I could never do in Saskatchewan.
⁴ I didn’t realize it existed and was so big!
Janet Leonard, Thomas Magnuson, and Allison Benner. And then there is Chris Coey: “Mad Dog” says thanks for all of the technical help and for all the laughs.

Finally, there is an extensive group of people I’ve met outside of the UVic environment (some at UofC, some during my travels abroad, and some during my sojourns at UBC) that I wish to thank for their generosity and support (either direct or indirect), insights, and/or trail-blazing: Darin Flynn, John Archibald, Martha McGinnis, Anthony Bowers, Erin Baillie, Donald Derrick, Sydney Fels, Ian Stavness, Ling Tsou, Peter Anderson, Ho Beom Kwon, Doug Pulleyblank, Chenhao Chiu, Matt and Andrew Richards, Zeki Majeed Hassan, Lise Crevier-Buchman, Chakir Zeroual, Rachid Ridouane, Marzena Żygis, Ken-Ichi Sakakibara, Peter Birkholz, Phil Rose, Amanda Miller, Sharon Rose, Brad Story, Ingo Titze, Fariborz Alipour, my Ling 183 students, and Michael Gira. I send a salutational growl to all of you.

_No vocal folds were harmed in the making of this dissertation._
Dedication

For Carly: my best buddy – and cute as a bamboo shoot, too!
Chapter 1

INTRODUCTION

The human epilarynx comprises the laryngeal structures that form a roughly tube-shaped space immediately above the vocal folds, as depicted in Figure 1.1. Physiologically, the epilarynx is an essential part of the mechanism that protects the airway leading to the lungs, but it also plays many important roles in speech production, which all relate to its position within the vocal tract – in between the vocal folds and the tongue.

Figure 1.1: Basic epilarynx anatomy. See §2.1.1 for a more detailed anatomical definition of the epilarynx.

Understanding the epilarynx is relevant to many functions of the larynx in general, such as respiration, swallowing, effort exertion and parturition, coughing and
throat clearing, vocalization, speech, and singing (Hillel, 2001). The main focus here is its role in speech, and our understanding of the linguistic functions of the epilarynx is still in its infancy (as noted in Miller, 2012, p. 38). There is, however, a considerable volume of empirical data showing activity of the epilarynx in connection with a wide range of sounds in language, including consonants (such as laryngeals [particularly glottal stop], pharyngeals, and secondary articulations associated with places of articulation, such as pharyngealized consonants and ejectives), vowels (such as pharyngealized vowels or certain vowels in some tongue-root harmony languages), and phonation types (creakiness, whisperiness, and harshness [including growling]), which has consequences for tone and intonation systems.

The epilarynx also plays a crucial function in the earliest stages of speech development: for all infants, phonetic learning starts with manipulation of the epilarynx in its constricted state and gradually expands into unconstricted productions and eventually into the oral domain (Bettany, 2004; Esling, Benner, Bettany, & Zeroual, 2004; Benner, 2009). The epilarynx also is a variable in paralinguistic functions, such as signaling emotional states. In acoustic terms, its location between the glottal source and the vocal tract filter enables it to play a unique role in controlling the aero-acoustic coupling of these components (Titze, 2008). The epilarynx important in the production of a wide range of voice qualities used in singing. It helps the voice compete with an orchestra by producing the singer’s formant, and it can actively narrow to produce certain vocal styles such as belting or twang (Sundberg, 1974; Yanagisawa, Estill, Kmucha, & Leder, 1989; Honda, Hirai, Estill, & Tohkura, 1995; Schellenberg, 1998; Titze, 2008; Honda, Kitamura, Takemoto, et al., 2010; Esling & Edmondson, 2011). It also is an
essential asset that performers and ventriloquists use to create different voices for their characters (Painter, 1986, p. 330) Vibration of the epilarynx occurs in a vast array of musical styles, from blues and jazz to hard rock and death metal (Borch, Sundberg, Lindestad, & Thalén, 2004; Eckers, Hütz, Kob, et al., 2009), and is an important feature of many ethnomusical and ethnovocal practices such as throat singing (Nattiez, 1999; Levin & Edgerton, 1999; Lindestad, Södersten, Merker, & Granqvist, 2001; Bailly, Henrich, & Pelorson, 2010). Despite the awareness of these linguistic and paralinguistic uses of the epilarynx, a fully articulated theory of its nature in anatomical and physiological, phonetic and phonological terms is still lacking.

The purpose of this dissertation is to develop a theory of the epilarynx in speech. This theory is built upon the work of Esling (1996, 1999, 2005), Edmondson & Esling (2006), and those that played an important role in laying the ground work for this research to develop, including Gauffin (1969; 1972; 1977; see Lindblom, 2009), Fink (1956, 1962, 1974a, 1974b, 1975), Catford (1977a, 1977b, 1983), Traill (1986), and Painter (1986, 1991). This theory has three parts describing three epilarynx functions: the first is epilarynx vibration, the second is the relationship between the epilarynx and the vocal folds, and the third is the interaction of the epilarynx with the supralaryngeal vocal tract. Each of these functions must be considered in the study of the phonetics and phonology of sounds involving the lower vocal tract in their production, sometimes referred to as post-velars or gutturals. The scope of the influence of the epilarynx in speech is much larger, however, since it has more subtle, and sometimes not so subtle, interaction with vowel quality and prosody that likely materializes on a regular basis in the speech of all talkers and representing all language groups. Thus, the broad aim of this
work is to bring us closer to the goal of understanding these more general uses of the epilarynx in speech and beyond.

1.1 The epilarynx as a speech production system

Before we delve into the issues surrounding the role of the epilarynx in speech, it will be useful to consider the context in which we are studying it. In this dissertation, I start with the conventional assumption that the characterization of speech sounds is stratified into three layers of information, each progressively more abstract (Laver, 1994, pp. 27–54; Pierrehumbert, 2003, p. 178; cf. Smolensky, 2006). The most concrete level is the domain of anatomy and physiology, which relates directly to the physical, organic, observable world. Layered on this is the phonetic domain, which maps physical properties and processes onto a finite set of sound categories relevant to human languages. Finally, there is the domain of phonology, which organizes the continuous domain of phonetics into a discrete and arbitrary code that is the basis of spoken communication (Hall, 2009, p. 26). Adjacent information layers show dependencies: phonetic analysis cannot proceed without understanding the physical world, and phonology likewise relies upon phonetics to predicate anything meaningful about the systemic nature of sounds.

In this dissertation, the focus is on the analysis of speech sound production. Each layer of information is important: anatomo-physiological, phonetic, and phonological domains all must be considered to understand the role that a particular part of the vocal tract plays in the formation of speech sounds and why those sounds behave the way they
do in language. The approach taken here modularizes the vocal apparatus into a hierarchically organized set of interacting speech production systems. A speech production system comprises a collection of anatomical structures which work together in producing a particular sound feature or set of speech sounds. For example, we might consider the lips to constitute a speech production system — the labial system. A simple analysis would maintain that the labial system is composed of the upper and lower lip and the musculature that moves these structures. The labial system is phonetically versatile: we have fine control over labial aperture and shape, allowing for diverse strictures to form an array of speech sounds and features, such as rounding. This system is also phonologically relevant: labial sounds serve a contrastive function in the majority of the world’s languages. There are many such systems in the vocal tract — the vocal folds, velum, and tongue tip being obvious examples; in fact, the vocal tract itself qualifies as a system of speech production, one composed of numerous subsystems that operate together to generate the speech signal.

The purpose of this dissertation is to argue that the epilarynx is one of these speech production systems. There is abundant evidence that shows that various parts of the epilarynx, such as the ventricular folds, epiglottis, or aryepiglottic (AE) folds, are phonetically active in many languages. However, to claim that the epilarynx operates as a speech production system is not trivial: its many components have relatively wide spatial distribution within the vocal tract and there are numerous dependencies in its operation: larynx height, tongue retraction, epiglottis and aryepiglottic action, ventricular fold action, vocal fold position, and pharynx constriction all influence and/or are influenced

---

5 Although the divisions between modules is likely more apparent than actual and biased by one’s cultural perspectives.
by what the epilarynx does in one way or another. It could be argued that only the vocal folds and tongue matter in producing the speech sounds associated with the epilarynx, such as pharyngeals, and as will be discussed, this is indeed the approach taken by many researchers in phonetics and phonology. Previous analyses also posit sub-structures of the epilarynx as speech production systems, such as the epiglottis (Laufer & Condax, 1979, 1981), aryepiglottic folds (Esling, 1996, 1999), or ventricular folds (Catford, 1977a). There is, thus, no a priori motivation to believe that the collection of components comprised by the epilarynx constitute a functionally unified mechanism spanning the anatomo-physiological, phonetic, and phonological levels. Nevertheless, this is the claim put forth here: at all of these levels, the notion of the epilarynx retains its coherence as a system associated with a specific speech production potential.

In building this account, all three layers characterizing speech sounds will be considered, along with the relationships between layers. Consideration begins at the anatomo-physiological layer, the foundation of which is established in Chapter 2. A great deal is known about the structures of the epilarynx and how they function physiologically in processes such as effort closure and swallowing. These facts cannot be ignored or glossed over: the very structure of the mechanism plays a major role in determining its phonetic function, and patterning in the phonological layer reflects the anatomo-physiological configuration of the epilarynx. Likewise, the phonetic system adapts or repurposes the fundamental, life-supporting physiological behaviour of the epilarynx, as Lindqvist/Lindqvist-Gauffin/Gauffin observed in the 1970s, but in regard to the entire laryngeal mechanism (Lindqvist, 1969; Lindqvist-Gauffin, 1972; Gauffin, 1977). I am assuming then that anatomy and physiology have considerable explanatory power in the
phonetic and phonological layers, and thus it is highly useful to have clarity on what is known about the epilarynx in terms of the anatomo-physiological layer.

The concern in this dissertation is ultimately about speech sounds and their patterning, so the account cannot stop at the anatomo-physiological layer. The question becomes a matter of what these body parts do in speech. Unlike when Gauffin first published his ideas in the 1970s, a great deal more is now understood about the phonetic contributions of the parts of the epilarynx in the production of a wide variety of speech sounds. A considerable body of work exists which convincingly demonstrates the speech activity of the components the epilarynx and strongly suggests they work as a unit: this is what we expect if the epilarynx constitutes a system of speech production. The aim here is to rely heavily upon previous phonetic empiricism in making the case for the epilarynx but also to contribute to this research by a series of instrumental investigations primarily featuring cardinal phonetic productions. I will also provide an outline of the new directions that empirical phonetic research needs to take based on the predictions and observations made in this dissertation.

The phonological layer represents the last step in the account of the role of the epilarynx as a speech production system. I make the assumption that phonology is subject to phonetic grounding, that is to say it is not entirely “substance free”, contra Hale & Reiss\(^6\) (2000a, 2000b, 2008). With this assumption in hand, the discussion becomes

\(^6\) Hale & Reiss (2000a, 2008, p. 175) compare the phonological system to the visual system and draw on the famous “Kanizsa triangles” (Kanizsa & Gerbino, 1982) visual illusion to make their point. The visual system can generate a percept of a triangle from a graphic of three circles, each missing a triangular wedge at just the right spot, and each centered such that they form the vertices of a triangle (a). There is no explicit triangle in such a graphic, yet our visual system generates one (from a cognitive representation of a triangle).

My interpretation is that, while Hale & Reiss are correct about saying there is a need for a system that interprets the raw data and supplies a triangle percept, they go too far in denying the connection
focused on the major division drawn between the laryngeal and supralaryngeal components in most phonological models. Regardless of the choice of formalism or approach, this segregation arises. In Feature Geometry, the two categories of features are separated by virtue of hierarchical organization: laryngeals are placeless (Clements, 1985; Sagey, 1986; Steriade, 1987), i.e. LARYNGEAL is a node of the geometry not dominated by that which constitutes place of articulation. In Kehrein and Golston (2004), laryngeal features are prosodic constituents (subordinate to onsets, rhymes, codas, etc.), not segmental ones. In Articulatory Phonology, Borroff (2007) comes close to drawing a connection between glottal activity and lingual activity, but defaults in favour of rendering extraglottal gestures in glottal stop as Tongue Root gestures, with the assumption that reinterpration of these gestures is merely a matter of changing the labeling in the formalism. Miller (2012) presents a feature emergence account of laryngeal features, which, although not discounting the possibility for the supraglottal activity of the laryngeal structures, strongly focuses instead on vocal fold level activity.

All of these models reflect a bipartite paradigm that has dominated mainstream phonological conceptualization: glottocentrism-linguocentrism (or GCLC for short; also...

between the stimulus and the resulting representation (also see Hawkins, 2010, p. 65): true, there is no distinguishing between the edges or inside of the triangle and the background of the figure (both are white), but the triangle percept is not arbitrarily related to the stimulus either. Rotating one of the circles so that the wedge-shaped gap is in the wrong place diminishes the illusion of a triangle (b). In phonological terms, the representation is not arbitrarily related to the phonetic structure of sounds, whether articulatory, acoustic, auditory, or any other modality, even though higher order constructs (onsets, syllables, segments, features) can arise from this structure.
see Gick, 2011). What this means is that, in correspondence with the laryngeal and supralaryngeal domains, the primary concern of phonologists is the behaviour of the glottis (formed by the vocal folds) and the tongue, respectively. This approach neglects the epilarynx, which, in physical terms, occupies the space in between the vocal folds and the tongue. The argument I make here is that the gap can no longer be ignored in the phonology: these structures are coupled, and this fact is manifest in the phonologies of many languages. Yet, on account of GCLC, these facts have not been integrated into phonological explanation.

It is not for lack of attention to the evidence that suggests this intimate lingual-laryngeal connection either. The idea was put forth by Czaykowska-Higgins (1987) predating, by a couple of years, the “guttural” research bubble of the 1990s (Hayward & Hayward, 1989; McCarthy, 1991, 1994; Davis, 1995; Halle, 1995; Rose, 1996; Zawaydeh, 1999). Much of the discussion was framed with the available conceptual tools at the time, which (as we will be discussed in §2.2.1), reflected available empirical evidence. The tongue root takes center stage in this discussion and is given a heavy explanatory burden, one that it still carries in many subsequent accounts (Paradis & LaCharité, 2001; Shahin, 2002; Bin-Muqbil, 2006). A turning point is evident in Shahin (2011a); she makes the observation that “guttural” phonology needs to catch up to developments in phonetics (which are outlined in Sections 2.2.2, 2.2.3, and 2.2.4). This dissertation is in part intended to help the process along.

Although this dissertation explores issues of how these three levels of analysis (anatomo-physiological, phonetic, and phonological) are connected, the work is not intended to resolve the debate entirely, especially since no detailed consideration of the
cognitive domain is given. Rather, this dissertation is designed to examine how each of the three behaviours posited in the theory of epilarynx in speech – vibration, vocal fold interaction, vocal tract interaction – are characterized across the three levels of analysis. Suggestions are made about the possibilities for organization that explain certain systematicity associated with epilarynx function in speech, but a more comprehensive account must be given that considers many more aspects of speech than are possible to do in this work, including but not limited to acoustic-auditory relationships, perception, cognitive processing, sensation and motor-control, sociolinguistic factors, deeper aspects of the physics of fluid motion within the vocal tract and the biomechanics of articulation and the aero-mechanical coupling between these systems, and higher order structure of phonology including prosodic organization, and interfaces between phonology and syntax, pragmatics, discourse, and so forth.
1.2 \textit{Dissertation hypotheses and outline}

This dissertation is intended to combine the developments in anatomical, physiological, phonetic, and phonological understanding of lower vocal tract sound production (reviewed in the preceding sections of this introduction) into a theory of the epilarynx in speech. The work proceeds from the vantage point established by articulatory models developed by Esling and his colleagues (notably Jerold Edmondson and Jimmy G. Harris): the Laryngeal Articulator Model (Esling, 2005) and the Valves Model (Edmondson & Esling, 2006), and assumes that, while further empirical observation is always required, the observations they have made (along with those before and after) form a solid empirical foundation upon which a theory of the epilarynx can be constructed. Preliminaries to the theory of the epilarynx in speech are provided Chapter 2: these include a review of epilarynx anatomy, physiology, and the previous phonetic research.

The dissertation then focuses on three key topics of epilaryngeal function in speech (and related phenomena such as singing): (1) epilarynx vibration; (2) the interaction of the epilarynx with the vocal folds; and (3) the relationship between the epilarynx and the supralaryngeal vocal tract. These three topics form the general structure of the dissertation, and each in turn is explored from a phonetic orientation (Chapters 3-5) and then from a phonological orientation (Chapter 6-7). In developing the theory, several small production studies are presented in complement to computational modeling and arguments derived from phonetic and phonological theoretical understanding.

Each of the phonetic chapters (Chapter 3, Chapter 4, and Chapter 5) investigates, in turn, the hypotheses associated with the three key topics of epilarynx functioning listed
above. These hypotheses arise from consideration of the previous research pertaining to the epilarynx and research that has subsequently challenged or expatiated on that which came earlier (Chapter 2). Then, in Chapter 6 and 7, the phonological issues surrounding each of the three topics are used to further address the hypotheses forming the scientific core of the theory of the epilarynx in speech.

The first hypothesis concerns epilarynx vibration (Chapter 3; Chapter 7, §7.1). The sound produced by epilaryngeal vibration, impressionistically called “growling” (Rose, 1989), is conventionally associated with pathological phonation (as it often is used to compensate for vocal fold pathology; see Crevier-Buchman, Pillot-Loiseau, Rialland, et al., 2012). As a consequence of its pathological stigma, research on the epilarynx vibration in non-pathological speech is considerably underdeveloped. This dissertation takes steps towards addressing this lack of research. The hypothesis proposed and examined in the dissertation is as follows: (1) all forms of epilarynx vibration tend towards the same basic effect (low frequency source structure), but (2) in speech, the tendency is for upper epilarynx (i.e. aryepiglottic) vibration. The prediction is that individuals will vary in the exact execution of epilarynx vibration, but the basic behaviour is categorical and, consequently, a suitable but rare basis for forming phonological contrast.

The second hypothesis concerns the interaction between the vocal folds and the epilarynx (Chapter 4; Chapter 7, §7.2). The hypothesis is that (1) the tissues of the lower epilarynx (i.e. the ventricular folds) compress into the vocal folds, and (2) this mechanism is a key basis for phonological interaction between the vocal folds and the

---

7 Which, by my impression formed from several years of listening to it and for it, is actually very commonly used in everyday speech, especially as a correlate of emphatically stressed utterances.
structures of the supralaryngeal vocal tract. The first part (1) of this hypothesis suggests that the ventricular folds perturb the dynamics of the vocal folds, changing and ultimately inhibiting their oscillatory behaviour through mechanical contact. As a corollary, this vocal-ventricular fold coupling is a key mechanism in producing “constricted” phonation types (i.e. those phonatory qualities with increased epilaryngeal engagement), and it is predicted that in normal speech these qualities will tend to exhibit some form of lower epilarynx involvement (minimally). The second part (2) of the hypothesis predicts that sounds which have relatively more epilaryngeal stricture will be more likely to exhibit the effects of vocal-ventricular fold coupling. Thus, in phonological systems, glottal stop and constricted phonation types should show some bias or susceptibility towards interaction with other sounds which employ epilaryngeal stricture, such as pharyngeals, and relatively more open vowels, for which the epilarynx tends to be narrower, than with relatively closer vowels, where the epilarynx is somewhat wider.

The third hypothesis is about how the epilarynx relates to the rest of the supralaryngeal vocal tract (Chapter 5; Chapter 7, §7.3). A key assumption is that the epilarynx has intrinsic and extrinsic control systems: respectively, these are the constriction-inducing intralaryngeal musculature and the tongue-retraction–larynx-raising muscle system. The hypothesis is that these two control systems give the epilarynx partial independence from the supralaryngeal configuration, but some supralaryngeal configurations are more favourable than others for constricting the epilarynx. Most favourable is when the larynx is raised and the tongue is retracted. The implication is that there are a range of possible vowel-related effects of the epilaryngeal stricture. There are somewhat subtle effects: relatively open vowels are more susceptible to epilaryngeal
striction; there are also extreme effects: strong tongue retraction, larynx raising, and epilaryngeal stricture acoustically “closes” the hypopharynx and reconfigures the resonating spaces above such that any vowel quality is still possible but will be produced with raised larynx voice quality (and be “condensed” in acoustic vowel space). Moreover, for reasons explained in §5.4.2, this extreme configuration has a palatal stricture bias, which makes pharyngeal-palatal phonological patterning possible (§7.3.1).

The dissertation is concluded in Chapter 8 with an outline of the theory of the epilarynx in speech and the pathways it opens up to future research.
Chapter 2

EPILARYNGEAL PRELIMINARIES

This chapter provides the preliminaries serving to ground the issues explored in the rest of the dissertation. Three topics are covered: §2.1 defines the epilarynx in anatomical and physiological terms; §2.2 provides a survey of the literature that pertains to epilaryngeal function in speech; finally, §2.3 reviews popular phonetic nomenclature used to describe sound production in the lower vocal tract. The chapter is summarized in §2.4.

2.1 Anatomo-physiological aspects of the epilarynx

Before proceeding into the theoretical aspects of epilarynx function in speech, it is essential to consider the anatomical and physiological nature of the epilarynx: the anatomy is addressed in §2.1.1, which provides an operational definition of the epilarynx that will be essential in the remainder of the dissertation; §2.1.2 examines the basic physiological nature of the epilarynx by reviewing its non-speech related functions.

2.1.1 Epilarynx anatomy: Defining the epilarynx

This section provides a rigorous definition of the epilarynx through anatomical and geometrical considerations (with some overlap into physiological considerations made in the following section). Although the terms epilarynx and epilarynx/epilaryngeal tube are used throughout the literature, to my knowledge, no substantive definition has been provided for what exactly constitutes the epilarynx. There is evidently a need for
such a definition, as judged from Miller’s comment that the “boundaries between the larynx, epiglottis, and pharynx are fairly ill-defined” (2012, p. 36). Several examples illustrating why clarification is necessary come from Borroff’s discussion of the articulation of glottal stop (2007, pp. 77–81), which prominently features interpretation of data discussed by Esling (2005 inter alia): she says “[Pharyngeal stop involves] general tightening of the epilarynx, as well as stop-like constriction at the glottis, the ventricular folds, and the aryepiglottic folds, in addition to tongue root retraction.” (Borroff, 2007, p. 79). In this example, Borroff confuses the epilarynx as anatomically differentiable from the ventricular folds and aryepiglottic folds: the very structures that the epilarynx comprises. Another example from Borroff (2007) illustrating the need for greater precision in the definition of the epilarynx is the following statement: “by constricting the epilaryngeal tube and closing the ventricular folds…” (p. 78). Here the problem is more subtle: constricting the epilarynx could be interpreted as potentially involving closure of the ventricular folds, i.e. the ventricular folds are a component of the epilarynx that may or may not be “closed” depending on the state of the vocal folds below. It thus leads to an important question: should we consider the ventricular folds to be part of the epilaryngeal tube? If not, then what defines the bottom of this tube? Another matter is the implication that Borroff (2007, p. 78) makes that ventricular closure is simply medialization and contact of the margins of the ventricular folds (i.e. “ventricular closure”)\(^8\), but not contact

---

\(^8\) This is evident in her interpretation of what the ventricular folds are doing in glottal stop. Her explanation of their function is in terms of the transglottal pressure drop: she suggests that the ventricular folds nullify the pressure differential and thus cause phonation to stop. Her account is problematic for three reasons. First, complete ventricular fold adduction will not have a significant effect on intraoral pressure, so it will not necessarily nullify the pressure difference between subglottal and supraglottal spaces. Second, ventricular fold action in glottal stop does not always involve complete medialization closure. Third, phonation is indeed possible even with very narrow epilaryngeal stricture. Borroff does not mention the possibility mechanical interaction between the vocal folds and ventricular folds.
between the superior surfaces of the vocal folds and the inferior surfaces of the ventricular folds, arguably more important than ventricular adduction.

The vocal folds (VF in Figure 2.2; \(v\) in Figure 2.3) and glottis (the space they enclose) are overwhelmingly the focus of attention in speech research\(^9\) because they are the primary source of the voice and laryngeal activity in normal speech. The plane of the vocal folds and glottis defines two important acoustic spaces relevant to speech: the subglottal and supraglottal vocal tracts. The subglottal cavity retains relatively fixed dimensions during speech, but the supraglottal cavity undergoes complex deformations that give rise to the patterns of airflow regulation and acoustic resonance that make up the speech signal. In the acoustic abstraction, we think of these important lumina as resonating tubes, and this is a convenient conceptual starting point for discussing the structure that is the principle subject of this dissertation: the epilarynx.

To start, the terms *epilarynx* and *epilaryngeal tube*\(^{10}\) critically refer to both the physical structures it comprises and the space it encloses. If we think of the larynx and trachea as one long tube, then epilarynx is the supraglottal, tube-shaped, upper extension of this tube roughly 2 cm in length (for males). In the context of the vocal tract, the epilarynx is a tube that opens into the larger pharyngeal tube (Sundberg, 1974, p. 839). Together the vocal folds, epilarynx, and pharynx define the lower or posterior region of the supraglottal vocal tract (or *lower vocal tract*). A picture of this idealization is presented in Figure 2.1: the epilarynx is the upper part of the tube found within the

---

\(^9\) This sentiment was originally expressed by Painter (1986, p. 329), but, in my opinion, it still holds weight in 2013.

\(^{10}\) I do not distinguish between *epilarynx*, *epilarynx tube* (as Ingo Titze is fond of saying; see 2008), and *epilaryngeal tube*. A tube is a (roughly) cylindrical structure enclosing a (roughly) cylindrical space, so it does not make sense to reserve *epilaryngeal tube* to refer just to the space enclosed by the structures of the epilarynx. These terms are therefore synonymous in my view. To refer strictly to the space, I will use the terms *epilaryngeal lumen*, *epilaryngeal cavity*, or *epilaryngeal space*. 
inferior region of the pharynx. Two important levels of the epilarynx are noted in this diagram: the aryepiglottic fold level and the ventricular fold level (indicated by the dashed ellipse). The tubes can constrict: thus, there is pharyngeal constriction (1) and epilaryngeal constriction (2).

Figure 2.1: The epilarynx as a tube-in-a-tube. The tube-shaped epilarynx is found at the bottom of the pharynx tube; together these structures define the lower vocal tract. The action of these tubes are pharyngeal constriction (1) and epilaryngeal constriction (2).

In anatomical contexts, the epilarynx is often referred to as the *laryngeal vestibule* (Fink, 1975, p. 36; Painter, 1986, 1991; Zemlin, 1998, p. 117; Titze, 2008, p. 2734; also see Esling, Fraser, & Harris, 2005, pp. 386–387), but other names have been applied: for example, Sundberg (1974), Gauffin (1977) and Nolan (1983, p. 182) use the phrase *larynx tube*; Honda et al. (2010, p. 443) use *(supraglottic) laryngeal cavity*; Edmondson, Padayodi, Hassan, & Esling (2007, p. 2066) and Esling, Zeroual, & Crevier-Buchman...
(2007, p. 586) occasionally refer to the epilarynx as the *supraglottic tube*, despite co-occuring use of *epilaryngeal tube*. *Epilarynx* is employed in the present work because it is more precise than *laryngeal vestibule, larynx tube*, or *supraglottic tube*. *Laryngeal vestibule* does not provide a strong connotation of location, whereas *epilarynx* does by virtue of the *epi* - prefix (meaning ‘above’, ‘on’, ‘over’, ‘outer’, and so forth); furthermore, for some (e.g. Zemlin, 1998, p. 117) *vestibule* excludes the laryngeal ventricle (a cavity above the vocal folds), which, as will be discussed below, is too restrictive for our present purposes; others include the ventricle in the definition of the vestibule (Esling, Fraser, & Harris, 2005, pp. 386–387). *Larynx tube* is too vague: as has already been suggested, the entirety of the larynx encloses a tube-shaped cavity; the epilarynx is the upper cavity. Gauffin’s usage of *larynx tube* is admittedly more broad in that it also deliberately includes the vocal folds, but I exclude them from the definition of *epilarynx*. The term *supraglottic tube* is also too vague and runs the risk of being mistaken for the vocal tract proper. Further motivation to use *epilarynx* comes from the fact that it is increasingly being used in voice literature (Titze, 2008), and its use in the present work is done in part to align with this literature.

As the name implies, the epilarynx is part of the larynx, and the larynx is a complex framework of cartilages, ligaments, muscles, and folds of epithelial tissue that is situated at the top of the trachea (see Figure 2.2). Critical epilaryngeal structures include the ventricular folds and aryepiglottic folds, the laryngeal ventricle and vestibule, and the epiglottis and cuneiform cartilages, and the arytenoid cartilages behind and below with their corniculate cartilage extensions. The quadrangular membrane (not depicted) forms
the body of the aryepiglottic fold and is continuous with the ventricular fold (meaning there is no strict separation between the ventricular and aryepiglottic folds).

![Diagram of the larynx](image)

Figure 2.2: Anatomical sketch of the larynx. Sagittal section (a); posterior view (b). A = arytenoid cartilage; C = cricoid cartilage; T = thyroid cartilage; VF = vocal fold. Illustrations of laryngeal anatomy important to the epilaryngeal tube. Diagrams based on anatomical photos of the larynx found in Zemlin (2010).

In considering the anatomical geometry of the epilarynx, we must consider two spaces: the laryngeal vestibule and the laryngeal ventricle (Fink, 1975, p. 36; Palmer, 1993, p. 109; Zemlin, 1998, p. 117). Zemlin closely follows the structural definition of the vestibule laid out by Gray (2003, p. 644): the epiglottis serves as the anterior border, the aryepiglottic folds serve as the lateral borders, and the apices of the arytenoids and corniculate cartilages form the posterior border. The ventricle, on the other hand, is the cavity immediately above the vocal folds and bounded by the caudal surface of the ventricular folds. The vestibule communicates with this space superiorly, and the spaces
can be (arbitrarily) separated by reference to the plane defined by the margin of the ventricular folds. Such a bipartite division of laryngeal space is useful insofar as it correctly conveys that these two spaces become separated by the adduction of the ventricular folds. In aero-acoustic literature, however, the spaces are lumped together as composing the epilaryngeal cavity: for example, both Sundberg (1974) and Titze (2008) regard the base of the epilarynx to be defined by the cephalad surface of the vocal folds. Furthermore, Sundberg (1974, p. 840) and Honda et al. (2010) consider the epilarynx a “twin-tube resonator”: it is the combination of the ventricular and vestibular cavities, the latter stacked on the former. Accordingly, Honda et al. (2010) state that the epilarynx approximates the shape of a Helmholtz resonator: the body being the ventricle and the neck being the vestibule. The aero-acoustic conception of the epilarynx as comprising these two subsections (rather than excluding the ventricle) is useful because one of the key effects of epilarynx activity is to obliterate the ventricle cavity, which has important acoustic (Pepinsky, 1942; van den Berg, 1955; Heselwood, 2007) and mechanical consequences for laryngeal function.

At the upper boundary of the laryngeal vestibule is the *laryngeal aditus* (inlet or aperture). A strict anatomical definition of the upper boundary of the epilarynx can be borrowed from the definition of the aditus which is demarcated by the uppermost extent of the structures of the vestibule (i.e. the epiglottis and aryepiglottic folds) (Zemlin, 1998, p. 228). The fact that the blade of the epiglottis (the suprahyoid part of the epiglottis) extends well above the level of the arytenoid apices means that the roughly elliptical aditus has a nearly 90° curve around its short (lateromedial) axis approximately at half the height of the epiglottis. The functional plane of upper epilarynx, however, below
which the epilarynx is basically a complete tube structure, can be considered to intersect with the location where the aryepiglottic folds (and the embedded cuneiform cartilages) make contact with the tubercle of the epiglottis (see Figure 2.3). The remainder of the epiglottis rising above this plane does not form a complete tube shape with other laryngeal structures, even though the epiglottis can exhibit strong curvature around its longitudinal axis in some individuals. In fact, this upper projection of the epiglottis can come into contact with the posterior pharyngeal wall to form an additional tube-shaped epiglotto-pharyngeal space, which is continuous with and effectively extends the epilarynx.

Although the details of physiology will be discussed in the next section, it will be helpful at this point to consider the laryngoscopic view of the larynx in two of its postures, which correspond to two extreme epilarynx states. These are contrasted in Figure 2.3: image (a) shows the larynx in its fully open state, which is associated with inhalation. The vocal and ventricular folds are widely abducted and there is a large posteroanterior opening of the epilarynx (dashed white line). Images (b) and (c), on the other hand, show the larynx progressively more constricted states ((c) is associated with the preparatory state immediately before a voiced aryepiglottic trill [s]. The vocal folds and right ventricular fold are not visible in this image because there is nearly full contact between the aryepiglottic folds and the epiglottis. In this constricted state, the epilaryngeal tube essentially has become bifurcated into two apertures: one associated with the right aryepiglottic fold and one with the left\(^\text{11}\). Regardless of epilaryngeal state

\(^{11}\) Stuart (1982) describes the epilaryngeal tube in this configuration as a “triradiate fissure in the form of a squat ‘T’”. Fink (1975: 86) also chooses to describe the epilaryngeal tube opening as “T” shaped during this configuration. In the laryngoscopic view of the larynx, where the “T” shape appears upside down (in
in the image series, the pharynx remains relatively unconstricted by comparison. The piriform fossae\textsuperscript{12} (pf) are depressions on either side of the larynx and they are formed in part by the aryepiglottic folds. These spaces extend the pharyngeal tube well below the level of the upper epilaryngeal border. Thus, the pharynx does not simply blend into the epilarynx, but rather continues further down and terminates at the bottom of the piriform fossae (hence the notion that the epilarynx is a tube within a tube). The epilarynx is within the laryngopharynx, but highly independent from it.

\textsuperscript{12}Latin for ‘pear-shaped ditches’.
Figure 2.3: Laryngoscopic views of the epilaryngeal tube in an unconstricted state (a), in a partially constricted state (b), and in an even more constricted state prior to voiced aryepiglottic trilling [ʢ] (c)$^{13}$. White, double-headed arrow = posteroanterior dimension of the epilarynx; dotted outline = epilaryngeal aperture; ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; i = interarytenoidal gap; k = arytenoid apex/corniculate tubercle; m = inner mucosal surface of epilarynx; pf = piriform fossa; ppw = posterior pharyngeal wall; v = vocal fold. (n.b. images (a) and (b) are frames from a different video than (c) so the laryngeal heights of these states cannot be directly compared.)

It has been claimed here that the epilarynx can be abstracted as a tube, but further comment should be made on this concept regarding the geometry. The idea is that the epiglottis constitutes its anterior half, and, at a first approximation, the aryepiglottic folds and the cuneiform tubercles define its posterior half. Topologically, however, the

---

$^{13}$ The high-speed laryngoscopic videos were obtained with the help of Dr. Lise Crevier-Buchman (and her research team) at Sorbonne-Nouvelle Paris III/CNRS-LPP-UMR 7018 research site located in l’Hôpital Européen Georges Pompidou, Paris. More details can be found in Moisik, Esling, & Crevier-Buchman (2010).
epilarynx is not strictly a hollow cylinder as the term *tube-shaped* implies; rather, it is a cylinder with a split running part way down the posterior border, the two free edges of which separate at the midline during vocal fold abduction (as in Figure 2.3a). One could therefore specify that the posterior border epilarynx is actually comprised of the upper part of the arytenoids, their corniculate cartilage extensions, and the interarytenoidal mucosa bridging these structures. Both structures are important in understanding epilarynx function. The cuneiform tubercles are an essential site of contact during epilarynx stricture (as in Figure 2.3b), but the corniculates and arytenoids can participate in epilaryngeal vibration, so it is important to incorporate them into the definition of the epilarynx. The approach here is to define the *arytenoidal cartilage complex*, the set of cartilages comprising the cuneiforms, upper arytenoids, and corniculates and their respective mucosal sheathings. This complex and the aryepiglottic folds which project from it serve as the structural basis for the posterior border of the epilarynx.

As discussed above, since the epilarynx comprises the ventricle, it is reasonable to define the lower boundary of the epilarynx as being the upper surface area of the vocal folds; therefore, the ventricle is included as part of the epilarynx. The ventricular folds are the lower most mobile part of the epilarynx and together with the ventricular space they define the *lower epilarynx*. The upper boundary of the epilarynx is delineated by a quasi-ellipse formed by tracing a path from the apex of the epiglottis down either side of its curved blade to its lateral margins and then proceeding along the crest of the aryepiglottic folds to the cuneiform tubercles: this region marks the *upper epilarynx*. There is no strict division between *lower epilarynx* and *upper epilarynx* (they are continuous with each other) but these terms will be used in this dissertation when
describing epilarynx function. In a neutral laryngeal posture, the plane of the lower boundary is orthogonal to the longitudinal axis of the human body, but the plane of the upper boundary is at an oblique angle roughly defined by the craniolateral orientation of the aryepiglottic folds.

So far the definition of the epilarynx has been in reference to its anatomical structures and abstract geometry, but it is also important to consider its larger anatomical context as well (as was illustrated in Figure 2.1). Critically, the epilarynx is distinct from the laryngopharynx, which is defined as the lower pharynx and is superiorly bounded by the hyoid bone and extends down to the level of the sixth cervical vertebra (Zemlin, 1998, p. 274). Major anatomical features of the laryngopharynx include the piriform fossae\(^\text{14}\), which flank the epilarynx and are partly formed by its aryepiglottic folds, and the upper esophageal sphincter, which is found posterior to the larynx. Painter (1986, p. 331) states that the volume of the piriform fossae cannot be actively enlarged, but action of the inferior pharyngeal constrictor muscles, posteroanterior expansion of the epilarynx, or raising the larynx can actively reduce their volume (for more discussion of the piriform fossae, see Dang & Honda, 1997; Honda, Kitamura, Takemoto, et al., 2010).

To summarize, the epilarynx is defined as the set of structures comprising the ventricular folds, aryepiglottic folds, epiglottis, and arytenoid cartilage complex. It encloses two spaces, the laryngeal ventricle and the laryngeal vestibule. Thus the epilarynx encloses a tube-shaped space (hereafter *epilaryngeal lumen, cavity or space*)

---

\(^{14}\) The following correlations hold between the larynx tube and the piriform fossae for various hypopharyngeal configurations (Honda, Kitamura, Takemoto, et al., 2010; also see Painter, 1986, p. 331): during modal phonation both the piriform fossae and epilaryngeal space are open; during inhalation, the entire larynx tube expands and unfolds while the piriform fossae are compressed; finally, during whisper the epilarynx narrows and the piriform fossae reduce in size.
beginning immediately above the level of the vocal folds and terminating at the laryngeal aditus. The epilaryngeal tube is nested within and independent from the enclosing pharyngeal tube. The lower epilarynx is defined as the ventricular region, comprising the ventricle cavity and the ventricular folds above; the upper larynx is the aryepiglottic region and vestibular cavity. These regions are anatomically continuous with each other, but it will be useful to proceed with this division as the theory of the epilarynx in speech is developed.

2.1.2 Epilarynx physiology

This section surveys the fundamental biological functions of the epilarynx. Although, as Traill (1986, p. 130) observes, speech functions do not necessarily overlap with these basic functions. Nevertheless, it is instructive to review them in order to compare them to the functions that do occur in speech. Figure 2.4 depicts the functional planes of the larynx; the concept of laryngeal planes is a useful abstraction to understand the epilarynx and is employed in the literature (e.g. Edmondson & Esling [2006, p. 162]). There are three planes (or levels) of activity depicted in Figure 2.4: the glottic, ventricular, and aryepiglottic planes. As depicted in the figure then, the epilarynx comprises the ventricle and ventricular plane and the vestibule and the aryepiglottic plane above. Critically, the epilarynx starts immediately above the glottic plane: it does not include it. Axes of activity in each plane are depicted with opposed arrows. The vocal and ventricular folds primarily act along the lateromedial dimension. However, this does not mean that their action is strictly medialization: one of the key contentions of this work is that the vocal folds and ventricular folds can in fact come into contact, effectively
obliterating the ventricle, and this is indicated by the opposed arrows in between these planes. Action in the aryepiglottic plane operates along a posteroanterior dimension; this plane is depicted obliquely to convey the fact that the posterior border of the epilarynx must rise to meet the lowering anterior border. Figure 2.4 depicts the longitudinal axis of the epilarynx as vertical in orientation, but in reality, this axis is generally not vertical, and in fact changes orientation during speaking. Generally, it is at an incline that approximates the longitudinal axis of the epiglottis (similar to what is represented in Figure 2.1).

Figure 2.4: Functional planes of the larynx. Arrows indicate the axes of activity in each plane, the opposed vertical arrows between the ventricular and glottic plane indicate that these planes make contact. The dashed lines inscribed on the ellipses indicate lines of stricture. Dimensions: x-axis is lateromedial; y-axis is inferosuperior; z-axis is posteroanterior.
The morphology and action of the human epilarynx reflect the evolutionary origin of the larynx itself as a mechanism for fixation of the thoracic volume\(^{15}\), protective closure of the airway, and expulsion of foreign particles trapped in the airway. The glottis alone does not fulfill these functions: the epilarynx is required (Negus, 1949). The fixation mechanism likely arose in arboreal primates which use brachiation as a primary means of locomotion. Hermetic larynx closure prevents air escape and therefore stabilizes the thoracic volume to support high loads placed on the pectoral girdle by the muscles of the upper limbs (Hayama, Honda, Oka, & Okada, 2002). The low position of the larynx relative to the velo-pharyngeal port in humans makes us vulnerable to potentially fatal aspiration of food and other substances, particularly during swallowing; the epilarynx is part of the closure mechanism that protects us from this liability. Finally, epilarynx closure during cough and throat clearing allows for the build-up of sufficiently high intrathoracic pressure; upon abrupt opening the resultant high air velocity efficiently expels any intrusive substances within the airway or epilarynx space (Hillel, 2001).

As suggested in previous research, the larynx itself is a multi-valve structure (Negus, 1929, 1949, pp. 100–112; Fink, 1975, p. 12; Edmondson & Esling, 2006): the vocal folds, ventricular folds, aryepiglottic folds, and epiglottis all participate in the valve-like closure of the larynx. Thus, these valves all serve to regulate airflow, but it is the supraglottal laryngeal valves associated with the epilarynx which play a role in closure for air trapping to support high intrathoracic pressure. Often the larynx is

\(^{15}\) Note that volume fixation is tantamount to a fixation of the mass of air in the lungs. Laryngeal airway closure prevents the air mass from flowing, thereby changing, and thus inhibits changes to the thoracic volume. The focus here is on volume fixation because many of the associated functions depend on the thoracic musculature (such as those connected with the forelimbs) having a fixed anchor point when contracting. However, in the case of building up intra-thoracic pressure for expulsion, the key parameter is air mass fixation so that thoracic volume reduction results in increased pressure.
analogized to a sphincter (Pressman, 1954; Traill, 1986), but Fink (1974a) argues that this view is both inappropriate and oversimplifying. The argument is that, while the larynx may have its phylogenetic origin as a circular sphincter of the trachea (citing Negus, 1929), its form and mechanism in modern humans is better characterized by the term *plication* – folding and buckling of the laryngeal tissues, especially the mucosal covering. Furthermore, anatomical sphincters of the human body\(^\text{16}\) have annular morphology and maintain closure: they only open momentarily to allow passage of matter. On the other hand, the larynx is morphologically complex and maintains patency of the airway, only closing momentarily to prevent the passage of matter (Fink, 1974a, pp. 124–125). To this end, Fink (1974b) also argues that the elastic structures of the larynx (especially the pre-epiglottic body, hyoepiglottic ligaments, epiglottis, cuneiforms, corniculates, conus elasticus, and quadrangular membrane) store potential energy during closure to assist in re-opening.

Working with this plication model, Fink (1974a) outlines four principle laryngeal functions that can be interpreted as forming a continuum of laryngeal (and epilaryngeal) folding depicted in Figure 2.5: respiration (a & b), phonation (c), effort closure (d), and swallowing (e).

\(^\text{16}\) e.g. those of the gastrointestinal tract (esophagus, stomach, bladder, bile duct, anus and so forth).
Figure 2.5: Fink’s continuum of laryngeal folding (lateral view). Traces based on lateral x-ray sketches found in Fink (1974a, 1975). States illustrated: inspiration (a); expiration (b); modal phonation at 318 Hz (c); effort closure (d); swallowing (e). “4” marks the fourth cervical vertebra; Dark gray region = laryngeal ventricle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; v = vocal fold. Note: hyoid = hyoid bone; thyroid = thyroid cartilage; cricoid = cricoid cartilage.

The first two functions, respiration (a & b) and phonation (c), lie on the unconstricted end of the spectrum: the epilaryngeal lumen is expanded to prevent obstruction of airflow or interference to vocal fold vibration. The second two functions, effort closure (d) and swallowing (e), fall on the constricted (narrowed) side of the spectrum. Illustrations in Figure 2.6 show coronal plane images of laryngeal and epilaryngeal folding in respiratory states (Figure 2.6a & b; c.f. Figure 2.5a & b) and effort closure (Figure 2.6c, c.f. Figure 2.5d).
In respiration, particularly during inspiration, the larynx, which obstructs airflow into the lungs must be expanded to reduce its airway resistance (Negus, 1949, p. 63). In the inspiratory phase, then, this expansion is accomplished by lowering the larynx and trachea: this causes vertical stretching of all the soft tissues within the larynx, and, importantly, elongates and thins the structures of the epilarynx (and pre-epiglottic body) (Fink, 1974a, p. 78; Sundberg, 1974; Painter, 1986, p. 329). The result is anti-plication:
the larynx tube is smoothed, the ventricle is widely opened and may effectively merge with the rest of the tube, and there is passive abducting displacement of the arytenoids, which follow the lateral and downwards motion of the connecting tissues. In expiration, elastic recoil returns the larynx to a moderately folded state (the vocal folds and ventricular folds medialize), increasing airway impedance and slowing the rate of flow out of the lungs with the effect of delaying the inspiratory reflex\textsuperscript{17} and allowing more time for gas exchange (Negus, 1949, pp. 63–64).

In phonation (understood here to mean non-pathological, neutral or modal vibration of the vocal folds), the primary folding is of the vocal folds, although the ventricular folds passively and moderately bulge towards the midline (relative to inspiration) following arytenoid medialization; lifting of the ventricular folds caused by vertical tension in the aryepiglottic folds helps to enlarge the ventricle, which is essential to allow for free and large amplitude displacement of the vocal folds when they vibrate (Fink, 1974a, pp. 125–126).

In effort closure (coughing, straining, and so forth) the bulging of the folds becomes substantial to the point that the entire laryngeal tube is narrowed and eventually closed off entirely, creating a hermetic seal (Fink, 1974a, pp. 126–127; Painter, 1986, p. 330). Critically, and as evident in tracing (c) of Figure 2.6, in addition to strong adduction of the vocal folds, the ventricular folds adduct and descend upon the vocal folds, eliminating the ventricular space and resulting in contact of the cephalad surface of the vocal folds and caudal surface of the ventricular folds (Fink, 1974a, p. 127). Fink claims that larynx-hyoid approximation (driven by the thyrohyoid muscles) is essential in aiding

\textsuperscript{17} Without valvular control (as in post-tracheostomy patients) the rate of respiration increases (Negus, 1949, p. 64).
the reduction of the vertical dimension of the larynx: it serves to bring the vocal folds and ventricular folds close together, causes backwards bulging of the epiglottis (citing Fink, 1956) and also assists in passive bulging of the aryepiglottic folds, which make contact with the epiglottis (especially in the vicinity of the [cuneiform and epiglottic] tubercles of these structures).

At the extreme end of the folding continuum is swallowing (Fink, 1974a, p. 127; also see Painter, 1986, pp. 329–330). Essentially, this configuration adds additional epiglottal folding to the three-layered seal characterizing effort closure (vocal, ventricular, and aryepiglottic folding). Fink suggests there are two types that vary by larynx height: in lowered swallow (used for continuous drinking of a liquid when it is not possible to digest all of the fluid at once, which means that some fluid must reside in the piriform fossae between swallow events), the epiglottis remains upright, but its infrahyoid extent lowers into the ventricular folds, an effect caused by compression and bulging of the pre-epiglottic body in front of the epiglottis driven by thyrohyoid approximation (also see Negus, 1949, p. 92; Reidenbach, 1997). This effect similarly occurs during effort closure. The other type of swallow involves larynx raising and advancement of the hyoid bone (although there is still approximation of the thyroid cartilage to the hyoid bone). In this case, on account of extreme larynx raising, the epiglottis tends to fold over the laryngeal aditus, effectively adding a fourth layer of closure to the entire laryngeal system.

An important consideration in this continuum is the role of larynx height in facilitating the folding or unfolding of the larynx. In general, lowering unfolds the larynx while raising causes folding, but thyrohyoid approximation adds a second layer of raising
potential to the overall mechanism, meaning that the larynx can be globally lowered (by the infrahyoid strap muscles) and still exhibit larynx-hyoid approximation, thanks to the thyrohyoid muscles (also see Shin, Hirano, Maeyama, Nozoe, & Ohkubo, 1981, p. 173). This can be seen in Figure 2.5 by comparing (d) (effort closure/strain) and (e) (raised swallowing). In this example of effort closure, the larynx is relatively low, but there is increased thyrohyoid approximation; the swallowing case shows similar thyrohyoid approximation, but now the larynx is globally raised. The significance here is that epilarynx stricture is possible regardless of global larynx height, since thyrohyoid approximation is sufficient to produce folding of the laryngeal structures. It is likely the case, however, that global larynx raising enhances the vertical compaction of the laryngeal structures more so than larynx lowering, which might elongate them vertically. Another noteworthy feature of the constricted (Figure 2.5d & e) states is that the gap between the cricoid arch and the thyroid cartilage increases (stretching the median cricothyroid ligament); this suggests that the cricoid cartilage is rotated such that the upper border of the lamina (which carries the arytenoids) is advanced towards the retracted epiglottis, an effect which also possibly facilitates closure.

Although it is tempting to think of this four-way closure mechanism as a special, super-redundant adaptation in humans serving to protect the larynx during swallowing on account of its descended position\(^\text{18}\), Negus (1949, pp. 98–99) claims that a simple

\(^{18}\) An intriguing evolutionary side note about the descended larynx in humans comes from Negus (1949, pp. 25–28). Unlike our simian cousins, our muzzle is comparatively shortened, but the size of the tongue has not changed. Our large tongue has, consequently, been forced partly into the pharynx (the tongue is oral in gorillas, orangutans, and so forth, being accommodated by the large oral cavity) and taken on a curved shape; thus humans have an oral-pharyngeal tongue. Negus claims that we retain a large tongue for the purposes of mastication (not for manipulating oral-pharyngeal resonance). Recession of the snout occurred with a transition from heavy reliance on olfaction (which benefits from an elongated nasal passage to extend the duration of time air passes over the olfactory organs and from obligatory nasal respiration) to
muscular sphincter is all that is really required to protect the airway. This is evident for species, such as the lung fish, which when under water must prevent inundation of the airway. Thus, some comment is warranted on other reasons why the larynx is elaborated as a set of complex folding structures analogous to valves. Understanding the functional motivation for elaboration of the larynx as a complex set of valves or folds in humans is instructive in understanding the nature of the epilarynx structures in speech.

The key property of our larynx that makes the epilarynx necessary is the fact that we have upturned vocal folds (Pressman, 1954, p. 222; Fletcher, 1993, p. 2173), as illustrated in Figure 2.7a. Airflow directed towards the lungs causes an overpressure above the adducted vocal folds, but since they are upturned, the force exerted by this pressure pushes the vocal folds together (a1): thus, (even weakly) adducted vocal folds are good at preventing air from entering the lungs. If the overpressure comes from below, however, then they will be inclined to open (a2). The opposite relations would hold if the vocal folds were downturned (b).

John Ohala (personal communication) doubts Negus’s explanation on the grounds that we should expect infants to have snouts at the stage of development when the larynx is ontogenetically high. This view assumes that ontogeny necessarily recapitulates phylogeny. There could be functional reasons that account for the initially high larynx position in infants (such as, possibly, enhancing cry).
Figure 2.7: Upturned/inlet (a) vs. downturned/outlet (b) valves. $u =$ airflow; $p =$ pressure acting on the valve (Pressman, 1954, pp. 222–224).

Since sphincters do not provide the same aero-mechanical advantages for maintaining closure, the inlet-valve-like (see Figure 2.7a) thyroarytenoid folds (i.e. vocal folds) are an adaptation present in all animals making use of forelimbs in grasping, climbing, or clinging (Negus, 1949, pp. 102–108). In all cases, the mechanical purpose is to fix the trunk for maximal energy efficiency in the action of the muscles controlling the forelimbs, such as the pectorals (lest energy be lost to expansion of the thorax in their contraction). A relevant example is brachiation (using the arms for locomotion while suspended, as found in arboreal species of primates such as lemurs, spider monkeys, and gibbons): strain on the thoracic cage by sudden suspension during brachiation causes expansion of the ribs. Through plural linkage, this would cause the lungs to expand and thereby take in air. However, the inlet valve (i.e. the vocal folds) prevents air from entering the lungs. Furthermore, the vocal-fold inlet-valve mechanism acts in concert
with abdominal muscle activity to prevent rib expansion. With the ribs fixed, pectoral muscle contraction will exert pull on the humerus, making this type of locomotion more energy efficient, especially since vocal fold adduction is an energetically inexpensive manoeuvre (Negus, 1949, p. 106).

The asymmetry in the efficiency in maintaining airway closure of the two different types of valve mechanisms – inlet and outlet (Figure 2.7) – is demonstrated for cat vocal folds. Cat\textsuperscript{19} vocal folds are upturned (i.e. they form an inlet valve, see Figure 2.7a) and can support 25 times more overpressure from above than from below the folds (Negus, 1949, p. 110). In species with ventricular folds, such as humans, the amount of overpressure supportable from below is much higher (Negus, 1949, p. 61).

Although human infants\textsuperscript{20} have a remarkable ability to support their own weight by clinging, humans have little need for arboreal modes of transit. Nevertheless, there is still a need for supporting high intrathoracic pressure during effort-closure-related functions such as exertion (to compress the viscera, as in parturition) or throat clearing and coughing, which involve the sudden release of this pressure (Hillel, 2001). This is the function Negus identifies for the downturned, adductable ventricular folds in the human larynx (1949, p. 111). Based on the argumentation just provided (using cat vocal folds), that the upturned, inlet-valve type vocal folds do not efficiently entrap air within the lungs, it is reasonable posit this air-entrapment function (to maintain high intrathoracic

\textsuperscript{19} The feline larynx lacks ventricular folds.
\textsuperscript{20} There are significant anatomical and physiological differences between the vocal tracts of human infants and those of human adults. The following discussion is focused on adult physiology (and the physiology of other animals, where relevant).
pressure) as one of the key biological functions of the ventricular folds\textsuperscript{21}. However, the fact that they cannot adduct independently of the vocal folds means that they do not act alone to perform this function and that the laryngeal ventricle does not inflate to support positive pressure to push the ventricular folds together and prevent the escape of air from the lungs. The other possibility is for the ventricular folds to counteract the upwards and outwards force of an overpressure below the vocal folds by coming into physical contact with the vocal folds. The addition of upper epilaryngeal stricture (posteroanterior narrowing at the aryepiglottic plane and epiglottal descent) and general vertical compaction likely support the ventricular folds in this function.

All the above discussion is not to suggest that the epilarynx plays no protective function; on the contrary, the aryepiglottic folds and epiglottis are essential in reducing the likelihood that foreign matter will enter the laryngeal airway. Negus suggests that aryepiglottic folds may even be more efficient in this role than lateral epiglottic folds (which are not found in humans). Furthermore, the aryepiglottic folds (which could be whimsically thought of as the shower curtains of the larynx; Mark Tiede, personal communication) provide a steep wall to allow for moderate pooling of fluid within the piriform fossae during a continuous swallowing manoeuver (Negus, 1949, p. 92; Fink, 1974a, p. 172). Furthermore, Negus asserts that the human epiglottis is degenerate from an evolutionary perspective as its main function is in smell not swallowing\textsuperscript{22}. However,

\textsuperscript{21} Note that the ventricle secretes mucous which functions as a lubricant for the larynx and vocal folds in particular. This fact will become important in §4.3.

\textsuperscript{22} The epiglottis is often in contact with the soft-palate or even intranarial (obligating nasal breathing) in terrestrial mammals that critically depend on olfaction, especially while eating or hunting prey (which requires mouth opening). Even more extreme coupling between the epiglottis and the nasal cavity is found in many marine mammals (such as dolphins and whales). Herbivorous animals have lateral food channels shaped by lateral epiglottic folds to help convey semi-liquid matter safely past the larynx over long duration; carnivorous animals which swallow large, unmasticated boluses whole make use of cricothyroid
the human epiglottis still sees secondary use for swallowing\textsuperscript{23} even in animals where it fulfills the function of enforcing nasal breathing for olfactory purposes. One might say that in humans this secondary swallowing function has now become primary or at least very important.

Our final physiological consideration concerns the musculature responsible for manipulating the configuration of the epilarynx\textsuperscript{24}. Recall that there are two planes of epilaryngeal activity to consider: the ventricular and the aryepiglottic planes (see Figure 2.4). With regard to the ventricular plane, there are two key motions which reduce the ventricular space: lateromedial adduction and inferosuperior approximation to the vocal folds\textsuperscript{25}. Figure 2.8 depicts two views of some of the musculature found within the larynx associated with epilaryngeal stricture (and certain other muscles to provide context, such as the cricothyroid muscles). Interdigitations make strict identification of muscle fiber association difficult, so these diagrams depict connections that are identified in the literature as relevant to epilarynx narrowing and ventricular adduction and lowering. What is referred to by one name by an anatomist may correspond to another closely related muscle group: for example, such may be the case with the ventricularis, which may be a projection of the external thyroarytenoid and possibly thyroepiglottic muscles. With these considerations, it is possible to describe two layers of musculature (depicted

\textsuperscript{23}John Ohala (personal communication) points out that surgical removal of the epiglottis has more severe consequences for swallowing than it does for speech.

\textsuperscript{24}For a more detailed survey of the overall laryngeal musculature, the reader is directed to Zemlin (1998, pp. 121–137) and, for a more recent literature survey see Moisik (2008).

\textsuperscript{25}As such, although the term \textit{ventricular adduction} is most prevalent in the literature, it is useful to apply a label \textit{ventricular incursion} (defined in Edmondson & Esling, 2006, p. 159) to emphasize the complex motion of the ventricular folds and their potential for direct contact with the vocal folds.
in Figure 2.8a), one primarily associated with the ventricular plane (dashed lines capped by black circles; 1, 2, 3) and one primarily associated with the aryepiglottic plane (solid lines capped by black circles; 5, 6, 7). Both act on the epiglottis, but the first is more inclined to lower the root of the epiglottis at the anterior extent of the epilarynx, likely impacting and displacing the adipose and glandular tissue in and around the ventricular folds and peri-epiglottic space (M. M. Reidenbach, 1998a, p. 412, 1998b; Reidenbach, 1997), resulting in medial bulging and lowering of the ventricular folds (particularly at their anterior extent), as illustrated in Figure 2.8b\(^26\). The second contributes to the downward motion, but it is also in a position to retract the epiglottis and to cause the tubercle to approach the rising and advancing cuneiforms. Both sets also show connection to the interarytenoid muscles. The ventricular plane group has fibers that wrap around the arytenoids and possibly interdigitate or continue as the transverse fibers of the interarytenoid muscles. In the aryepiglottic plane group, the aryepiglottic muscle extends the oblique division of the interarytenoid muscles, and there are possible interdigitations with the lateral cricoarytenoid muscles. Both groups are in a position to draw the arytenoids forward, and possibly even cause torque on the cricothyroid joint that will bring the upper border of the cricoid lamina and the arytenoid cartilage complex along with it towards the descending and retracting epiglottis.

\(^{26}\) Tongue retraction may help displace this adipose tissue (forming the pre-epiglottic body and continuous with the tissue in the anterior epilarynx) downwards, aiding in compression in the anterior epilarynx. In MRI images, distortion of the pre-epiglottic body is apparent during low vowels (such as [a]).
Figure 2.8: Intralaryngeal musculature associated with the epilarynx. (a) lateral view; (b) coronal section corresponding to long dashed line in (a) and based on images in Reidenbach (1998a). f: ventricular (false) fold; v: vocal fold. Dashed lines with black circles (1, 2, 3): muscle-fiber paths for muscles primarily acting on the ventricular plane; Solid lines with black circles (5, 6, 7): muscle-fiber paths for muscles primarily acting on the aryepiglottic plane.

The exact mechanism of ventricular fold adduction remains controversial, but there is little doubt that it occurs and that it is highly dependent on the vocal folds being adducted. It is also clearly not the case that ventricular fold adduction mandatorily occurs when the vocal folds adduct: ventricular fold adduction follows vocal fold adduction, suggesting recruitment of additional muscle activity and other passive sources of medialization arising from compaction of the epilarynx in general. The cricomucosal and cricoepiglottic muscles (extensions of the lateral cricoarytenoid muscle into the
quadrangular membrane and epiglottis, respectively; see Thumfart, Platzer, Gunkel, Mauer, & Brenner, 1999, pp. 40–41) and thyroepiglottic muscles are suggested as agonists of ventricular adduction. These muscles cause a medializing force on the ventricular folds when they contract and narrow the epilaryngeal space. This narrowing is achieved by drawing the base of epiglottis and quadrangular membranes downwards and by raising the cricoid upwards. The effect is to bring the arytenoid cartilage complex closer to the epiglottic tubercle (Kimura, Sakakibara, Imagawa, et al., 2002; Imagawa, Sakakibara, Tayama, & Niimi, 2003, p. 2). The connection of the cricomucosal and cricoepiglottal fibers to the lateral cricoarytenoids helps to explain the connection between vocal fold adduction and ventricular adduction.

Following arguments by Fink and Demarest (1978), Painter (1986, pp. 330–331) proposes that the external thyroarytenoid muscles, which are situated lateral to the ventricular folds and project upward into the body of the aryepiglottic folds, towards the epiglottis, and medially in the form of the ventricularis muscle, are key drivers of ventricular fold adduction. Reidenbach (1998a) discusses the so-called ventricularis muscle specifically and suggests it helps to cause both medialization and descent of the ventricular fold. Esling et al. (2007, p. 585) draw similar conclusions, but they suggest that the thyroarytenoid branches are key, especially to connect higher level stricture of the epilarynx with basic ventricular adduction observed in certain phonation types exhibiting epilaryngeal stricture (such as creakiness); they also cite Réthi (1966), suggesting that the muscle system formed by the stylopharyngeus, aryepiglottic, and oblique interarytenoid muscles correlates with ventricular medialization, suggesting the

27 This muscle is formed from branches of the external thyroarytenoid muscle (M. M. Reidenbach, 1998a).
larynx height and epilaryngeal narrowing systems are intimately connected. Early electromyographic data of glottal closure (in glottal stop) supports the interpretation that the external thyroarytenoid acts in complement to the vocal fold adductors (Hirano & Ohala, 1969).

Already it is apparent that ventricular control is not independent of either the vocal fold system or the general mechanism for complete epilaryngeal stricture. It is convenient to see it as an intermediary stage between full vocal fold adduction and total closure of the epilarynx, dependent on the posteroanterior narrowing mechanism of the epilarynx and therefore functionally a part of the epilarynx system with overlapping relation to vocal fold adduction. This view is supported by Painter (1991, p. 456), who observes that epilarynx stricture occurs in two stages: the first involving moderate fronting and medialization of the cuneiform tubercles with concomitant posteroanterior narrowing of the vestibular space and, provided that the vocal folds are adducted, ventricular fold adduction; the second step involves considerably more posteroanterior narrowing of the vestibular space such that the cuneiform tubercles now are very near or in contact with the retracted epiglottic tubercle, resulting in a 90° bend in the aryepiglottic fold (with the angle centered at the cuneiform tubercle) giving the impression of T-shaped or closure of the inlet to the larynx (Negus, 1949, p. 86).

Musculature of the final stage is similar to the mechanisms employed for the ventricular level, but presumably involves increased activity. Some support comes from Hillel’s (2001; see also Faaborg-Andersen, 1957) EMG data which indicate increased activity of the thyroarytenoid muscles for activities associated with both effort closure (coughing and throat clearing) and swallowing in complement to lateral cricoarytenoids
and interarytenoids. Furthermore, the height mechanism evidently plays an increasingly important role, especially the thyrohyoid muscle (already noted above in explaining the decreased distance between the thyroid cartilage and hyoid bone seen in effort closure and swallowing). Two effects can be attributed to this muscle: first the thyroid cartilage will rotate anteriorly such that its oblique line becomes more vertically oriented (an effect also caused by anterior digastric and geniohyoid contraction; Griesman, 1943, p. 19). Without further action, these changes could cause increased tension on the vocal folds, but muscular attachments to the posterior margins of the cricoid cartilage (via cricomucosal, cricoepiglottal, and thyroarytenoid muscles\(^{28}\)) will bring the arytenoids closer to the internal notch of the thyroid, mitigating any tension increasing effect on the vocal folds from longitudinal stretching (although contraction of the thyroarytenoids will build longitudinal tension within the body of the vocal folds). More importantly, thyrohyoid contraction acts on the epiglottis as well, which is connected to both structures: raising its base (connected to the thyroid) and lowering its medial body (connected to the hyoid bone via the lateral hyoepiglottic ligaments) will put pressure on the epiglottic tubercle to retract (Fink & Demarest, 1978; Painter, 1986, p. 331). Thus, epiglottis retraction is not solely driven by lingual retraction (although lingual retraction does push the upper body of the epiglottis towards the posterior pharyngeal wall; Painter, 1986, p. 331). The thryoepiglottic muscle assists the lowering of the epiglottic tubercle and plausibly causes tissues associated with the epiglottis to compress into the ventricular folds below (and thus the thyroepiglottic muscle is a plausible member of the general protective closure mechanism as pointed out by Lindqvist-Gauffin; 1972, p. 2).

\(^{28}\) Possibly even a superior thyroarytenoid muscle, although this muscle is not universally found in humans (Zemlin, 1998, p. 133).
The aryepiglottic muscles are sometimes suggested to play a role in narrowing the vestibular space (Lindqvist-Gauffin, 1972, p. 3; Rose, 1989, p. 242; Thumfart, Platzer, Gunkel, Mauer, & Brenner, 1999, p. 41) and possibly even help to retract the epiglottis (Zemlin, 1998, p. 134) or assist in drawing the arytenoids anteriorly (Vilkman, Sonninen, Hurme, & Körkkö, 1996, p. 89), but they are not substantial muscles and may not even connect with the epiglottis in all cases, or in any case (M. M. Reidenbach, 1998b). Furthermore, it is important to remember that once the posteroanterior narrowing of the epilarynx space starts to occur, the aryepiglottic folds begin to swing from a posteroanterior orientation to a lateromedial one, mitigating any posteroanterior influence that the aryepiglottic muscles might have on the structures of the upper epilarynx. These muscles could still assist in epilaryngeal narrowing by causing concentric bulging of the aryepiglottic folds (Painter, 1986, p. 331, 1991, p. 457). More interesting perhaps is their connection to arytenoid adduction on account of their nature as projections of the oblique interarytenoid muscles (and considered by some to be the aryepiglottic part of the oblique interarytenoid muscles instead; see Thumfart, Platzer, Gunkel, Mauer, & Brenner, 1999, p. 41). Contraction of the aryepiglottic muscles could serve to stiffen the aryepiglottic folds and thus ensure patency of the epilaryngeal airway during respiration and minimize the likelihood of aryepiglottic vibration during modal phonation.

Pharyngeal musculature is also sometimes imputed to be important in laryngeal constriction (Shin, Hirano, Maeyama, Nozoe, & Ohkubo, 1981; Keyser & Stevens, 1994, p. 231), or is loosely mentioned in connection with it. The pharynx likely plays a role in

---

29 As corroboration of this speculation, the author has observed apparent contraction of the aryepiglottic muscles during normal phonation. Their action seems to have an effect on the orientation of the cuneiform cartilages and produces a narrowing of the aryepiglottic fold.
causing an increase in global larynx height: indirectly, on account of the generalized raising effect on the pharynx caused by the salpingopharyngeus and stylopharyngeus, and directly, on account of the palatopharyngeus (which inserts into the posterior margin of the thyroid cartilage), and the oblique (downward sloping) orientation of the medial and inferior pharyngeal constrictor muscles (Negus, 1949, p. 91), which connect to the hyoid bone and thyroid and cricoid cartilages, respectively. These muscles likely are important in influencing vocal fold tension via the cricopharyngeus (Vilkman, Sonninen, Hurme, & Körkkö, 1996) and palatopharyngeus, and in the absence of epilaryngeal stricture, that effect is likely an increase in glottal pitch.

2.2 Observations of the epilarynx in data and theory

The epilarynx has been an empirical blind spot in speech research. A conspiracy of factors have led to it being overlooked by many speech researchers and to it ultimately not being recognized in mainstream phonetic or phonological theory. The factors leading to this blind spot stem from the different challenges that imaging technologies present to the researcher interested in examining sound production in the lower vocal tract: it has been and continues to be a relatively inaccessible location to study.

In §2.2.1, I examine the difficulties associated with epilaryngeal observation in x-ray and laryngoscopy, which helps to explain the omission of the epilarynx in phonetic and phonological theory. Next, in §2.2.2, early sources in the literature that confirm the activity of the epilarynx in speech-related behaviour are surveyed. Developing out of this early work is the research of John Esling and his colleagues, which constitutes the topic

30 Which, incidentally, relates to the low position of the larynx in relation to the pharynx in humans.
of §2.2.3. Finally, more recent research both confirming the articulatory phonetic model established by Esling and identifying controversies concerning the epilarynx are discussed in §2.2.4.

2.2.1 Epilaryngeal observation difficulties: Focus on x-ray and laryngoscopy

Imaging and measuring lower vocal tract articulation is a long standing challenge in phonetic research. It is very difficult, if not entirely impossible, to take direct measurements of the structures of the larynx. This is true for devices such as the Carstens Articulograph, which tracks tissue movement using adhesive electromagnetic sensors; likewise, recording muscle activity using electromyography (EMG) is highly invasive, presents the difficulty of obtaining a sufficiently isolated recording of the desired muscle activity, and, especially in the case of the finer muscles of the epilarynx, is prone to failure (Hillel, 2001; Sakakibara, Kimura, Imagawa, Niimi, & Tayama, 2004). Electroglottography (EGG) does not easily allow vibrations of the different parts of the larynx to be distinguished. Visual means of observing epilarynx behaviour, such as x-ray, magnetic resonance imaging (MRI), laryngoscopy, and laryngeal ultrasound can more easily obtain data of epilarynx behaviour. However, there are still challenges facing research using these instruments, and the effect of these challenges is reflected in early (and highly influential) research on sounds made in the laryngeal and pharyngeal regions of the vocal tract. Since the data of these studies forms the empirical foundation of modern phonetic and phonological theories, the consequence is the propagation of the failure to identify the epilarynx as an integral speech production component of the lower vocal tract.
X-ray (usually cineradiography, i.e. video x-ray), popular in speech research from the 60s up to the mid-80s, provides a two-dimensional projection of vocal tract structures. On account of the predominantly ventral-dorsal orientation of the vocal tract in the human body, midsagittal x-ray imaging is typical. In the midsagittal view, it is possible in this view to image the sagittal profile of the epilaryngeal tube. The question is, then, what factors caused it to be overlooked? In early linguistic x-ray (approximately prior to 1970), image intensifier technology placed a strict limit on the diameter of the safely imageable area: only a portion of the vocal tract could be photographed, as Delattre (1971, p. 132) comments “if the lips showed, the pharynx did not, if the tongue tip showed, the tongue root did not.” Such a limitation might account for some images where the laryngopharynx is outright absent from tracings, as depicted in Figure 2.9a (Bgažba, 1964; Ladefoged, 1964; Gaprindašhvili, 1966; Ladefoged, DeClerk, Lindau, & Papçun, 1972), precluding evaluation of epilarynx state (Esling, 1996, p. 69; Edmondson & Esling, 2006, pp. 179–180). With limited imaging real-estate, it would make sense to focus on the tongue, which exhibits large, easily observable displacements. Delattre further notes that at the time he conducted his seminal x-ray study in 1971, new image intensifying technology was recently made available, meaning that the imageable x-ray footprint was no longer an impediment to visualizing the epilarynx in action. However, another problem can be identified in a large number of x-ray tracings (Delattre, 1971; Lindau, 1975, 1978; Hess, 1998): often only the outline of the laryngopharynx is

---

31 In a sagittal x-ray, one might picture the pharynx as a horse’s leg: the epilarynx is the hoof, and exhibits a characteristic anterior slant, just as the horse hoof does. Alternatively, one might picture the lower vocal tract as a high-heeled boot with the piriform fossae and post-cricoarytenoid space being the heel and the epilarynx being the foot proper.

32 This statement is not entirely accurate. For example, x-ray images in Russell (1931) and Negus (1949, pp. 147–148, figs. 160, 161) do show the entire vocal tract. Despite lacking optimal structural registration, these images suggest that technology is not the only factor hindering the study of the epilarynx.
indicated (see Figure 2.9b). Many such tracings show a line following the contour of the epiglottis down to the level of the vocal folds, where it changes directions according to the orientation of the vocal folds and is then continued all the way to the posterior pharyngeal wall, where it ascends. This type of trace permits evaluation of larynx height, but it omits the posterior border of the epilarynx, which precludes evaluation of posteroanterior epilaryngeal narrowing.

Figure 2.9: Different styles of x-ray tracing. (a) pre-1970s x-ray tracing that omits the epilarynx entirely; (b) tracing showing only the laryngopharynx; (c) tracing that includes an outline of the epilarynx but no clear indication of the aryepiglottic folds or ventricular folds (symbolized with the question marks). All images are retracings of tracings from the listed sources.

This empirical limitation of x-ray data is critical because the models of the speech production capacity of the lower vocal tract which were formulated during the second

\[33\] Lindau’s (1975, pp. 34–36) suggestion that the bottom of the trace of the x-ray view of the vocal tract (its lowest visible point) represents “the top of the shadow of the hyoid bone” (p. 36; see Point F, p. 34) indicates a misunderstanding of the anatomy of the lower vocal tract (or a monstrously long epiglottis).
half of the 20th century drew heavily upon and reflect these early x-ray observations (especially in regard to pharyngeals and pharyngealization). The action of the tongue takes primacy in these data, which justifies the use of the expression *linguo-centrism* to describe these accounts. Accounts of data from several of the language groups possessing lower vocal tract sounds can be characterized this way34: the languages of the Caucasus (Dzhejranishvili, 1959; Bgažba, 1964; Gaprindašhvili, 1966; Catford, 1983, p. 348; Ladefoged & Maddieson, 1996, pp. 170, 308), Tungusic language Even (Novikova, 1960; Ladefoged & Maddieson, 1996, p. 307), Niger-Congo and Nilo-Saharan languages with so-called tongue root vowel harmony (Ladefoged, 1964; Ladefoged, DeClerk, Lindau, & Papçun, 1972; Painter, 1973; Lindau, 1975, 1978; Casali, 2008, p. 506), and Arabic (Al-Ani, 1970, 1978; Delattre, 1971; Ali & Daniloff, 1972; Ghazeli, 1977; Boff-Dkhissi, 1983). Most of the tracings of these data omit the epilarynx profile entirely, although there are some exceptions as illustrated by Figure 2.9c. Contemporary phonetic texts reproduce these observations (Laver, 1994; Ladefoged & Maddieson, 1996) thereby reinforcing the focus that is placed on the action of the tongue both in phonetic and phonological research. For example, the feature [Advanced Tongue Root]/[ATR] (Stewart, 1967)35 reflects the message conveyed in this research: the tongue is in control of sound production in the lower vocal tract.

Although the emphasis in this research is on actions of the tongue, particularly of the tongue root, it is possible to identify some features of epilarynx stricture in x-rays. There are x-ray data tracings that do include the profile of the epilarynx (as illustrated in

34 Much of the data is incorporated into typological surveys of post-velar and pharyngeal articulation (e.g. see Bessell, 1992; Hess, 1998).
35 John Ohala (personal communication) attributes the “apparent discovery” of ATR, the phenomenon, to Ladefoged (1964) but acknowledges Stewart as the coiner of the term *ATR*. 
Figure 2.9c). This type of trace allows for evaluation of posteroanterior narrowing and vertical compaction of the epilarynx. Despite this, many features of epilarynx state still cannot be detected using this type of trace, such as the lateromedial dimension, the relationship between the ventricular folds and vocal folds, and the configuration of the aryepiglottic folds. Even in x-ray tracings that omit the epilarynx it is possible to identify epilarynx state by the contour of the lower epiglottis and evidence of larynx raising. Thus, it is not the case that the existence of epilarynx activity is wholly unknown in this research, and, in certain instances (e.g. Hess, 1998), its presence is noted as a possible feature of some sounds. Generally, however, focus falls back on lingual position, being more visible and easy to interpret in x-ray data. A notable example of this comes from Ladefoged & Maddieson’s *Sounds of the World’s Languages* (1996, pp. 311–313); in their discussion of strident vowels, they observe that the epilaryngeal stricture seen in x-ray data of Anthony Traill’s (1985, 1986) imitations of !Xóõ strident vowels does not occur for the pharyngealized vowels in the same language or any other Khoisan language and is ostensibly absent from the x-ray data of pharyngealized vowels in the languages of the Caucasus. Their claims are based on x-ray data in which it is far easier to see action of the tongue than that of the epilarynx. These observations then propagate in the literature: an example is Pulleyblank’s (2006, p. 17) interpretation of Ladefoged & Maddieson’s account that stridency, as in Khoisan languages, is a specialized tongue configuration, which illustrates the linguocentric bias characterizing much of the phonetic and phonological literature, even when there is some x-ray data suggesting more is at play.
Fibreoptic laryngoscopy is another popular method for collecting visual data of lower vocal tract – particularly laryngeal – function. There are two major impediments to visualizing epilarynx function using this approach. First, the epiglottis visually occludes most of the lower anterior section of the epilarynx, and it can entirely obstruct the view of the epilarynx if it is sufficiently retracted; this makes [i] the ideal vowel for laryngoscopic examination because advancement of the pharyngeal portion of the tongue helps draw the epiglottis forward, expanding the viewable area of the epilarynx (Williams, Farquharson, & Anthony, 1975, p. 305). For open vowels, the tongue and epiglottis eclipse the laryngeal aperture; this is especially the case for [a], during which the mesopharynx is very narrow. This is evident in Gauffin & Sundberg’s (1978, pp. 160–161) study of pharyngeal constrictions in vowels. Their data show that, for [a] and [ɑ], the lateral margins of the epiglottis actually make contact with the latero-posterior aspect of the pharyngeal wall. Williams et al. (1975, p. 305) note that epiglottopharyngeal distance can reduce to 5 mm for [a]. Painter (1986, p. 331) also notes that epiglottis retraction narrows the pharynx proper for the [ɑ] vowel. Its tendency to approach and even make contact with the posterior pharyngeal wall during such vowels makes the epiglottis a general “nuisance” (Painter, 1986, p. 331) for the phonetician wanting to see the vocal folds. More extreme pharyngeal narrowing found in consonants

---

36 Early laryngoscopic visualization involved the use of a dental mirror. Fiberscopes first appeared in 1957, but the true watershed in this research came in 1968 when flexible transnasal laryngoscopy was invented (Sawashima & Hirose, 1968; Edmondson & Esling, 2006, p. 160). It was at this time that Jan Gauffin (then Jan Lindqvist) would visit the Research Institute of Logopedics and Phoniatrics at the University of Tokyo (collaborating with Osamu Fujimura, Masayuki Sawashima, Hajime Hirose, and others) and produce some of the first evidence and theory on the use of the speech functioning of the epilarynx (Lindqvist, 1969; Lindqvist-Gauffin, 1972; Lindblom, 2009).

37 They estimate the narrowest cross-sectional pharynx area for these vowels at two camera heights (level-1 is just below the uvular; level-2 is at the inferior edge of the second cervical vertebra): [ɑ] has areas of 1.7 cm² (level-1) and 0.3 cm² (level-2) and [a] has areas of 1.5 cm² (level-1) and 0.3 cm² (level-2).
such as pharyngeals only serves to make laryngoscopic observation of the epilarynx more challenging.

The second impediment to laryngoscopic imaging of epilarynx activity is the limited impression it gives of the vertical dimension. The vertical dimension is one of the key axes of epilarynx action. Particularly important is the apposition of the vocal folds and ventricular folds when there is vertical epilarynx reduction. In laryngoscopy, one can only speculate whether vocal-ventricular fold contact occurs or not. Visual cues such as ventricular medialization and the shadows cast by the ventricular folds on the vocal folds help to indicate the distance, but they cannot be used conclusively to judge the distance between these structures.

Another matter concerns the elevation of the larynx. Often researchers rely upon the relative scaling of the laryngeal structures to assess larynx height. Brightness changes are also used to this end based on the assumption that there is greater illumination of the larynx and stronger light reflection when it is closer to the camera. Such impressionistic observations of larynx-height changes are complicated by the fact that camera height relative to the height of the larynx is typically not constant or even known during laryngoscopic examination. The distance between the camera and the larynx will have a significant impact on apparent brightness of the laryngeal structures. Also problematic is that apparent brightness can increase when the angle of a laryngeal surface changes and thereby reflects more or less light – even though this may not involve a change to larynx height.

Great strides continue to be made in imaging technology, such as MRI and more recently, real-time MRI (rtMRI; see e.g. Lammert, Proctor, & Narayanan, 2010), which,
in the future, will be pivotal in illuminating epilarynx behaviour in speech and its relation to the rest of the structures of the vocal tract. However, there are still examples in this work which reflect the earlier tendency in x-ray to neglect the epilarynx or conflate it with the laryngopharynx. For example, Tiede’s (1996) MRI study of so-called tongue root harmony in Akan and tense/lax vowels in English combines measurement of the epilarynx with that of the piriform fossae. In Kröger et al. (2004), mid-sagittal traces of MRI neglect many of the details of epilarynx structure evident in the raw data, which consequently suggest that the epilarynx is wider for the vowels they are documenting than it really is. A more recent study by Honda et al. (2010) of the hypopharyngeal/laryngopharyngeal cavities (the epilarynx and piriform fossae) shows the utility of MRI in studying how the epilarynx is configured during basic postures sustained over time (they examine inspiration, humming, and whispering). With continual advances in rtMRI, 3D visualization of time-dependent changes to epilarynx configuration will soon be possible.

2.2.2 Previous evidence and theory for the speech activity of the epilarynx

As §2.1.2 indicates, laryngologists such as Negus (1929, 1949), Pressman (1954), Griesman (1943), Fink (1956, 1962, 1974a, 1974b, 1975) and numerous others have contributed to a theory of the basic, life-supporting functions of the epilarynx, and the functioning of its components. Although its speech functioning is not undocumented in this work, it is generally treated as peripheral. There are some early observations\(^\text{38}\) that

\(^{38}\) John Ohala (personal communication) comments that Fant (1960) gave some consideration to the sagittal conformation of the epilarynx.
laryngeal constriction in speech production involves epilarynx narrowing: for example, Ringgaard (1960, 1962) describes x-ray of stød in Western Jutland Danish and notes that “both the true and the false vocal chords (sic) are pressed firmly together and sinus Morgagni is quite obliterated” (1962, p. 206). However, no coherent phonetic theory of the contribution of the epilarynx or its various components had been articulated in research prior to 1969.

The first attempt at theorizing about the role of the epilarynx in speech is creditable to a series of related reports by Jan Gauffin (Lindqvist, 1969; Lindqvist-Gauffin, 1972; Gauffin, 1977; see also Lindblom, 2009) and culminating in his dissertation (1972). Gauffin’s model was based on observations of his own speech production using the latest developments in fiberscopic laryngoscopy at the time (Sawashima & Hirose, 1968). In his model, he posits that the phylogenetically-basic functions of the larynx – breathing, phonation, and protection – are exapted into speech as breathing position, voicing position, and protective closure (Lindqvist-Gauffin, 1972, p. 2) each independent and freely combinable with the others, and each possessing its own dedicated physiological mechanism, those being, respectively, the glottal abduction mechanism, the pitch control mechanism, and the laryngealization mechanism. Such a model directly challenges the one-dimensional, abduction-adduction continuum proposed by Ladefoged (1971), i.e. voiceless – breathy – murmur – lax voice – tense voice – creaky voice – creak – glottal stop (similarly, see Figure 2.10).
Gauffin suggests a two-dimensional system, where laryngealization cross-cuts all stages along the abduction-adduction continuum, giving possibilities such as laryngealized-breathiness (whisperiness); see Figure 2.11. An important connection he makes is the interaction between the pitch control system and laryngealization: the latter mechanism induces low pitch values with increased closed quotient and the occurrence of irregularities in the glottal pulse. Gauffin suggested that vocal fold adduction induced by the laryngealization mechanism is the cause.

It is the mechanism of laryngealization and its effect of producing the protective closure state that relate directly to the idea of the epilarynx as a speech production
mechanism in the present work (although Gauffin does not strictly refer to the epilarynx). The closure that Gauffin observed, which occurred during his own production of a Swedish glottal stop, was complete closure of the epilarynx in complement to vocal fold adduction (see, e.g. Lindqvist-Gauffin, 1972, p. 2), and his description highlights the coaptation of the tubercles of the cuneiforms and that of the epiglottis in forming a hermetic seal: one suitable for protection but applied with various degrees of engagement to speech production.

Gauffin’s work did not go unnoticed. In Halle and Stevens’ monumentally influential “note” on laryngeal features, there is a marginal effort (1971, p. 59) to countenance Gauffin’s findings by suggesting the (reduced) protective closure mechanism is the realization of the non-syllabic correspondent of their glottal closure feature (although they place emphasis on the ventricular folds as the mechanism responsible), but they ultimately left the matter to future research. Gauffin’s work was also not the sole source of evidence showing the relevance of the epilarynx in speech. Roach’s (1979) report that laryngoscopic and x-ray imaging of his own productions of glottalized /p t k ʃ/ in English reveal a laryngeal closure produced at the vocal, ventricular, and aryepiglottic levels, casting doubt on the appropriateness of the term “glottal closure” used in describing these sounds, and directly matching the mechanism described by Gauffin, and similar to observations of laryngeal constriction in Danish (Ringgaard, 1960, 1962) and to reports of potential ventricular fold involvement in

39 Although, in a commemororative publication, Lindblom (2009) intimates that (forty years after Gauffin’s original work on the subject and following Gauffin’s passing in 2008) Gauffin’s model did not receive enough attention to significantly advance standard phonetic theory of the larynx beyond Ladefoged’s (1967) one-dimensional, abduction-adduction model, likely because of the work’s limited distribution (Lindblom, 2009, p. 151).
glottalization in American English (Fujimura & Sawashima, 1971). Laryngoscopic evidence (Iwata, Sawashima, Hirose, & Niimi, 1979) shows that glottalization on final stops /p t k/ in Fukienese (Taiwanese) involves moderate epilarynx activity (ventricular fold adduction and posteroanterior narrowing) with a concomitant drop in F0 (by 20 to 30 Hz, a fact which the authors point out is predicted by Gauffin’s model). The epilaryngeal engagement decreases with increased speaking rate (towards basic vocal fold adduction). Iwata et al. suggest that the ventricular fold component functions to reinforce glottal closure in order to prevent or halt vibration (p. 76-77).

Work not strictly focused on glottalization and glottal closure in speech but rather more broadly on phonation type, voice quality, and pharyngeal articulation shows how Gauffin’s mechanism applies to far more speech phenomena than that which Gauffin himself originally recognized. Laminagraphic x-ray evidence of creaky phonation has shown that the ventricular folds descend upon and compress into the vocal folds (Allen & Hollien, 1973); likewise, creaky and whispery phonation are shown by Sawashima and Hirose (1980) to involve epilaryngeal narrowing. Laufer & Condax (1979, 1981) were among the first to observe pharyngeals laryngoscopically; their data, which are for Hebrew, do not directly show epilaryngeal stricture due to obstruction by the epiglottis, but they correctly inferred that an aryepiglotticoepiglottal constriction was occurring. The tendency for the epiglottis to retract towards the posterior pharyngeal wall without concomitant lingual retraction (especially for [h]) led to their analysis that the epiglottis acts independently as the articulator of pharyngeal sounds.

of the lower vocal tract in raised, neutral, and lowered larynx voice quality settings during [ə]. He observes that raised larynx voice quality does not only involve an effective shortening of the vocal tract but also concomitant deformation to the epilaryngeal tube. He describes the tube as “‘folding’ to accommodate the raising [of the larynx]” (p. 182). Despite the effort to maintain the [ə] configuration constant for all laryngeal states he examined, there is obvious narrowing of the pharynx concomitant with the raising of the larynx in the raised larynx voice quality.

Painter (1973, 1986, 1991) has made especially important contributions to understanding the diverse configurations the epilarynx can assume that are available for the production of a broad range of sound categories based on laryngoscopic observations (of phonetically trained productions). He suggests the following occurrences of epilarynx activity in language:

- Arabic emphatics and ['ain] … pharyngealization in Caucasian languages … glottalization (i.e. secondary glottal stricture) in certain American Indian languages and Yorkshire English … consonantal laryngealization (i.e. secondary creaky voice) in certain West African languages … implosives and ejectives, which involve vertical laryngeal movement … creaky voice, either as tone related in Vietnamese, or segmental as in Danish … vowel harmony in certain West African languages that have two sets of vowels, one of which is characterized by a narrowed pharynx. (Painter, 1986, p. 330)

Traill (1985, 1986) documents phonetic and phonological systems of Khoisan language !Xóó, which has become recognized for its so-called sphincteric phonation type. Traill presents laryngoscopic observations of a native speaker’s production (1986, p. 124) and x-ray evidence of his own phonetic imitations of the sound (1986, p. 126; also see Ladefoged & Maddieson, 1996, p. 331). Traill uses the term sphincteric to refer to constriction of the epilarynx, particularly its upper margin (involving aryepiglottico-
epiglottal approximation), and vibration of these structures (Traill suggests the epiglottis and arytenoids\textsuperscript{40}); the label \textit{sphincteric} is somewhat ironic because Traill regards each of the three physiological planes of the larynx as constituting a sphincter of the larynx, yet, in !Xóõ, the vocal folds and (by inference) the ventricular folds are abducted and supposedly not vibrating (although acoustic evidence indicates that onset and offset stages of the sphincteric mode involve vocal fold vibration).

Traill (1986, pp. 129–130) is directly critical of Gauffin’s observations (calling his production of glottal stop as “most unusual” [p. 129] and ultimately idiosyncratic) and his model, stating that the !Xóõ data show that laryngeal behaviour in speech must be more refined than that predicted by a model suggesting it is attributable to “basic vegetative” (p. 130) and phylogenetically ancient mechanisms of laryngeal function. Traill’s argument is that !Xóõ speakers do not fully engage the protective mechanism (although this is reminiscent of Gauffin’s notion of “reduced protective closure”; see Gauffin, 1977) and, rather, balance muscular forces in such a way as to permit narrowing of the upper part of the protective mechanism (i.e. the upper epilarynx) but abduction of its lower part (i.e. the vocal folds and the lower epilarynx). Such a mode of phonation, while not explicitly identified by Gauffin, is arguably compatible with Gauffin’s model given that it implies a continuum of laryngealization (or epilarynx engagement) and independent vocal folds which may abduct during laryngealization to yield a whispery state\textsuperscript{41} (rather than a breathy state).

\textsuperscript{40}Traill (1986, p. 125) incorrectly labels the cuneiform tubercles as the arytenoids. His description is strongly suggestive that aryepiglottic vibration is occurring, although this is never said explicitly.

\textsuperscript{41}I would suggest that sphincteric phonation as Traill describes it is nothing more exotic than whispery harsh voice (with growl [aryepiglottal and/or epiglottal trilling]).
In any case, it is clear that such a mode is not compatible with the Ladefogedian one-dimensional model; this is clear in Gordon and Ladefoged’s (2001, pp. 391, 401) somewhat puzzling treatment of “sphincteric” phonation in !Xôô (which they refer to as “strident phonation”): it is discussed under the heading of creaky phonation, and, despite awareness of non-glottal mechanisms of source generation, it is still understood in terms of a function of the glottis (and not, for example, the epilarynx). The one-dimensional, abduction-adduction model does not help us understand the !Xôô phonation system and conceptually confines us into reducing the larynx to just a glottis (the essence of glottocentrism). It may be that this confinement led Gordon and Ladefoged to conclude their survey of linguistic phonation types by remarking that “If the Xôô (sic) did not exist, and someone had suggested that [sphincteric phonation] could be used in a language, scholars would probably have said that this was a ridiculous notion” (Gordon & Ladefoged, 2001, p. 401).

Catford (1977a, 1977b, 1983) has suggested numerous sounds, particularly those found in Caucasian languages, are produced by the action of the components of the epilarynx. He identifies the possibility for a ventricular stop, which he analyzes as a “strong glottal stop”: a sound that uses ventricular reinforcement of the lower, primary closure at the glottis formed by the vocal folds. Such a sound is said to occur in Chechen, Nakh, and some Dagestanian languages (Tsez and Dargi) with “some constriction of the pharynx”, and he claims it contrasts in these languages with an unreinforced (i.e. lacking the ventricular component), “weak” glottal stop (Catford, 1977b, p. 289), which is often realized as creakiness. Catford retracts this claim of two types of glottal stop somewhat in

\[42\] It must be granted that they do in fact admit it to be “somewhat of an oversimplification” (Gordon & Ladefoged, 2001, p. 384).
a later work (1983) on the basis of Laufer & Condax’s (1979) argument that the epiglottis is an active articulator and retreats to the position of viewing the strong glottal stops as involving epiglottotarytenoidal stricture. In his “Mountain of Tongues” paper (1977b), Catford does not explain what evidence he used to make his assessments about the articulation of the “strong” and “weak” glottal stops in the first place, but it is clear that x-ray (lacking a trace of the epilarynx, as in Figure 2.9a) forms a basis for his later judgments (1983). In this later work, he postulates the existence of “deeper” pharyngeals. These are said by Catford to use “contraction of the pharynx just above the larynx”. He further describes them as “raucous … sounds, articulated with noisy participation of the epiglottis” and suggests that “they would be regarded as pathological if they were produced by a person speaking English” (p. 346). (One can infer that the impressionistic label deeper may, in addition to a sense of depth within the vocal tract, partially reflect an impression of very low or “deep” pitch produced by the “raucous” vibration of the components of the epilarynx).

Also important is the connection Catford reports between the pharyngealized vowels found in Caucasian languages and (general phonetic) rhotic vowel quality. Catford specifically reflects on Kodzasov’s remark that pharyngealized vowels found in some Caucasian languages (such as Tsez and Udi) give the impression of being “r-coloured” to American listeners (Catford, 1983, pp. 347–349). Catford (1983, pp. 349–350) examines x-ray evidence (which omits the epilarynx) showing a strong epiglottotarytenoidal closure. He then speculates that this is the basis of the so-called pharyngeal stop (e.g. found in Chechen) often reported in the literature (Catford, 1983, p. 347, also see 1977b, p. 289).

---

43 Catford, who earlier claimed that the epiglottis was not capable of independent movement (1977a), concedes to Laufer and Condax the possibility that it indeed might be capable of moving independently of the tongue and essential in producing what he labels epiglottotarytenoidal closure. He then speculates that this is the basis of the so-called pharyngeal stop (e.g. found in Chechen) often reported in the literature (Catford, 1983, p. 347, also see 1977b, p. 289).
pharyngeal narrowing and a lingual configuration reminiscent of the so-called bunched R in American English with “bulging” in the palatal and lower pharyngeal regions and a conspicuous dip in the mid-dorsum near the uvular region (and roughly in the region of middle genioglossus fibers, although this is not suggested by Catford). As pointed out above, Painter had suggested that some West African languages constrict the epilarynx in producing the set of vowels involving a narrowed pharynx (1986, p. 330). Tiede’s (1996) MRI study of this contrast in Akan and comparison with the tense/lax vowel contrast in English lends support to the suggestion. While Tiede concludes that the difference between the two sets involves total pharyngeal volume and not just larynx height and tongue root position (in agreement with Lindau’s (1975, 1978) earlier assessment of such vowel harmony systems), some details of the study indicate that the epilarynx is playing a critical role. Most notably, English and Akan look comparable with respect to the configuration of the mesopharynx for the two vowel sets, but Akan is uniquely distinguished by its manipulation of the posteroanterior dimension (which Tiede calls pharyngeal depth) of the hypopharynx to differentiate the vowels, and this space is mainly a function of epilarynx volume (Tiede, 1996, p. 415 & fig. 8).

The work reviewed in this section constitutes the departure point for a new direction that soon followed in research showing the activity of the epilarynx, and this is the subject of the next section. Specifically, the mid-1990s marks the start of a Renaissance of Gauffin’s ideas and considerable growth in the empirical evidence

---

44 Sonya Bird (personal communication) has observed by means of lingual ultrasound a similar lingual configuration in non-rhotic infact vocalization.

45 One oversight in the comparison between these languages is the fact that larynx height is relevant in the Akan contrast but not the English one. This means that axial scan levels of the pharynx (defined with respect to the lowest point within the vallecula as a reference point), which are directly compared across speakers in the study, do not correspond to the same level in the different languages.
showing the action of the epilarynx in normal speech behaviour. Lindblom (2009), in his paper commemorating Jan Gauffin’s work on laryngeal mechanisms in speech, points out that this Renaissance is largely attributable to the work of John Esling and his colleagues.

2.2.3 Esling and colleagues

In 1996, Ladefoged & Maddieson’s phonetic epic, *The Sounds of the World’s Languages*, was published. Despite its survey of many of the empirical contributions from the sources reviewed in §2.2.2, the understanding of the phonetics and phonology of sounds produced in the lower vocal tract had evidently not progressed far beyond the popular notion of treating the larynx as independent from the rest of the vocal tract, i.e. the glottocentric-linguocentric view. Pharyngeals and related sounds were still conceived of as lingual sounds involving retraction of the tongue root to produce a constriction between the suprahyoid body of the epiglottis and the posterior pharyngeal wall. These views persisted, despite the growing body of evidence to the contrary that the epilarynx constitutes the primary constriction in these sounds (and numerous other related ones). Even in Ladefoged & Maddieson’s text, several of these facts are acknowledged, yet their model of place of articulation does not incorporate these observations: as illustrated in Figure 2.12, *pharyngeal* (15) is a constriction located in the mesopharynx produced by the upper part of the tongue root and the posterior pharyngeal wall, *epiglottal* (16) involves narrowing between the epiglottis and the posterior pharyngeal wall, and *glottal* (17) is simply the very lowest point (see Ladefoged & Maddieson, 1996, p. 14). Such a model does not mark a significant departure from early sentiments about what pharyngeal articulation is. For example, Hockett (1958, p. 66; as quoted in Esling, 1996, p. 65) states
that “a complete closure can be made in the lower pharyngeal region, by drawing the root of the tongue back against the back wall of the passage.” In Delattre’s x-ray study of pharyngeal articulation, a similar understanding of pharyngeal is expressed: “a pharyngeal articulation is one in which the root of the tongue assumes the shape of a bulge and is drawn back towards the vertical back wall of the pharynx to form a stricture” (Delattre, 1971, p. 129). One of Ladefoged’s introductory texts on phonetics reinforces this view: “Pharyngeals are produced by pulling the root of the tongue back toward the wall of the pharynx” (Ladefoged, 1993, p. 163). The choice Tiede (1996) makes to lump the epilarynx volume with that of the piriform fossae also reflects a paradigmatic emphasis on the pharynx and treatment of the epilarynx as simply a part of the laryngopharynx. Likewise, Hess (1998) lumps the epilarynx into the notion of laryngopharynx.

---

46 The changes that happen to the spaces of the hypopharynx (the epilarynx and piriform fossae) during various articulations do not even linearly correlate with each other, as raising the larynx will narrow both spaces, while lowering the larynx will cause epilarynx expansion but narrowing of the piriform fossae (Honda, Kitamura, Takemoto, et al., 2010).
Figure 2.12: Place of articulation redrawn from *Sounds of the World’s Languages* (based on Figure 2.1 and 2.2 in Ladefoged & Maddieson, 1996, pp. 12–13). Details of articulatory categories associated with the anterior part of the vocal tract are omitted (for space reasons; see Ladefoged & Maddieson, 1996, p. 14). The question mark indicates that the epilarynx is not represented in this model: there is the epiglotto-pharyngeal stricture and glottal stricture but no aryepiglotto-epiglottal stricture and the status of the ventricular folds is unknown.

Esling’s (1996) laryngoscopic study of phonetic productions of laryngeal and pharyngeal sounds marks a return to Gauffin’s work, but Esling’s approach is much broader and involves a number of refinements not articulated by Gauffin. For example, Gauffin (1972) regarded glottal stop as the extreme end point of laryngealization, involving closure of the larynx from vocal folds to aryepiglottic folds; Esling’s glottal
stop is conceived of as a gesture involving vocal fold adduction and supporting ventricular fold adduction, while his interpretation of pharyngeal stop is a stricture that is aryepiglottic in nature (in addition to vocal and ventricular fold closures below). Esling considers the articulation of pharyngeal sounds in general to be aryepiglottic in nature — in fact, he asserts that the aryepiglottic folds are the active articulator — and participation of the epiglottis and tongue is regarded as passive, contrary to Laufer & Condax’s (1979, 1981) argument that the epiglottis is the active articulator of these sounds. Esling emphasizes that the aryepiglottic folds rise up and forwards to meet the base of the epiglottis rather than just the epiglottis retracting to cover the arytenoids; the narrowing of the pharynx is also evaluated to be of secondary importance. Tongue retraction and larynx raising are suggested to be default settings associated with producing a tightening of the so-called aryepiglottic sphincter mechanism; Esling suggests, however, that aryepiglottic constriction is possible with any particular larynx height to achieve different auditory results. Pharyngeal articulation is interpreted in his framework as involving three possible constriction locations: the upper pharynx is the site of uvulars, the mid pharynx corresponds to pharyngeals produced with a raised larynx setting and aryepiglottic constriction, and the lower pharynx is the site of pharyngeals produced with a lowered larynx setting and aryepiglottic constriction.

Three details Esling (1996) observes regarding this mechanism are particularly relevant for this dissertation: first, the aryepiglottic folds can be set into vibration (as evinced by Traill’s work), which suggests they may have a phonatory function; second, Esling notes the inherent relationship between pharyngeals and laryngealization by pointing out the often reported detail that pharyngeals involve “glottalized” (p. 71, a
comment by Pierrehumbert in Honda, Hirai, Estill, & Tohkura, 1995) phonation; and third, pharyngealized or raised larynx voice quality of the sort described by Laver (1980, pp. 24–29) and Nolan (1983, pp. 182–187) is a consequence of the engagement of the aryepiglottic sphincter mechanism described by Esling. Thus, all together, the mechanism acts like a source of sound in its own right, impacts the vocal fold sound source, and contributes to the peculiar quality of voice (partially attributable to a shortening of the vocal tract because of the raised larynx setting) associated pharyngealized vowels in Caucasian languages, tense register in languages such as Bruu and Mpi, and West African languages with harmony systems (1996, pp. 80–81).

In a follow-up paper, Esling (1999) gives a critical examination of the categories pharyngeal and epiglottal, and, once again, based on observation of his own carefully controlled phonetic productions, he concludes that the distinction is spurious: Esling asserts that pharyngeals and epiglottals are all fundamentally aryepiglottic in nature. He suggests the following definition of pharyngeal articulation: “[sounds that involve] retracting the tongue root (with the attached epiglottis) towards the back wall of the pharynx, raising the larynx and approximating the cuneiform cartilages of Wrisberg within the aryepiglottic folds to the base of the epiglottis, bringing the folds parallel with the coronal line of the epiglottis” (p. 369). Larynx height is shown to be a freely manipulable parameter: pharyngeal and epiglottal sounds are demonstrated at raised, neutral, and lowered settings of the larynx (although no quantification is provided to confirm the height of the larynx in the different cases; see Esling, 1999, pp. 360–362). In this model, the supposed pharyngeal/epiglottal distinction in Agul (Ladefoged & Maddieson, 1996, pp. 37–39; also see Catford, 1983, p. 346) actually is a manner contrast
at the aryepiglottal place of articulation involving the use of aryepiglottic trilling in the case of the epiglottals.

While Esling’s study of cardinal phonetic possibilities was much in the spirit of Painter’s (1986, 1991) work and the work of Gauffin before him, empirical evidence showing activity of the epilarynx (Esling’s aryepiglottic/laryngeal sphincter mechanism) in the speech production of naïve speakers of languages containing lower vocal tract sounds was still needed to buttress Esling’s model. This empirical foundation was established through an extensive series of collaborations which helped to document the suspected action of the epilarynx in diverse linguistic phenomena of the lower vocal tract. Published highlights\(^{47}\) are summarized in Table 2.1 (video data are available online, see Esling & Harris, 2003a; for additional images see Moisik, 2008, Chapter 4): this work, which documents over twenty languages representing fourteen language families, illustrates epilarynx involvement in sounds such as laryngeals (glottal stop in comparison to glottal fricative), glottalized segments (glottalized resonants and consonants with glottal reinforcement), pharyngeals, vocal/tonal registers (tense/lax, harsh, breathy, faucalized), and the so-called ATR contrast.

---

\(^{47}\) There is a large volume of unpublished laryngoscopic material available for viewing in the Speech Research Laboratory in the Department of Linguistics at the University of Victoria. This database includes additional data for the languages listed in Table 2.1 and includes data for Danish, Cantonese, Korean, and Mandarin.
Table 2.1: Laryngoscopic natural language data collected by Esling, Edmondson, Harris, Carlson, Zeroual, Crevier-Buchman, and collaborators. Phenomena are phonetic but most have phonemic status.

<table>
<thead>
<tr>
<th>Family</th>
<th>Language</th>
<th>Phenomena</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Austronesian</strong></td>
<td>Amis</td>
<td>hʔ hʔ nhŋgʔq</td>
<td>(Edmondson, Esling, Harris, &amp; Huang, 2005; Edmondson &amp; Esling, 2006)</td>
</tr>
<tr>
<td><strong>Cushitic</strong></td>
<td>Somali</td>
<td>nʔ q̚q̚</td>
<td>(Edmondson &amp; Esling, 2006; Edmondson, Padayodi, Hassan, &amp; Esling, 2007)</td>
</tr>
<tr>
<td><strong>Gur</strong></td>
<td>Kabiye</td>
<td>ATR</td>
<td>(Esling, 2005; Edmondson &amp; Esling, 2006; Edmondson, Padayodi, Hassan, &amp; Esling, 2007)</td>
</tr>
<tr>
<td><strong>Interior Salish</strong></td>
<td>Nlaka’pamux</td>
<td>hʔ hʔ</td>
<td>(Edmondson, Esling, Harris, &amp; Wei, 2004)</td>
</tr>
<tr>
<td><strong>Kwa</strong></td>
<td>Akan (Asante/Fante/Twi)</td>
<td>ATR</td>
<td>(Edmondson, Padayodi, Hassan, &amp; Esling, 2007)</td>
</tr>
<tr>
<td><strong>Nilotic</strong></td>
<td>Bor Dinka</td>
<td>VR (V V V ñ)</td>
<td>(Edmondson &amp; Esling, 2006)</td>
</tr>
<tr>
<td><strong>Semitic</strong></td>
<td>Tigrinya</td>
<td>hʔ hʔ</td>
<td>(Esling, 2003; Esling &amp; Harris, 2003b)</td>
</tr>
<tr>
<td><strong>Semitic</strong></td>
<td>Moroccan Arabic</td>
<td>h h ŋ</td>
<td>(Zeroual, 1999, 2000; Zeroual &amp; Crevier-Buchman, 2002; Edmondson, Esling, Harris, &amp; Huang, 2005, p. 2)</td>
</tr>
<tr>
<td><strong>Semitic</strong></td>
<td>Palestinian Arabic</td>
<td>h h ŋ</td>
<td>(Esling, 2003)</td>
</tr>
<tr>
<td><strong>Semitic</strong></td>
<td>Iraqi Arabic</td>
<td>hʔ h ŋ ŋʔʔ</td>
<td>(Edmondson, Padayodi, Hassan, &amp; Esling, 2007; Hassan, Esling, Moisik, &amp; Crevier-Buchman, 2011)</td>
</tr>
<tr>
<td><strong>Sino-Tibetan</strong></td>
<td>Tibetan</td>
<td>tone, chanting</td>
<td>(Esling, 2002a)</td>
</tr>
<tr>
<td><strong>Sino-Tibetan</strong></td>
<td>Mandarin</td>
<td>tone</td>
<td>(Moisik, Lin, &amp; Esling, 2010; Moisik, Esling, Bird, &amp; Lin, 2011; Moisik, Lin, &amp; Esling, forthcoming)</td>
</tr>
<tr>
<td><strong>Tai</strong></td>
<td>Thai</td>
<td>hʔ</td>
<td>(Esling &amp; Harris, 2003b)</td>
</tr>
<tr>
<td><strong>Tibeto-Burman</strong></td>
<td>Bai</td>
<td>VR (V V ñ) (tense/lax)</td>
<td>(Edmondson, Esling, Harris, Li, &amp; Lama, 2001; Esling, 2002b, 2002a, 2005; Esling &amp; Edmondson, 2002; Esling &amp; Harris, 2003b; Edmondson, Padayodi, Hassan, &amp; Esling, 2007; Edmondson &amp; Esling, 2006)</td>
</tr>
<tr>
<td><strong>Tibeto-Burman</strong></td>
<td>Yi</td>
<td>VR (V V ñ) (tense/lax)</td>
<td>(Edmondson, Esling, Harris, Li, &amp; Lama, 2001; Esling, 2002b, 2002a, 2005; Esling &amp;</td>
</tr>
</tbody>
</table>
An important concept emerging from this work is the notion that there are two basic states of the larynx that are definable in terms of the setting or configuration of the structures of the epilarynx: these are the constricted and unconstricted settings of the larynx. Laryngeal behaviours can be grouped into these two categories based on laryngoscopic diagnosis of the position of the various components of the epilarynx. The constricted/unconstricted parameter is orthogonal to the abduction/adduction parameter of vocal fold activity, although these intersect in the state associated with laryngeal closure. (Pitch constitutes a third parameter, somewhat orthogonal to the other two but interacting in several important ways.) Furthermore, while the speech functioning of the constricted/unconstricted setting is not exactly comparable to basic, life-supporting laryngeal mechanisms such as inspiration, effort closure, and swallowing, there is parallelism: the constricted setting closes off the laryngeal airway and the unconstricted setting works to open it. Esling and Harris (2005) provide an illustration of canonical states (based on laryngoscopic observation of laryngeal behaviour in phonetic

48 [\text{\textasciitilde{V}}] is equivalent to underscore double tilde elsewhere.
productions and natural language productions), and their model is summarized in Table 2.2.

Table 2.2: Unconstricted and constricted laryngeal states and vocal fold parameters (based on Esling & Harris, 2005).

<table>
<thead>
<tr>
<th></th>
<th>unconstricted</th>
<th></th>
<th>constricted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no vibration</td>
<td>vibration</td>
<td>no vibration</td>
</tr>
<tr>
<td>abducted</td>
<td>breath / [h]</td>
<td>breathy voice</td>
<td>whisper / [h]</td>
</tr>
<tr>
<td>adducted</td>
<td>pre-phonation</td>
<td>modal voice</td>
<td>[? ?]</td>
</tr>
<tr>
<td>longitudinal tension</td>
<td>high pitch (towards falsetto)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are two conceptual models that provide a framework for understanding the laryngeal behaviour observed in the laryngoscopic data. The first is Esling’s (2005) Laryngeal Articulator Model (LAM) which broadly divides the vocal tract into two regions, an oral vocal tract and a laryngeal vocal tract, as depicted in Figure 2.13. The division point falls along an imaginary plane that intersects with the upper pharynx (in the uvular region and corresponding loosely with the oropharyngeal isthmus), divides the tongue into two sections (creating anterior and posterior sections, the latter of which approximately corresponds with the concept of the tongue root). Following Esling (1999), pharyngeal and epiglottal place categories are reinterpreted as referring to constriction at the level of the aryepiglottic folds. Modifications to voice quality associated with this location are made possible by the larynx height and tongue retraction parameters.
The key insight of Esling’s model is the connection it establishes between vowel quality and laryngeal quality. Traditional models of vowels posit a quadripartite system (high-low and front-back) governed by the movement of the tongue: this type of model is the essence of the linguocentric view. On the other hand, in the LAM, the laryngeal articulator is understood to be interwoven into the vowel system through the retraction parameter. Retraction characterizes the articulation of open-mid–to–open back vowels [ɑ ɒ ɔ] (which Esling calls “retracted vowels” or “pharyngeal vowels”; p. 40): Esling predicts that these vowels are “inherently susceptible to increasing degrees of laryngeal
constriction” (2005, p. 40) on account of the synergistic relationship between retraction, raised larynx configuration, and pharyngeal/epiglottal stricture. (Critically, this is not to say that they inherently involve aryepiglottic stricture of the sort found in pharyngeal/epiglottal consonants\(^{49}\); rather it is a matter of bias.) The LAM provides other sets of vowels organized by lingual action: the “raised” set covers the mid-close–
to–close back vowels \([u \ u \ o \ ɤ]\) and the “fronted” set covers the front vowels \([i \ i \ y \ y \ e \ ø \ æ \ æ \ a \ æ]\). Lowness is also redefined. Esling claims that, on account of the vaulted shape of the palate, tongue raising is not required for producing high front vowels. Furthermore, the jaw height parameter operates more effectively to manipulate the height quality of the front vowels than it does for the “back” vowels. This is because the mass of the tongue lies further from the temporomandibular joint (the main rotational axis of the jaw) when fronted than it does when either retracted or raised. Esling claims that lowness for back vowels is more influenced by laryngeal constriction than jaw height. In fact, Esling suggests that, at least phonetically, using the label *back* to describe the so-called “back vowels” mischaracterizes\(^{50}\) the raised and retracted sets and blurs the special synergistic relationship between retracted vowels and pharyngeal/epiglottal articulation.

The second conceptual model derived from the empirical foundation is Edmondson & Esling’s (2006) valves of the throat (hereafter the Valves Model). This work outlines six valves defining articulatory possibilities within the laryngeal vocal tract (*sensu* LAM). These valves are: the vocal folds (Valve 1), the ventricular folds (Valve 2),

\(^{49}\) Shahin confuses these details in summarizing Esling’s model (2011a) by assuming Esling means that all “[−ATR]” (i.e. (“tongue root”) retracted or “lax”) vowels, e.g. \([i \ e \ ø \ ø]\), involve aryepiglottic stricture and are therefore “pharyngeal” vowels. The reader should be careful to observe that this is an incorrect interpretation of the LAM.

\(^{50}\) Hence the title of Esling’s (2005) paper: *There are no back vowels: the Laryngeal Articulator Model.*
the aryepiglottic folds and base of the epiglottis (Valve 3), tongue/epiglottis retraction to the pharyngeal wall (Valve 4), larynx height (Valve 5), and pharyngeal narrowing (Valve 6). The work attempts to delineate the operation and interaction of the valves in the production of speech sounds of the lower vocal tract (see Table 2.1). There are three essential observations resulting from this work. First, Edmondson & Esling (2006, pp. 176, 183) demonstrate vibrations of the aryepiglottic folds associated with low tones of the harsh register in Bai and of the epiglottis associated with voiceless pharyngeal consonants in Somali. Second, they posit that the principal action of the ventricular folds is to compress into the vocal folds, resulting in changes to vocal fold vibration that lead towards arresting their motion: they call this mechanism ventricular incursion. Third, they connect changes in vowel quality associated with vocal register and with ATR-type harmony to voice quality (both phonatory quality and global quality involving vocal tract length). They suggest a constricted quality corresponding with epilarynx narrowing that involves more retracted vowel quality and constricted voice-and-phonatory qualities, mainly raised larynx voice, harsh voice, and/or creaky voice. In relation to this, they demonstrate a secondary degree of pharyngeal/epiglottal stricture co-occurring with aryepiglottic stricture which they label linguo-epiglottal-pharyngeal on account of extreme tongue retraction to the point of contact between the epiglottis and posterior pharyngeal wall.

The work of Esling and that of his colleagues has played a significant role in developing the phonetic model of how the larynx operates in speech. Initial observations of careful phonetic productions made using laryngoscopy were later corroborated by evidence from studies of numerous languages. This work emphasizes the role of the
aryepiglottic folds as the active articulator forming the primary stricture in pharyngeal/epiglottal sounds and suggests that there is a categorical opposition between constricted and unconstricted laryngeal settings that plays an important role in ATR systems and so-called tense/lax vocal register contrasts. In all of this work, the epilarynx is (for the most part) implicitly acknowledged, but Esling’s analysis focuses more on the component parts than the whole mechanism. His work has outlined some of the key hypotheses regarding epilarynx function, and its relationship with other vocal tract structures. These hypotheses are summarized in Table 2.3.

Table 2.3: Major hypotheses arising from the research of Esling and colleagues.

<table>
<thead>
<tr>
<th>hypothesis</th>
<th>details</th>
</tr>
</thead>
<tbody>
<tr>
<td>aryepiglottic primacy</td>
<td>pharyngeal/epiglottal sounds always have primary aryepiglottic constriction; secondary linguo-epiglotto-pharyngeal stricture can be added to this basic stricture</td>
</tr>
<tr>
<td>quasi-phonatory device</td>
<td>vibration of the aryepiglottic folds can be used as trilling enhancement to pharyngeals or as part of harsh voice quality</td>
</tr>
<tr>
<td>harshness</td>
<td>typically involves perturbation to vocal vibration caused by ventricular incursion, but aryepiglottic trilling is a possible exponent</td>
</tr>
<tr>
<td>ventricular incursion</td>
<td>during ventricular fold adduction, the caudal surface of the ventricular folds compresses into the cephalad surface of the vocal folds; this impedes vocal fold vibration</td>
</tr>
<tr>
<td>(vocal-ventricular fold contact)</td>
<td></td>
</tr>
<tr>
<td>glottal stop is not just glottal</td>
<td>glottal stop involves varying degrees of ventricular incursion to arrest or prevent vibratory motion of the vocal folds</td>
</tr>
<tr>
<td>larynx height</td>
<td>aryepiglottic constriction is possible with any larynx height setting, but a raised larynx setting is most natural</td>
</tr>
<tr>
<td>the tongue root and the larynx</td>
<td>retracted vowels (such as [a ɔ]) are most prone to engaging aryepiglottic constriction</td>
</tr>
<tr>
<td>constricted/unconstricted</td>
<td>narrowing of the larynx constitutes a basis of phonological contrast, and it is used in producing ATR and vocal register</td>
</tr>
</tbody>
</table>
2.2.4 Beyond Esling and colleagues

The work of Esling and colleagues has received both support and criticism from various sources. Work associated with the three topics of particular importance for this dissertation (epilarynx vibration, the relationship between the vocal folds and the epilarynx, and the relationship between the epilarynx and the rest of the vocal tract) are reviewed below.

Vibration/trilling of the aryepiglottic folds is perhaps the least explored mechanism, but there is acoustic and electroglottographic research on Khoisan languages confirming that “epiglottalization” of vowels (Miller, Brugman, Sands, et al., 2009; Miller, 2007) is a real linguistic phenomenon, as Traill argued persuasively for !Xóõ (1986). To date, however, no investigation has evaluated the distribution of trilling or growling in speech and related processes, although several studies suggest it is prevalent in popular and ethnic vocal stylization in performance contexts (Borch, Sundberg, Lindestad, & Thalén, 2004; Eckers, Hütz, Kob, et al., 2009; Tsai, Wang, Wang, et al., 2010; Crevier-Buchman, Pillot-Loiseau, Rialland, et al., 2012).

Both the LAM and the Valves Model posit an interaction between the vocal folds and the ventricular folds involving the descent of the latter upon the former and a presumed influence to the vibratory behaviour of the vocal folds, i.e. the effect Edmondson & Esling (2006) call ventricular incursion. There is considerable evidence that ventricular incursion plays a role in the production of glottal stop and glottalization (as in glottal reinforcement or in the glottalized resonant series found in many languages of the Pacific Northwest), laryngealization (creaky phonation), “tense” or “harsh” vocal
registers, pharyngeals and pharyngealization, and possibly in ATR phenomena. Yet matters of its function and distribution in speech still require research.

Work by Edmondson, Chang, Hsieh and Huang (2011) shows that glottal reinforcement of word-final stops in Taiwanese and Hakka involves ventricular incursion in addition to vocal fold adduction. In Taiwanese, however, only the older participants showed signs of ventricular incursion; no discernible ventricular incursion occurred for their younger participants (although it is not possible to judge in laryngoscopy whether changes to the vertical relationship between the vocal and ventricular folds occurred for these speakers). The details of the secondary epilaryngeal contribution to glottal reinforcement also varied by participant. For one of the elder Taiwanese participants, ventricular adduction was mainly anterior; for the other, there was engagement at both the ventricular and aryepiglottic levels. In the case of the Hakka participants, one showed anterior ventricular medialization accompanying glottal closure, while the other used full adduction of the ventricular folds (involving contact along their entire longitudinal extent). Their data show that, in these languages, glottalization of final stops involves specific strategies that vary by speaker (and possibly by generation in Taiwanese): to the basic glottal adduction one can add ventricular incursion and, subsequently, full posteroanterior narrowing of the epilarynx. Edmondson et al. (2011) suggest that the involvement of the ventricular folds in glottal reinforcement is to facilitate stop production by inhibiting vocal fold vibration through bracing. They suggest that the use of this mechanism is not surprising since vocal-fold geometry makes the vocal folds more inclined to open and possibly vibrate when subjected to increased subglottal air pressure (see §2.1.2, especially Figure 2.7).
Laryngoscopic observations reported by Brunelle et al. (2010) of tone “glottalization” in Northern Vietnamese provides evidence that ventricular incursion and other supraglottal gestures of the larynx accompanying the glottalized tones may be a matter of individual variation (the “laryngealized” [p. 148] tones they study are as follows: tone B2 is mid falling, tone C1 is low falling, and C2 is rising and interrupted). They report that nearly half of their participants (4 of 10) visibly use ventricular incursion for tone B2 (several variations are described ranging from anterior ventricular fold contact, complete contact, concomitant posteroanterior shortening of the epilarynx, and tilting of the epiglottis [such that there is anterior movement of the suprahypoid part and posterior movement of the infrahyoid part]). Speakers show split behaviour for production of tone C1: some use constricted states involving features like epiglottis tilting and ventricular incursion\(^{51}\), while others use more unconstricted laryngeal states and light breathiness. Finally, they report that most speakers do not use any ventricular incursion for tone C2; of the two speakers that do, irregular vibrations are said to occur prior to the onset of ventricular incursion. This leads them to suggest that partial ventricular incursion does not necessarily entail perturbation to phonation.

Unfortunately, as the authors concede (Brunelle, Nguyễn, & Nguyễn, 2010, p. 156), the laryngoscopic data they present are not of high quality – having high image noise and being low in contrast, giving a poor registration of the laryngeal structures – and there is no illustration of the absence of ventricular incursion on the glottalized tones for comparison. Nonetheless, they use this data to directly assess the vocal-ventricular fold contact hypothesis and conclude the “extreme” cases are instances of optional and

\(^{51}\) An extreme ventricular incursion is noted to result in large glottal jitter and drop in F0 (but with continuous vocal fold vibration).
individual variation. Moreover, they conclude that there is no systematic use of ventricular incursion (or other epilaryngeal mechanisms) in Northern Vietnamese tone, although glottal constriction (at the vocal folds) is always present.

Another contentious matter concerns the relationship between the epilarynx and the rest of the (supralaryngeal) vocal tract. In the Valves Model (Edmondson & Esling, 2006), there is a suggestion of dependencies in the operation of the valves, but the exact details are not given. It is suggested, for example, that engagement of Valve 2 (the ventricular folds) is dependent on engagement of Valve 1 (the vocal folds) but that Valve 2 can optionally engage with Valve 3 (the aryepiglottic folds). In fact, their model has been interpreted by others as suggesting a hierarchical organization amongst the valves which tends to require the spatially-lower valves to be engaged before the spatially-higher valves (e.g. see Brunelle, Nguyễn, & Nguyễn, 2010, pp. 149–150): for example, Valve 4 (linguo-epiglottopharyngeal stricture) can only fully engage once Valve 3 (aryepiglottophiglottal stricture) has engaged. The valve conception may overemphasize the independence of the mechanisms of the laryngeal vocal tract, and it is not certain whether the valve metaphor applies equally well to all of the mechanisms that Edmondson & Esling (2006) discuss (e.g. larynx raising [Valve 5] and pharyngeal narrowing [Valve 6] are anatomically and physiologically quite distinct from the vocal, ventricular, and aryepiglottic folds [Valves 1, 2, and 3]). Nor is it clear that the valves are even entirely distinct entities in all cases: for example, Painter points out that larynx height and pharyngeal narrowing are clearly related through the palatopharyngeus
muscles\textsuperscript{52}, which, in helping to lift the larynx (to help raise pitch), begin to adduct (Painter, 1986, pp. 331–332, 1991; Zemlin, 1998, p. 270). The oblique orientation of the lower and middle pharyngeal constrictor muscles helps to raise the pharynx as it constricts; by physiological chain-linkage, this causes the larynx to rise as well (Zemlin, 1998, pp. 275–277)\textsuperscript{53}.

Entangled in this debate is the function of the epiglottis and the tongue root. Laufer & Condax (1979, 1981) argued that the epiglottis is an active articulator independent of the tongue, which is typically thought to push the epiglottis during retraction. Their observation that the epiglottis appears to tilt on its own accord (in a way similar to that observed in Brunelle, Nguyên, & Nguyên, 2010; also see Heselwood & Al-Tamimi, 2006; Heselwood, 2007, p. 2) seems to support the notion that it can move independently of the tongue in speech (although pharyngeals can also be realized with enough tongue retraction such that its sides come into contact with the pharynx). Esling, however, repeatedly asserts the aryepiglottic folds are active and any action of the tongue and attached epiglottis is of secondary importance (1996, 1999, 2005; also see Edmondson & Esling, 2006). Nonetheless, pharyngeal articulations, which (very likely always) involve the aryepiglottic constriction described in this work, are also very often observed to have epiglottal retraction sufficient enough for epiglotto-pharyngeal contact, and tongue retraction is very often a concomitant of this. The posture used for pharyngeal consonant production may also have the (seemingly) paradoxical effect of biasing a

\textsuperscript{52} Painter (1986, pp. 331–332) points out that the lateral pharyngoepiglottic folds shorten as pitch increases, which is also likely attributable to the palatopharyngeus muscles.

\textsuperscript{53} The stylopharyngeus muscle can also transitively raise the larynx (i.e. by lifting the pharynx), but, because these muscles have their origin on the styloid process of the temporal bone, they dilate the pharynx rather than narrow it.
somewhat palatal stricture of the tongue, which, as Catford remarks (1983), possibly gives it a rhotic quality.

Despite these unresolved issues, the conception of pharyngeal/epiglottal articulation outlined in the Laryngeal Articulator Model and the valves of the throat models has assisted recent research in more deeply understanding pharyngeal sounds, as in Arabic. For example, Heselwood (2007) argues for a “tight approximant” variant of the Arabic ‘ayn based on acoustic and laryngographic (i.e. electroglottographic [EGG]) evidence of speakers from eleven different countries\(^{54}\). The production of this ‘ayn variant is suggested to involve extreme laryngeal constriction with concomitant vocal-ventricular fold compaction (although no direct evidence supports this). A key acoustic feature of these sounds is a band-pass effect centered around \(1000\) Hz (and spanning roughly \(1000\) Hz) – the locus of the converging first and second formants caused by pharyngeal narrowing (as in \([a]\), but more extreme). There is superimposition of this acoustic effect over the shift in formant frequencies associated with the retracting lingual and narrowing pharyngeal configurations. In order to explain the articulation responsible for the band-pass effect, Heselwood appeals to Esling’s model of laryngeal articulation and suggests that the tight approximant must involve considerable narrowing and compaction of the structures of the epilarynx, along with narrowing of the pharynx. X-ray imaging of this ventricular compaction during geminate ‘ayn in Qatari Arabic (Bukshaish, 1985, pp. 290, 305; images reproduced in Heselwood, 2007, p. 4) shows this is indeed a possibility. Heselwood suggests the obliteration of the ventricle causes a

\(^{54}\) Laryngoscopic data showing /a/ context pharyngeal consonants /h Ω/ in Jordanian Arabic helps support Heselwood’s analysis (see 2007, p. 2, fig. 1): in these data, the epiglottis is retracted to the extent that it makes contact with posterior pharyngeal wall but there is a gap between the tongue and epiglottis, exposing the valleculae; the data do not provide a clear visualization of what is happening below the epiglottis.
loss of its low-pass filtering effect (the ventricles are analyzed as Helmholtz resonators by Pepinsky, 1942, p. 32; also see van den Berg, 1955; citation from Heselwood, 2007, p. 12) and suggests pharyngeal narrowing as the cause of attenuation of frequencies above the pass-band (a characteristic of the acoustic relationship between vowels and homorganic approximants in general). Heselwood also points out that auditory impression of the sound is that of a stop, which is remarkable since there is continuous acoustic energy during the sound.

2.3 Nomenclature and taxonomy of lower vocal tract sounds categories

This dissertation deals extensively with sounds that, as part of their production, involve constriction somewhere in the posterior region of the vocal tract. This region is bounded inferiorly by the glottis and superiorly by the oropharyngeal isthmus and velopharyngeal port. Thus, the domain encompasses the laryngeal and pharyngeal parts of the vocal tract, which will often be referred to collectively throughout the dissertation as the lower vocal tract.

The tradition of phonetic description comes with a set of labels that are part of the taxonomic classification of speech sounds produced with a constriction somewhere in the lower vocal tract. There are labels which are associated with the pharynx, such as pharyngeal and pharyngealization, labels that are associated with the larynx, such as glottal, laryngeal, glottalization, laryngealization, laryngeal tension (e.g. Butcher & Ahmad, 1987) and laryngeal constriction, and there are labels which fall somewhere in between, such as epiglottal, epiglottalization, retracted, and retracted tongue root. There is also uvular and uvularization, which denote constriction in the upper pharynx (near the
uvula and possibly involving its vibration). In more phonologically-oriented literature, one will find the cover terms *guttural* and *post-velar* to group all of these types of sounds (and suggesting their union as a natural class in phonological analysis, see Chapter 6).

One difficulty in the use of these labels is that, unlike those used for articulations of the anterior vocal tract, such as *palatalization* and *labialization*, terms such as *glottalization*, *laryngealization*, *epiglottalization*, and *pharyngealization* intersect with the taxonomic description of phonatory quality. Phonatory changes (i.e. the generation of audible acoustic energy by the larynx) correlate with many of the states implied by these labels, but the labels may also imply non-phonatory events or nil-phonation, *sensu* Laver (1994, pp. 187–189), such as various types of laryngeal stops (most often glottal stop): for example, Painter claims that pressed phonation can occur with *glottalization*, *laryngealization*, and *pharyngealization* (1986, p. 330), which implies that *glottalization* can involve stop or vibratory events. Roach (1979), however, applies *glottalization* strictly to stop events (at all functional levels within the larynx, i.e. in the vocal, ventricular, and aryepiglottic planes). In the work of Gauffin (1972, p. 5), *laryngealization* is defined as a general term encompassing states of laryngeal vibration, such as tense or creaky voice, and extreme laryngealization at the point when vibration ceases is defined as glottal stop. The terms *glottalization* and *laryngealization* are used interchangeably by Dilley, Shattuck-Huffnagel, and Ostendorf (1996) and signify both vibratory and stop events at the larynx, but *glottalization* takes more general scope. Similarly, Michaud (2004, p. 120) uses *glottalization* as a cover term to encompass both laryngeal stop and vibratory events, and this usage is adopted by Brunelle, Nguyẽn, & Nguyẽn (2010, p. 147). Esling (1999) and Edmondson & Esling (2006, pp. 169–171; also
see Esling & Harris, 2003b, 2005), like Roach (1979), strictly reserve glottalization to indicate laryngeal stop events and apply laryngealization to phonatory events (either involving creakiness or harshness, see below for more discussion). In the Salishanist and Wakashanist tradition, the term glottalized resonant (e.g. Kinkade, 1967, p. 228; van Eijk, 1997, p. 4) is typically the term used to refer to oral approximant articulation with concomitant glottal stop or creakiness\(^{55}\), but occasionally laryngealized resonant is encountered (Bird, 2011; Thompson & Thompson, 1992, p. 4).

In terms of accuracy and transparency, the labels mentioned above may be problematic and lead to misconceptualization of the structural configurations they imply. For example, the labels epiglottal and epiglottalization may place too much emphasis on the role of the epiglottis itself in the articulation of the sounds denoted by the IPA symbols [ʔ ʕ n]. These labels do not unambiguously convey the nature of the constriction involved (similar arguments are made in Traill, 1986, p. 129, f.n. 2): it is unclear if they denote aryepiglotticoepiglottal or epiglottopharyngeal stricture. Both uses are found in the literature, as illustrated by Ladefoged & Maddieson (1996, p. 13), who claim that epiglottal corresponds to epiglottopharyngeal stricture (as in Delattre, 1971), not aryepiglotticoepiglottal stricture\(^{56}\). Terms such as pharyngeal and pharyngealization may be too vague: the pharynx is a large region. Some attempt has been made to assert that generalized pharyngeal narrowing or expansion do in fact have linguistic functions (Lindau, 1975, 1978; Ladefoged & Maddieson, 1996, pp. 306–313); on the other hand,

---

\(^{55}\) A glottalized resonant “/R’/”, such as /w’m/, is realized as [ʍ], [ʍʰ], [ʕʍ], and so forth.

\(^{56}\) Although aryepiglotticoepiglottal stricture is a possible referent of the label epiglottal following the active-passive convention of naming place of articulation categories (e.g. lamino-alveolar or apico-alveolar become simplified as alveolar) and assuming the aryepiglottic folds are active and the epiglottis is passive (as in Esling, 1999, p. 364). Furthermore, with linguo-epiglottopharyngeal stricture, the term pharyngeal implies the posterior pharyngeal wall is the passive articulator using the active-passive nomenclature.
many have suggested regional terms for pharyngeal constriction, most common of which are *upper pharyngeal*, for uvular-type constrictions, and *lower pharyngeal*, for pharyngeal-type constrictions (Catford, 1983, p. 346; Czaykowska-Higgins, 1987; Hess, 1998).

Also important for the present purposes is consideration of the tradition of labeling of different qualities of sound produced by manipulation of the way the vocal folds vibrate (more generally, the way the larynx vibrates), and manipulation of global resonance quality caused by change in larynx height. In regard to phonation, many of the terms imply non-default status, i.e. they are distinguishable (in some sense) from *modal phonation*, which is a baseline or default phonation mode. This too, however, is an abstraction or idealization, and it is not a simple matter to develop an operational definition of modal phonation. It is nevertheless a very useful term and one that is very frequently used: it enables the phonetician to differentiate specialized vibratory modes, or types of *non-modal phonation*, that can be linguistically distinctive in language (Gordon & Ladefoged, 2001). Gerratt & Kreiman (2001) provide an interdisciplinary evaluation of the taxonomic description of non-modal phonation. They evaluate three categories of non-modal phonation: (1) breathiness (alternatively *slack voice, murmur*); (2) creakiness (alternatively *creak, creaky voice, vocal fry, pulse register, laryngealization, glottalization*; 2001, p. 377)\(^{57}\), and (3) harshness (alternatively or including *tense* and *pressed voice, growling, laryngealization*). Their conclusion, based on the perceptual evidence they present, is that the modal vs. non-modal dichotomy is misleading; rather, it

\(^{57}\) Esling (1999, p. 354) and Edmondson & Esling (2006, p. 171) note that creakiness and harshness are closely related categories and are often collapsed into a single category in the phonetic literature (usually laryngealization).
seems that people perceptually differentiate modal phonation from two different types of non-modal qualities. One non-modal quality groups creakiness and harshness (which themselves are perceptually distinct from each other) and differs categorically from modal phonation; the other is breathiness, which differs along a phonatory continuum from modal quality.

Two terms for larynx height setting, which will be useful for the dissertation but also show overlap with both place of articulation terms and phonation terms, are *raised larynx* and *lowered larynx* voice qualities (Sundberg & Nordström, 1976; Laver, 1980; Nolan, 1983, pp. 182–187). Esling (1999, p. 367) connects the states implied by *raised larynx voice* with *pharyngealization* as equivalent in articulation (both involving a raised larynx and laryngeal constriction with strong tongue retraction) but differing in glottal pitch: the former being high in pitch and the latter being low in pitch. Similarly, Edmondson & Esling (2006, p. 172) claim a parallel relationship holds between the states implied by *lowered larynx voice* and *fauicalization/fauicalized voice* (such that both states involve strong larynx lowering and a tendency towards partial vocal fold abduction) except the former generally occurs with low pitch and the latter with high pitch. Thus, both of the terms for larynx-height relate to phonation because of their connotation of a tendency towards a particular pitch level. Furthermore, the partner terms *pharyngealization* and *fauicalization* indicate that changes to more than just the elevation of the larynx occur, such as narrowing or stretching of the pharynx and associated structures (i.e. the faucal pillars [palatoglossal and palatopharyngeal arches]; see Laver, 1980, pp. 56, 72).
Based on the above discussion, one should exercise caution in the interpretation of phonetic and phonological nomenclature used to describe lower vocal tract sounds. Many of the terms have an anatomical basis, but this basis may not adequately convey what really occurs in the production of these sounds. Most of the reviewed terms denote sounds (or components of sounds in the case of terms for secondary articulations) which exhibit both localized and distributed changes to vocal tract configuration and concomitant changes to phonatory quality. The purpose of the above survey highlights issues associated with terminology which is used extensively in this dissertation. Wherever necessary, terminological matters are addressed in detail, and, when appropriate, new terminology is introduced where previous terms fail to adequately convey new understanding derived from phonetic and phonological considerations. (Definitions of important terms used as they are used in this dissertation are in Appendix B.)

2.4 Chapter summary: Epilaryngeal preliminaries

This chapter has defined the epilarynx, reviewed the history of ideas surrounding its phonetic existence and use in speech, and laid out some of the terminological issues in describing speech contributions of the lower vocal tract.

In summary, the epilarynx can be thought of as a tube nested within the enclosing pharyngeal tube. The epilarynx comprises the supraglottal laryngeal structures, most significantly the ventricular folds, aryepiglottic folds, and epiglottis, which together form two important epilaryngeal spaces, the laryngeal ventricle and the vestibule. Anatomically and physiologically the epilarynx has two levels, one in the ventricular
plane and one in the aryepiglottic plane, and it appears that the larynx is well-endowed with intrinsic musculature for producing constriction in each of these planes.

Early research posited a role for the epilarynx in forming laryngeal closure, but, possibly because of difficulties in imaging this region of the vocal tract, these ideas did not resurface until Esling and colleagues provided substantial evidence for active participation of the “aryepiglottic sphincter mechanism” in the production of pharyngeal consonants. Despite this evidence, much remains to be discovered about epilaryngeal function in speech and related contexts, and, consequently, the speech-significance of the mechanism is not without controversy.

The remainder of this dissertation is devoted to expanding upon what is known about the speech-functioning of the epilarynx (and related matters) through imaging, modeling, and theoretical considerations. The goal is to establish a theory of the phonetic and phonological role of the epilarynx. This theory has three parts related to epilaryngeal function, which have direct reflexes in the phonological domain. These are: epilaryngeal vibration (Chapter 3; Chapter 7, §7.1), the relationship between the epilarynx and the vocal folds (Chapter 4; Chapter 7, §7.2), and the relationship between the epilarynx and the supralaryngeal vocal tract (Chapter 5; Chapter 7, §7.3).
Chapter 3

EPILARYNGEAL VIBRATION

This chapter addresses the source-related capabilities of the epilarynx from anatomo-physiological, anthropogenic, and phonetic perspectives. A major theme of this chapter is the complexity of the mechanism of epilarynx vibration. Despite this complexity, however, it is argued that regardless of what part of the epilarynx is vibrating, or how it is vibrating, it always generates a harsh, “growly” auditory quality.

The discussion starts with some theoretical considerations of the phonetic nature of epilarynx vibration as phonation-like and trilling-like and a survey of what is currently known about epilarynx vibration, which includes a system of classification of epilaryngeal aperture and vibration (§3.1). The next section (§3.2) is a study of high-speed laryngoscopy video of voiced and voiceless aryepiglottic trilling variants of Iraqi Arabic pharyngeal consonants. Finally, §3.3 presents a simulation of aryepiglottic vibration using a lumped element model and compares this to existing models of simultaneous vocal and ventricular fold vibration. A chapter summary is provided in §3.4. The issue of epilaryngeal vibration in phonological contexts will be addressed in §7.1.

3.1 The epilarynx as a source mechanism

Numerous structures within the vocal tract can engage in self-sustaining oscillation. This is true for the lips, tongue tip, and uvula, and each structure underlies IPA designations for corresponding phonetically relevant “trilling” sounds produced by
their oscillation, i.e. [br̩], respectively. Other mechanisms include the buccal mucosa$^{58}$, the soft palate proper, which oscillates during snoring, and the upper esophageal sphincter$^{59}$ (Singer, 1988), which can be used as an alternative to a glottal source to compensate for certain voice pathologies. None of these mechanisms receive IPA symbolization as they are not attested in normal, non-pathological speech. The epilarynx is an interesting case because it turns out that every part of it can vibrate: the ventricular folds, aryepiglottic folds, and epiglottis can all engage in self-sustained vibration, and vibration of the vestibular mucosa and even the arytenoid cartilage complex are also possible; the IPA representation is limited to (voiceless) [h] and (voiced) [ɬ], which conventionally denote epiglottal vibration, but the use in this dissertation will be expanded to any kind of epilaryngeal vibration, which is in the spirit of Esling (1996).

The epilarynx and esophageal sphincter are distinct from the other mechanisms because of their low position within the vocal tract; the significance being that, like the vocal folds, they are predisposed to function as vocal tract sources, i.e. they can excite all of the essential resonances of the vocal tract for producing intelligible speech: they can generate a source while allowing the other articulators to be relatively free to carry out their ordinary functions. The esophageal sphincter, not being employed in natural language is of less phonetic relevance than the epilarynx, which has a surprisingly large number of phonetic, phonological, and paralinguistic functions. Unlike the vocal folds, however, epilarynx vibration cannot operate efficiently as a source mechanism without substantial alteration to the neutral vocal tract configuration: usually narrow epilarynx

$^{58}$ It can be used in producing the voice of Donald Duck (although a “tight” voiceless uvular trill also suffices).
$^{59}$ Also known as the crico-esophageal sphincter.
stricture is required. Consequently the configuration of the rest of the vocal tract will likely become compromised, as will be discussed further in Chapter 5.

3.1.1 *Phonetic classification of epilarynx vibration*

It was noted in the previous section that many vocal tract structures can engage in self-sustaining oscillation\(^60\). The vocal folds also exhibit this property, which is fundamental to voice-source production. The phonetic function of this vibratory behaviour is not the same in all cases, however, and typically phoneticians classify these vibratory behaviours as either trilling or phonation.

The standard view is that the term *trilling* is applicable to the lips, tongue tip, and uvula and the term *phonation* properly describes the phonetic function of vocal fold vibration. Given the location of the epilarynx in the vocal tract – very low within the pharynx and immediately above the vocal folds – we might wonder whether it should be phonetically regarded as a phonatory mechanism or a trilling mechanism. From an aero-acoustic and mechanical standpoint, the difference between the terms is irrelevant: there is quasi-periodic (or possibly irregular) vibration that leads to an unsteady, pulsatile airflow capable of acoustically exciting nearby resonators. These mechanisms are thus all acoustic *sources*. The difference in terminology does, however, matter from a phonetic and ultimately phonological standpoint: trilling mechanisms have highly restricted use in speech, mainly fulfilling a consonantal role; on the other hand, the vocal fold phonatory

\(^{60}\) The body will continue to oscillate, overcoming its internal resistance, provided there is continuous energy input in the form of aerodynamic driving forces.
mechanism is used pervasively in speech: it provides the acoustic energy for vowels and serves as the basis for voicing contrast.

Epilaryngeal vibration is somewhat anomalous in the context of phonetic theory, since it does not clearly classify strictly as phonation or trill. Since awareness of epilaryngeal vibration in phonetic and phonological contexts has not yet penetrated conventional theory, (to my knowledge) no one has ever formally stated an operational definition to classify vocal tract vibrations as phonation or trill\textsuperscript{61}. This lack of definition in what makes trilling trilling and what makes phonation phonation reflects the lack of a need to distinguish clearly between these categories, up until now. There is no confusion when one talks about vocal fold vibration as phonation\textsuperscript{62} on one hand and, say, tongue tip vibration as trilling on the other. The epilarynx, however, blurs the boundaries between phonation and trilling, both phonetically and phonologically (see §7.1.3), and it will therefore be useful to develop an operational definition.

To decide how to interpret epilarynx vibration, then, let us consider the following five proposed criteria to differentiate between trilling and phonation behaviours: (1)
frequency control, (2) mode control, (3) vocal tract location, (4) simultaneity, and (5) articulatory interference.

Criterion (1) concerns frequency (or pitch) control. The vocal folds, which are said to phonate, are unlike any other self-sustaining oscillator in the vocal tract in regard to the range of frequencies at which they can be made to vibrate, and this control is important in speech in obvious ways, such as in tone or intonation. In contrast, we have fairly good control of the frequency of bilabial vibration, but this pales in comparison to the capabilities of the vocal folds and has not been demonstrated to be linguistically relevant. Moreover, it is much harder to control the rate of uvular vibration, and this is true for the epilarynx as well. Thus, phonatory-like structures have excellent and linguistically relevant frequency control; trilling-like structures do not.

Criterion (2) is quality manipulability or vibratory mode control: again, the vocal folds are extremely flexible in regard to the different types of vibratory patterns that can be produced (modal, breathy, creaky, and so forth) and again, these are linguistically relevant. The lips also exhibit some flexibility, and again, none of which is phonetically relevant. The other mechanisms, including the epilarynx, are further restricted in the distinct qualities they can produce. So we can say that good vibratory mode control that is linguistically relevant makes for a phonation-like mechanism.

Criterion (3) concerns the location of the oscillatory mechanism for the purpose of vocal tract excitation: how much of the vocal tract proper is below/behind the mechanism and how much is above/ahead of it. In the case of the vocal folds, the entire vocal tract lies above; in the case of lips, the vocal tract is entirely behind. This factor has significant consequences for how effectively these mechanisms can excite the resonances of the
vocal tract. The epilarynx is comparable to the vocal folds in this regard, but the other mechanisms are primarily oral-cavity exciters. Furthermore, the presence/absence of vocal fold vibration (and its timing) is unique among vibrating mechanisms in being responsible for the speech-fundamental voiced/voiceless distinction. No other vibration in non-pathological speech contributes such a possibility, and it is even possible to have voiced and voiceless epilaryngeal vibration. Thus, being “low down” in the vocal tract, makes a structure more phonatory-like in its phonetic nature.

Criterion (4) also concerns the location of oscillation, but what matters now is whether additional oscillators can engage upstream of the mechanism. The significance of this has to do with the fact that downstream oscillators will tend to entrain\(^ {63}\) aero-acoustically with upstream ones; upstream oscillators also determine the amount of available aerodynamic driving force available for downstream ones. The former will influence how a mechanism will oscillate, and the latter will influence whether a mechanism will oscillate. The vocal folds can aero-acoustically influence all other oscillating structures in the vocal tract (with the exception of the upper esophageal sphincter), which is critical for aerodynamic conditions involved in voiced/voiceless trills. Glottal flow serves as a limiting factor for downstream oscillators: insufficient glottal flow means there will be insufficient pressure build-up to surpass the vibratory threshold pressure to engage oscillation. The degree of aero-acoustic coupling, which will

\(^{63}\) Entrainment is the tendency for independent but coupled oscillators to exert mutual influence on each other’s frequency and phase of oscillation. Technically speaking, Izhikevich & Kuramoto (2006) reserve the term entrainment for 1:1:…. frequency locking, which they take to be the most general term for “locking of oscillators”. In the literature at large, however, entrainment is used as the general term although some sources simply use synchronization (Guevara & Glass, 1982; Aronson, McGehee, Kevrekidis, & Aris, 1986; Schuster & Wagner, 1989; Heath, 1998; Bennett, Schatz, Rockwood, & Wiesenfeld, 2002; Clayton, 2012; Giver, Jabeen, & Chakraborty, 2011).
determine the degree of entrainment, is a function of the distance from the vocal folds\textsuperscript{64}. There is also the possibility for mutual influence, which means that a downstream oscillator can influence the dynamics of an upstream oscillator, such would be the case if tongue tip vibration had any non-negligible impact on vocal fold dynamics. The magnitude of the effect is likely strongly related to the distance between the oscillators. Unlike other vibrating structures, the epilarynx through its own vibration is most predisposed to influence vocal fold vibration, an effect which “phonetically couples” epilaryngeal vibration to the voice source in that this property of vocal fold perturbation helps to make epilaryngeal vibration audible and thereby suitable as a phonatory mechanism. No other vibrating mechanism exhibits this coupling property in any phonetically significant way\textsuperscript{65}.

Criterion (5), articulatory interference, relates to how much the vocal tract configuration is compromised by making the necessary adjustments to produce the vibration, and consequently, once vibration is engaged, what other concomitant articulations are possible. Phonatory action is less compromising to articulation while trilling actions are more compromising. In tongue tip and uvular trills, the tongue body is preoccupied by forming the trills, so articulating vowels becomes compromised. In

\textsuperscript{64} The reason being that pulsatile flow produced by the upstream oscillator which drives the downstream oscillator is subject to a time delay as it travels along the length of the vocal tract. For any system of coupled oscillators, the longer the time delay in the coupling, the more possible synchronization frequencies there are (Schuster & Wagner, 1989). This means that the oscillators have more “freedom” to vibrate at their own natural frequencies and exhibit less accommodation in phasing. Furthermore, since the pulsatile structure of the airflow coming from the upstream oscillator will dissipate as the air moves along the vocal tract, the coupling strength will also diminish (and effectively becomes a DC source). Apart from aero-mechanical coupling, even purely mechanical vibrations of upstream oscillators will also be able to transfer more energy to downstream oscillators that are closer because, again, the vibrations of the upstream will be dissipated by tissue damping.

\textsuperscript{65} Note that this criterion focuses on aerodynamic coupling between two oscillators separated in space by some distance. Other forms of interaction, such as the aerodynamic voicing constraint can occur between the vocal folds (as oscillators) and downstream constrictions, but such effects are of a different nature than that focused upon in this criterion.
bilabial trills, the tongue is free, but any vowel produced will be obscured by the narrow stricture at the lips. The vocal folds do not greatly compromise other articulatory possibilities, being highly independent in their action, which makes them canonically phonatory-like. The epilaryngeal vibration is more phonatory-like in this regard since it is possible to produce any vowel with epilaryngeal vibration. It may be that certain types of epilaryngeal vibration involve greater compromise (see §3.1.3), making these more trilling-like by this criterion.

Using these criteria, epilarynx vibration classifies as a quasi-phonatory (see Esling, 1996, p. 82), quasi-trilling mechanism. According to criteria (1), (2), and (4) it appears to share more in common with other structures said to “trill”; according to criteria (3) and (5) epilaryngeal vibration seems more like vocal fold vibration in being implicated in “phonation”. (The phonological significance of the phonation-like vs. trilling-like distinction will be considered in §7.1.3). Perhaps the most important property that lends to its near phonatory nature is its position deep in the vocal tract immediately above the vocal folds, criteria (3). Whether it ever engages during the production of oral trills (e.g. [ɾ]) in language is a doubtful, but open question, even though this is entirely possible from a physiological stand point. The limitations on pitch and quality are severe enough to limit its phonetic uses in comparison to the vocal folds. This does not mean, however, that it is rarely used in the wider context of speech and vocalization. In fact, as the next section indicates, examples abound.
3.1.2 A survey of paralinguistic and linguistic uses of epilarynx vibration

It is the impression of the author that epilarynx vibration is in fact quite common, but it is clearly a function of particular circumstances. It commonly occurs in shouting and “aroused” speech (Sakakibara, Fuks, Imagawa, & Tayama, 2004) or speech expressing “excessive emotions” (Tsai, Wang, Wang, et al., 2010). One need only visit the local soccer field to hear near continuous engagement of epilarynx vibration of one form or another as players tussle, toss, and bellow about in the hurly-burly of the sport: it could be a function of the amount of machismo on display (Ohala, 1996). Children employ the epilarynx frequently in rambunctious play or if sound effects are being produced, like when they simulate the roar of a monster or dinosaur. Excited or emphatic natural speech provides ample examples of sporadic use of epilarynx vibration, but performance, voice acting, and singing are perhaps the most fertile sources for examples of continuous epilaryngeal vibration. The voices of Jim Henson and Frank Oz provide copious examples: Cookie Monster, Miss Piggy, Animal, and Star Wars’ Yoda are some favourites. Bill Cosby’s Fat Albert uses a whispery voiced growl. Frank Welker of 80s children’s cartoon fame contributed many memorable examples of growling as well, such as Inspector Gadget’s Dr. Claw or Megatron from the Transformers cartoon. A subtype of villains in Japanese Anime, a style of cartoon with international popularity, features epilarynx vibration in addition to a suite of other articulatory properties (such as raised larynx, jaw and lip protrusion, jaw lowering, and lingual retraction) which correlate with negative character traits, such as ugliness and disloyalty (Teshigawara, 2003). A scene from Inuit film, Atanarjuat: The Fast Runner (2001), shows the confrontation between two shamans, one invoking a polar bear spirit, the other invoking a walrus spirit; distinct
auditory qualities of epilarynx vibration seem to differentiate these two spirits. Moisik (forthcoming) demonstrates how American performers, especially comedians, use epilaryngeal vibration as part of general harsh voice quality in sociophonetic capacity to generate and contrast stereotypically racialized characters.

In singing and ethnomusical contexts, there is a diverse range of styles that employ epilarynx vibration. Jazz, blues, and many early rock and roll musicians frequently exploit epilarynx vibration as part of their stylistic repertoire: Louis Armstrong, Koko Taylor, ‘Scatman’ Crothers, Howlin’ Wolf, Muddy Waters, Little Richard, and Captain Beefheart are just a couple examples (some of which probably indicate voice pathology); emulation of these voices may explain the presence of epilarynx vibration in the singing styles used by Tom Waits, Björk, and numerous other contemporary pop artists with strong blues and jazz roots. Singing in the death metal genre employs conspicuous growling as part of a beastly or “brutal” persona; it is a notable example of persistent engagement of epilarynx vibration (most likely aryepiglottic) (Eckers, Hütz, Kob, et al., 2009). Borch et al. (2004) provide visual evidence that the so-called “dist” (distorted) tones in heavy rock singing (exemplified by AC/DC and Whitesnake) engages semi-irregular vibration of the supraglottic (ventricular region) mucosa. Ethnic throat singing styles and voice games also feature chants or modes which engage simultaneous vocal and ventricular fold vibration: chanting examples include the Kargyraa mode of Tuvan and Mongolian Khöömei/Khöömij ‘throat’ style, and Tibetan Dzo-ke ‘deep chant’ (Lindestad, Södersten, Merker, & Granqvist, 2001; Sakakibara, Imagawa, Konishi, et al., 2001; Esling, 2002a; Imagawa, 2002).

66 My impression is that the polar bear spirit is produced with vocal-aryepiglottic vibration, while the walrus is produced with vocal-ventricular vibration.
Mongolian musical tradition also employs vibration of the aryepiglottic folds, as in the case of *Urtyn Duu* ‘Long Song’ trilling ornamentations (Crevier-Buchman, Pillot-Loiseau, Rialland, et al., 2012). Vocal games featuring epilarynx vibration include Inuit *katajjait* or ‘throat-games’67, Ainu *reikutkar*, and Chukchi *pič eynen* among other variations found amongst Siberian groups, including the Koriak, Even, Evenk, and Tungusic peoples (Nattiez, 1999). Simultaneous vocal and ventricular fold phonation occurs in the *bassu* voice of traditional Sardinian *A Tenore* singing (Henrich, Lortat-Jacob, Castellengo, Bailly, & Pelorson, 2006; Bailly, Henrich, Webb, et al., 2007). Sakakibara et al. (2004) suggest several additional sources of “growl”: Brazil’s carnival lead voices, pop vocalist Elza Soares, country singers Bruno and Marrone, Japan’s Enka (a pop singing style as exemplified by Harumi Miyako), the *Kakegoe* (at the onset of phonation) of Noh percussionists, and South Africa’s Xhosa people use it in *umngqokolo*, a traditional vocal technique. “Growling” is a key feature of the *jing* role in Chinese opera, particularly the *jia-tzi* or openly aggressive variant. The *jing* role actors dress in flamboyant costumes and face paint and adopt aggressive or childish personas (Tsai, Wang, Wang, et al., 2010).

There are a few notable examples in the literature which suggest epilarynx vibration serves a sociolinguistic function in marking group identity or by being a feature of a particular register or regional variety. An example is Harold Paddock’s (1977) all-too-brief comment on Newfoundland’s *roach*, a “low, growly register” spoken by “certain men” around the northeast coast of the province. The use of *growly* here

---

67 Some variants feature raised larynx voice, which is suggestive of the use of vocal-aryepiglottic phonation.
provokes the interpretation that there is epilarynx vibration, but apparently this has never been investigated further. The other case is Zhenhai, the Ningpo variety of Wu Chinese (Rose, 1989). Among other possibilities, phonetic low tone targets can be realized with growling; Rose speculates, partially based off his own imitations of this variant, that the physiological mechanism is likely aryepiglottic vibration. The use of this phonetic variable has apparently been identified and enregistered (sensu Schilling–Estes, 2004) as a stereotype of the Ningpo variety indexing hot-headedness: “better to argue with a person from Suchow than converse with someone from Ningpo” (p. 230). Dinghai Wu, spoken in roughly the same region, apparently shows similar tendency towards growled low tones (Phil Rose, personal communication). These accounts are tantalizing and demand further investigation using imaging techniques such as laryngoscopy.

The Zhenhai and Dinghai examples also serve to illustrate one of the possible linguistic applications of epilarynx vibration: it is a phonetic exponent of tonal contrast. Jianchuan Bai, which will be discussed further below, also exhibits this mechanism in relation to tone: epilarynx vibration can be found on mid-to-low tones in the “tense” or “harsh” register (Edmondson, Esling, Harris, Li, & Lama, 2001; Edmondson & Esling, 2006, p. 176). However, in Bai, it is more clearly part of a phonologically established vocal-register system; in Zhenhai it is a phonetic variant occurring alongside whisper and whispery voice realizations.

It was observed in §3.1.1, that epilaryngeal vibration is quasi-phonatory in nature. This role is illustrated in vocal-register systems which incorporate a phonatory register with epilarynx vibration. The best (and perhaps only) example, comes from the Khoisan-group languages, such as !Xóõ (Traill, 1986), Ju’hoansi (Miller, 2007), and N|uu (Miller,
Brugman, Sands, et al., 2009). Traill provides the most direct evidence for this in the form of laryngoscopic video of single native speaker and published x-rays of his own productions. These data show that the so-called “sphincteric phonation” involves abducted vocal and ventricular folds and vibration primarily of the “arytenoids” (although he probably is referring to the cuneiform tubercles as the observed structures are said to vibrate against the epiglottic tubercle) and vibration of the tip of the epiglottis – if the production is “forceful” (Traill, 1986, p. 125). Traill does not mention the aryepiglottic folds (which are not fully visible in his laryngoscopic images since these are obtained from a deep camera-position within the pharynx), but it is probable that there was also vibration of these structures. Following the taxonomy proposed by Esling & Harris (2005), which posits whisper/whispery voice as vocal fold abduction with narrowing of the epilarynx above (also see Honda, Kitamura, Takemoto, et al., 2010), we can infer that the overall state being described by Traill is one of whisper or whispery voiced epilaryngeal vibration. This is supported by Traill’s observations from visual and acoustic data (1986, pp. 125–128), which indicate abducted vocal fold vibration occurs initially in sphincter vowels, but it is irregular and gradually diminishes towards the medial phase of the vowel. As the vocal fold vibration gives way, all that remains is purely voiceless vibration of the upper epilarynx (Traill emphasizes that the ventricular folds do not participate, which is consistent with the observation that they are abducted just as the vocal folds are). Corresponding with this is the absence of the clear harmonic structure in the spectrum, unlike that which is observed for modal and breathy registers,

These are noted to be comparable in configuration to that observed in (unpublished) poor-quality x-rays of a native speaker’s productions (Traill, 1986, p. 125).
and the presence of considerable noise with a weak and irregular ~50 Hz oscillation, which Traill attributes to the upper epilarynx (1986, p. 125).

Consistent with the dual classification in §3.1.1 of epilaryngeal vibration as quasi-phonatory, quasi-trilling, the epilarynx has also been implicated in trilling-like behaviour. Esling (1999) proposes that trilling occurs as a manner enhancement on pharyngeal consonants, and suggests several sources: in 1934, Jones claimed Somali pharyngeals exhibit “rapid vibration” of epilaryngeal tissue (Jones, 1934, pp. 8–9; Esling, 1999, p. 353): this claim is substantiated by recent laryngoscopic evidence (Edmondson, Padayodi, Hassan, & Esling, 2007), which is illustrated in the next section; Esling convincingly argues that the supposed pharyngeal-epiglottal fricative contrast in Burkikhan Agul (Ladefoged & Maddieson, 1996, pp. 37–38) reduces to a manner contrast, which involves epilarynx vibration for sounds formerly identified as epiglottal fricatives, but only frication for those identified as pharyngeal fricatives (this is discussed further in §7.1.2); paralinguistically, Esling reports that epilarynx vibration enhances pharyngeals /h/ and /ʕ/ in sea-lion related elements of Nuuchahnulth story-telling (1996); Esling also speculates (1999, p. 354) that Butcher and Ahmad’s (1987, p. 166) use of laryngealization to characterize Iraqi Arabic pharyngeals might be a mislabeling in some instances for what is actually epilarynx vibration, or perhaps coarsely represents the phonetic variation characterizing these sounds, which includes epilarynx vibration. Esling’s speculation about this possibility that “laryngealization” in Iraqi Arabic might actually be (on occasion) epilaryngeal vibration has been supported by Edmondson, Padayodi, Hassan & Esling (Edmondson, Padayodi, Hassan, & Esling, 2007, p. 2066).
and will be further substantiated with high-speed laryngoscopic evidence (see §3.2) which shows that epilarynx vibration, in this case aryepiglottic, indeed can occur.

Despite these impressions that epilarynx vibration occurs quite frequently in ordinary speech, the research documenting and describing its occurrence and phonetic nature is still relatively sparse. The next section analyzes the known cases of non-pathological\textsuperscript{69} epilarynx vibration to summarize what we know about what parts of the epilarynx can vibrate.

3.1.3 A tour of the different types of epilaryngeal vibration

The epilarynx exhibits numerous possibilities for localized oscillation\textsuperscript{70}. The literature (much of which was already introduced in §3.1.1 and §3.1.2) documenting cases of non-pathological epilaryngeal vibration is not particularly extensive, but does help to justify the analysis provided here. The following visual evidence is taken into consideration: vocal-ventricular\textsuperscript{71} modes (Lindestad, Södersten, Merker, & Granqvist, 2001; Borch, Sundberg, Lindestad, & Thalén, 2004; Bailly, Henrich, Webb, et al., 2007; Bailly, Henrich, & Pelorson, 2010); vocal-aryepiglottic and voiceless aryepiglottic modes

\textsuperscript{69} See Crevier-Buchman et al. (2012) for an example of pathological occurrence of epilaryngeal vibration associated with surgical intervention requiring removal of the lower larynx and vocal folds.

\textsuperscript{70} The discussion here is primarily the studied opinion of the author based on normal and high-speed (500 fps) laryngoscopic data collected by Dr. John Esling in the Speech Research Lab at the University of Victoria over the past (nearly) twenty years and with collaborative assistance from Dr. Lise Crevier-Buchman’s team at the Laboratoire de Phonétique et Phonologie (UMR 7018, CNRS/Sorbonne-Nouvelle) in Paris. Acknowledgement must also be given to Dr. Ken-Ichi Sakakibara and his colleagues such as Dr. Leonardo Puks, who have made important and fascinating observations regarding epilaryngeal vibration in singing styles.

\textsuperscript{71} By vocal-ventricular it is meant that the vocal folds and ventricular folds are independently and simultaneously vibrating. This does not imply then that there is mechanical coupling between the vocal and ventricular folds. The term vocal-ventricular fold coupling will be used to indicate the different case of the ventricular folds pressing into the vocal folds and mechanically coupling with them. Vibration is possible in this latter case but it will be considered a form of vocal fold vibration (as opposed to vocal-ventricular fold vibration).

Based on the above, it seems that each of the main components of the epilarynx (ventricular folds, aryepiglottic folds, and epiglottis) can vibrate and each has a distinct auditory quality and distribution of uses in normal, non-pathological voice production. Ventricular fold vibration appears to be restricted to a strictly vocal-ventricular mode: ventricular vibration without concomitant vocal fold vibration is impossible by hypothesis, or at least not reported for non-pathological speech\textsuperscript{72}. Claiming this restriction requires the assumption of the fact that ventricular fold adduction is not possible in non-pathological cases without vocal fold adduction. From the literature (reviewed in §3.1.2), we can conclude that this pattern occurs mainly in ethnomusical contexts, most well-known perhaps in throat singing styles, for which there is ample empirical evidence (see §3.1); no evidence has been brought to bear on whether it occurs in linguistic contexts. Aryepiglottic fold vibration is attested in linguistic contexts, however, and it also seems to occur extensively in paralinguistic contexts (as the survey in §3.1.2 demonstrates). Although this remains speculative, it seems to be the case that aryepiglottic vibration is the most widely distributed form of epilarynx vibration. Finally, there is epiglottal vibration (see below), which has only been observed without\textsuperscript{73} vocal fold vibration (i.e. in voiceless pharyngeals, e.g., see below) and with simultaneous

\textsuperscript{72} To constrain scope, epilaryngeal vibrations of the pathological larynx are not considered.

\textsuperscript{73} Whether voiced epiglottal vibration is possible is still an open question, but it is at least physiologically plausible. However, the reduced airflow caused by concomitant vocal fold vibration may make voiceless epiglottal vibration physiologically unfavourable.
aryepiglottic vibration. It is the author’s opinion that epiglottal vibration is the most uncommon form of epilaryngeal vibration (across all contexts, i.e. speech and general vocalization).

Although appealing in simplicity, the three primary mechanisms of epilarynx vibration – vocal-ventricular, aryepiglottic, and epiglottal – require further elaboration. From observations of various individuals, it is clear that individual anatomical variation, variation in production strategies, and the general viscoelasticity of the epilaryngeal mucosa and associated cartilages, make it possible to vibrate nearly any part or localized region of the epilarynx. Thus, each of the three basic types of epilaryngeal vibration is characterized by rather wide range of air-channel topologies and associated vibration patterns. Figure 3.1 provides illustrations of this variation from a superior (laryngoscopic) view; Figure 3.2 provides supplementary illustrations of epilaryngeal vibration regions in the mid-sagittal plane.

Topology of the epilaryngeal air-channel when constricted can potentially become very complex. The exact topology will depend on individual anatomical variation, the forcefulness of epilaryngeal constriction, and whether the vocal folds are adducted or not (although all of the topologies shown can occur regardless of vocal fold state). It should be noted that asymmetry is probably the norm and a given asymmetry in topology can apply to either side (as far as we know).

We will start by considering epilaryngeal aperture topology during constriction as depicted in Figure 3.1, (a) to (f). Forceful constriction results in a “tight” configuration such that the cuneiform tubercles press into the epiglottic tubercle, air flows into the lateral channels formed by the bodies of the aryepiglottic folds, and it is possible for one
(a) or two (b) such channels to form. A third, central channel (c) sometimes forms, especially if the vocal folds are (partially) abducted. If cuneiform positioning is rather more posterior in orientation (attributable to individual anatomical variation), a centrolateral channel can result on one side (d) when the epilarynx constricts; if both cuneiform tubercles fail to make contact with the epiglottic tubercle, then one large, irregularly shaped channel results (e), or a single central channel can form if the constriction is very narrow but the cuneiforms still do not make complete contact with the epiglottic tubercle (f). (Note that the epiglottis can be more or less retracted, and the long-axis epiglottal curvature varies from one individual to the next, both of which will influence the nature of the air-channel of the epilarynx, and its communication with the posterior pharyngeal wall).
Figure 3.1: Variation in epilaryngeal aperture topology and vibration patterns (superior view). ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; p = peri- (and inter-)arytenoidal mucosa; v = vocal fold. Solid gray shaded regions = apertures; diagonal-lined regions = vibration region. Arrows indicate outward (expanding) motion during vibration. Configurations can be more or less symmetric. With the exception vocal-ventricular fold vibration, all patterns can occur with concomitant vocal fold vibration.
Figure 3.2: Variations on epilaryngeal vibration patterns (mid-sagittal view, n.b.: an unconstricted configuration is used to illustrate regions where vibration is occurring, normal epilaryngeal vibration is constricted). ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold. Dash-outlined transparent gray region = vibration region. With the exception vocal-ventricular fold vibration, all patterns can occur with concomitant vocal fold vibration.

We will now consider the different vibrational patterns as depicted in Figure 3.1, (g) to (n), and Figure 3.2, (a) to (g). The ventricular folds (see Figure 3.1, (g) & (h);
Figure 3.2 (a)) primarily show lateromedial displacement during vibration (Lindestad, Södersten, Merker, & Granqvist, 2001, p. 83; Borch, Sundberg, Lindestad, & Thalén, 2004, p. 151; Bailly, Henrich, & Pelorson, 2010, p. 3217), but a “billowing” vibration involving a horse-shoe-shaped region of ventricular, inner aryepiglottic, and epiglottal mucosa (see Figure 3.3) is attested by Esling (2002a) for the Tibetan Dzo-Ke or “deep chant”. This manoeuvre seems to involve the expansion of the ventricle through larynx lowering (evident in the sequence of selected laryngoscopic images in Figure 3.3) to allow air to “billow out” the anterior and lateral mucosa within the epilarynx.

![Figure 3.3: Voiced, “billowing” vibration of the inner aryepiglottic mucosa (dashed outline) in Tibetan Dzo-Ke “deep chant”. (Select frames showing sequence). Data obtained from the University of Victoria laryngoscopy video database (Esling, 2013). Related data appear in Esling (2002a).](image)

As Esling observes (2002a), the Tibetan Dzo-Ke epilaryngeal configuration vaguely resembles the harsh, low-tone in Bai (2006, p. 176; also see Edmondson, Esling, Harris, Li, & Lama, 2001), but the former is the result of carefully practiced manipulation, while we can safely assume the configuration found for the Bai speaker represents a non-deliberate, phonetically natural setting. For the Bai case (Figure 3.4), there is a medial aperture formed by ring-like configuration (dashed line) of the inner mucosa of the aryepiglottic folds and epiglottis and it appears to be this mucosal lining
that vibrates with roughly concentric displacement (see Figure 3.1, (j), and Figure 3.2, (b)). The Dzo-Ke and Bai examples are not quite vocal-ventricular and not quite “canonically” aryepiglottic: they fall somewhere in between.

Figure 3.4: Voiced vibration of the inner aryepiglottic mucosal “ring” (dashed outline) in Bai (/tei⁵.21/ ‘flag’) (Select frames showing sequence; the two leftmost frames show how the larynx looks as it engages the constriction, which is fully achieved in the middle frame). Data obtained from the University of Victoria laryngoscopy video database (Esling, 2013). Related data appear in Edmondson & Esling (2006, p. 176; also see Edmondson, Esling, Harris, Li, & Lama, 2001).

The “canonical” aryepiglottic pattern is characterized by the formation of two lateral air-channels on either side of the laryngeal midline. The displacement involved is nearly perpendicular to that of the vocal folds during vibration, although, as we have already discussed, the topology can be much more complex than this.

The region of maximum tissue displacement in aryepiglottic vibration can vary in terms of vertical depth at which vibration occurs. It seems that vibration situated lower down within the aryepiglottic aperture is possible, and involves the inner aryepiglottic mucosa (Figure 3.1, (i) to (j), and Figure 3.2, (b)). An example is found in Sakakibara et al. (2004): the aryepiglottic apertures are relatively patent and vibration appears to occur
low within the recesses they form (see Figure 3.5); the tissue oscillating is close to the ventricular fold edge, but is clearly not the same as vocal-ventricular vibration (as described above).

Figure 3.5: Lower, inner aryepiglottic mucosa vibration – “growling” (select frames). ae = aryepiglottic fold; c = cuneiform tubercle; et = epiglottic tubercle; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall. Data found in Sakakibara et al. (2004).

Vibration of the upper aryepiglottic body and border is quite common (Figure 3.1, (j) to (k), and Figure 3.2, (c) to (d)). Characteristically unlike vocal-ventricular vibration, a single aryepiglottic aperture exhibits a strong tendency towards variability in period and amplitude, and the two apertures can simultaneously exhibit different phases and frequencies, which suggests they are weakly coupled.

In Moisik, Esling, & Crevier-Buchman (2010), and consistent with the observations of Iraqi Arabic trilling in §3.2, the exact nature of aryepiglottic vibration is clearly a function of laryngeal tension, both at the glottal level and at the epilaryngeal level. They attest that cuneiform orientation has a profound influence on how the aryepiglottic aperture vibrates. Complete contact between the cuneiform and epiglottic

74 Sakakibara et al. (2004) incorrectly label the cuneiforms in their Figure 4 as the arytenoids. This is a common mistake encountered in much of the literature (and in undergraduate speech physiology classes).
tubercles forces the associated aryepiglottic fold to come into contact with the epiglottis; if the cuneiform is oriented such that its apex is pointing away from the epiglottic tubercle, then the aryepiglottic fold is prevented from coming into contact with the epiglottis. An example can be seen in Figure 3.6: the participant’s left-side cuneiform is further away from the epiglottic tubercle and, consequently, the left aryepiglottic fold never completely closes (as in the fourth frame from the left; this same asymmetry actually characterizes the participant in Figure 3.5, and, likewise the left aryepiglottic aperture never closes). It is possible that mucosal adhesion plays a significant role in determining the frequency of oscillation (cf. Ayache, Ouaknine, Dejonkere, Prindere, & Giovanni, 2004), as when there is complete contact between the aryepiglottic fold and epiglottis, the frequency tends to be halved; visible strands of mucous connecting the structures during aperture opening confirm that mucous is present.

Figure 3.6: Voiced aryepiglottic trilling (Select frames). ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold. Data obtained from the University of Victoria laryngoscopy video database (Esling, 2013). Related data appear in Moisik, Esling, & Crevier-Buchman (2010).
Moisik et al. (2010) also describe how tension influences aryepiglottic trilling. When the laryngeal tension is increased, the left cuneiform is brought closer to the epiglottis, and, consequently, both folds come into contact with the epiglottis. Whether or not the vocal folds are adducted clearly influences the tension in the system: adducted vocal folds result in higher aryepiglottic frequency in general.

Also of interest is that more than just the aryepiglottic fold proper can vibrate: the components of the arytenoid cartilage complex can also vibrate (Figure 3.1, (l) to (m), and Figure 3.2, (e) to (f)). In Moisik et al. (2010), during a voiceless aryepiglottic trill production, violent vibration of the cuneiform is also seen – and it has the appearance of a rubber rod being shaken back and forth (see Figure 3.7). (The “applauding” configuration in Figure 3.1m is illustrated in §3.2.2, Figure 3.17).

![Figure 3.7](image.png)

Figure 3.7: Voiceless aryepiglottic trilling with left-side cuneiform oscillation (Select frames). ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; k = corniculate tubercle. Data obtained from the University of Victoria laryngoscopy video database (Esling, 2013). Related data appear in Moisik, Esling, & Crevier-Buchman (2010).

Finally, the epiglottis can exhibit a posteroanterior vibratory mode localized to the upper, free body of the cartilage (Figure 3.1, (n), and Figure 3.2, (g)). Visual evidence of
this from of vibration in Somali, as documented by Edmondson & Esling (2006, p. 183; also see Edmondson, Padayodi, Hassan, & Esling, 2007), is illustrated in Figure 3.8. Edmondson et al. (2007) describe this as voiceless “vibration extending up through the supraglottic tube” (p. 2066), and it is evident that this is not “pure” epiglottal vibration as both aryepiglottic (small arrow) and epiglottal oscillation (large arrow) are evident in Figure 3.8. It is unknown if an isolated epiglottal trill can occur, or if it is always accompanied by other epilaryngeal vibrations. Traill’s (1986) imitations of !Xóõ are the only other known report of epiglottal vibration, and he notes that other concomitant “arytenoidal” vibrations also occur. Traill doubts the epiglottis can vibrate as rapidly as 50 Hz (1986, p. 128), and, consistent with this assertion, the trilling in Somali seems to be around 30-40 Hz.

The shape of the epiglottis probably plays a role in determining how and whether it will vibrate. One variation frequently observed is posteriorly concave curvature: the reduced area orthogonal to the posteroanterior dimension should correspond with a diminished tendency for epiglottis vibration as less aerodynamic force will act in a retracting direction. Nevertheless, it is suspected that, given the generally high elasticity of the elastic cartilage forming the epiglottis, other vibrations should be possible, such as vibration of the lateral margins.
Figure 3.8: A voiceless epiglottal trill in Somali (/haniin/ ‘testicle’) (Select frames showing sequence). Large arrow = epiglottal vibration; small arrow = left aryepiglottic fold vibration. Data obtained from the University of Victoria laryngoscopy video database (Esling, 2013). Related data appear in Edmondson & Esling (2006, p. 183; also see Edmondson, Padayodi, Hassan, & Esling, 2007).

It is unknown exactly what types of composite modes of vibration of the basic components of the epilarynx can occur. As pointed out above, in Traill’s (1985, 1986) description of sphincteric register in !Xóõ, the epiglottis and arytenoids are said to vibrate but there is no vocal fold vibration. A vocal-ventricular-aryepiglottic mode has never been attested. It is an open question as to what restrictions apply to co-oscillation of epilaryngeal structures. Esling et al. (2007, p. 586) suggest that during voiced aryepiglottic trilling, only the vocal and aryepiglottic folds vibrate and not the ventricular folds.

Although epilaryngeal vibration typically occurs in the constricted configuration (as in all of the above cases), it is possible for unconstricted epilaryngeal vibration to occur. Videos produced by Phil Rose (personal communication) of his imitations of growling variants of tones in Dinghai Wu show vibration of the entire right aryepiglottic fold and its cuneiform tubercle, but there is not as much posteroanterior constriction of the epilarynx.
In summary, from what is known, the following general statements can be made. Purely vocal-ventricular phonation has only been observed in singing-type contexts (i.e. throat singing); it has never been observed in linguistic contexts, although inner mucosal vibration of Bai resembles this mode. Vocal-ventricular vibration, if it relies on larynx lowering, may be regarded as a highly specialized setting which goes against the natural tendency for the larynx to raise during epilaryngeal constriction. Epilaryngeal vibration in linguistic contexts is, thus, apparently restricted to the upper epilarynx (i.e. the aryepiglottic folds, associated mucosal tissue, and, less commonly, the epiglottis), although much more documentation is needed as it is sparse at best.

It is also fair to say that individuals will vary considerably in exactly how epilaryngeal vibration is executed, and we should not expect to see much in the way of discrete correlation between localized oscillations and certain linguistic functions (such as the difference between phonatory and trilling functions). We might tentatively suggest a coarse epilaryngeal-vibration taxonomy that classifies vocal-ventricular vibration as non-linguistic, aryepiglottic vibration as either phonation-like or trilling-like depending on linguistic context, and epiglottal vibration as the most trilling-like of all since it requires rather extreme epiglottis retraction and thus preoccupies the tongue. The problem is that our data is far too limited to assert such functional correlations.

A safer conclusion is that, while the details vary, the auditory result does not stray too far from harsh phonatory quality, or maybe more appropriately, “growling”. The next section will demonstrate the acoustic correlates of known cases of such growling, for which we can only guess at what exactly is vibrating (although our money would probably be safe on a bet of something “epilaryngeal”).
3.1.4 Epilaryngeal vibration acoustics: Agul and N|uu

Several researchers have made a connection between epilaryngeal vibration (“growl”) and harsh voice quality (Rose, 1989, p. 237; Esling & Harris, 2005). Laver (1980, pp. 126–132, 141–156; also see Scherer, 1986, p. 152) defines harsh voice quality by high variation in the duration (jitter) and intensity (shimmer) of the pulse cycle associated with the voice source and says that it may be acoustically characterized by noise and an increase in signal intensity. Gerratt & Kreiman (2001) identify the following acoustic properties for supraperiodic non-modal phonation, which they interpret as characterizing harsh voice quality: (a) period-doubling, (b) amplitude modulation, and (c) low harmonics-to-noise ratio. Properties (a) and (b) imply periodic or quasi-periodic damping of the glottal source causing subharmonic structure in the source spectrum. Property (c) suggests this amplitude modulation is highly irregular in nature, which will result in high noise content and weak harmonic structure (or “harmonic smearing”).

Before we examine non-physiologically-correlated acoustic data, it is useful to consider data where acoustic and physiological vibrations associated with known epilaryngeal vibration, in this case aryepiglottic vibration, are correlated. High-speed laryngoscopic video data in Figure 3.9 illustrate aryepiglottic vibration at glottal F0s of (a) 100 Hz and (b) 200 Hz, respectively. Each figure contains three time series plots: the electroglottograph (EGG) signal (top), the acoustic waveform (middle), and a

---

75 The reader should be careful here not to confuse harshness and creakiness. Gerratt & Kreiman argue that, while the two phonatory qualities are grouped together in their own perceptually-defined class of non-modal phonation, creakiness acoustically differs from harshness by having an “extremely low” F0 and increased between-pulse vocal-tract damping (2001, pp. 375–376). In parallel to their view, this dissertation views harshness and creakiness as two varieties of “constricted” phonation (i.e. phonation types produced with epilaryngeal constriction).

76 These data were obtained by directly following the methodology in Moisik et al. (2010). Figure 3.9 is published in Moisik (forthcoming).
kymographic image of AE fold displacement (bottom). These figures illustrate the correlates of AE vibration: vertical lines match (nearly but not perfectly) synchronized events in the three signals and the arcs in (b) indicate the imprint of the AE pulse event in the waveform and identify the amplitude modulation of the glottal pulses that results from the AE pulse (note: only a few pulses are indicated with arcs, but the effect extends throughout the figure).
Figure 3.9: High-speed laryngoscopy data (500 fps) of voiced aryepiglottic vibration\(^{77}\) at 100 Hz (a) and 200 Hz (b) glottal F0. Laryngoscopic still frame images are on the left (R = right AE fold; L = left AE fold); solid lines in laryngoscopic frames indicate kymographic pixel strip. Dashed lines (a) = glottal closure, period-doubled acoustic output, and phases of the AE aperture (closing and opening). Arrows (b) = (some) glottal pulses in EGG and acoustic signals; arcs iconically illustrate amplitude modulation of glottal pulse found throughout; dotted lines (b) = left AE aperture opening. (*n.b.*: kymograph in (b) only shows AE fold displacement; aperture open when fold displacement peaks in kymographic image.)

\(^{77}\) Figure appears in Moisik (forthcoming).
Through entrainment, an approximate\textsuperscript{78} aryepiglottal-to-glottal (AE:G) pulse ratio can be identified, which varies depending on the specific configuration of the larynx (e.g. the amount of longitudinal tension on the vocal folds). Acoustically, strong entrainment results in amplitude modulation of the voice source, with period-doubled phonation occurring if the AE:G pulse ratio is approximately 1:2 (Figure 3.9a). Smaller pulse ratios yield variations of this (e.g. in Figure 3.9b, the AE:G pulse ratio is approximately 1:5). One can therefore generalize that aryepiglottic vibration can cause amplitude modulation of the glottal source. In the frequency domain, this is associated with the presence of subharmonic structure stemming from the low frequency aryepiglottic periodicity relative to the periodicity of the glottal pulse.

As described in the previous section, the epilarynx has complex geometry and can form tortuous air-channel topology. These features ought to bias irregular aryepiglottic vibration and result in a rather noisy acoustic signature. As would be expected, this is especially true if the vibration is voiceless at the glottal level (Moisik, Esling, & Crevier-Buchman, 2010). In either case, (sub-)harmonic structure associated with the aryepiglottic pulse may tend to be rather weak due to period-to-period variation in pulse shape and amplitude.

What follows is a set of illustrations drawn from various sources which are suspected to involve epilaryngeal vibration. Unlike the data just discussed, it cannot be known for certain in these data what type of epilaryngeal vibration each case represents, but the general acoustic pattern typically involves highly variable amplitude modulation.

\textsuperscript{78} It must be emphasized that the pulse ratio is approximate since the tendency of the aryepiglottic folds towards irregular motion means that the pulse timing will not be consistent from one moment to the next. A higher degree of entrainment between co-oscillating structures will result in a higher exactness of the pulse ratio (towards an integer-multiple relationship).
corresponding with a noisy spectrum and subharmonic structure (if the sound is voiced). The claim is that for each case suspected to involve epilaryngeal vibration, there is a conspicuous amplitude modulation that matches the physiologically-correlated data in Figure 3.9.

Each case shows the acoustic waveform and transcription (top) and a Hamming windowed spectrum associated with the waveform (in cases of comparison, comparable waveform durations were used). An amplitude envelope has manually been traced over each waveform to indicate the amplitude modulation (of the glottal pulse in voiced cases, or of noise, in voiceless cases) suspected to be associated with epilaryngeal vibration. The first few harmonics (h) and subharmonics\(^{79}\) (s) are marked in the spectra when appropriate and easily discernible.

\[\text{As a reminder, harmonics and subharmonics are frequency domain correlates of (quasi-)periodicity in a time-varying signal. A perfectly periodic (sinusoidal) vibration has a single “harmonic” defined by the frequency and amplitude of the sinusoid. A quasi-periodic vibration, such as the voice source will be comprised by many harmonics, the amplitudes of which can be obtained from the Fourier Transform and depend on the shape of the quasi-periodic signal (the more it is sinusoidal, the weaker the strength of upper harmonics). The first harmonic is said to be the fundamental and defines the fundamental frequency or F0 of the vibration. Complicating matters is the possibility for a “primary” vibration to become modulated in various ways. The type relevant here is amplitude modulation which means that, instead of each period of the primary vibration being roughly equivalent in amplitude, the amplitude of each period changes over time. The amplitude modulation has its own frequency, which will be lower than that of the main vibration, if the source of modulation is entrained or synchronized with the main vibration (it does not have to be so). If these conditions prevail, then there is yet another periodicity of the signal, i.e. the amplitude modulation, and it is harmonically related to that of the primary vibration. Frequency analysis will resolve these periodicities as a secondary harmonic structure which is likely to be weaker in amplitude than the harmonics associated with the primary source (but this really depends on the nature of the signal). If the amplitude modulation is not strongly regular (which characterizes aryepiglottic vibration), then there will be weak correlation of periodic signals with sinusoids and the resulting frequency analysis looks noisy or random, or might appear to have “smeared” harmonic structure.}\]
Figure 3.10: Voiced epilaryngeal vibration in Burkikhan Agul ‘bridges’. Analysis based on freely available data found in The UCLA Phonetics Lab Archive (“Aghul,” n.d.).

The first set of examples is drawn from recordings of two speakers of Agul (“Aghul,” n.d.), one representing the Burkikhan dialect (a female; Figure 3.10, Figure 3.11, and Figure 3.12) and the other representing the Tpig dialect (a male; Figure 3.13) (a comprehensive transcription of this data is in Appendix C, and these data will be discussed further in §7.1.2).

Figure 3.10, illustrates a ~2:1 (potentially AE:G) amplitude modulation during [§]. The onset and offset of this modulation is gradual, allowing for us to identify confidently the envelope associated with the amplitude modulation as alternating pulses associated with the vocal fold source during [u] become successively more damped. The
spectral profile confirms the presence of subharmonic structure concomitant with the
harmonic structure of the vocal fold source.

Figure 3.11: Voiceless epilaryngeal vibration in Burkikhan Agul ‘wheys’. The left
column shows an ordinary voiceless pharyngeal fricative; the right column shows an
alternate production of the same word which arguably involves epilaryngeal vibration.
Analysis based on freely available data found in The UCLA Phonetics Lab Archive
(“Aghul,” n.d.).

Voiceless pharyngeals produced by the same speaker are illustrated in Figure
3.11. This example illustrates two alternate productions of the /h/ in the same word for
‘wheys’, one of which is a fricative and one of which involves ostensible epilaryngeal
vibration, as judged from the amplitude modulation of the fricative noise (and auditory
impression). Since the spectrum picks up harmonic structure from the flanking vowels,
comparison of the fricative (left) and “trilled” (right) productions reveals the presence of
subharmonic structure only in the trilled production. The brace (right, “growling” waveform) indicates an example of the amplitude modulation period, which is estimated to be 20 ms; this corresponds to a frequency of 50 Hz (the approximate frequency of the first subharmonic ($s_1$)).

Figure 3.12 shows epilaryngeal vibration following the release of [ʔ], although there is a delay before the amplitude modulation becomes fully manifest. The quality of the vowel preceding the onset of the epilaryngeal vibration (under the curly brace) is auditorily harsh with some raised larynx voice quality, but it lacks “growl”. For comparison, the vowel from the second syllable of the word is shown, and it is completely void of any sign of amplitude modulation, and, correspondingly, its auditory quality is judged to be modal with even some slight breathiness. Again, we note the presence of subharmonics in the first syllable (left) and overall weakened or “smeared” harmonic structure in comparison to the second syllable (right) and infer that its cause is amplitude modulation of the glottal source by epilaryngeal vibration.
Figure 3.12: Voiced epilaryngeal vibration following [ʔ] in Burkikhan Agul ‘to cry’. The left column shows epilaryngeal vibration in the first syllable; the right column provides the second (modal) syllable for comparison. Analysis based on freely available data found in The UCLA Phonetics Lab Archive (“Aghul,” n.d.).

The final Agul example (Figure 3.13) is from the Tpig dialect\(^{80}\), which illustrates a word-final [ʕ]. The preceding vowel (left) has well defined harmonic structure, but this is lost once the pharyngeal is encountered (right). The production is relatively short, but we can still identify a roughly 2:1 (possibly AE:G) amplitude modulation, as the envelope indicates.

\(^{80}\) The data for the Tpig dialect lack any examples of voiceless epilaryngeal vibration, but there are plenty of voiced instances.
Figure 3.13: Voiced epilaryngeal vibration following Tpig Agul ‘crest’. The left column shows epilaryngeal vibration in the vowel preceding the final pharyngeal consonant for comparison with the data in the right column, which illustrates epilaryngeal vibration in the pharyngeal consonant. Analysis based on freely available data found in The UCLA Phonetics Lab Archive (“Aghul,” n.d.).

The second set of examples is drawn from data of a (female) N|uu speaker collected by Amanda Miller and colleagues (see Miller, Brugman, Sands, et al., 2009). This example set compares “epiglottalized” vowels with modal vowels: [ḛ] and [e] are compared in Figure 3.14 and [o̰] and [o] are compared in Figure 3.15. In both cases, the amplitude modulation is strongest in the first half (top left) of the vowel and diminishes considerably towards the second half; in reflection of this, the corresponding subharmonic structure is very crisply defined in the first half of the vowel and disappears in the second half, where only strong harmonic structure can be observed. The modal
(possibly somewhat breathy) vowels (lower row of plots) show no sign of the subharmonic structure.

We can conclude from these cases that, first of all, even without visual evidence, we can confidently identify epilaryngeal vibration from the acoustic (and auditory) evidence. This is because we have strong reason to believe that one of the main effects of epilaryngeal vibration is amplitude modulation (of a harmonic or noisy sound source). Thus, while not exclusively generable from epilaryngeal vibration, subharmonic structure is nonetheless a very useful indicator, especially in combination with other evidence (such as auditory quality). It must rank above other quantitative indices such as jitter, harmonics-to-noise ratio (HNR), open/closed quotient, spectral tilt (for a summary, see Gordon & Ladefoged, 2001), since these are not as uniquely correlated with aero-acoustic changes imparted by epilaryngeal vibration but rather more broadly group non-modal phonation; for example, low HNR is shared by breathy, creaky, and epiglottalized phonation and decreased spectral “slope”81 (as measure by H1 – H2) is correlated with creaky and epiglottalized vowels (Miller, 2007, p. 79).

81 To qualify mathematically as a slope (rise over run), one should take the forward difference of harmonic amplitudes (i.e. H2 – H1 not H1 – H2) and divide by the frequency interval separating the harmonics.
Figure 3.14: Voiced epilaryngeal vibration in N|uu ‘to fly’ (top row of plots) and modal voice in ‘to insert’ (bottom row of plots) for comparison. The left column shows the first half of the vowel and the second half appears in the right column. Analysis based on freely-available data (Miller, Brugman, Sands, et al., 2009).
Figure 3.15: Voiced epilaryngeal vibration in N\uu ‘chameleon’ (top row of plots) and modal voice in ‘man’ (bottom row of plots) for comparison. Left column shows first half of vowel; second half appears in right column. Brace (upper left) shows amplitude modulation period, (~66 Hz). Analysis based on freely-available data (Miller, Brugman, Sands, et al., 2009).
3.2 High-speed laryngoscopy of aryepiglottic trilling in Iraqi Arabic

This study examines aryepiglottic vibration in Iraqi Arabic voiced and voiceless pharyngeals using high-speed laryngoscopic observation. Since the vibration is in the context of pharyngeal consonants, aryepiglottic vibration will be referred to here as aryepiglottic trilling. The classification of the vibration as trilling is based on analogy with classification of uvular vibration as trilling in the context of uvular consonants. However, this matter should be kept in mind during the discussion. Four trills are qualitatively evaluated using acoustic, electroglottographic (EGG), kymographic data, and epilaryngeal aperture estimate techniques. Quantitative evaluation of trilling frequency (and duration) indicate a greater degree of laryngeal constriction for voiced trills. The vibration clearly involves mucosal wave transmission, and it is suggested that this causes a phase shift amongst the driving forces acting on the aryepiglottic folds, resulting in more irregular vibration than for the vocal folds. Finally, the vibration “bleeds” into neighbouring vowels, which can be interpreted as a coarticulatory effect. We might wonder at what point we should begin to call aryepiglottic vibration a phonatory mechanism.

Although Arabic /h ʕ/ have often been labeled voiceless and voiced pharyngeal fricatives, their aryepiglottic or trilling components have not previously been attested. Since the early Arab grammarians, there has been disagreement over their precise place

---

82 This study was originally published as Hassan, Esling, Moisik, & Crevier-Buchman (2011). The data were also presented (by John Esling) at the 2012 biennial colloquium of the British Association of Academic Phoneticians (BAAP) (March 28th, 2012; Leeds, England). Data collection was conducted by Zeki Majeed Hassan, John Esling, and Lise Crevier-Buchman. All text, analysis, and interpretations – except where noted – are attributed to Scott Moisik with assistance from John Esling.

83 The following research and some of the writing underlying the next two paragraphs was contributed generously by Zeki Majeed Hassan. Minor alterations to the original text have been made for this dissertation by Scott R. Moisik. Its original form can be seen in Hassan et al. (2011).
and manner of articulation. Sibawayh placed them in the middle part of the throat (Al-
Nassir, 1993), implying between the glottis and uvula. They are also placed generally in
the pharynx for different varieties of Arabic, e.g. Tripoli Libyan (Laradi, 1983), Sudanese
(Adamson, 1981), Qatari (Bukshaisha, 1985), Iraqi (Laufer & Baer, 1988), and Lebanese
and Palestinian Arabic (El-Halees, 1985; Pettorino & Giannini, 1984). While “epiglottal”
is the IPA auditory label for “more extreme” pharyngeals (Ladefoged & Maddieson,
1996), recent research has shown the pharyngeal/epiglottal category of sounds to be
produced by epilaryngeal constriction with tongue retraction (Edmondson, Padayodi,
Hassan, & Esling, 2007; Esling, 2005). Heselwood (2007) notes the importance of
epilaryngeal tube shape beneath the stricture point and of larynx-height variations. Esling
(1999) notes that both pharyngeals and epiglottals occur at the aryepiglottic sphincter
point of stricture (i.e. within the aryepiglottic plane), and sounds which have been
described auditorily as epiglottal, “deeper” or “more extreme” than pharyngeals, are
associated either with aryepiglottic trilled varieties of the simple fricative or approximant
or with the default raised-larynx posture of the laryngeal sphincter, with retraction of the
tongue.

Different manners have been reported for different Arabic dialects and
phonological contexts. Early studies classify /ʕ/ as fricative (Cantineau, 1960; Gairdner,
1925; Ghazeli, 1977; May, 1981), whereas it is classified as stop by Al-Ani (1970, 1978)
for Iraqi Arabic (thought to be an occlusion at the ventricular folds), for Sudanese
(Adamson, 1981) and for Qatari (Bukshaisha, 1985). An approximant variant has also
been reported for /ʕ/ (Laufer & Baer, 1988), a glide (Heselwood, 1992) and most recently
a tight approximant (Heselwood, 2007). A trilled variant based on Sibawayh’s term
The primary aim of the current study is to confirm the occurrence of aryepiglottic vibration in Iraqi Arabic, which was originally suggested by Edmondson et al. (2007, p. 2066). While the present study cannot resolve the debate surrounding place and manner of production of pharyngeal consonants in Arabic dialects, it does (re-)assert that aryepiglottic vibration can occur as a speech sound, and it encourages similar investigations in languages such as !Xóô and Agul, which are suspected to exhibit epilaryngeal vibration in their sound system.

3.2.1 Methodology

Native Iraqi Arabic productions of voiced and voiceless aryepiglottic (AE) trills were examined using high-speed laryngoscopy. The trills are qualitatively analyzed and

---

84 It may be that vibration of the aryepiglottic folds is most easily initiated in intervocalic context. The same is probably the case for other types of trills: often one can hear a schwa-like vocalic onset to trills (this actually occurs for /h/ in the Iraqi Arabic word /rahːːl/ ‘to travel a lot’ discussed in this study). Furthermore, in the Burkikhan Agul data, supposedly word-initial pharyngeals with audible trilling are often preceded by a brief [a] (see Appendix C).
illustrated using video frame montages and composite plots containing EGG, acoustic, kymographic (see Švec & Schutte, 1996), and aperture estimate signals. Quantitative evaluation is made of trilling frequency and duration.

The participant is an adult male native speaker of Iraqi Arabic, born and raised in Basra. Before performing high-speed filming and capture, preliminary normal-speed laryngoscopic observations were made of the speech items suspected to be produced with AE trilling to confirm that trilling did indeed occur.

The data consist of four Iraqi Arabic words containing trilled variants of the /h/ and / kullanılan consonants in intervocalic context (/aCi/). These words can be grouped into two pairs by morphological relation; each pair is minimally differentiated by consonant length: /raħiːl/ ‘travel’ ~ /raħħiːl/ ‘travel a lot’ and /saʕiːd/ ‘happy’ ~ /saʕʕiːd/ ‘make people happy’.

The system used to capture the trilling sequences was an SL Kamera 500 connecting a rigid oral endoscope to a Weinberger SpeedCam Lite interface (Erlangen, Germany), set to a frame rate of 500 Hz with a resolution of 256 × 256 pixels.

Throughout the recording session, the audio signal of the trills was obtained using a head-mounted AKG C410 microphone, positioned 4 cm from the lips, with a 45° angle from sagittal orientation, sampled at 44100 Hz, 16-bit resolution. Larynx periodicity was obtained with a Glottal Enterprises EG2-PCX EGG with the surface electrodes placed externally over each thyroid lamina. The audio and video signals were digitized using ANAVOX (custom software; Vannier-Photelec, Antony, France). A time-to-live signal was used to ensure synchronization of the audio with the video signal. Uncompressed audio and video files (exported from ANAVOX) were processed and analyzed using MATLAB.
Because of the camera positioning, only the participant’s right aryepiglottic fold was fully visible, and, thus, this was the only fold to be examined.

The reader will observe that the EGG signals are unstable during the trills and contain numerous discontinuities and “dead zones”. This was probably due to change in larynx height during the pharyngeals. Despite this, the EGG signals are included in the results. This is partially to present a complete picture of the data, but it is also because these signals do contain regions of clarity, usually during modal voicing, which provide a rough index of laryngeal state.

Not shown in the figures, but another feature of the analysis is the extraction of the AE aperture signal, which is an estimate of the area function of the corresponding AE aperture. This signal is calculated by summing the difference between 255 (full white in 8-bit grayscale) and the grayscale value of each pixel in the kymographic strip; the result is then normalized as a percentage. The aperture signal was used to estimate the AE frequency by means of a fundamental frequency algorithm based on peak detection.

3.2.2 Results and discussion

Duration and frequency measurements for all four trill tokens are presented in Table 3.1 along with a narrow transcription, qualification of constriction degree based on visual appearance of the epilarynx, and references to the corresponding figures for each trill. These figures contain the following data: a 25-frame montage, the kymographic strip (white line), and the first frame of the sequence indicates the right aryepiglottic (ae) fold, the cuneiform cartilages (c), and the epiglottic tubercle (et); below this is a set of four plots showing (from top to bottom) the audio waveform, spectrogram, kymographic trace
(Kymo), and the electroglottograph signal (EGG); two arrows mark the start (left) and stopping (right) points of the sequence shown in the montage; a transcription is provided above the waveform to help guide interpretation of the data; finally, moments of interference in the EGG signal are marked with curly braces.

Table 3.1: Aryepiglottic trill duration and average frequency.

<table>
<thead>
<tr>
<th>word</th>
<th>transcription</th>
<th>duration (ms)</th>
<th>Hz</th>
<th>relative constriction</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>/rahiːl/</td>
<td>[raʰiːl]</td>
<td>59</td>
<td>87</td>
<td>moderate</td>
<td>Figure 3.16</td>
</tr>
<tr>
<td>/raḥhiːl/</td>
<td>[ʔraʰhiːl]</td>
<td>211(^{85})</td>
<td>78</td>
<td>least</td>
<td>Figure 3.17</td>
</tr>
<tr>
<td>/saʃid/</td>
<td>[saʃiːd]</td>
<td>69</td>
<td>130</td>
<td>high</td>
<td>Figure 3.18</td>
</tr>
<tr>
<td>/saʃʃid/</td>
<td>[saʃʃiːd]</td>
<td>93</td>
<td>107</td>
<td>greatest</td>
<td>Figure 3.19</td>
</tr>
</tbody>
</table>

Aryepiglottic trilling was evident in all recordings, but the geminates exhibited the greatest degree of aryepiglottic fold displacement possibly because there was more time for the aryepiglottic folds to build up oscillatory momentum (although laryngeal airflow may have increased for these productions). One important difference between the voiced and voiceless trills was the degree of epilaryngeal constriction. An impressionistic ranking of degree of constriction from greatest to least would be: [ʰː], [ʰ], [ʃ], [ʃː]. The difference is partly reflected in trilling frequency, although [ʃ] is more rapid than [ʃː], which may be due to the greater mucosal wave that occurs for the latter. Esling et al. (2007) make similar observations for trills performed by a Moroccan phonetician.

\(^{85}\) The geminate voiceless trill is atypically long because these were recorded separately and at a slower speech rate.
Also worthy of note is the EGG signal, which exhibits moments (curly braces) of interference while aryepiglottic trilling is occurring or being engaged. Changing laryngeal configuration, especially larynx raising, during trill production is the most probable cause for this effect. Once the larynx returns to a more neutral position, the signal often resumes (as in the latter parts of the /i/ vowels). Sometimes the EGG signal appears to register the AE vibration, as in Figure 3.17 (the moment in between the braces).

The speaker has a single-channel, centro-lateral aperture (Figure 3.1e) in most cases because the cuneiform tubercles (c) do not come into contact with the epiglottic tubercle (et). but in [ʰː] the configuration approaches a dual-channel, centro-lateral configuration (Figure 3.1e). Furthermore, while vibration pattern is typically of the aryepiglottic fold (and its inner mucosa), in [ʰː], the posterior arytenoid cartilage complex becomes engaged in oscillation (giving the appearance of applauding hands). Thus, the air-channel is partly bifurcated during this trill.
Figure 3.16: Singleton voiceless aryepiglottic trill [h] in /rahiːl/ ‘to travel’.
Figure 3.17: Geminate voiceless aryepiglottic trill $[\mathbf{h}]$ in /rahhi:l/ "to travel a lot".
Figure 3.18: Singleton voiced aryepiglottic trill [ʕ] in /saːiːd/ ‘happy’.
Figure 3.19: Geminate voiced aryepiglottic trill [ː] in /saːʃiːd/ ‘make people happy’.
In all cases, the aryepiglottic vibration overlaps with neighbouring vowels, but the effect varies by trill. This can be seen both in the acoustic signals (waveforms and spectrograms) and in the physiological data (i.e. the kymography; the EGG does not provide a clear indication of when aryepiglottic trilling starts or stops). The greatest overlap is observed in [ʃː] (Figure 3.21), and both preceding and following vowels were affected. The least “co-articulatory bleed” is observed in [ʰː] (Figure 3.17). The voiced trills in general seem more continuous in transition into and out of the aryepiglottic vibration, for the voiceless trills, the vibration starts quite abruptly but tends to blend somewhat into the following vowel. Anticipatory narrowing of the aryepiglottic aperture is visible in all cases and extends even into the preceding consonant ([s] or [r] depending on which example).

The classification of these vibrations as trilling is justified on the grounds that aryepiglottic vibration is not spontaneously occurring in just any context but rather localized to a narrow distribution within the phonetic system involving the pharyngeal consonants. The question is whether the exact same physiological mechanism labeled as trilling in this context would be appropriately deemed as such if it spontaneously occurred, say, during an arbitrary vowel in a passage of excited speech. In the latter case, we would be inclined to say it was phonation. What is intimated here is that linguistic function will play a key role in how the ambiguous epilaryngeal vibration gets phonetically classified; these matters will be considered further in §7.1.3.

Finally, auditory evaluation indicates that, even though trilling occurs, one could easily confuse the voiceless trills for voiceless pharyngeal fricatives and the singleton voiced trills for a pharyngeal stop or “pharyngeal tap”, even though several aryepiglottic
pulses are evident (Figure 3.18). The most trilling-like by auditory impression is the voiced geminate [ːː]. This interpretation is strikingly similar to the observation made by Heselwood (2007, pp. 24–25) that the “tight ‘ayn”, despite evidently being an approximant as judged from acoustic data, is very much stop-like when judged auditorily. The harsh quality imparted to the flanking vowels by the co-articulatory bleed during the pharyngeals seems to qualify as an “edge effect” (Heselwood, 2007, pp. 24–25) which is likely perceptually useful to speakers in distinguishing pharyngeal from laryngeal place of articulation.

In terms of trilling biomechanics, we can infer that trilling frequency is correlated with the degree of laryngeal tension (or stiffness), particularly in the aryepiglottic folds. It could be that the higher rate of vibration in the voiced trills is attributable to entrainment with the vocal folds, not just increased tissue stiffness. In fact, sympathetic vibrations of the aryepiglottic tissue with the vocal folds below during the voice trills is even evident well before and after the trilling vibration proper in all cases. However, once aryepiglottic vibration begins in earnest, the pulses are always highly irregular.

All trills observed here involve the upper aryepiglottic mucosa, but the fact that the apices of the arytenoids also participate in the oscillation during the voiceless trill [ːː] (Figure 3.17), but not in the other trills, underscores the possible variation in the nature of trilling (as discussed in 3.1.3). Despite such variation, auditory impression does not markedly vary – both voiceless trills still sound very much like “raucous” pharyngeal fricatives.

Another important biomechanical feature visible in the high-speed frames is a mucosal wave that radiates outwards across the surface of the aryepiglottic fold. Based on
vocal-fold body-cover theory (Hirano, 1975; McGowan, 1992; Story & Titze, 1995), which emphasizes the importance of the vocal-fold mucosal wave in self-sustaining oscillation, we can reason that the same role is played by the mucosal wave on the aryepiglottic folds (this will be discussed in more detail in §3.3), but it is also possible that, if its phasing is not favourable, it could plausibly hinder AE vibration or skew vibration towards irregularity. In fact, the data do indicate that the our participant’s aryepiglottic trilling is highly irregular in nature, as expected based on previous observations of controlled phonetic productions of aryepiglottic trills (Moisik, Esling, & Crevier-Buchman, 2010). This irregularity can be associated with harsh voice quality and also probably has perceptual functions in pharyngeal consonant discrimination.

3.2.3 Summary of high-speed laryngoscopy of AE trilling in Iraqi Arabic

This study attests that aryepiglottic trilling occurs as a phonetic variant of Iraqi Arabic pharyngeal consonants /ʕ/ and /ћ/, regardless of phonological duration (either singleton or geminates). The trills appear to co-articulate with neighbouring vowels and phonatory quality is harsh, but the actual consonants do not auditorily differ very much from stop and fricative congener. Aryepiglottic biomechanics in trilling-like (and presumably in phonatory-like) contexts are dynamically complex and lead to irregular pulse periods. This work strongly encourages future research which confirms or disconfirms the presence of epilaryngeal vibration as it has been reported to occur in various languages, such as !Xóõ, Zhenhai Wu, and Agul. It is possible that epilaryngeal vibration has a wider distribution in speech than previously thought, and documenting its occurrence will have a significant impact on how we approach modeling the voice source.
in phonetic and phonological theory, especially since the epilarynx appears to play a dual role as phonatory mechanism and trilling mechanism.

### 3.3 A two-trap-door model of aryepiglottic vibration

So far this chapter has explained the phonetic theory behind epilarynx vibration, discussed its applications and its variations in speech and vocalization, and presented empirical evidence attesting its occurrence as a phonetic variant of Iraqi Arabic pharyngeal consonants. This section documents a model\(^{86}\) of the dynamic interaction between the vocal folds and the epilarynx that helps explain the voice-source role played by the epilarynx.

The purpose of creating the model is to perform an analysis by synthesis of the essential properties of aryepiglottic vibration, determine what it has in common with other forms of epilarynx vibration, and, from this basis, clarify what epilarynx vibration contributes to the voice source in general. Aryepiglottic vibration, as observed in trilling-like and phonatory-like behaviour, is the reference case because it appears to be the most phonetically relevant form of epilarynx vibration (vocal-ventricular vibration being unattested and epiglottal trilling being less common). A model of aryepiglottic vibration can be compared to existing models of vocal-ventricular vibration.

From the outset, there is no reason to expect that aryepiglottic and vocal-ventricular vibration behave fundamentally identical as source mechanisms. The available observations point to at least one major difference: vocal-ventricular vibration

\(^{86}\) This model was originally presented at the 7th International Conference on Voice Physiology and Biomechanics, June 6th, 2010. See Moisik, Bowers, Esling, & Crevier-Buchman (2010).
is more periodic than aryepiglottic vibration (whether voiced or voiceless). They also appear to have different uses, the former being primarily associated with throat-singing variations and the latter appearing in linguistic contexts. Furthermore, the anatomical-morphological differences between the ventricular and aryepiglottic folds is significant and justifies a different model design from those proposed for the ventricular folds. Another consideration is the fact that, unlike vocal-ventricular phonation, an aryepiglottic model must simulate both voiced and voiceless conditions, as both are attested in linguistic phenomena.

Despite these differences, aryepiglottic trilling shares something fundamental in common with ventricular vibration: modulation of the vocal-fold source. As Gerratt & Kreiman (2001) argue, this is a key perceptual feature of harsh voice, in addition to noise content. The implication is that epilarynx vibration, despite its many physical variations, maps exclusively onto the harsh-voice-quality category, which has applications in several phonological systems found throughout the globe.

3.3.1 Previous models of epilaryngeal vibration

Although modeling of the vocal folds has already been very thoroughly explored, much less attention has been paid to modeling of epilarynx vibration; however, a couple of models do exist. Sakakibara et al. (2001; Imagawa, Sakakibara, Tayama, & Niimi, 2003) modeled vocal-ventricular phonation using a 2×2-mass model (two masses per set of folds). This model lacks mechanical coupling between the vocal and ventricular
masses and uses a persistent, cylindrical space, 5 cm$^2$ in cross-sectional area and 1.6 cm\(^8\) in height to approximate the ventricle, which separates the two pairs of folds/oscillators. The purpose of the model is to simulate phonatory modes in Tuvan and Mongolian Khöömei/Khöömij for the purpose of singing synthesis.

The inclusion of the ventricle in a model for throat singing is supported by the observation of larynx lowering in throat singing modes involving vocal-ventricular vibration (Esling, 2002a), although no evidence has been brought to bear on whether this assumption is truly justified. X-ray imaging of Tuvan throat singing styles in Levin and Edgerton (1999, p. 85) does not provide a very clear image of the laryngeal structures, although the authors’ tracing suggests there is some separation between the vocal and ventricular folds in all of the different throat singing styles. In any case, the purpose of larynx lowering for these modes may in fact be to ensure separation between the vocal and ventricular folds (see 1974a); contact between these structures could be a hindrance to vibration of either set of folds.

Sakakibara et al. (2004) apply the (vocal-ventricular) throat singing model to the task of modeling growling phonation caused by vocal-aryepiglottic vibration, but the physiological realism of this is dubious: the presence of a persistent ventricular space is very unlikely when aryepiglottic vibration occurs. From the videofluoroscopic imaging in (discussed further in §5.1), it is apparent that the infra-aryepiglottic space is compliant, but it is unlikely that this space is the ventricle since the extreme vertical compaction of the epilarynx obliterates the ventricle (i.e. there should be vocal-ventricular contact);

\(^8\) Sakakibara et al. (2001) state they use a ventricle height of 16 cm, which is probably a mistake. Physiological measurements (e.g. Agarwal, Scherer, & Hollien, 2003) indicate that even 1.6 cm far exceeds typical ventricle height (i.e. the separation between the vocal and ventricular folds, which is more near ~5 mm). It is possible that they intended 1.6 cm as the height of the epilarynx itself, which is more realistic.
observations by Fink (1974a) and by Heselwood (2007) support this view; furthermore, a similar interpretation, that “growling” with the aryepiglottic folds involves vocal-ventricular fold contact and obliteration of the ventricle, could be made of the x-ray traces found in Sakakibara et al. (2004).

Detailed modeling of the translaryngeal aerodynamics through the vocal and ventricular channel using both physical replicas and computational modeling demonstrate the aerodynamic influence of the ventricular folds on vocal fold oscillation (Bailly, Pelorson, Henrich, & Ruty, 2008; Bailly, Henrich, & Pelorson, 2010). The Bailly et al. (2010) model will be a useful point of comparison with the present aryepiglottic vibration model, so it is worth summarizing its main properties and its correspondence to observations of actual productions of throat singing with vocal-ventricular vibration. A static ventricular aperture changes the pressure conditions within the glottis; if the aperture is not too narrow\(^8\), it supports phonation (lower phonation threshold pressure), otherwise it begins to act as a hindrance. Bailly, Henrich, & Pelorson (2010) present a two-mass model of the vocal folds aerodynamically coupled to a pre-specified model of ventricular aperture. The data for this parameter were obtained from measurement of high-speed laryngoscopic video of a professionally trained singer producing an imitation of the vocal-ventricular singing style based on auditory approximation to “Asian throat singing” (p. 3213). Like the Sakakibara et al. (2001) model, their model includes a persistent ventricle.

Figure 3.20 depicts a summary of the key data of the Bailly, Henrich, & Pelorson (2010) study. The production data for the ventricular fold aperture (dashed line) indicate

\[^8\] Specifically, if the distance between the ventricular folds is greater than or equal to the glottal width (Bailly, Pelorson, Henrich, & Ruty, 2008, p. 3306).
that the full ventricular fold cycle is double the glottal period, which accounts for the corresponding subharmonic structure characteristic of growl, but the ventricular folds also exhibits a moment of partial closure half way through the cycle. The amplitude of the vocal fold aperture (dotted line) exhibits period alternation, but it is slightly delayed with respect to the measured ventricular fold aperture. The larger of the two pulses is synchronized with the moment of partial ventricular fold closure, and this may be the cause of failure for the ventricular folds to close at this point (due to increased input pressure to the ventricular folds at this point). The smaller pulse occurs during the moment of complete ventricular fold closure, which is consistent with other empirical observations, such as Lindestad et al. (2001). The simulated vocal fold aperture (solid line), however, predicts increased glottal area during the ventricular fold aperture closure phase, contrary to the pattern exhibited by the empirical data. Like the empirical data for the vocal fold aperture, it too is out of phase, but in the opposite direction (i.e. lagging behind the ventricular fold aperture). The authors concede this as a limitation of their model (Bailly, Henrich, & Pelorson, 2010, p. 3221). They suggest that it stems from not modeling collision forces, although it could be attributable to the fact that they only modeled the vocal folds (the ventricular fold aperture area function is used as time-varying input to the fluid simulation) not both the vocal and ventricular folds. Not shown in Figure 3.20 are the pressure variations immediately below the ventricular folds. The model predicts that 100% pressure recovery occurs at ventricular closure moments, which the authors suggest effectively nullifies the transglottal pressure drop and consequently perturbs vocal fold vibration.
Figure 3.20: Trace of simulated and empirical results for vocal-ventricular phonation. Solid line: simulated vocal fold aperture; dashed line: measured ventricular aperture; dotted line: measured vocal fold aperture. Data traced from Bailly, Henrich, & Pelorson (2010, p. 3219). FF = ventricular (false) fold; VF = vocal fold.

Moisik’s (2008) interactive 3D larynx model, intended primarily for visual illustration of laryngeal articulation, implements a biomechanical model of the aryepiglottic folds based on Titze’s (1973) (henceforth T73) mathematical model of the vocal folds. Extending the T73 description, each aryepiglottic fold is modeled as a spring-mass lattice. Geometric configuration (determined by cuneiform positioning) is a manipulable parameter used to simulate the asymmetric modes observed in high-speed laryngoscopy data (Moisik, Esling, & Crevier-Buchman, 2010). This 3D larynx model also includes the original T73 vocal folds. From the description of the model, voiced and voiceless aryepiglottic trilling simulation should be possible, but only voiced trilling is discussed89. Moisik also does not describe any details regarding oscillatory interaction between the vocal folds and aryepiglottic folds. However, to the author’s knowledge, this is the only computational model that explicitly represents the vibration of aryepiglottic

89 The truth being that voiceless trilling never worked: the aryepiglottic folds inflated like balloons.
folds. As mentioned above, Sakakibara et al. (2004) extend their vocal-ventricular model to the task of simulating vocal-aryepiglottic phonation, but although this draws an important parallel between the two different types of epilarynx vibration, it is an abstraction that does not take under consideration the unique properties of the aryepiglottic folds as an oscillating system.

The aryepiglottic vibration model presented below matches the empirical description of aryepiglottic vibration presented in the first half of this chapter with the low-dimensional, lumped-element mathematical analysis afforded by spring-mass physics and aero-acoustic considerations of the laryngeal airway.

3.3.2 Principles of design for the aryepiglottic vibration model

Modeling proceeded according to the following considerations: (1) to study dynamical interaction between the vocal and aryepiglottic folds, the model should include a dynamical model of the vocal folds, and (2) the aryepiglottic folds, like the tongue tip, can be approximated as trapdoors to simplify the mathematical description of their dynamics. For (1), Story & Titze’s (1995) model (henceforth ST95) was selected as the model of the vocal folds, because it is adequately described for reproducibility, and it benefits from being a true body-cover model of the vocal folds (Story, 2002). Airflow

---

Following Hirano (1974, 1975), the vocal folds can be described as a layered structure, consisting of a cover layer, which comprises the mucosa epithelium and lamina propria (and its superficial, intermediate, and deep layers) and a body layer, consisting of the (internal) thyroarytenoid muscle. In lumped-element formulation, it is convenient to use this as the basis for abstracting the distribution of tissue properties characterizing vocal fold biomechanics. The significance is that, as the vocal folds vibrate, energy transmission occurs in the form of a traveling wave (mucosal wave) across the vocal fold “cover” and this is thought to be essential to self-sustaining oscillation. In fact, earlier one mass models (e.g. Flanagan & Landgraf, 1968), which do not adequately represent the body-cover relationship and cannot fully represent the mucosal wave, fail to engage in self-sustaining oscillation unless they are coupled to a an inertive vocal
considerations were made with consultation of T73 since the present formulation used an electrical equivalence circuit rather than wave-reflection mechanics to model vocal tract aerodynamics.

The benefit of (2) is that a well-defined biomechanical model is readily available: McGowan (1992) (henceforth M92). M92 was taken to form the mathematical foundation for the aryepiglottic fold model. Admittedly, the morphological similarity between the tongue tip and the aryepiglottic folds is somewhat tenuous: on the most gross level, both structures are oriented roughly parallel, rather than perpendicular to the direction of airflow, but, as discussed in §3.1.3, the morphology of the epilarynx is complex in ways that the tongue aperture is not. Perhaps most important is the fact that AE trilling involves two oscillating fold structures, whilst the tongue is singular in nature. In any case, lumped-element modeling, such as M92 and that proposed here, involves a high-order of abstraction of the spatial distribution of mass (as compared to, say, finite element models), so no matter what choice is made, some detail specific to the subject of modeling will be lost.

The M92 mechanical abstraction represents the tongue tip as a rigid trapdoor with an upstream compliance associated with the yielding walls of the oral cavity. According to McGowan (1992, p. 2903), the compliance allows for the breaking of symmetry in energy exchange between the air and the tongue tip, one of several properties that enables self-sustaining oscillation. Another possibility for breaking energy-exchange “symmetry” tract: tissue velocity and intraglottal pressure must be in phase (Story, 2002, p. 199); the addition of a second mass does represent the mucosal wave and can engage in self-sustaining oscillation even when not coupled to an inertive vocal tract above, but it cannot represent the coupling between the cover and the body (it is really just a “cover-only” model), which makes it difficult to manipulate the parameters responsible for changing the mode of vibration. A body-cover model, however, does exhibit the ability to engage in self-sustaining oscillation, even when not coupled to a vocal tract. The cost is a slightly more detailed lumped element model.
is to increase the degrees of freedom of the oscillating body. For example, the simplest approximation of the vocal folds as a single-mass system requires acoustic loading to support self-sustaining oscillation; adding another degree of freedom (or using a rotational degree of freedom instead of a translational freedom) removes the requirement for acoustic loading (also see f.n. 90). The explanation is that an additional degree of freedom allows for asymmetrical opening and closing motions, and this introduces delayed feedback into the system. In physiological terms, this characterizes the mucosal wave, which corresponds to a convergent-divergent glottal opening-closure pattern. Delayed feedback (a nonlinear effect) in the system helps to reduce its internal resistance and thereby support self-sustaining oscillation (Titze, 2006, pp. 350–351). McGowan does not include an additional mass on his tongue tip model on the basis that there is no evidence for a lingual mucosal wave.

Since M92 is a single degree-of-freedom system, it relies upon the compliant space upstream of the tongue trapdoor to introduce delayed feedback in the system and promote self-sustaining oscillation (McGowan, 1992; also see Titze, 2008). In simple mechanical terms, it breaks the symmetry in energy exchange between the volume velocity and the velocity of the body in motion. In (aero-acoustic) impedance terms, this space introduces compliant reactance into the system governing airflow; the ultimate result is the introduction of a phase delay in aero-mechanical forces, which favours oscillation. During opening, stored energy is released as positive pressure impinging on the trapdoor, but this does not happen immediately because of the time constant of the compliance. This time lag also applies to “charging” of the compliant space during closure, such that positive pressure does not build up too quickly, allowing elastic
restoring forces to return the trapdoor to its closed position. Thus, there is effectively a push-pull force regime that favours the opening and closing oscillatory motion of the trapdoor/tongue tip. A similar analysis applies to the capacitance of air in the subglottal space for modal vibration of the vocal folds (Fletcher, 1993; Titze, 2008). If the acoustic impedance characteristics of the air spaces immediately upstream and downstream of the oscillating structure do not lead to synchronization of the aerodynamic forces with this push-pull regime, self-sustaining oscillation will be hindered or may not occur at all as internal and external energy losses will drive the system towards temporal stability (i.e., in terms of non-linear dynamics, motion of the system will decay towards a point attractor).

In the case of the aryepiglottic folds, there is evidence for a yielding wall compliance of the epilarynx space upstream of the aryepiglottic folds (Figure 3.21); there is also evidence for propagation of a mucosal wave across the surface of the aryepiglottic folds suggesting multiple degrees of freedom characterize its motion (see §3.1 and 3.2, respectively). Unlike the oral-pharyngeal cavity in McGowan’s tongue-tip trilling model, however, the epilaryngeal “plenum” is a small space and overlaps significantly, if not entirely, with the aryepiglottic fold body. In fact, it is quite difficult to discern from the videofluoroscopic evidence (see §5.1) the exact extent of the compliant space, but it appears that it minimally involves the lower mucosa of the epilarynx, particularly the surfaces under and around the tubercle of the epiglottis. It is probable that, in reality, compliant yielding occurs for all mucosal surfaces of the epilarynx (i.e. ventricular, epiglottic, and aryepiglottic). The approach taken here, then, was to model the yielding wall compliance explicitly as the mechanical system for the aryepiglottic folds itself,
which is an abstraction extending from the very base of the epilarynx (i.e. immediately above the vocal folds) up to the aryepiglottic margins.

\[
\begin{array}{cccccc}
29 \text{ mm}^2 & 32 \text{ mm}^2 & 42 \text{ mm}^2 & 39 \text{ mm}^2 & 51 \text{ mm}^2 & 48 \text{ mm}^2 \\
\end{array}
\]

Figure 3.21: Estimated epilarynx mid-sagittal area during voiced aryepiglottic trilling (videofluoroscopy data are discussed further in §5.1). Glottal F0 is 100 Hz; Black dotted line = measured area; white outlines = hyoid bone and cricoid-arytenoid cartilages; white dashed line = epiglottis surface contour. Analysis done with ImageJ; metric values based on estimated scaling conversion of pixel area values, by obtaining pixel scale of reference distance on participant in video.

Figure 3.22 depicts, the mechanical model of the aryepiglottic folds comprising a trapdoor component, which represents the aryepiglottic fold body, and a conventional (damped) spring-mass component, representing the mucosa of the fold, particularly at its upper, free margin. Although the model includes left and right folds for the vocal and aryepiglottic systems, which allows for asymmetrical simulation, the present study examines conditions of perfect symmetry only (for simplicity sake, although this probably never occurs in reality). Each set of folds has adducted and abducted configurations. Adducted vocal fold (pre-phonatory) position follows the ST95 model; the abducted configuration uses an average glottal area based on values suggested by

\[91\] These data were originally presented at the 2010 Annual Conference of the Canadian Acoustical Association (CAA). See Moisik & Esling (2010).
McGowan (1992, p. 2905): $4.5 \times 10^{-5}$ m$^2$. For simulating aryepiglottic trilling, the aryepiglottic folds were assumed to touch the epiglottic tubercle, completely sealing the laryngeal aditus, leaving zero area. In non-trilling simulation, the cross-sectional epilaryngeal area formed by fully abducted aryepiglottic folds in the model is $2.37 \times 10^{-4}$ m$^2$, which is narrower but agrees well with Titze (2006, 2008), in which the epilarynx cross-sectional area is $3 \times 10^{-4}$ m$^2$ at maximum opening.

Mechanical properties used for the vocal folds fall within those suggested in ST95 and Story (2002). Physical dimensions and biomechanical properties for the aryepiglottic folds were fixed by analogy to values used for the ST95 and M92 models: the aryepiglottic mucosa masses are based on the values given for the vocal fold mucosa (cover) in the ST95 model; the trapdoor values follow from the M92 trapdoor parameters. Table 3.2 and Table 3.3 contain these parameters along with a comparison to parameters stated in the aforementioned sources. Figure 3.23 shows the aryepiglottic fold geometry illustrating the dimensions provided in these tables.

Table 3.2: Aryepiglottic fold mucosa parameters in comparison those to ST95 (see Appendix D for abbreviation meanings).

<table>
<thead>
<tr>
<th>Model</th>
<th>$w_m$ (cm)</th>
<th>$h_m$ (cm)</th>
<th>$l_m$ (cm)</th>
<th>$\rho$ (g/cm$^3$)</th>
<th>mass (g)</th>
<th>$k$ (N/m)</th>
<th>$d$</th>
<th>$\zeta$</th>
<th>$\eta$</th>
<th>$k_c$ (N/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST95/ VF mucosa</td>
<td>0.03</td>
<td>0.15</td>
<td>1.0</td>
<td>1.02</td>
<td>0.01</td>
<td>3.5 – 80</td>
<td>0.035</td>
<td>0.4</td>
<td>100</td>
<td>0.5 – 2</td>
</tr>
<tr>
<td>AE mucosa</td>
<td>0.1</td>
<td>0.15</td>
<td>1.5</td>
<td>1.02</td>
<td>0.01</td>
<td>10</td>
<td>0.008</td>
<td>0.4</td>
<td>100</td>
<td>50</td>
</tr>
</tbody>
</table>
Figure 3.22: Mechanical model of the aryepiglottic folds. Parameters: \( A \) = aryepiglottic aperture area; \( d \) = damping; \( k \) = stiffness; \( m \) = mass; \( T \) = trapdoor; \( \theta \) = trapdoor rotation angle. Subscripts: \( \text{GL} \) = lower glottis; \( \text{GU} \) = upper glottis; \( \text{B} \) = vocal fold body; \( \text{AL} \) = left aryepiglottic fold; \( \text{AR} \) = right AE fold; \( \text{CL} \) = left AE trapdoor-mucosa coupling; \( \text{CR} \) = right AE trapdoor-mucosa coupling.

Deviations in parameters between the M92 trapdoor and the AE trapdoor reflect conformation of trapdoor dimensions to mucosa dimensions and general approximation of aryepiglottic fold. The main difference between these structures is that the aryepiglottic fold length/breadth is assumed to be roughly half the distance between the posterior-most inner surfaces of the thyroid laminae, which is roughly 3 cm (Titze, 2006,
p. 8). Also, the mucosa width (or cover depth in ST95) was assumed to be greater than
the very thin vocal fold cover layer, but not as great as the rather sizeable tongue tip,
therefore a value in between these two reference points was chosen (somewhat arbitrarily as
0.1 cm). The trapdoor height is comparable to M92, but made slightly larger, assuming
the epilarynx is about 2 to 3 cm in length along its longitudinal axis but that the vibrating
part of the aryepiglottic fold does not span this length.

Table 3.3: Aryepiglottic fold trapdoor parameters in comparison to M92 (see Appendix D
for abbreviation meanings).

<table>
<thead>
<tr>
<th>Model</th>
<th>( w_{td} ) (cm)</th>
<th>( h_{td} ) (cm)</th>
<th>( l_{td} ) (cm)</th>
<th>( \rho ) (g/cm(^3))</th>
<th>( I ) (g·cm(^2))</th>
<th>( k/I )</th>
<th>( d^* )</th>
<th>( \eta )</th>
<th>( k_v/I )</th>
</tr>
</thead>
<tbody>
<tr>
<td>M92 tongue tip</td>
<td>0.5</td>
<td>1</td>
<td>0.6</td>
<td>1.0</td>
<td>0.1</td>
<td>30 – 240</td>
<td>0.5 – 0.6</td>
<td>400</td>
<td>–</td>
</tr>
<tr>
<td>AE trap door</td>
<td>0.1</td>
<td>1.2</td>
<td>1.5</td>
<td>1.0</td>
<td>0.17</td>
<td>450</td>
<td>0.6</td>
<td>400</td>
<td>50</td>
</tr>
</tbody>
</table>

The choice of stiffness and damping parameters were determined through
experimentation by starting at values proximal to those listed in the ST92, M92, and
Story (2002, p. 200) and manipulating them as required until results approaching
expectations for aryepiglottic fold vibration were achieved.
Equations of motion for the trapdoor and mucosa of the aryepiglottic folds, and the vocal fold masses directly follows the ST95 and M92 models. The reader is directed to those sources for the equations; however, the coupling between the aryepiglottic fold mucosa and trapdoor does merit comment. Both the trapdoor and mucosa mass were constrained to a single degree-of-freedom each: the trapdoor is restricted to its axis of rotation; the mucosa mass is confined to motion in the posteroanterior dimension. The coupling spring produces a torque on the trapdoor and a force on the mucosa mass. The vector representing coupling spring displacement was used in the calculation of the
torque on the trapdoor and the component of this displacement in the direction of the mucosa mass’ displacement was used to calculate the coupling force acting on the mucosa mass.

The electrical-analog flow-circuit (Figure 3.24) shows the aerodynamic components of the model. The capacitive element representing the epilaryngeal space ($C_E$) is a function of the epilaryngeal volume defined by the position of the aryepiglottic folds. In the interest of simplicity, and following McGowan’s approach, the model does not include details of acoustic propagation within the vocal tract on the assumption that frequencies of interest are below 200 Hz\textsuperscript{92}. Rather, the vocal tract is represented by a single resistive element $R_{VT}$. The model deviates from the ST95 model in regard to the calculation of glottal flow and pressure. The motivation here was to retain parallelism with the McGowan model, but since the glottis is static in M92 model, the glottal circuit elements in the current model are those used in the T73 model for viscous ($R_{GLv}$ and $R_{GUV}$), Bernoulli junction (kinetic) pressure ($P_{BGJ}$), and turbulent ($R_{GLt}$ and $R_{GUt}$) losses and air inertance ($L_{GL}$ and $L_{GU}$) within the glottal channel. These same elements have been implemented in the epilaryngeal loop of the circuit and follow a similar naming schema ($R_{TDv}$, $R_{EMv}$, $R_{TDt}$, $R_{EMt}$, $L_{TD}$, and $L_{EM}$).

\textsuperscript{92} A similar simplification is adopted in Bailly, Henrich, & Pelorson (2010, p. 3215).
Figure 3.24: Equivalence circuit the aryepiglottic model corresponding to Figure 3.22. C = capacitance; L = inductance; R = resistance; P = pressure source; U = volume velocity
Subscripts: G = glottis; GL = lower glottis; GU = upper glottis; E = epilarynx; TD = trapdoor; EM = epilaryngeal mucosa; BGJ = (Bernoulli) pressure at the glottal junction; BEJ = (Bernoulli) pressure at the epilaryngeal mass junction; P_{SG} = subglottal pressure; R_{VT} = vocal tract resistance.

Table 3.4 shows formulae used for calculating the circuit elements representing the properties of the air spaces enclosed by the mechanical components in the model. Subglottal pressure P_{SG} is set at 800 Pa for all simulations, and the resistive loading of the vocal tract is 1.96×10^{6}.

Table 3.4: Circuit parameter formulae for airflow in the aryepiglottic model

<table>
<thead>
<tr>
<th>Liljencrants (1985)</th>
<th>Titze (1973)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) air compliance</td>
<td>c) viscous resistance</td>
</tr>
<tr>
<td>( C = \frac{V}{\rho c^2} )</td>
<td>( R_v = 12\mu l^2 \frac{h}{A^2} )</td>
</tr>
<tr>
<td>b) air inertance</td>
<td>d) turbulent losses</td>
</tr>
<tr>
<td>( L = \frac{A}{h} )</td>
<td>( R_t = 0.875 \frac{P_B}{</td>
</tr>
</tbody>
</table>
The geometry used to calculate the variable volume of the epilaryngeal tube as the aryepiglottic folds oscillate is illustrated in Figure 3.25. When the epilarynx is fully closed, no volume exists at the level of the mucosa mass; the minimal volume (a) is a triangular prism formed between the epiglottic tubercle and aryepiglottic trapdoor; when partially open (b), a rectangular prism volume is added to the triangular one for the space up to the limit of the trapdoor and within the space enclosed by the mucosa mass. Fully open (c), the spaces become purely rectangular prisms.

Figure 3.25: Areas enclosed during simulated phases of aryepiglottic oscillation: (a) fully closed, (b) partially open, and (c) fully open.

Pressures within the epilaryngeal air space are calculated according to Figure 3.26. These equations for the vocal folds can be found in Titze (1973) and are applied by analogy to the aryepiglottic folds. Forces are likewise calculated as in Titze (1973) (and M92 for the trapdoor torque).
Figure 3.26: Pressure changes in the epilarynx. Formulation based on T73 and M92; $K_R$ = pressure recovery coefficient (Ishizaka & Flanagan, 1972); $L_E$ = lumped epilaryngeal inductance; $P_E$ = input pressure to epilarynx; $P_{E1}$ to $P_{E4}$ = pressures acting on the aryepiglottic trapdoor and mucosa; $P_{BJ}$ = pressure loss at a “Bernoulli” junction (kinetic pressure loss); $R_E$ = lumped epilaryngeal resistance; $U_E$ = epilaryngeal airflow; $U'_E$ = time derivative of epilaryngeal airflow.

As in T73 and the ST95 model, solution of the differential equations governing the system was non-simultaneous. The model relies upon MATLAB’s ode23s function for numerical integration, which works well for stiff systems and was more successful than the ode45 routine. Stiffness arises in the present model from the inductive and capacitive elements within the glottal and epilaryngeal air spaces operating alongside the relatively low frequency spring-mass components of the mechanical model. McGowan’s (1992, p. 2906) RK4 approach uses a $10^{-6}$ s time step to achieve stability; with ode23s,
stability in the present model is achieved with an adaptive time step determined by the function, which on average is about $10^{-5}$ s.

3.3.3 Simulation results

This section first illustrates model performance with the following simulations: phonation without aryepiglottic vibration (Figure 3.27), voiceless aryepiglottic vibration (Figure 3.28), and voiced aryepiglottic vibration (Figure 3.29). These simulations vary exclusively in regard to initial configuration of the vocal and aryepiglottic folds. The first simulation (Figure 3.27) demonstrates the normal/modal behaviour of the vocal folds and generally mirrors the behaviour of the ST95 model.

Figure 3.27: Simulation of vocal fold vibration (no aryepiglottic vibration). Upper plot: right side vocal fold mass displacement (solid line = lower mass; dashed line = upper mass; dash-dotted line = body mass). Lower plot: glottal volume velocity.

In the second simulation (Figure 3.28), the vocal folds are abducted but the aryepiglottic folds begin adducted to the epiglottic tubercle: the result is voiceless
“trilling”. The frequency response of the aryepiglottic folds as they are parameterized is 53 Hz, which agrees with empirical observations (see earlier in this chapter).

Figure 3.28: Simulation of voiceless aryepiglottic “trilling”. Upper plot: angular displacement of the right side aryepiglottic trapdoor (left axis, solid line) and translational displacement of the mucosa mass (right axis; dashed line). Lower plot: glottal volume velocity.

The third simulation adds vocal fold adduction, which yields voiced aryepiglottic “trilling”. The frequencies of both oscillating systems become lowered: glottal F0 drops by 2 Hz while aryepiglottic F0 drops 9 Hz. The glottal volume velocity is now amplitude modulated in correspondence with the aryepiglottic pulse; the harmonic relationship between the vibratory frequencies of the two sets of folds causes the amplitude modulation pattern to appear as a period alternation. Curiously, it is during the aryepiglottic closure phase that the larger of the two glottal pulses occurs. The phasing of the vocal fold mucosa also becomes perturbed: during aryepiglottic closure the phase
difference between vocal fold cover masses decreases, but increases during the open phase of the aryepiglottic pulse.

Figure 3.29: Simulation of voiced aryepiglottic “trilling”. Top plot: vocal fold mass displacements (see Figure 3.27); Middle plot: angular displacement of the right side aryepiglottic trapdoor (left axis, solid line) and mucosa mass (right axis; dashed line) displacement. Bottom plot: glottal volume velocity.

The amplitude modulation generates subharmonic (s) structure in the (glottal) voice source spectrum, as evident in Figure 3.30c. These subharmonics roughly correspond with the harmonics (h) generated by voiceless aryepiglottic vibration (Figure 3.30b); the harmonics in Figure 3.30c roughly correspond with those produced in the case of modal vocal fold vibration without aryepiglottic oscillation (Figure 3.30a). Thus, the subharmonics in Figure 3.30c are due to the amplitude modulation of the glottal source by the aryepiglottic folds. It should also be noted that the harmonics produced by
voiceless aryepiglottic trilling in the model are much stronger than those which occur in reality, as Moisik et al. (2010) observe, since the increased volume flow will generate substantial turbulence and consequent noise in addition to general irregularity of AE vibration, which cannot be easily modelled with a lumped-element model.

Figure 3.30: A comparison of glottal source spectra. Spectrum (a) corresponds with normal phonation (Figure 3.27) Spectrum (b) corresponds with voiceless aryepiglottic trilling (Figure 3.28); Spectrum (c) corresponds with voiced aryepiglottic trilling (Figure 3.29). Harmonics are marked (h) and subharmonics are marked (s).

Additional simulations, shown in Figure 3.31 and Figure 3.32, help to further demonstrate two important features of aryepiglottic trilling: perturbation to the aryepiglottic pulse (Figure 3.31) and larger scale amplitude modulation of the glottal pulse (Figure 3.32).

The first additional simulation involves adjustment of vocal fold parameters to produce a lower rate of vocal fold vibration (Figure 3.31), at around 65 Hz. The result is a stronger frequency mismatch between the two pairs of folds: not only is there the expected pulse alternation of the vocal folds, but, in this case, the aryepiglottic folds
exhibit frequency perturbation (compare period-duration of lines $a$ and $b$ in Figure 3.31 – without perturbation, the periods would be equal). This demonstrates the mutual influence that the structures have on each other because of their aero-acoustic coupling. Furthermore, during every second aryepiglottic pulse, there is glottal leakage (thin arrow) during the glottal closure phase caused by a premature abductory motion of the lower-mucosal-mass of the vocal fold cover (thick arrow), and this results in an increased cover-mass phase delay.

![Figure 3.31: Simulation of voiced aryepiglottic trilling at a G-F0 of 65 Hz. Upper plot: right side vocal fold mass displacement (solid line = lower mass; dashed line = upper mass; dash-dotted line = body mass). Middle plot: angular displacement of the right side aryepiglottic trapdoor (left axis, solid line) and mucosa mass (right axis; dashed line) displacement. Lower plot: glottal volume velocity.](image)

The second additional simulation (Figure 3.32) involves manipulation to vocal fold parameters such that glottal frequency is increased to 124 Hz. Instead of a period
alternation pattern (a strong-then-weak glottal flow pulse pattern), amplitude modulation now occurs over a larger set of glottal pulses (approximately 3, indicated by the thick, dashed gray line which impressionistically represents the amplitude modulation envelope and corresponds with the aryepiglottic pulse). Evidently, the anharmonic relationship between glottal and aryepiglottic frequencies causes drifting of the amplitude modulation across the glottal pulse sequence.

Figure 3.32: Simulation of voiced aryepiglottic trilling at a G-F0 of 124 Hz. Upper plot: right side vocal fold mass displacement (solid line = lower mass; dashed line = upper mass; dash-dotted line = body mass). Middle plot: angular displacement of the right side aryepiglottic trapdoor (left axis, solid line) and mucosa mass (right axis; dashed line) displacement. Lower plot: glottal volume velocity.

3.3.4 Discussion

The model can simulate aryepiglottic trilling in the range of 40 to 60 Hz, with phase lag values between the trapdoor and mucosa mass ranging from 24 – 64°. From
experimentation, it was apparent that the major factor determining the frequency of the system is the stiffness of the trapdoor spring.

While aryepiglottic vibration is typically reported at these frequency values, it is possible for rates as high as 90 to 100 Hz to occur, depending on the glottal frequency and the tightness of the configuration (Moisik, Esling, & Crevier-Buchman, 2010). The higher frequency trilling likely involves far less effective mass than that which characterizes the aryepiglottic model presented here. Specifically, as more posteroanterior narrowing occurs, a greater portion of the aryepiglottic fold body mass is prevented from oscillating (from contact and compression forces acting on the tissue). Gradually, only the upper mucosal border of the aryepiglottic fold oscillates. This is the case for the Moisik et al. (2010) study and the study of the Iraqi Arabic data in §3.2.

Prior evaluation (Moisik, Esling, & Crevier-Buchman, 2010) of high-speed laryngoscopy with simultaneous EGG data indicates that the changing aryepiglottic aperture during voiced aryepiglottic trills influences the dynamical behaviour of vocal fold vibration: the tendency is for the vocal folds to oscillate in an alternating pattern of rapid and delayed closures that depends on the state of the aryepiglottic aperture above the vocal folds. This is illustrated in Figure 3.33: comparing the EGG with the aryepiglottic aperture reveals a close match between the simulated (see, e.g., Figure 3.29) and real trills. In reality, one can infer that during aryepiglottic closure, the glottal opening phase is extended and closure happens very rapidly (thin arrow corresponding with the sharp peak in the EGG signal), suggesting tighter alignment of the upper and lower cover regions of the vocal fold mass. During the aryepiglottic opening phase, the glottal pulse shows a delayed closure pattern suggesting a de-phasing of the cover and a
shorter glottal aperture during this part of the cycle. The implication is that the glottal flow will be increased during aryepiglottic closure and decreased during aryepiglottic opening, which the model confirms. Furthermore, it is evident that in the opening phase of the aryepiglottic cycle, the vocal fold cover masses are less in phase than they are during aryepiglottic aperture closing. This too matches the pattern seen in the EGG data. Thus, there is evidence that the model is physiologically realistic. A key difference, however, is that real aryepiglottic trilling is highly irregular, but the model never deviates from this basic pattern. Simulation of turbulent effects and pulse irregularity require much more sophisticated modeling.

Figure 3.33: Vocal fold closure alternation pattern during voiced aryepiglottic trilling from data analyzed in Moisik et al. (2010). Top: EGG; Middle: kymograph of the right aryepiglottic fold; Bottom: aperture of the right aryepiglottic fold. Bold arrow: delayed glottal closure; Thin arrow: rapid glottal closure; Dashed vertical line: aryepiglottic closure; Solid vertical line: aryepiglottic opening.
Contrary to what is typically observed for aryepiglottic vibration, the model cannot simulate irregular motion very effectively (Edmondson, Esling, Harris, Li, & Lama, 2001; Moisik, Esling, & Crevier-Buchman, 2010). Some period perturbation did become manifest when physical properties of the aryepiglottic and vocal fold systems were somewhat incommensurate (Figure 3.31), but generally the aryepiglottic motion in the model was highly regular, and consequently, harmonic/subharmonic structure (depending on whether the simulation was voiceless/voiced) was unrealistically strong and well defined (cf. §3.1.4).

Sakakibara et al. (2004) observe that the acoustic power of subharmonics in growl (vocal-aryepiglottic phonation in their study, i.e. voiced aryepiglottic trilling) is weaker relative to kargyraa (vocal-ventricular phonation in Tuvan and Mongolian Khöömei singing style); this might be attributable to oscillatory irregularity that characterizes aryepiglottic fold dynamics. In comparison to kargyraa, which exhibits surprisingly regular/periodic ventricular pulsing at typically half the glottal frequency, and vocal-ventricular vibration in general (Tsai, Wang, Wang, et al., 2010, p. 210), aryepiglottic oscillation fluctuates between quasi-periodicity and aperiodicity (Edmondson, Esling, Harris, Li, & Lama, 2001; Moisik, Esling, & Crevier-Buchman, 2010). Part of the distinction between the two types of vibration (aryepiglottic and vocal-ventricular) is the fact that aryepiglottic oscillation is characterized by topologically complex possibilities for air-flow channels (the left and right aryepiglottic apertures; see §3.1.3) which can vary in time – all of which leads to greater complexity in the system in comparison to the vocal-ventricular mode. Strong periodicity in the flow signal corresponds with stronger harmonics in the frequency domain since there are strong sinusoidal correlations with the
signal; irregularities or aperiodicities result in weak harmonic structure or harmonic smearing since particular sinusoidal components are poorly correlated with the signal and consequently the acoustic energy is dissipated over larger bands of sinusoids.

Ventricular oscillation has been observed to pass through a transient, onset phase of irregular vibration, but it eventually locks into the oscillatory regime of the vocal folds (Fuks, Hammarberg, & Sundberg, 1998; Lindestad, Södersten, Merker, & Granqvist, 2001) and vibrates at 1/2, 1/3, or equal to the glottal F0 (Sakakibara, Fuks, Imagawa, & Tayama, 2004). This frequency locking results from strong entrainment (synchronization tendency) of the two oscillating systems (the vocal folds and ventricular folds) which are coupled to a common aerodynamic source, i.e. the laryngeal air stream. The aryepiglottic folds have the potential to become entrained with the vocal folds, in the case of voiced aryepiglottic trilling, during which point there is a degree of periodicity in the system. These instances, however, are relatively uncommon in comparison to the more irregular vibratory occurrences. Thus, the time evolution of the (real) system erratically shifts in and out of moments of vocal fold entrainment.

The difference in entrainability between the ventricular and aryepiglottic systems might directly relate to their geometry. The ventricular folds are oriented parallel to the vocal folds, undergo displacement in the same direction, and lie immediately above them; on the other hand, the spatial orientation of the aryepiglottic folds and their displacement are roughly perpendicular to those properties of the vocal folds, and the aryepiglottic folds are also further away. Furthermore, the air channel between the vocal and ventricular folds is less tortuous, and thus we would predict the flow will exhibit less turbulence. On the other hand, the bifurcated or even trifurcated flow channel of the
aryepiglottic region is likely to induce considerable turbulence. The assumption here is that turbulence, in conjunction with the geometric factors, will promote irregularity in the oscillation of the aryepiglottic folds through aero-mechanical coupling, hindering entrainment with the glottal pulse. A further factor leading to irregularity could be the mucosal adhesion between the epiglottic and aryepiglottic surfaces (Esling, Zeroual, & Crevier-Buchman, 2007, p. 586; Moisik, Esling, & Crevier-Buchman, 2010). If the adhesive forces are strong enough, they could interfere with consistent patterns of separation between these surfaces during trilling.

At the phonetic level, it appears that aryepiglottic trilling is linked with the production of low tone targets (Rose, 1989; Edmondson & Esling, 2006). This could stem from several factors, including the inherently low frequency nature of epilaryngeal vibration, the presence of subharmonics in the case of voiced epilaryngeal vibration, and the perception of voices with epilaryngeal vibration as low pitched (Teshigawara, 2003). The model predicts that, through aerodynamically mediated entrainment, the vibratory frequency of the vocal folds will drop during aryepiglottic vibration. Even though this effect is small in the case of the model, if it occurs in real AE vibration, then this too could help explain the distribution of aryepiglottic trilling with low tone productions. The other phonetic purpose concerns the study of pharyngeal consonants, for which aryepiglottic trilling is attested as variant of both voiced and voiceless types in Iraqi Arabic (see §3.2) and is distinctive in relation to non-trilled pharyngeal consonants in languages such as Burkikhan Agul. In the case of Iraqi Arabic, there is evidently a difference between voiceless and voiced trills in terms of laryngeal tension and this is reflected in the oscillatory frequency of aryepiglottic trilling in the two cases. Lower
tension in the voiceless case corresponds with a lower oscillatory frequency. Presumably the cause of increased tension in the voiced case is partially associated with vocal fold adduction, but it may reflect the generally “tight” variant of the Arabic approximant (Heselwood, 2007).

The aryepiglottic-vibration model can be compared to simulation of vocal-ventricular vibration. The result suggest that all forms of voiced epilarynx vibration lead to a very similar effect: amplitude modulation of the glottal pulse. Furthermore, in light of the results from the vocal-ventricular coupling model (which will be presented in §4.3), we can also state that even mechanical interaction can lead to this effect. Stepping back then, what we have is a biomechanically robust system for generating a class of effects groupable as auditorily harsh and “growly” (Gerratt & Kreiman, 2001). The tendency towards upper epilaryngeal vibration in phonetic contexts reflects the nature of the sounds being produced: they are pharyngeal, which are formed with substantial epilaryngeal stricture. It is doubtful that vocal-ventricular vibration of the form observed in throat singing could occur in the context of these sounds: the larynx is too constricted.

Throat singing seems to require trained and deliberate adjustment of the epilaryngeal mechanism to “free-up” the ventricular folds from the vocal folds by lowering the larynx to allow them to freely vibrate. This unusual posture – both constricted (narrow lower epilarynx) and anti-constricted (lowered larynx) – may explain why one must undergo training to master the throat singing technique.

Accurate and precise prediction of the time course of aryepiglottic vibration would require advanced computational resources and modeling. Given that the effects of aerodynamic turbulence on account of structural tortuosity and its aero-mechanical
interaction are still poorly understood, it is doubtful that we will be making these predictions any time soon. Furthermore, aryepiglottic vibration can occur with or without vocal fold vibration, giving rise to voiced and voiceless phonetic variants.

The observation that the mechanical system is inherently predisposed to generating amplitude modulation of a glottal source, if one is present, is important for the auditory and phonological domains too. The robustness of the mechanism to produce a perceptually distinctive (Gerratt & Kreiman, 2001) phonatory quality makes it a good candidate for incorporation into phonological systems, such as its use in sphincteric or epiglottalized register in !Xôô and other Khoisan-group languages, even if this is not particularly common. The generation of subharmonics and the tendency towards lowered glottal F0 also help to explain why this phonation type occurs as a correlate of low tone targets, as part of the harsh register in Bai and as a phonetic variant of low tone realizations in Zhenhai Wu, and related dialects. These issues are addressed further in §7.1 and §7.1.4.

From a physical standpoint, a key feature of aryepiglottic trilling is the way the varying aryepiglottic aperture influences the acoustic loading on the vocal folds, leading to destabilization from their normal oscillatory behaviour, which is what is observed in the empirical studies. It is suspected that the changing aryepiglottic aperture causes the degree of acoustic coupling between the vocal folds and the rest of the vocal tract to be a causal factor for the destabilization of vocal fold oscillation. This relates directly to Titze’s (2008) nonlinear source-filter coupling theory, where the degree of interaction between the vocal fold source and the vocal tract filter is directly controlled by the cross-sectional area of the epilaryngeal tube. In Titze’s model, however, this area is assumed to
remain static in time, while in the case of aryepiglottic trilling, this area is continually in flux.

3.3.5 **Summary of the aryepiglottic vibration model**

The model presented here formalizes the low degree-of-freedom, lumped-element mechanics of the laryngeal oscillating system, which, in this simplified case, comprise only the vocal folds and the aryepiglottic folds but not the ventricular folds. It is based upon McGowan’s (1992) model of tongue-tip trilling and lumped element models of the vocal folds and air way (Titze, 1973; Story & Titze, 1995). The mechanical conceptualization of the aryepiglottic folds in the present model parallels the trapdoor system specified for the tongue-tip in McGowan’s model, the key difference being that there are two trap doors instead of one: the aryepiglottic folds are paired structures.

The model is an improvement over that presented in Moisik (2008) as it allows for evaluation of both voiced and voiceless conditions, and actually simulates aero-acoustic coupling between the vocal and aryepiglottic folds. However, the present model can also be greatly improved upon. Measurement of the biomechanical properties of the aryepiglottic folds has received very little attention, and the parameters here were defined through analogy with other structures. In the aero-acoustic domain, a more realistic pressure loss model that accounts for turbulent losses would be an improvement, but even better would be a detailed account of turbulence generation within the narrowed epilaryngeal airway. To address the time-varying impact on the non-linear source-filter coupling effects identified by Titze (2008), a more realistic acoustic propagation model
needs to be used. In the interest of simplicity, only a simple resistance was used. Future research will need to address these matters.

The voiceless case demonstrates the self-sustained oscillation of the aryepiglottic folds; considerable turbulence should arise based on the high levels of volume flow through the glottis required to drive aryepiglottic vibration, and this turbulence will severely hamper the acoustic efficiency of the aryepiglottic folds as an independent source. The more complex case in which the vocal folds oscillate simultaneously with the aryepiglottic folds suggests there is aero-acoustic interactivity between the pairs of folds. The key feature is the amplitude modulation of the glottal pulse: this is what characterizes the source nature of the aryepiglottic folds. The interpretation then is that the vocal and aryepiglottic systems form a complex source-generating mechanism; a similar conclusion is offered by Lindestad et al. (2001) for simultaneous vocal and ventricular fold vibration. It has yet to be established whether epiglottal vibration produces similar effects, or whether voiced epiglottal vibration is even possible, and so this is a point for future research to consider.

3.4 Chapter summary: Epilaryngeal vibration

This chapter has contributed anatomo-physiological, phonetic, and biomechanical observations to a phenomenon that has hitherto received very little attention – epilaryngeal vibration. There is some early work hinting at what might be occurring in “growling”, and a vocal-ventricular vibration mode has been clearly established to occur in “throat singing” traditions, but no theory of the overall vibratory capacity of the epilarynx in non-pathological speech has ever been put forth. The broad implication of
this theory is that amplitude modulation and subharmonic generation (impressionistically “growling”, auditorily “harsh voice”) is a fundamental effect that epilarynx vibration has on the glottal voice source.

A key phonetic problem is classification of epilaryngeal vibration: is it phonation or trilling? A rubric for classifying vocal tract vibrations was developed to provide a means to distinguish between trill and phonation. The interpretation, based on this rubric, is that it is a bit of both, which is attributed to its unique location in the vocal tract between the “phonating” and the “trilling” structures. The phonetic context it occurs in can help guide the classification – its occurrence in register (as in !Xóõ) is more phonation-like, and its occurrence with pharyngeals (as in Iraqi Arabic) is more trilling-like. It will be argued in §7.1.3 that the classification has phonological significance.
Chapter 4
THE EPILARYNX AND THE VOCAL FOLDS

When the epilarynx undergoes sufficient narrowing, its tissues start to compress together. There is compaction between the surface of the epiglottis and the ventricular folds below, and more importantly there is evidently cranial-caudal apposition of the ventricular folds against the vocal folds. The observation of ventricular involvement in glottal stop, glottalization, and laryngealization is not new and actually appears in a surprisingly large number of places (Lindqvist, 1969; Lindqvist-Gauffin, 1972; Allen & Hollien, 1973; Gauffin, 1977; Iwata, Sawashima, Hirose, & Niimi, 1979; Roach, 1979; Laver, 1980, pp. 122–123; Painter, 1986, 1991). Despite this evidence, it has not penetrated mainstream phonetic or phonological theory of these sounds. Furthermore, the connection between vocal-ventricular contact and more general compaction of the epilarynx, clearly articulated by Lindqvist-Gauffin (1972), has received only superficial attention until recently (see Lindblom, 2009).

This chapter builds on the previous research in two ways. First, vocal-ventricular contact is visualized using laryngeal ultrasound in §4.1. This study demonstrates function of the lower border of the epilarynx in glottal stop as a means to dampen and arrest vocal fold vibration. In §4.2, simultaneous laryngoscopy and laryngeal ultrasound (SLLUS) data for glottal stop and creaky voice are presented. Finally, an in-depth analysis of the mechanical influence of vocal-ventricular contact on vocal fold dynamics is presented in §4.3. The chapter is summarized in §4.4.
4.1 Visualizing vocal-ventricular contact with laryngeal ultrasound

The goal of this study\textsuperscript{93} is to present a methodology of how laryngeal ultrasound may be used to bring more evidence to bear on ventricular involvement in sounds such as glottal stop. The “glottal” in glottal stop implies that the sound symbolized as [ʔ] simply involves contact of the \textit{rima glottidis} – the variable “chink” delimited by the medial surfaces of the vocal folds running from the anterior to posterior commissure and divided longitudinally into membranous/ligamentous and cartilaginous regions. To qualify phonetically as a glottal stop, this contact should result in the is cessation of vocal fold vibration and correlate with silence in the acoustic signal. The larynx is endowed with muscles that can both abduct (posterior cricoarytenoids) and adduct (lateral cricoarytenoids and interarytenoids) the vocal folds, effectively widening or narrowing the glottis, respectively. This is where the account of glottal stop usually ends in traditional phonetic and phonological theory (e.g. Halle & Stevens, 1971).

Catford outlines an extended model of laryngeal stops based on his experience with Caucasian languages (e.g. 1977b). In several languages, multiple laryngeal stops ostensibly coexist within the phonological system, and thus Catford acknowledged two zones of laryngeal control in his classic \textit{Fundamental Problems in Phonetics} (1977a, p. 163): glottal and ventricular (the latter defined by the ventricular folds). Catford points out that ventricular fold constriction “almost certainly involves simultaneous glottal stop” and perhaps an increased constriction of the lower pharynx.

\textsuperscript{93} The original laryngeal ultrasound study was presented at Ultrafest V (March 21\textsuperscript{st}, 2010, New Haven, Connecticut); see Moisik (2010). The theoretical discussion (and Figure 4.11), methods for SLLUS, and the SLLUS data (Figure 4.7, Figure 4.8, and Figure 4.9) for glottal stop and the two variants of creaky voice are published in Esling & Moisik (2012). Text, analysis, and interpretations are attributed primarily to Scott R. Moisik with assistance from John Esling.
Extensive laryngoscopic observation reported by Esling and colleagues (see §2.2.3) reveals that the vocal-ventricular closure that Catford described is found not just in Caucasian laryngeal stops but actually commonly occurs in ordinary glottal stop production, which corroborates the scattered observations of the early studies noted above (see the introduction to Chapter 4). Esling et al. (2007, p. 585) go as far to suggest that “‘glottal stop’ cannot be uniquely glottal but is at least glottal-ventricular” and ventricular involvement can even occur in “lax” varieties of glottal stop and in creaky phonation, all of which is accompanied by moderate anteroposterior narrowing of the epilarynx. As Esling (1999, p. 353) observes, however, the nature of the interaction suspected to occur between the vocal folds and ventricular folds during this manoeuver, particularly in regard to the effect on the overall vibratory mass, has not yet been determined.

Part of the difficulty in coming to understand this vocal-ventricular fold interaction has been in simply visualizing the contact. With the exception of x-ray imaging studies (e.g. Griesman, 1943; Ringgaard, 1962; Allen & Hollien, 1973; Fink, 1974a) (not all of which are strictly phonetic in nature), most of the empirical evidence comes from laryngoscopic observations. As discussed in §2.2.1, while laryngoscopy provides visualization of axial-plane action, the impression it provides of vertical relations is limited at best. From the x-ray evidence, it is known that the vocal folds can actually come into contact with the ventricular folds during constricted postures (such as the production of creaky voice), which mostly obliterates the ventricle in the process.

Although still relatively new in linguistic research, ultrasound has primarily been used to study the position and motion of the tongue in speech and best practices and
special techniques for analysis have started to emerge in the literature (e.g. Stone, 2004; Mielke, Baker, Archangeli, & Racy, 2005; Davidson, 2006). The larynx has received far less attention using ultrasound. Some early attempts were made with M-mode ultrasound to study vocal fold vibration (Mensch, 1964; Hertz, Lindström, & Sonesson, 1970). Although not strictly linguistic in nature, researchers have recently used colour Doppler imaging to study the surface mucosal waves of the vocal folds (Shau, Wang, Hsieh, & Hsiao, 2001) and ventricular folds (Tsai, Shau, & Hsiao, 2004).

Medical imaging studies (e.g. Harries, Hawkins, Hacking, & Hughes, 1998; Loveday, 2003; Sonies, Chi-Fishman, & Miller, 2002) attest that laryngeal ultrasound can be used to image a wide array of laryngeal structures, such as the hyoid bone, strap muscles, pre-epiglottic space, and the epiglottis, thyroid, arytenoid, and cricoid cartilages. According to this literature, the vocal folds mainly appear as darkened regions (due to acoustic scattering in muscle tissue), while the ventricular folds are usually visible, and the laryngeal vestibule itself generates bright echoes due to the presence of air pockets. Acoustic “blind spots” generated from the laryngeal air column partially obscure the arytenoids, posterior commissure, piriform fossae, aryepiglottic folds, and the hypopharynx. Calcification/ossification of laryngeal cartilages can also negatively impact laryngeal imaging with ultrasound.

The work presented here demonstrates that ventricular component of laryngeal articulation can be fruitfully studied with ultrasound. The focus is on glottal stop production, and while the lateral-medial action of the vocal folds does not image well, the vertical changes are imageable, and the descent of the ventricular folds, even in conventional glottal stop, is one of the striking features of this data.
4.1.1 Laryngeal ultrasound methodology

Laryngeal structure and articulation during glottal stop is examined within the coronal plane using laryngeal ultrasound. The examination was conducted with a portable LOGIQ e R5.0.1 system and a convex 8C-RS probe (both manufactured by General Electric Corporation). The probe pulse frequency was 10 MHz, which allowed for optimal resolution of laryngeal structures. The field of view was consistently set to 120°. The system is pre-calibrated for measurement in the imaging plane; a ruler on the image allows for a pixel-to-mm scaling factor to be determined. A Sennheiser ME66-K6 shotgun microphone was used to record audio, digitized at 44100 Hz (16 bit), using an M-Audio Mobile Pre-Amp as an external sound card. The video of the ultrasound machine was captured using an XtremeRGB video card at 30 fps (uncompressed, 8-bit greyscale, 1024 × 768 pixels) and both signals were integrated and manually checked for alignment using Sony Vegas Pro (version 8.0b).

To conduct the laryngeal ultrasound, the ultrasound probe was applied manually to the participant’s right thyroid lamina near the laryngeal prominence (Figure 4.1a & b) and orientated to obtain a coronal image of the larynx.
Figure 4.1: Illustration of ultrasound probe placement. In (a), the gray elliptical region shows the approximate probe target area on the neck; (b) is a side view illustrating probe positioning on the neck (with the hand kept clear for visibility of probe location); (c) illustrates how the probe was held; (d) shows a side view of thumb anchoring and index finger contact on the neck; (e) illustrates probe positioning from the front.

In an effort to achieve the best probe stability possible, the following approach was used. During each examination, the participant was seated in an examination chair equipped with a head rest to help provide stabilization. The probe was held such that the examiner’s thumb was free (Figure 4.1c) to anchor on the side of the participant’s neck.

---

94 This figure appears in Moisik, Esling, & Lin (forthcoming).
while the index finger was rested lightly on top of the probe (on the orientation notch) for orientation stability and to press lightly against the participant’s upper neck (below the chin) to help minimize probe migration (Figure 4.1d & e). Before elicitation commenced, the participant was instructed to produce an [i] vowel at a normal pitch so that the vocal folds could be located and centered in the ultrasound view. Occasionally, probe readjustment was required. After readjustment, the participant was instructed to swallow and to produce glides into high and low pitches to ensure that there was still a reasonable registration of larynx height change.

Using the above methodology, five trained phoneticians and one lay-person (three males and three females; five between 25-30 years of age, one, a female, at 45 years of age) were instructed to produce the sequence [iʔi].

4.1.2 Results

Structural registration in the coronal plane (Figure 4.2) for all participants consistently allowed for the ventricular folds to be identified, which can be attributed to their hyper-reflectivity (Loveday, 2003), but general structural visibility and image quality varied by individual, with males registering better than females, probably because of variation in tissue density and structural orientation (Loveday, 2003; Sonies, Chi-Fishman, & Miller, 2002).
Figure 4.2: Structural registration in coronal laryngeal ultrasound. (a) laryngeal ultrasound image; (b) schematic. AE = aryepiglottic fold; FF = ventricular (false) fold; P = probe location; VF = vocal fold.

The 45-year-old female exhibited the poorest registration, which may be due to increased ossification of laryngeal cartilages. Vocal fold vibration during phonation is easily detected as flickering in the video image, and the vocal ligament and medial edge of the vocal fold produced a moderately bright echo. Overall, however, it was found that the vocal fold produced generally poor registration, especially within the area corresponding to the body (i.e. thyroarytenoid muscle). Using phonation to locate the vocal folds and thereby establishing a frame of reference greatly facilitates interpreting the laryngeal ultrasound image.

Due to poor registration of the vocal folds and the inability to judge the location of the vocal fold opposite to the side being imaged because of the laryngeal air column, it is difficult to ascertain glottal state purely from the coronal laryngeal ultrasound image. The ventricular folds did image well across subjects, although, once again, lateral-medial motion was difficult to judge without a clear idea of the location of the laryngeal midline.

---

95Figure appears in Moisik, Esling, Bird, & Lin (2011).
Despite this, vertical changes in ventricular fold position were manifest, and this means that the vertical dimension of laryngeal constriction can be observed. The most important finding related to this is that across all subjects the ventricular folds gave the impression of compressing into the superior surface of the vocal folds. Without crisp resolution of the vocal fold, it is difficult to know whether contact is actually made between the vocal folds and the ventricular folds, but it seems probable given the extent of motion observed and the fact that the ventricle height is around 2-3 mm in modal phonation (Agarwal, Scherer, & Hollien, 2003, pp. 101, 105), which means that there is not a great deal of distance separating the structures even under ordinary conditions. Figure 4.3 illustrates this for a male participant. The corresponding spectrogram is consistent with glottal stop: there are no formant transitions into the consonant and slight creakiness precedes and follows the sound.

Figure 4.3: Coronal laryngeal ultrasound of glottal stop. FF = false (ventricular) folds; VF = vocal folds; P = probe.

---

96 Figure appears in Moisik, Esling, Bird, & Lin (2011).
4.1.3 Discussion

This research corroborates the early accounts of vocal-ventricular fold contact occurring in glottal stop, and suggests that this contact probably occurs in laryngoscopic observations even though it is not directly visible. It also seems possible that the contact might occur even without complete medialization of the ventricular fold. It is difficult to judge from the laryngeal ultrasound how much medial motion of the ventricular fold occurs, but the impression is that the motion is largely downwards. As discussed in §2.1.2, we can attribute this downwards motion to the action of the external thyroarytenoids (M. M. Reidenbach, 1998a; Reidenbach, 1997).

As Lindqvist-Gauffin (1972) argued, the physiological mechanism involved in the production of laryngeal sounds bears strong resemblance, if not its provenance, to life-supporting functions of the larynx. The larynx can be used to trap air in the lungs for the purpose of building intrathoracic pressure. Such a manoeuvre is necessary in coughing, but also plays a role in providing support for increasing pressure within the abdominal viscera (Negus, 1949, p. 111). The vocal folds acting alone are ill-suited to air trapping: the concave downwards, “gothic arch” shape of the subglottal space results in lateral aero-mechanical force on the folds when subglottal pressure rises (see §2.1.2, Figure 2.7) – good for glottal opening during phonation, but not optimal for building up thoracic pressure. The optimal shape would be convex downwards: air pressure would act in a medial, closing direction in this case, minimizing muscular effort required to maintain closure. In fact, for animals where laryngeal/tracheal inundation is a threat (such as creatures that dwell in an aquatic environment) or which need to prevent air from entering the lungs, the “thyroarytenoid folds” (i.e. vocal folds) form an upturned, pointed
or convex shape (Negus, 1949, pp. 98–100). In the human, whose upturned vocal folds reflect arboreal ancestry, downturned ventricular folds are present to assist in closure during air-trapping by buttressing the vocal folds. Vocal-ventricular fold contact is therefore an intrinsic part of laryngeal closure functioning to prevent air passage and, as a consequence, vibration. Such an account helps to clarify why ventricular incursion might occur in speech during glottal stop, although its engagement is not always directly observed using laryngoscopy (Edmondson, Chang, Hsieh, & Huang, 2011), particularly as speech rate increases (Iwata, Sawashima, Hirose, & Niimi, 1979). It could be that ventricular fold participation in glottal stop sequentially progresses from vertical vocal-fold contact (not visible in laryngoscopy) and then to medialization of the ventricular folds (which is visible in laryngoscopy). It is also possible that, with sufficient medial compression, vocal fold adduction alone is all that occurs in some cases. Future studies will be required to determine the status of ventricular incursion as intrinsic to glottal stop (and laryngealized phonation), but its occurrence in some glottal stops cannot be contested, so it warrants explanation.

Although the data in the study presented above involves artificial phonetic productions, the laryngeal ultrasound technique represents a valuable new way to collect data for these sounds, and improves over laryngoscopy in several ways: it is non-invasive, relatively inexpensive (unlike MRI), and it provides information about the vertical component of the constriction occurring in these sounds. Such a technique could be valuable, and could help to conclude more decisively whether vocal-ventricular contact is occurring, which cannot always be determined successfully with laryngoscopy (cf. Brunelle, Nguyễn, & Nguyễn, 2010).
4.2 SLLUS observation of lower epilarynx function and larynx height

The study in §4.1 demonstrates the vertical component of ventricular fold involvement in glottal stop production. The study here probes deeper into vertical dimension of constricted laryngeal sounds. Simultaneous laryngoscopy and laryngeal ultrasound (SLLUS) are applied in the examination of glottal stop and two different varieties of creaky voice production which vary by larynx height setting. The goal here is to correlate laryngeal state changes visible in laryngoscopic view with larynx height changes and ground these mechanisms in the functioning of the epilarynx.

It is known that larynx height has several speech-related roles. Hardcastle (1976, p. 69) posits two functions of larynx raising: swallowing-type closure (i.e. constriction) and glottalic initiation (either egressive, as in ejectives, or ingressive, as in implosives). The correlation of larynx raising with pitch raising is also well known (e.g. Honda, 2004). Less often considered is how these three different functions of larynx raising relate to each other. In the present study we will focus on the role of larynx height in pitch in comparison to constriction (neglecting glottalic initiation).

The closure in swallowing is an extreme form of (epi-)laryngeal constriction; and it is unsurprising that larynx raising would assist in such a gesture. It is less obvious that larynx raising should be important in a sound such as glottal stop. In standard phonetic and phonological theory “constricted glottis” is purely a vocal fold function. It should be independent of larynx height. Adductive force driven by the interarytenoid muscles maintains cartilaginous glottal closure and the lateral cricoarytenoid muscles manipulate the medial compression or membranous compression of the vocal folds (Laver, 1980, p. 109). Various phonetic possibilities emerge from this model. Glottal stop is produced
with strong adductory force and medial compression. If the vocal folds are set into vibration in this strongly adducted state, then only the anterior, membranous portion of the vocal folds vibrate (the posterior, cartilaginous glottis being firmly closed by the adductive force). Catford pointed out, however, that during this “anterior voice” (as he called it) the “whole upper part of the larynx may be constricted to some extent”, going on to infer that “it appears as if the arytenoidal constriction essential for anterior phonation is part of a general sphincteric constriction of the (upper) larynx”.

The vertical component of constriction is intimated in Catford’s statement, but its consequences are not fully explored. It has left the epilaryngeal (upper larynx) component dangling as a mechanical byproduct of the fundamental vocal-fold-level mechanism. The approach in the present work is to expand upon Catford’s suggestion that glottal constriction is actually a function of a more holistic action of the larynx, one which must take into consideration the physiological nature of the epilarynx. Once again, we are reminded of Lindqvist-Gauffin’s (Lindqvist, 1969; Gauffin, 1972; Lindqvist-Gauffin, 1972; Lindblom, 2009) assertion that these mechanisms are phylogenetically anchored and not intrinsic to speech, but are rather exapted into speech. The view adopted in the present work is consistent with Gauffin’s conceptualization of laryngeal function in speech.

Sounds such as glottal stop and laryngealized phonatory quality (i.e. creakiness) are generally not thought to involve any intrinsic laryngeal height component since they are conventionally considered glottal-only mechanisms. The model that they engage a

---

97 Catford suggests that anterior voice is equivalent to “tight”, “hard”, “sharp”, and “tense” phonatory qualities reported in various places. Also, he provides the following sociophonetic tidbit: anterior voice is characteristic of phonatory qualities heard in North German and in North East Scotland (Aberdeenshire and Banff).
larger-scale mechanism involving the epilarynx, however, predicts that, if anything, larynx height ought to help produce these sounds since an important part of their production is vocal-ventricular fold contact (as we saw in the previous section).

One difficulty facing this type of research is that quantifying larynx height during speech is not a trivial task (Honda, 2004). Attempts have been made using thyro-umbrometry (Ewan & Krones, 1974; Ewan, 1979; Sprouse, Solé, & Ohala, 2010), but this requires that participants have visible laryngeal prominences, and, since the measurement depends on observing a specific location on the larynx, the technique is subject to being confounded by rotational movements of the larynx. MRI can be used successfully on static postures (e.g. Honda, Hirai, Masaki, & Shimada, 1999) and real-time MRI is certainly promising (Nissenbaum, Kirsch, Halle, et al., 2002; Narayanan, Bresch, Ghosh, et al., 2011), but both techniques remain prohibitively expensive.

In this study, simultaneous laryngoscopy and laryngeal ultrasound (SLLUS) is used to obtain information about larynx state from the laryngoscopic data and information about larynx height from optical flow analysis of the laryngeal ultrasound data. With this approach it is possible to cross-correlate laryngeal postures visible in laryngoscopy with changes in vertical height of the larynx. Using this technique, the relationship between larynx height and laryngeal constriction in glottal stop and creaky voice is demonstrated.

4.2.1 SLLUS methodology

The phonetic production data consist of ordinary glottal stop, “relaxed” creaky voice, i.e. without larynx height suppression, and creaky voice produced with deliberate
and forceful larynx height suppression (i.e. larynx lowering). All productions are in [i] context and numerous productions of each were attempted. The participant is the author (SM). Representative tokens were selected for presentation here.

To conduct the SLLUS technique, a standard laryngoscopic examination was performed, and the attending physician was seated in front of the participant, while the laryngeal ultrasound examination was performed simultaneously by approaching the participant from the side. The laryngeal ultrasound was conducted by following the procedure outlined in §4.1.1.

The laryngoscopy equipment is a Kay 9106 rigid endoscope fitted with a 28mm wide-angle lens to a Panasonic KS152 camera. The ultrasound equipment is a portable LOGIQe R5.0.1 system with a straight-line 12L-RS probe (both manufactured by General Electric Corporation). The probe pulse frequency was 10 MHz, which allowed for optimal resolution of laryngeal structures. Two audio signals were also recorded (at 44100 Hz, 16 bit) for the purposes of signal synchronization and analysis. One was captured along with the ultrasound video using a computer running Sony Vegas; the other was recorded using the camcorder. All signals were integrated, carefully aligned through manual inspection, and segmented into isolated tokens using Sony Vegas. F0 traces were obtained using the STRAIGHT algorithm (Kawahara, de Cheveigné, & Patterson, 1998); it should be noted that STRAIGHT failed to register the creaky voice sections of our recordings.
4.2.2 *Optical flow analysis*

Change in larynx height was observed using laryngeal ultrasound and quantified by means of an optical flow algorithm based on a block-wise, absolute differences method. Gradient methods for optical flow were avoided since ultrasound data do not meet assumptions of smoothness in the brightness pattern of an image sequence (Horn & Schunck, 1981).

The optical flow field is calculated as follows. Each frame from the video sequence is broken down into an analysis grid. For each node in the grid, an analysis block centered about the grid node is obtained from the current frame, and from the next frame an expanded region of pixels three times the size of the analysis block is obtained. To obtain the flow vector for this frame pair, the current pixel block is “swept” across the next pixel block; the algorithm is in some respect like a two-dimensional convolution of images, but takes the sum of the absolute differences between the current and next pixel blocks rather than their products. The formula for calculating this is

\[
D := \sum_{j=1}^{m-k} \sum_{i=1}^{m-k} |N_{i:i+k,j:j+k} - C|
\]

where \( C \) is a \( k \times k \) matrix for the pixel block of the current frame, \( N \) is an \( m \times m \) matrix for the expanded pixel block of the next frame, and \( D \) is an \( m-k \times m-k \) of the resulting sum of matrix differences\(^98\). The indices of the global minimum in \( D \) (with the center entry defined as the origin of the index-coordinate space) are taken to reflect the vertical and horizontal components of object displacement vector imaged by the pair of frames.

\(^98\) Note that for notational convenience, the ‘:’ is used to define submatrices of \( N \) and the vertical bar notation indicates entry-wise absolute value of the difference matrix, *not* its determinant.
The negative of $D$ is thresholded at 95% of its brightness range producing $D_{\text{thresh}}$; the ratio of ones to zeros in $D_{\text{thresh}}$ defines a weighting $\sigma$ that represents the accuracy of the result (smaller ratios indicate greater accuracy). The optical flow field for a pair of frames constitutes the set of all velocity vectors obtained over the field defined by the analysis grid. A video sequence of $f$ frames will yield $f - 1$ velocity fields.

Vertical larynx displacement was calculated as follows. For each velocity field, the weighted average of all its vertical components was obtained using the $\sigma$ values as weights. The results constitute a discrete representation of the vertical velocity. At this point the velocity information is in units of pixels/time, and conversion to mm/s was performed. The ultrasound ruler (in units of cm), which is superimposed on all ultrasound data by the ultrasound machine, was used to determine the pixel-to-mm scaling factor, which is applied to the velocity function. The velocity function must then be numerically integrated (using the trapezoidal method\textsuperscript{99}) to obtain change in larynx height over time. All data analysis was carried out using MATLAB (R2009a).

\textsuperscript{99} This was done using the MATLAB (R2009a) function \texttt{cumtrapz()}. 
Figure 4.4: Validation of the optical flow algorithm. Frames 1 (a) and 2 (b) from the test case video showing a movement of the metal bar approximately 4 mm to the left. The corresponding results from the optical flow analysis of this frame pair are shown in the form of a vector field plot (c).

Validation of the optical flow algorithm was conducted on control data of a metal bar sliding 11.14 cm along a ruler (see Figure 4.4). The velocity of the metal bar was calculated with manual measurements and with the algorithm (Figure 4.5). The normalized RMS error between these two velocity functions is 12.17% and numerical integration of the velocity data obtained from the algorithm yields 11.35 cm, for an error of 1.8%, which is taken here as an acceptable level of error relative to a manual analysis.
Figure 4.5: Velocity and position of the metal bar in the test case as determined by manual measurement (actual; solid line) and by the optical flow algorithm (dashed line). The negative values indicate movement to the left in millimetres.

4.2.3 Qualitative analysis of laryngoscopy

Laryngoscopy data is qualitatively evaluated with reference to three visual indicators of laryngeal constriction (cf. Painter, 1986; Esling & Harris, 2005). Figure 4.6 presents two of these indicators: (i) the posteroanterior dimension of the epilaryngeal tube, and (ii) the positioning of the ventricular folds. The third indicator is the elevation of the larynx. In laryngoscopy, changes to larynx height appear as a change in scale of the laryngeal structures; if the larynx elevates, it generally appears larger and vice versa for larynx lowering (Kagaya, 1972).
Figure 4.6: Indicators of laryngeal constriction. ae = aryepiglottic fold; c = cuneiform tubercle; et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold. (i) dashed line = anteroposterior diameter of the epilaryngeal tube; (ii) opposed arrows = ventricular fold medialization.

4.2.4 SLLUS results

Results shown in Figure 4.7, Figure 4.8, and Figure 4.9 for the SLLUS data consist of composite plots comprising audio, F0 trace, and larynx height (vertical larynx displacement) plots, with select laryngoscopy frames below the plots and the temporal locations of these marked on the plots with vertical lines. In the case of Figure 4.8, a spectrogram is included (at the top).
Figure 4.7: SLLUS data for glottal stop. ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold.

An illustration of the SLLUS data for glottal stop is in Figure 4.7. Frame 20 shows modal phonation at the mid-point of the first vowel. The larynx is unconstricted, the vocal folds (v) are adducted, and there is “full glottal” vibration – apparently unimpeded by the ventricular folds (f), which are laterally retracted. The epiglottic

---

100 The audio data shows a high noise floor; this is due to the laryngoscopy and ultrasound equipment.
101 That is, vibration of the membranous and cartilaginous portions of the vocal folds.
tubercle (et) is not visible, only the apex of the epiglottis is (e). Frames 24, 28, and 32 then show the progressive adduction/medialization of the ventricular folds during the glottal stop; some anticipatory ventricular fold adduction is evident at the very end of the vowel (frame 24) and full adduction is attained early during the onset phase of the glottal stop (i.e. prior to the temporal mid-point). This adduction evidently continues to increase in degree of medial compression (frame 28) until offset occurs where release of adduction must take place in anticipation of the upcoming vowel. Also notable is that ventricular fold contact happens first at the anterior edges of the ventricular folds and progressively involves the posterior surfaces as constriction degree increases. During the ventricular fold adduction, there is a reduction in the posteroanterior dimension of the epilarynx which correlates with a slight rise in larynx height. Ewan reports larynx raising for Thai [ʔ] (1979, pp. 73–80) and Roach observes that larynx raising and epilarynx closure occur in his production of glottally-reinforced stops in English (1979, p. 2). The change is on the order of 1.24 mm and occurs in about 0.275 s. Somewhat surprisingly, it continues to increase into the offset phase of the glottal stop, peaking at frame 32, and then rapidly decreases to its level during the following vowel and slowly continues its descent.

Figure 4.8 illustrates the SLLUS data for creaky phonation, which was gradually engaged during the midpoint of an extended [i] sequence. During this production larynx height was not deliberately suppressed (unlike Figure 4.9). The larynx raises by 3.38 mm over the course of 1.100 s. The spectrogram is included in this image to show how F2
exhibits arc-like raising (dashed-underline) in correspondence with the larynx raising\textsuperscript{102}. Note that the raising can be discerned in the laryngoscopy frames as a slight enlargement of the appearance of the structures. Ventricular fold medialization once again is present, although unlike the glottal stop, complete adduction of the ventricular folds does not occur: throughout the creaky section of the performance there is a posterior gap at the ventricular level.

Posteroanterior narrowing of the epilaryngeal tube is comparable with, if not greater than, that which occurs for the glottal stop (Figure 4.7), but the degree of arytenoidal adduction (as judged by the position of the cuneiform and corniculate tubercles) does not increase appreciably during the production (the same is true of the glottal stop production). It seems as if roughly maximal adductory displacement of the arytenoid complex (and vocal folds) is achieved even in the modal configuration (i.e. the adducted phonatory state), as in frame 20.

---

\textsuperscript{102} The target vowel for these sequences was [i], but the quality tends to be centralized because of the presence of the laryngoscope in the oral cavity.
Figure 4.8: SLLUS data for creaky voice without height suppression. ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold. (n.b.: creaky voice not detected by STRAIGHT).
Figure 4.9: SLLUS of creaky voice with deliberate height suppression (larynx lowering). ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold. (n.b.: creaky voice not detected by STRAIGHT).

The production of creaky voice in Figure 4.8 can be compared with that in Figure 4.9, which represents an attempt to produce creaky phonation while deliberately suppressing the height of the larynx (by forceful larynx lowering). While it was found to be possible to produce creak in this lowered larynx state without any obvious ventricular fold adduction, there is a clear state change that occurs with the onset of creakiness at
frame 50. There is a very slight elevation of the larynx at this point which is accompanied by a substantial reduction of the posteroanterior dimension of the epilaryngeal tube; this continues to increase in degree throughout the creaky region (frame 65). Furthermore, at the transitional point (frame 50), the vocal folds abruptly switch vibratory mode from one which allows for loose vibration along their entire longitudinal extent, to one allowing only for limited lateral displacements confined to a short anterior (membranous) section along their longitudinal extent (compare frames 20 and 35 with 50 and 65). Adduction of the arytenoid complex does not appear to appreciably change, and, somewhat surprisingly, the cuneiforms even appear to separate somewhat (compare frame 35 and 50), which, given that the larynx is lower in frame 50 is not expected to occur on the basis of camera perspective alone: rather, we might expect them to appear even closer together. The answer for why this happens might have to do with increased activity of the aryepiglottic muscle during the engagement of constriction in this state.

4.2.5 Discussion

It was observed above that glottal aperture was once thought to be the critical dimension of sounds thought to involve a “constricted glottis”. Figure 4.10 presents the axial plane view of the forces acting to shape the longitudinal dimension of vocal fold vibration. The left column depicts forces thought to be shaping the glottis, and the right column illustrates the vibratory profile of the vocal folds depicted as a black region (think of it as a snapshot taken at the moment of maximal glottal area during vibration).
Figure 4.10: Longitudinal nature of the effect of the epilarynx on the vocal folds. Solid lines = structure edges visible in laryngoscopy; dashed outlines = cartilage structures; a = arytenoid cartilage; ae = aryepiglottic fold; c = cuneiform cartilage; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold.

Modal voice (Figure 4.10a) involves the “full glottis”, i.e. the cartilaginous and membranous regions of the vocal folds are free to vibrate. Note that this state is essentially fully adducted and relies on activity of the lateral cricoarytenoid and
(especially the transverse) interarytenoid muscles (Hillel, 2001). This is not to say that more muscle force is available, but the tissues manipulated by the arytenoids are in full contact. We might suspect that further adductory engagement of the IAs will begin to engage epilaryngeal stricture in the ventricular and aryepiglottic planes according to the muscle-chain mechanisms described in §2.1.2 (especially Figure 2.8).

The states in Figure 4.10b and Figure 4.10c respectively are creaky voice produced in raised larynx and lowered larynx conditions, corresponding with the productions illustrated in Figure 4.8 and Figure 4.9 (respectively). Both qualify as “anterior voice” (only the membranous vocal folds vibrate) LCA activity must increase to hinder vibration of the cartilaginous vocal folds, but this does not account for the engagement of the epilarynx. We can infer that further recruitment of the external TA and oblique IA (along with the AE) muscles is occurring. The mechanism is engaged even with active height suppression (i.e. lowering of the larynx by approximately 5 mm). Furthermore, whether raised or lowered, both states involve low pitched creaky phonation, yet the entire larynx appears to be under considerable tension, and we can infer that posteroanterior narrowing is driven, in part, by the external and internal TA muscle branches (see §2.1.2), which will have the effect of raising vocal fold stiffness. The question then becomes – how is the low pitch achieved?

An answer comes from considering overall laryngeal articulation. In addition to glottal aperture control, there are two principal articulatory axes of the larynx: the axis of height and the axis of constriction. The former is causally associated with changing the volume of the pharyngeal cavity; the latter denotes action of the mechanism that induces stricture of the epilarynx. Both mechanisms have an impact on the pitch control system.
Larynx height positively correlates with pitch: larynx lowering results in cricoid rotation favouring reduction of vocal fold tension (Ohala, 1972; Honda, Hirai, Masaki, & Shimada, 1999); larynx raising increases vocal fold tension through anterior traction on the hyoid-thyroid complex and through thyrohyoid contraction in conjunction with contributions from other factors such as internal vertical tension and tracheal pull (Ohala, 1972; Honda, Hirai, Estill, & Tohkura, 1995; Vilkman, Sonninen, Hurme, & Körkkö, 1996). It is assumed that larynx height is relatively slow acting compared to the longitudinal pitch-control mechanism.

Laryngeal constriction is a function of three primary physiological components that act synergistically (Esling, Zeroual, & Crevier-Buchman, 2007): tongue retraction (which is partly responsible for posterior displacement of the epiglottis; although, see §2.1.2), larynx raising, and contraction of the thyroarytenoid muscle complex and possibly the aryepiglottic muscle (Painter, 1986). Pitch is thought to be impacted by concomitant ventricular incursion of epilaryngeal stricture (Laver, 1980): the ventricular folds are thought to impinge upon the upper surfaces of the vocal folds and mechanically couple with them. The suspected increase in oscillating mass should result in lowered frequency response of vocal fold vibration.

Figure 4.11 provides a depiction of these relationships; the arrows indicate the direction of causality. Larynx height plays a dual role in laryngeal control: it acts agonistically in the mechanisms for pitch control and for laryngeal constriction. As discussed in Esling & Harris (2005), the primary pitch control mechanism stands in antagonistic relationship to the laryngeal constriction mechanism; this is because engagement of the laryngeal constrictor acts along the posteroanterior dimension to
narrow the epilarynx, this is opposite to the posteroanterior widening of the epilaryngeal aperture caused by engagement of the cricothyroid muscles.

![Diagram](image)

Figure 4.11: The relationship amongst larynx height, constriction, and pitch control\(^\text{103}\).

When unopposed by the pitch raising mechanism, larynx raising has the effect of reducing the vertical dimension of the epilarynx (see §2.1.2). If all else is held constant, the vocal folds will approach the ventricular folds, which may be in descent because of active constriction forces (M. M. Reidenbach, 1998a; Reidenbach, 1997) and compaction from tissues above. The opposite effects are expected in the case of larynx lowering, which Fink (1974a) described as driving anti-plication of the larynx (or laryngeal “unfolding”).

---

\(^{103}\) With regard to pitch, this diagram can be interpreted as follows: larynx height positively correlates with the pitch mechanism (raising the larynx helps the pitch mechanism to raise pitch and lowering the larynx assists in pitch lowering); likewise, larynx height positively correlates with constriction (raising the larynx assists in compacting the larynx while lowering the larynx causes vertical expansion and separation of laryngeal tissues). The pitch mechanism (which expands the anteroposterior dimension of the larynx) is diametrically opposed to the constriction mechanism (which contracts the anteroposterior dimension). Engaging both at the same time results in the compromised configuration associated with harsh voice at high pitch (Esling & Harris, 2005).
The basic mechanical biases have been schematized in Figure 4.12: the figure shows an idealization of the relationship between the vocal and ventricular folds (viewed in coronal section; only half of the larynx is depicted) as the larynx is lowered (a) and raised (c) relative to the neutral height (b). Larynx lowering (a) stretches and elongates the vertical dimension of the larynx – thereby reducing or counteracting laryngeal constriction; the effect should bias phonation towards breathiness as there is a lateral force on the vocal folds that draws them away from the glottal midline. This lateral force results from the tendency for tissue to thin as it is stretched along an axis, and it is well described and illustrated by Fink (1974a). Larynx raising (c) results in a collision between the two sets of folds, which are thus forced to buckle towards the glottal midline.

Figure 4.12: The effect of larynx height on the relationship between the vocal folds (v) and ventricular (false) folds (f) seen from coronal section of the larynx. Dotted line = glottal midline.

---

104 Figure appears in Moisik, Esling, & Lin (forthcoming).
as the laryngeal space becomes more compacted in the vertical dimension; this effect ought to bias phonation towards harshness, tenseness, or creakiness depending on subglottal pressure and other factors.

In general, the figure illustrates the laryngeal configuration tendencies or biases for the different height settings, *ceteris paribus* (especially the vocal fold parameters); it must be remembered that these are not absolute relationships. The demonstration in Figure 4.9 shows that, even when larynx height is actively decreased, laryngeal constriction appears to engage and, nonetheless, still probably involves vocal-ventricular fold contact, but the degree to which the epilarynx can narrow is hindered, and this is manifest in the fact that ventricular fold medialization is slight by comparison to the non-height suppressed production in Figure 4.8. In fact, it is still entirely possible to constrict the epilarynx in the lowered-larynx setting (Esling, 1999)\textsuperscript{105}. The other aspect that must be remembered is that larynx height is not the sole determinant of laryngeal constriction: lingual/epiglottal retraction and the intrinsic constriction mechanism also matter a great deal in determining the final laryngeal state.

### 4.2.6 Summary: Lower epilarynx function and larynx height

This study has provided an illustration of the relationship between narrowing of the upper larynx (as Catford put it) or epilarynx in connection with larynx height. Sounds such as glottal stop and creaky phonation considered to have “glottal constriction” appear to involve much more than a purely vocal-fold mechanism, which corroborates numerous earlier reports. This is the first time, however, that larynx height has been quantified

\textsuperscript{105} The author is in possession of SLLUS data which confirm this possibility.
using independent equipment, rather than drawing inferences from laryngoscopic video, as in Brunelle et al. (2010). The data in the present study are limited to phonetic productions, but the SLLUS technique can be readily applied to future research, or, if a less invasive study of larynx height is desired, then the results here confirm that optical flow quantification of larynx height data can be performed using only laryngeal ultrasound.

Larynx raising seems to be an intrinsic part of the epilaryngeal constriction occurring in the “constricted glottis” sounds, and it is suggested to be especially important in regulating the relationship between the vocal and ventricular folds, which has important dynamic consequences for the overall vibrating system. Furthermore, the study carries the implication that the role of “glottal” constriction may have been overemphasized in previous phonetic models: there is vocal fold adduction (as in prephonation; see Esling & Harris, 2005), and apparently medial compression is relevant to controlling which part of the vocal folds can vibrate, but, in keeping with proposals by Catford and Lindqvist-Gauffin before him, the creakiness and closure mechanisms seem to critically rely on a general mechanism of laryngeal closure, and glottal closure is merely the first step in a larger, relatively slow-acting sequence involving narrowing, compression, and collapse of the epilaryngeal air-way and associated tissues.

4.3 Modeling vocal-ventricular fold coupling

Physiological and phonetic studies suggest that at moderate levels of epilaryngeal stricture the ventricular folds impinge upon the vocal folds and influence their dynamical behaviour, which is thought to be responsible for constricted laryngeal sounds. In this
study\textsuperscript{106}, we examine this hypothesis through biomechanical modeling. The dynamical response of a low-dimensional, lumped-element model of the vocal folds under the influence of vocal-ventricular fold coupling was evaluated. The model was assessed for F0, cover mass phase difference, jitter, and shimmer. Case studies of simulations of different constricted phonation types and of glottal stop illustrate various additional aspects of model performance.

Simulated vocal-ventricular fold coupling causes a lowering of F0 and perturbs the mucosal wave. It is also shown to cause and reinforce irregular patterns of oscillation, and it can enhance laryngeal closure in glottal stop production. The effects of simulated vocal-ventricular fold coupling are consistent with sounds observed to involve epilaryngeal stricture and apparent contact between the vocal folds and ventricular folds. This supports the view that such contact actually occurs in such sounds, and, furthermore, suggests that sounds such as creaky voice, harsh voice, and glottal stop, are intrinsically epilaryngeal rather than glottal in nature.

4.3.1 Introduction: The larynx – more than just the vocal folds

Research on laryngeal behaviour in normal and disordered voice production and in speech emphasizes the vocal folds as the cause of diverse phenomena: vibration with slight abduction produces breathy phonation. Tight adduction is responsible for glottal stop (Ladefoged, 1971; Gordon & Ladefoged, 2001) and phonatory onset or offset with

\textsuperscript{106} This study has been submitted to the \textit{Journal of Speech, Language, and Hearing Research} for publication; see Moisik & Esling (under review; invited). The research was originally presented at The 8\textsuperscript{th} International Conference on Voice Physiology and Biomechanics (July 6\textsuperscript{th}, 2012; Erlangen, Germany); see Moisik & Esling (2012).
“hard glottal attack” (Hess, Verdolini, Bierhals, Mansmann, & Gross, 1998; Zemlin, 1998, p. 102) or “pressed” phonation (Stevens, 1998, pp. 82–85), if there is concomitant vibration. Asymmetrical vibration causes diplophonia or biphonation, meaning that one vocal fold oscillates at a different frequency from the other (Neubauer, Mergell, Eysholdt, & Herzel, 2001; Eysholdt, Rosanowski, & Hoppe, 2003). Organic, structural, or functional problems of the laryngeal musculature and tissues underlie disordered voice (Bridger & Epstein, 1983; Hess, Verdolini, Bierhals, Mansmann, & Gross, 1998; Voigt, Döllinger, Braunschweig, et al., 2010; Kelchner, Brehm, de Alarcon, & Weinrich, 2012). As Painter observed (1986, p. 329), however, far less attention has been given to the study of the voice-production functions of the epilarynx.

Laryngeal sounds produced when the epilarynx is constricted can be impressionistically characterized as “tight”, “metallic”, or even “strangulated”. Sounds produced in this state are referred to as “constricted” here. Constricted phonation can be broadly classified as a subtype of nonmodal phonation, but specific labels vary by tradition and discipline (Gerratt & Kreiman, 2001). Here the focus is on two types of speech-related laryngeal behaviour: (1) constricted phonation types, which includes creaky voice (or vocal fry/pulse register/laryngealized voice) and harsh voice (or tense/pressed voice), and (2) glottal stop (an arrest or prevention of vocal fold vibration resulting in silence).

Although the vocal folds are of obvious importance to voice production, the contention here is that, to fully understand laryngeal behaviour, we need to understand the contributions of the epilarynx to voice production. The literature contains numerous examples of how the epilarynx contributes. For example, Van den Berg (1955, p. 63)
proposed that obliteration of the laryngeal ventricle by downward movement of the ventricular folds causes a loss of the low-pass filtering effect associated with this space. Sundberg (1974; also see Yanagisawa, Estill, Kmucha, & Leder, 1989; Honda, Kitamura, Takemoto, et al., 2010) demonstrated that the uncoupled resonance of the epilarynx causes the so-called singing formant. Titze (2008) argued that the epilarynx can, through constricting, aero-acoustically modify source-filter coupling, which influences vocal fold vibration and acoustic power. Bailly et al. (2008) claimed that the ventricular folds impose aero-acoustic effects on the vocal folds, such as increasing the pressure recovery above the glottis, which can suppress or perturb phonation and which is important in the process of vocal-ventricular phonation (also see Bailly, Henrich, & Pelorson, 2010).

several studies have shown that vibration of the epilarynx is responsible for “growly” voice used in singing styles and speech (Fuks, 1998; Lindestad, Södersten, Merker, & Granqvist, 2001; Borch, Sundberg, Lindestad, & Thalén, 2004; Sakakibara, Fuks, Imagawa, & Tayama, 2004; Eckers, Hütz, Kob, et al., 2009; Bailly, Henrich, & Pelorson, 2010; Moisik, Esling, & Crevier-Buchman, 2010; Tsai, Wang, Wang, et al., 2010; Crevier-Buchman, Pillot-Loiseau, Rialland, et al., 2012), which typically involves aero-acoustic damping of the concomitant glottal source or independent, quasi-periodic source production if the epilarynx vibrates alone.

Mechanical influence of the epilarynx: Vocal-ventricular fold coupling

Another possible role of the epilarynx in voice production – that it mechanically interacts with the vocal folds – has been postulated by several researchers, notably Allen
& Hollien (1973; also Hollien, 1974, p. 19), Laver (1975, 1980), and, more recently, Edmondson & Esling (2006) and Moisik & Esling (2011a); additionally, several others have hinted at this possibility (e.g. Gerratt & Kreiman, 2001; Agarwal, Scherer, & Hollien, 2003; Imagawa, Sakakibara, Tayama, & Niimi, 2003; Bailly, Pelorson, Henrich, & Ruty, 2008). The basic hypothesis emerging from these suggestions is that, during epilaryngeal constriction with concomitant vocal fold adduction, the ventricular folds, being pushed downwards and medially by the act of constricting the epilarynx, physically contact and press into the superior surfaces of the vocal folds, obliterating the ventricle and causing a change to vocal fold dynamics thought to be responsible for constricted sounds. Allen & Hollien (1973; also Hollien, 1974, p. 19) suggested that the ventricular folds mechanically load or couple with the vocal folds, resulting in damped vibration. Laver (1975, p. 224) proposed that the “ventricular folds become involved in the phonation of the true vocal folds by ... pressing down on the true vocal folds ... [and] ... combine to vibrate as more massive, composite elements.” Edmondson & Esling describe the apparent ventricular fold impingement on the vocal folds evident in their laryngoscopy data as “ventricular incursion”, which is defined as “partial covering and damping of the adducted glottal vocal fold vibration by the ventricular folds” (2006, p. 159). In the present work, the interpretation of ventricular incursion is that it only applies to situations involving contact between the vocal folds and ventricular folds and consequent interference with normal vocal fold vibration: it does not apply if the volume of the ventricle has decreased because of a vertical adjustment to the vocal folds or ventricular folds; the term is taken to be nonspecific with respect to the extent of ventricular fold medialization.
This vocal-ventricular fold coupling (hereafter VVFC) hypothesis is supported by studies of epilarynx control and activity in basic physiological functioning and in speech. Physiological research suggests a complex combination of internal and external laryngeal mechanisms is responsible for epilarynx constriction (Griesman, 1943; Fink, 1974a; Painter, 1986, 1991; Reidenbach, 1997; M. M. Reidenbach, 1998a, 1998b; Sakakibara, Kimura, Imagawa, Niimi, & Tayama, 2004; Esling, Zeroual, & Crevier-Buchman, 2007). In constricting, the epilarynx undergoes posteroanterior narrowing, as during swallow or effort closure. The soft tissues of the larynx fold, buckle, and bulge together as the narrowing occurs (which is typically facilitated by raising the larynx). Since part of the mechanism that drives epilaryngeal constriction also causes medial and downward displacement of the ventricular folds, it is possible for these structures to compact into the vocal folds during epilaryngeal constriction, provided the vocal folds have adducted first (Esling, 1996).

*Phonetic considerations: Why glottal stop is not just “glottal”*

There is good reason\(^\text{107}\) to suspect that more is involved in the production of glottal stop than just cessation of vibration of the vocal folds. Still, such a simple and well-known sound has often eluded comprehensive explanation. Conventional wisdom dictates that vocal fold adduction and the concomitant increase in medial compression should be sufficient to produce glottal stop (Halle & Stevens, 1971; Stevens, 1998, pp.

\(^{107}\) With the exception of some editing and other minor contributions by Scott Moisik, the next three paragraphs are credited to John Esling
A comprehensive explanation needs to make clear what the role of the supraglottal laryngeal component of glottal stop is.

Languages can make a clear distinction between glottal stop and aryepiglottal stop, as found, for example, in the Wakashan language group known as Nuuchahnulth (Esling, Fraser, & Harris, 2005), but the distinction is produced at epilaryngeal level: in glottal stop, there is vocal fold adduction and ventricular fold engagement; in epiglottal stop, the closure additionally involves closure at the aryepiglottic level. Detailed taxonomic descriptions have also shed light on the laryngeal/pharyngeal relationships differentiating these postural states, and a distinction between glottal stop and the state of “prephonation” has been identified (Esling & Harris, 2005). When the vocal folds are adducted at the glottis, initiated by the lateral cricoarytenoid (LCA) muscles and with the subsequent assistance of the transverse interarytenoid (IA) muscle (Hillel, 2001), the result is a coming-together of the vocal processes of the arytenoids, usually with cartilaginous adduction for modal voice, leaving the glottis in the biconvex shape that has been labeled prephonation. In this state, the vocal folds do not touch at the midline, and their vibratory undulation only begins when a stream of air is applied through the opening. The degree of adduction present for prephonation (so labeled because it is the posture immediately prior to voicing onset) is not adequate to achieve the abrupt auditorily recognized quality of glottal stop. Further adductory forces must therefore be present to generate the “reduced protective closure” (Gauffin, 1977, p. 308) of glottal stop.

It has been argued in Esling (1996, 2005) that the production of glottal stop closure requires activation of the laryngeal sphincter mechanism to generate this reduced
closure in the form of posteroanterior epilaryngeal narrowing and to initiate the vertical compression by the ventricular folds necessary to make vibration of the vocal folds cease. A corollary of this argument is that glottal stop should be more efficient to produce in the raised-larynx condition than in a lowered-larynx condition, as is also the case with epiglottal stop. These deductions are supported by visual laryngoscopic, laminagraphic, and laryngeal ultrasound evidence comparing the actions of the laryngeal structures in performing these various laryngeal maneuvers, both in controlled phonetic conditions as well as from native-speaker subjects of some two dozen language backgrounds (Ringgaard, 1962; Lindqvist-Gauffin, 1972; Allen & Hollien, 1973; Iwata, Sawashima, Hirose, & Niimi, 1979; Painter, 1986, 1991; Esling, 1996, 1999; Clements & Osu, 2002; Esling & Harris, 2003b, 2005; Edmondson & Esling, 2006; Edmondson, Chang, Hsieh, & Huang, 2011; Esling & Moisik, 2011, 2012; Moisik, Esling, Bird, & Lin, 2011). This research strongly evinces a role for epilaryngeal stricture in glottal stop and related constricted phonation types such as creaky and harsh voice qualities.

The present study evaluates the plausibility and possible dynamical effects of the supposed mechanical interaction between the vocal folds and ventricular folds using a low-dimensional, lumped-element computational model of these structures. From the outset, the expectation is that contact between the vocal folds and the ventricular folds results in mechanical loading or coupling of these structures such that the dynamical response of the vocal fold system becomes dependent on the behaviour of the ventricular fold system. The dynamical effects expected to occur as a result of the coupling are as follows:
(1) The addition of the ventricular fold mass to that of the vocal folds increases the effective mass of the oscillating system as a whole, which will lower its frequency response;

(2) The resistive impedance of the ventricular-fold loading on the vocal folds will result in greater damping within the oscillating system, which will limit its ability to engage in self-sustained vibration and reduce the amplitude of vibratory displacements;

(3) The ventricular fold coupling introduces new degrees of freedom to the overall system, resulting in new modes of oscillation, increasing the oscillatory complexity, and inducing irregularity in the pattern of oscillation;

(4) Contact between the ventricular mass and the superior surface of the vocal fold cover will interfere with the transmission of the mucosal wave.

Increased mass (1), damping (2), oscillatory irregularity (3) and perturbation to the normal mucosal wave pattern (4) are all generally consistent with constricted phonation types. The general prediction is that, with coupling, the vocal folds should vibrate at a lower frequency, exhibit more difficulty in engaging self-sustained oscillation, and show oscillatory irregularity in terms of period and amplitude. Other factors, such as subglottal pressure and vocal fold configuration, determine the specific phonation type.

---

108 In general terms, impedance characterizes the transmission of energy within a system. It is defined as the ratio of input driving a system to its output response: in aero-acoustic terms, this is the ratio of pressure to flow (see Titze, 2008); in mechanical terms, it is the ratio of force to velocity. In any case, impedance has two parts, resistance and reactance. The resistive part of impedance is a direct loss of input energy; the reactive part involves the delayed response of the system to the input of energy, either by storing the energy (compliance) or by inertia (inertance).
These four principles are examined in the context of three core questions about vocal-ventricular fold coupling: (A) Does VVFC affect vocal fold dynamics and are the effects consistent with what is observed in constricted sound production? (B) What is the nature of the coupling in VVFC? (C) Can vocal fold activity alone (i.e. without VVFC) simulate constricted sounds, and, if so, what does this tell us about VVFC? More details about how these questions are evaluated are presented below after the model has been described.

4.3.2 **Method: The Vocal-Ventricular Fold Coupling (VVFC) Model**

**Model foundation**

The VVFC model has two components: a mechanical model of the vocal and ventricular folds and a one-dimensional, aero-acoustic model of the vocal tract. The mechanical model of the vocal folds is based upon the ST95 model (Story & Titze, 1995), which is a low degree-of-freedom body-cover model of the vocal folds. The body is represented by a large mass $m_b$ connected to the laryngeal wall by spring $k_b$ with damping $d_b$; the cover is represented by an upper-cover mass $m_u$ and a lower-cover mass $m_l$, coupled together by spring $k_c$, and each individually connected to the body mass $m_b$ via springs $k_u$ and $k_l$ and damping $d_u$ and $d_l$ (see Figure 4.13c). The ST95 model was used to model the vocal folds in the VVFC model because it is a true body-cover model (Story, 2002; Birkholz, 2011) and allows for effects on the mucosal wave to be predicted (for more details, see footnote 90). Body-cover modeling is important for the VVFC model because one of the hypothesized effects of VVFC is interference with the vocal-fold mucosal wave (also see Figure 4.14, which is discussed in further detail below).
In its original implementation, the ST95 model was driven by an aero-acoustic simulation of the vocal tract based on wave-reflection mechanics (Story, 1995; also see Titze, 2006, Chapter 6). For aero-acoustic simulation in the VVFC model, Birkholz’s (2005, Chapter 3) inhomogeneous transmission line model of the vocal tract was implemented. Birkholz’s model is based on the simultaneous solution of the transmission-line network equations at each time step. It was implemented here because of its stability and the ease with which it permits the mechanical and aero-acoustic systems to be coupled together.

Equations governing the mechanical motion of the system and the aero-acoustic state are sufficiently documented in other sources (Ishizaka & Flanagan, 1972; Story & Titze, 1995; Birkholz, 2005) and will not be repeated here. Numerical solution to the equations of motion for the mechanical model was conducted using the 4th order Runge-Kutta (RK4) method with a time step of $10^{-4}$ s. Following Birkholz (2005), solution to the system of equations for the aero-acoustic simulation was achieved via successive over relaxation ($\omega = 1.25$); a trapezoid-rule-based scheme was used to approximate the derivative and integral functions of the pressure and flow state variables (with $\theta = 0.53$; see Birkholz 2005). The model is bilateral, but only symmetrical behaviour was simulated. As a convention, all signals were obtained from the centroid locations of the right-hand side masses.

---

109 Test simulations were conducted at $10^{-4}$ s and $5 \times 10^{-5}$ s. The signal corresponding to the smaller time step had a slightly diminished amplitude and a slight phase delay, but the overall behaviour of vibration remained consistent, and the glottal flow signals were 94.06% correlated. Furthermore, a previous version of this model constructed using an entirely different aero-acoustic simulation (Titze, 2006, Chapter 6) produced very similar results to those presented here.
Modeling vocal-ventricular fold coupling

The key change to the ST95 model is the addition of mass $m_v$, which is connected to the laryngeal wall by stiffness $k_v$ and damping $d_v$. This mass represents the supraglottal laryngeal mass that is proposed to contact and couple the vocal-fold mass during VVFC. Although $m_v$ is primarily meant to represent the ventricular fold, it also represents the lumped mass of the base of the epiglottis and the tissue of the wall of the aryepiglottic fold, which is continuous with the ventricular fold, as depicted in Figure 4.13a & b.
Figure 4.13: Abstraction of VVFC into low-dimensional model. (a) midsagittal profile of larynx showing location (dashed line) of (b), which is an anterior coronal section of the larynx showing the anatomical distribution of model masses (dashed outlines) and an illustration of the compressive forces of ventricular incursion as a vector field (left side; single-headed arrows) and the representation of these forces as coupling-springs acting on the upper ($m_u$) and body ($m_b$) masses of the vocal folds (right side; double headed arrows). (c) schematic of the model. *n.b.*: In (b) the vocal folds are depicted as partly abducted (following the anatomical photo of the dissected larynx used to create the diagram), but we assume they are far more adducted in constricted states.
Lumped-element models are, by nature, considerably abstract in the way they represent structure. Because of this, the additional ventricular mass proposed here may seem similar in nature to that used in models of laryngeal disease or abnormal growths on the vocal folds (such as polyps, cysts, nodules, and so forth) (e.g. Koizumi & Taniguchi, 1990). There are, however, some key differences. First, models of vocal pathology involving an extra mass are typically bilaterally asymmetrical – the mass is added to one vocal fold only (Birkholz, 2011); in the VVFC model, the additional mass is bilaterally symmetrical. Second, pathological masses are typically only connected to one of the vocal fold “cover” masses; by comparison, each ventricular fold mass in VVFC model is connected to the ipsilateral laryngeal wall (and vocal fold during VVFC). Finally, unlike simulations of smallish growths on the vocal folds, the ventricular masses are large in vertical extent and in mass relative to the vocal folds; thus, the aero-acoustic and mechanical conditions are quite distinct from those in models of pathological growths on the vocal folds.

Several previous models represent the ventricular folds with two coupled masses; the purpose of these models, however, was to simulate vocal-ventricular phonation found in certain singing styles (Imagawa, Sakakibara, Tayama, & Niimi, 2003; Bailly, Pelorson, Henrich, & Ruty, 2008; Bailly, Henrich, & Pelorson, 2010). In general, it is conventional to model self-sustained oscillation of masses with at least two degrees of freedom to provide the delayed mechanical feedback required to support self-sustaining oscillation (McGowan, 1992; Titze, 2006, p. 351). The goal here, however, was not to represent such forms of vocal-ventricular fold vibration; thus, to maintain simplicity, each ventricular fold is represented by only a single mass.
The anatomical distribution of the lumped masses is represented in Figure 4.13b, which also illustrates the configuration of the VVFC springs (right side) and the compressive forces acting on the vocal-ventricular-epiglottal complex (left side). We assume that, in reality, there are three causes responsible for such compressive forces which drive adduction and descent of the ventricular folds and their consequent compression into the vocal folds: a downwards force from the epiglottis, likely driven by the cricoepiglottic and thyroepiglottic muscles (Kimura, Sakakibara, Imagawa, et al., 2002), a downwards force from the anteromedial component of the craniolateral extension of the thyroarytenoid (TA) muscle, and an adductory force from the posterolateral component of the same muscle system (M. M. Reidenbach, 1998a). Additionally, posteroanterior action of the epilaryngeal stricture mechanism will cause longitudinal compression of the ventricular folds and consequent concentric bulging, effectively contributing to their medial and downwards motion into the vocal folds.

Based on the above considerations, it seems likely that, in reality, the force exerted on the vocal folds by the ventricular folds has a vertical component; however, all masses in the VVFC model have one degree of freedom, which is in the lateromedial dimension: no motion in the vertical (inferosuperior) dimension was considered. This abstraction stems from the simplifying choices made in representing the vocal folds with a lumped-element approach and the aero-acoustic simulation of the vocal tract: introducing a vertical degree of freedom would complicate the calculation of glottal flow. (Since this is a first attempt at representing VVFC, the goal was to maintain simplicity.) This does not mean, however, that the model does not represent VVFC; it only means that the coupling here acts exclusively in the lateromedial dimension. The ST95 model...
makes a similar abstraction in its representation of the body and cover masses of the vocal folds: in the model, upper-cover mass $m_u$ and body mass $m_b$ move in the same dimension, but, in reality, the “cover” mass is spatially distributed medially and superiorly around the “body” (see Figure 4.13b). The mucosal wave travels across the vocal fold upwards and laterally, as depicted in Figure 4.14a. Thus, the cover has vertical and lateromedial influences as the mass is redistributed throughout the glottal cycle (if we constrain our attention to two dimensions). Figure 4.14b shows the same cycle but adds VVFC and depicts the horizontal component of the coupling force arising from the collision (col) between the mass in the mucosal wave and the ventricular fold mass above (dotted arrow).
Abstractions are intrinsic to lumped-element modeling. Not all of the details can be represented, but this is partly why these models are useful since they distill a complex scenario to its essential features. In both cases, the ST95 and the VVFC model, the detraction of abstraction is that fewer vibratory modes can be simulated than models with higher-dimensionality, but this does not make either simulation invalid – merely more spatially abstract.
A major consideration for model design is how the suspected coupling should be modeled. Two coupling forces are under consideration here: collision forces and mucous adhesion forces. It was assumed that contact between the vocal and ventricular folds constitutes collision, and, thus, this was always incorporated into the modeling of their coupling. Mucous-based coupling was also considered as a special case. Thus, two basic types of coupling examined here: one with mucous adhesion and one without. Since coupling was implemented using springs, two different VVFC springs were implemented: collision-and-mucous-type springs (push-pull; hereafter CM) and collision-only-type springs (push-only; hereafter CO). In both cases (CM and CO), a compressive restoring force arises during collision between the vocal and ventricular folds (the horizontal component of the collision force arising between $m_u$ and $m_v$ is depicted in Figure 4.14b), and this is modeled by following the ST95 approach to vocal fold collision (which is a common approach to representing collisions in many other low-dimensional vocal fold models; Birkholz, 2011). In the CM case, however, the pulling force was assumed to arise from mucous adhesion inhibiting tissue separation. This assumption is based on reports of mucous adhesion influencing vocal fold behaviour. For example, Ayache, Ouaknine, Dejonkere, Prindere, & Giovanni (2004) demonstrated that mucous-related adhesive forces which arise during the contact phase of the glottal cycle when the vocal folds press together can lower fundamental frequency and increase glottal contact. Since the ventricle is rich in laryngeal mucous glands (M. M. Reidenbach, 1998a; Agarwal, Scherer, & Hollien, 2003), it seems plausible that mucous might play a role in VVFC. Since the scale of these forces is unknown, the approach here was to model the adhesive forces in the form of the pulling influence of the CM spring when it is extended
past its equilibrium point. In contrast, the CO spring is made inactive (force and damping constants set to 0) when it is extended past its equilibrium point (representing separation of effective masses but not separation of the vocal and ventricular fold overall; see Figure 4.14b).

The diagram in Figure 4.13c shows two springs for VVFC: $k_{vb}$, which connects the ventricular fold mass to the vocal fold body-mass $m_b$, and $k_{vu}$, which connects the ventricular fold mass to the vocal fold cover-mass $m_u$ (see Figure 4.13c). Each of these springs is associated with its own damping, $d_{vb}$ and $d_{vu}$ respectively, which represents the loss of vocal fold mechanical energy resulting from coupling with the ventricular fold. Although the main point of contact in VVFC is between the mucosal surfaces of the vocal and ventricular folds, a ventricular fold mass $m_v$ to vocal fold body mass $m_b$ coupling spring $k_{vb}$ was included as an option to increase the coupling between the two sets of folds (hereafter VB coupling). This was implemented even though the vocal fold body is intended to represent the mass of the TA muscle tissue within the vocal folds and, hence, in reality, such mass cannot come directly into contact with that of the ventricular fold. It is supposed to represent the transfer of energy between the ventricular fold and the more lateral portion of the vocal fold (see Figure 4.13b). No such option was included for coupling the ventricular fold with the lower cover-mass of the vocal folds on the assumption that this tissue is further away from the ventricular fold mass than either the upper-cover or body masses are.
Model parameters

There are three main biomechanical parameter sets that are under consideration here. The first concerns the configuration, stiffness, and effective mass of the vocal folds. The second pertains to those same properties for the ventricular folds. The third parameter set concerns the stiffness and damping of the VVFC. Although an effort was made to adhere closely to parameter values given by other researchers, it was found that considerable experimentation was required to locate sets of parameters which enabled self-sustaining oscillation (as the goal was to study what happens to vibration with VVFC). Thus, the values that are presented in this paper reflect the choice to use self-sustained oscillation as a heuristic for parameter settings. Parameters used for particular simulation trials and the case studies are presented in their respective result sections (see Table 4.1 and Table 4.2).

For the vocal folds, the damping ratios were held constant ($\zeta_u$ is 0.4, $\zeta_l$ is 0.4, and $\zeta_b$ is 0.2, following the specifications of the ST95 model; $\zeta$ is related to damping coefficients by the equation $d = 2\zeta(mk)^{1/2}$), and $k_c$ (upper-lower mass coupling) was always 2 Nm$^{-1}$. Other parameters were manipulated according to the objectives of the simulation. Two conditions were explored in the context of the body-cover model of the vocal folds (Hirano, 1974, 1975): a “normal” (N) one and a “stiff-slack” (SS) one. In the normal condition, the vocal fold parameters fall within the range of parameter specifications provided by Story & Titze (1995) and Story (2002); the intention was to simulate a neutral configuration expected to produce modal-like phonation in the absence of VVFC. In contrast, the stiff-slack condition was used to simulate the vocal-fold
configuration suspected to occur during epilaryngeal stricture. It is characterized by an increase in body spring $k_b$ stiffness and concomitant decrease in the stiffness of the cover springs, $k_u$ and $k_l$, although more so for the upper cover spring (i.e. $k_u < k_l < k_b$). The upper-to-lower cover stiffness ratio was allowed to vary more widely than in the ST95 model, which uses a ratio of 0.7. At high levels of lower cover stiffness we assume that the lower cover begins to act as part of the body component (effectively reducing the vocal fold system towards a two-mass rather than three-mass configuration).

The stiff-slack condition is based on physiological observations (Hirano, 1975; Fujimura, 1981, pp. 279–280; Zemlin, 1998, p. 129) and supported by recent finite element modeling (Deguchi, Kawahara, & Takahashi, 2011) showing that unopposed TA contraction causes slackening of the vocal fold cover layer and a large increase in vocal fold body tension (as the muscle stiffens under contraction); activation of the TA during constricted phonation types like creaky voice and harsh voice, and during glottal stop production is consistent with the muscular requirements of epilaryngeal stricture (Esling, Zeroual, & Crevier-Buchman, 2007). Given the divergence of parameter values between the body and the cover, it was thought that the stiff-slack setting should produce some effects associated with creaky and harsh phonation types, such as oscillatory irregularities. The point was to determine whether adding VVFC with this configuration would produce any additional effects that might reveal why the suspected VVFC occurs if all that is required to produce constricted sounds is adjustment of the vocal fold configuration (as is implicit in, e.g., Stevens, 1998, p. 82).

According to Agarwal, Scherer, & Hollien (2003), the ventricular folds have higher tissue viscosity and lower mechanical stiffness than the vocal folds (also see
Bailly, Henrich, & Pelorson, 2010). Imagawa et al. (2003) used the following parameters for their two-mass ventricular fold model: mass $m_v = 260$ mg ($2 \times 130$ mg), stiffness $k_v = 50$ Nm$^{-1}$, and a damping ratio $\zeta_v = 0.4$. The Imagawa et al. (2003) values were used as baselines for experimentation within the VVFC model, but it should be noted that, unlike in the present study, the Imagawa et al. model does not represent VVFC. Thus, some divergence from these values was necessary because it was found that the different conditions for VVFC (CM, CO, and the activity of VB) and vocal fold configuration (N or SS) were highly sensitive to ventricular fold and vocal fold parameter values in regard to whether or not self-sustaining oscillation would occur.

Parameters for VVFC springs and damping ($k_{vb}$, $k_{vu}$, $d_{vb}$, and $d_{vu}$), were manipulated according to the needs of the particular simulation. Stiffness ranged from 0 Nm$^{-1}$ up to 3 times the value of the primary spring controlling the mass to which the representative spring was coupled (following the definition of collision springs in the ST95 model). In practice, values were used in between this range which allowed for self-sustaining oscillation to occur (see Table 1 and Table 2 for specific settings). Damping was typically determined by setting $\zeta_{vb} = 0.2$ and $\zeta_{vb} = 0.4$, although values as low as 0 and as high as 1.6 were used (again, based on whether self-sustaining oscillation could be achieved or not). For no VVFC, or to “turn off” coupling, stiffness and damping were both nullified.

Ventricular fold positioning during modal conditions was set to match the Story (1995) data for the first supraglottal tube section; under normal conditions (non-singing, non-pathological, etc.) the lateromedial distance separating the medial edges of the ventricular folds (the ventricular fold “gap”) is on the order of 0.5 cm (Agarwal, Scherer,
& Hollien, 2003), and the area-function values presented by Story (1995) and used in the VVFC model are consistent with this. For constricted conditions, the center of $m_v$ was set to be 2.3 mm from the glottal midline, which yields an initial ventricular fold separation distance of 0.6 mm (corresponding to an area of 0.28 mm$^2$ for the tube section of defined by the ventricular masses in the aero-acoustic model). This narrow initial cross-sectional area reflects laryngoscopic data of creaky and harsh phonation (Esling & Harris, 2005; Esling & Moisik, 2012), where ventricular fold medialization is so strong at the onset of phonation that the folds appear to be touching at their medial edges. In these data, it can be observed that some degree of increase in the lateromedial distance between the ventricular folds occurs during the course of phonation, but often the folds remain partially in contact, especially along their anterior extent.

**Aero-acoustic conditions**

The aero-acoustic model used in the simulations closely follows the formulation put forth by Birkholz (2005), but the exact configuration is modified for the VVFC model. Two vowel configurations were used, [i] and [a]; the area functions for these vowels were obtained from MRI-based data presented in Story (1995, pp. 308–309). The supraglottal vocal tract was divided into 35 tube sections and the subglottal tract was divided into 20 sections. All tube sections in the model were made to be 0.5 cm long, except for the two sections constituting the lower and upper glottis, which were set to be 0.2 cm in length. The supraglottal vocal tract is therefore 17.5 cm long. The laryngeal ventricle, piriform fossae, and nasal cavity were not modeled. Lung pressure was set
according to the needs of different simulations: generally, low values (200 – 400 Pa) were used to simulate creaky voice and higher values were used to simulate harsh voice (1000 Pa); in other cases, a default of 800 Pa was used.

The impedance of the basic tube sections in the transmission line model was calculated using the specifications outlined in Birkholz (2005): each tube section has air and wall impedance components. However, the glottal and ventricular sections are treated differently. Since wall impedance in the glottal and ventricular tube sections is modeled explicitly in the form of the mechanical models of the vocal folds and ventricular folds respectively, no wall impedance circuit components were used in the transmission line for these sections. Furthermore, while basic tube sections implement a frictional resistance based on the calculation of airflow resistance in circular tubes under steady, laminar flow, the glottal and ventricular sections employ “slit resistance” which models the resistance to steady, laminar flow through a narrow slit (modeled as an ellipse with high eccentricity) intended to represent the more triangular rather than circular cross-sectional shape of these sections (Van den Berg, Zantema, & Doornenbal, 1957; Ishizaka & Flanagan, 1972; see Birkholz, 2005, pp. 56–57).

**Model evaluation**

An noted above, the key questions under consideration here are (A) whether VVFC affects behaviour of the vocal folds, (B) what the nature of VVFC is, and (C) whether constricted sounds can be simulated without using VVFC, and, if so, what this tells us about VVFC.
Question (A) was examined by comparing the dynamical response of the vocal folds with and without VVFC. According to the VVFC hypothesis, effects such as (1) decreased fundamental frequency, (2) inhibition of self-sustained oscillation, (3) vibrational irregularities, and (4) perturbation to the mucosal wave are predicted to occur with VVFC applied. Such effects are assumed to be consistent with constricted phonation produced while epilaryngeal stricture is visibly present and the presumed vocal-ventricular contact is occurring, as evident in laryngoscopic imaging and so forth. With regard to glottal stop production, the application of VVFC is predicted to stabilize glottal stop, meaning that vocal fold arrest will be maintained during the targeted stop manoeuver even under continuous subglottal pressure.

Question (B) concerns the different options for implementing VVFC in a lumped-element model. Two types of coupling are examined, CM coupling, which represents coupling with both mucous adhesion and collision forces, and CO coupling, which represents only collision forces. Another possibility is whether VB coupling, i.e. coupling between \( m_b \) and \( m_v \) via \( k_{vb} \) and \( d_{vb} \), is active. (It is assumed that “cover” coupling, i.e. between \( m_v \) and \( m_u \) via \( k_{vu} \) and \( d_{vu} \), is always active to some extent during VVFC.)

Question (C) was formulated in response to the long-held assumption that sounds such as creaky and harsh voice and glottal stop should be attainable purely through manipulations to the vocal folds. To address (C), two types of vocal fold settings, normal/N and stiff-slack/SS, were examined with the presence and absence of VVFC in conditions suspected to produce creaky and harsh phonation and during the glottal stop manoeuver.
Model behaviour is examined in two parts. The first part compares the response of the model with different test conditions in terms of four response variables: F0 (glottal fundamental frequency), $\phi$ (vertical phase difference of the vocal fold cover masses), mean glottal jitter, and mean glottal shimmer. According to the VVFC hypothesis, coupling should decrease F0, change $\phi$, and increase mean glottal jitter and shimmer (which are measures of irregularity in frequency and amplitude of the glottal pulse). Each measure was obtained over a 100 ms window located at least 100 ms into the simulation to avoid contamination by transient responses at the onset of simulation. F0 was calculated using a peak finding algorithm on the glottal flow pulse, calculating adjacent peak time-differences to obtain glottal period, and then inverting to obtain the fundamental frequency. Vertical phase difference ($\phi$) between the upper and lower cover masses was calculated using the formula provided by Story & Titze (1995, p. 1256): $\phi = 360\tau/T$ ($\tau$ is upper-lower mass time delay, T is fundamental period). Mean jitter and mean shimmer were calculated using algorithms described by Bielamowicz, Kreiman, Gerratt, Dauer & Berke (1996).

To carry out the testing in the first part, two batteries of simulations were conducted. In the first battery, the response of the model was tested using combinations of the different conditions (no VVFC vs. VVFC, N vs. S vocal fold parameter settings, [i] vs. [ɑ] vowel contexts, and CM, CO, CM + VB, and CO + VB coupling). A total of 20 simulations were conducted (4 without VVFC and 16 with VVFC). In the second battery of simulations, coupling parameters (stiffness and damping) were adjusted for both the CM and CO coupling types using percentage-wise (10%) increments from 10% to a
maximum value for these parameters. Thus, for each coupling type, 10 simulations were conducted.

The second part examines five case study simulations: 4 of constricted phonation types (creaky-like and harsh-like) and 1 for glottal stop. The focus here is the comparison of model response with and without VVFC applied. To facilitate the exposition, some methodological and data-presentation details unique to the individual cases (such as specific parameter settings), have been presented alongside the simulation results.

4.3.3 Results: General overview of model performance:

Figure 4.15 illustrates model performance with all of the different coupling conditions (see Table 4.1 for specific parameters). Shapes and shading indicate the different conditions: without VVFC (black diamond); with VVFC: CM (square), CO (circle), with VB coupling (gray), and without VB coupling (white). Coupling strength between the ventricular and the upper-cover masses was necessarily different for the normal and stiff-slack contexts because the settings used for the former generally precluded phonation in the latter. Coupling had to be weak enough to allow the upper-cover mass $m_u$ to exceed the lower-cover mass $m_l$ in lateral excursion; otherwise oscillation would die out (with greater coupling strength). Despite this divergence in parameter settings, the same basic trends hold between the two coupling conditions for $F_0$ and $\phi$. 
Table 4.1: Biomechanical parameter ranges (mass in mg, stiffness in Nm\(^{-1}\); see Appendix D for abbreviation meanings)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Test Condition</th>
<th>VocF</th>
<th>VentF</th>
<th>VVFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.15</td>
<td>normal (N)</td>
<td>30</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>stiff-slip (SS)</td>
<td>20</td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>N, CM, +VB</td>
<td>30</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>N, CO, +VB</td>
<td>30</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>SS, CM, +VB</td>
<td>20</td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>SS, CO, +VB</td>
<td>20</td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td>4.16</td>
<td>CM sweep (N)</td>
<td>30</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>CO sweep (N)</td>
<td>30</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

(N = “normal” vocal fold parameters; CM = collision-and-mucous-type springs (push-pull); CO = collision-only-type springs (push only); VB = ventricular-fold–vocal-fold-body coupling; VocF = vocal fold; VentF = ventricular fold)

VVFC has a lowering effect on F0 (Figure 4.15, upper left), regardless of the coupling strength or the conditions of coupling, but the effect is more substantial with normal than stiff-slip vocal fold parameters. In general, the explanation is that VVFC increases the effective mass and thereby lowers the natural frequency of the entire system (although the particular frequency depends on the specific parameters used for the VVFC, the \( m_v \), and \( k_v \)). The frequency response is more sensitive to whether there is VB coupling or not in the stiff-slip context.
As predicted (regardless of coupling type), F0 lowers and \( \phi \) increases with VVFC; however, jitter and shimmer are not as clear, but they are included here for reference. The phase-difference pattern (Figure 4.15, upper right) is opposite to that of F0: VVFC increases the phase difference. In these examples, the phase shift induced by VVFC is on the order of 10 to 15°, but the effect is enhanced in the stiff-slack context. Interference with the phase-difference pattern is an important effect of VVFC because it can inhibit complete vocal-fold cover contact (i.e. closure at the upper and lower glottal
sections) during the glottal closure phase. In such cases, the lower-cover mass begins to abduct before the upper-cover mass has had a chance to close the upper glottis. A similar effect occurs if the upper and lower masses have widely divergent parameterization, and this is evident in the elevated phase difference levels for the stiff-slack condition in general.

Jitter and shimmer (Figure 4.15, lower plot row) proved to be problematic measures. Jitter is generally low regardless of condition (on the order of $10^{-3}$), which may reflect the low-dimensionality of the system. No obvious patterns resulted: some VVFC simulations increase jitter and others decrease it. The same is true for shimmer. Values for jitter and shimmer for CO-VVFC without VB coupling (white circles) stands out as being almost always elevated above those values for the control condition (black diamonds). The cause of these variations likely has to do with the effect that different acoustic loadings associated with the different vowels has on glottal flow pulse skewing, since these measures are taken relative to the local maxima in the glottal flow function. Slight changes in the shape of the pulse and location of the peak in flow can influence the result. Because of this lack of clarity, these measures will henceforth be omitted.

The behaviour of the system in the two different vowel contexts is not perfectly uniform, but generally vowel context did not make a noteworthy impact.
Figure 4.16: Model performance as a function of percentage-wise increasing VVFC Jitter and shimmer omitted. (N; [a] context). Squares: CM-VVFC; circles: CO-VVFC. As predicted (regardless of coupling type) F0 tends to decrease and \( \phi \) tends to increase with VVFC.

Figure 4.16 shows the effects of incremental, percentage-wise (10% increments starting at 10%; values at the maximum/100% are given in Table 1) increase of VVFC in the context of CM (squares) and CO (circles) coupling (VB coupling is used in both cases). Parameters are listed in Table 4.1 (those values listed as belonging to Figure 4.16). The important point here is that, regardless of the type of VVFC (CM or CO), the increase of coupling strength coincides with the expected effect along each of the measures: F0 decreases and \( \phi \) increases.

4.3.4 Simulation case studies: Overview

The overarching goal of the VVFC model was to simulate a given vocal fold state with and without VVFC. The selection of case studies presented below are handpicked
from a large battery of trial simulations because they represent the range of behaviour observed in the process of experimental simulation and illustrate some of the more interesting effects associated with VVFC. Not shown here are cases in which the application of VVFC prevented phonation from occurring or in which it did not have a discernible impact to the pattern of oscillation (not considering F0). Table 4.2 below lists the different simulations illustrated here along with the parameters applied for the vocal fold masses, the ventricular fold masses, and for their coupling (symmetry was maintained in all simulations).

Table 4.2: Specific biomechanical parameters used in the case studies (Pressure in Pa; mass in mg; stiffness in Nm$^{-1}$; see Appendix D for abbreviation meanings)

<table>
<thead>
<tr>
<th>#</th>
<th>Test Cond.</th>
<th>VVFC “effects”</th>
<th>VocF</th>
<th>VentF</th>
<th>VVFC</th>
<th>$\zeta_{vb}$</th>
<th>$\zeta_{vu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N, CM, VB</td>
<td>“growly harsh”</td>
<td>800</td>
<td>15</td>
<td>50</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>k$_u$</td>
<td>k$_l$</td>
<td>k$_b$</td>
<td>m$_u$</td>
<td>m$_l$</td>
</tr>
<tr>
<td>2</td>
<td>SS, CM, VB</td>
<td>“irregularity reinforcement”</td>
<td>200</td>
<td>4</td>
<td>30</td>
<td>400</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>SS, CO</td>
<td>“harsh voice at high pitch”</td>
<td>1000</td>
<td>20</td>
<td>150</td>
<td>400</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>SS, CO, VB</td>
<td>“stabilized creaky”</td>
<td>400</td>
<td>4</td>
<td>100</td>
<td>400</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>N → SS, CO, VB</td>
<td>“enhanced glottal stop”</td>
<td>800</td>
<td>15</td>
<td>100</td>
<td>→ 100</td>
<td>50</td>
</tr>
</tbody>
</table>

(N = “normal” vocal fold parameters; CM = collision-and-mucous-type springs (push-pull); CO = collision-only-type springs (push only); VB = ventricular-fold–vocal-fold-body coupling; VocF = vocal fold; VentF = ventricular fold)

In the first four case studies, a figure is provided which contains two sets of plots. The upper set of plots is for model response without VVFC and the lower set is for
model response with VVFC. Each plot set contains the following plots: a spectrogram (showing frequencies from 0 to 2500 Hz), a phase plane plot for the upper-cover mass (showing position vs. velocity), a plot of glottal flow which provides an indication of glottal aperture (upper right), and a time-series plot of the right-side mass positions. More details are provided in Figure 4.17.

Case study #1: “growly harsh” (N vs. N, CM, VB; Figure 4.17)

Simulation using the normal vocal fold settings with CM-VVFC results in a train of alternating glottal flow pulses or “period alternation” (compare arrow-1 and arrow-2 in Figure 4.17) – essentially a type of amplitude modulation of the glottal source and classifiable as supraperiodic, nonmodal phonation (Gerratt & Kreiman, 2001, pp. 366–367). In the no-coupling condition, F0 is 116 Hz and $\phi$ is 42°; phonation can be considered “modal voice” or “chest voice”. In the VVFC condition, F0 drops to 94 Hz (a difference of 22 Hz) and $\phi$ nearly doubles to 81°. The presence of subharmonics in the spectrum (e.g., in Figure 4.17, compare arrow-3 vs. arrow-4, which are both placed between the first and second harmonics) can be attributed to the amplitude modulation; this result is therefore consistent with a low, harsh voice quality (Laver, 1980; Esling & Harris, 2005) which might be described as “growly harsh”. The phase plot for the upper-cover mass in VVFC condition shows two orbits associated with the alternating glottal cycles, one with slightly greater displacement and velocity than the other. The displacement and velocity of the upper-cover mass are diminished by nearly a factor of two in comparison to the no VVFC condition, which can be attributed to increased
damping acting on the upper-cover mass. Inspection of the mass displacement time series for the VVFC condition reveals that the alternating-pulse pattern is present in the motion of the ventricular fold mass and somewhat less so in the body mass of the vocal folds. The lower-cover mass is least influenced in this way. The magnitude of the flow pulse is considerably diminished in the VVFC condition because of the diminished glottal aperture caused by the increased damping.
Figure 4.17: Plots for case study #1 (Ps = 800 Pa; [α] context; for parameters, see Table 4.2). No VVFC condition (upper plot set); VVFC condition (lower plot set). Plots (for each condition): spectrum (upper left); glottal flow (upper right); upper-cover mass phase plane plot (lower left); right-side mass positions (lower right). Lines in mass position plot: thin black line: upper-cover mass $m_u$; light gray line: lower-cover mass $m_l$; dotted line: body mass $m_b$; thick black line: ventricular fold mass.

Case study #2: “irregularity reinforcement” (SS vs. SS, CM, VB; Figure 4.18)

This simulation illustrates “irregularity reinforcement” (Figure 4.18) (note that VVFC damping was nullified to achieve self-sustaining oscillation in this case): without VVFC, the stiff-slack configuration exhibits a slight tendency towards irregular
oscillatory behaviour (see bracket-1 in Figure 4.18); this initial perturbation to vocal fold oscillation may reflect the influence of the transient response of the aero-acoustic system, which is more pronounced in the stiff-slack condition than when the vocal fold body-cover system is more uniformly stiff), but stable oscillation takes over suddenly just after the 200 ms mark (corresponding to the limit orbit, the dark ring marked by the arrow in the phase plot); with VVFC, this irregular tendency is “stabilized” into a period-alternating-type irregular pattern (bracket-2 demarcates one cycle in Figure 4.18).

Once again F0 decreases when VVFC applied (from 129 Hz to 103 Hz) and \( \phi \) increases (from 44° to 92°). The glottal flow pulse has increased in intensity somewhat with VVFC, but this is attributable to the absence of damping.

The original goal for this simulation was to attain a creaky-like behaviour. Despite the low glottal flow, irregularity in amplitude of the pulses, and the lengthened closed phase between the alternating pulses, the quality is still quite “growly” because of the strength of subharmonic structure (compare arrows 1 and 2 in the spectra of Figure 6, which show a subharmonic between the first and second harmonics in the lower plot [VVFC] and in the upper [no VVFC]).
Figure 4.18: Plots for case study #2 (Ps = 200 Pa; [i] context; for parameters, see Table 4.2). See Figure 4.17 details.

Case study #3: “harsh voice at high pitch” (SS vs. SS, CO; Figure 4.19)

This simulation (Figure 4.19) shows a high F0 achieved with stiff-slack parameters (but in this case, cover stiffness has been increased and effective mass of the vocal folds has been decreased; VVFC damping was again nullified to achieve self-sustaining oscillation). Without VVFC, the vibration approaches a falsetto-like mode, but the cover displacement is large in comparison to a true falsetto (Story & Titze, 1995, p. 1257), and, moreover, vocal fold contact still occurs. With VVFC applied, the model represents the physiologically conflicting combination of cricothyroid contraction (which
is required to increase longitudinal tension of the vocal folds but consequently expands
the posteroanterior dimension of the epilarynx) and epilaryngeal constriction, which is
characterized by posteroanterior reduction of the epilaryngeal space. Esling & Harris
(2005) classify this state as “harsh voice at high pitch”, and it is linguistically attested in
the “harsh” register of Bai (see Edmondson & Esling, 2006), a Tibeto-Burman language.

The results indicate that VVFC in this configuration introduces instability into the
glottal flow pulse; compare the phase plane plots of the two conditions: unlike the no
VVFC condition, the plot for the VVFC condition shows considerable cycle-to-cycle
variation. The difference in magnitude between these cases reveals that oscillation is also
much more restrained in amplitude of displacement with VVFC applied (even without
damping): this corresponds to a substantial drop in glottal flow. In the spectrum, the
VVFC case has weak harmonic structure that is nearly lost in the interharmonic noise
content; this too contrasts with the no VVFC condition, where the harmonics are very
strong (~50 dB above the interharmonic noise). Because of the absence of strong
subharmonic structure, the VVFC condition might be best characterized as a “noisy
voice” nonmodal variant (Gerratt & Kreiman, 2001, p. 372), but this is still consistent
with harsh voice (Laver, 1980).
Figure 4.19: Model state plots for Simulation #4 (Ps = 1000 Pa; [a] context; for parameters, see Table 4.2). See Figure 4.17 details.

Case study #4: “stabilized creaky” (SS vs. SS, CO, VB; Figure 4.20)

This case study (Figure 4.20) marks another attempt to simulate creaky voice, but this time at a very low F0. To represent the thickening up of the vocal folds in this condition, m_u and m_l were increased by a factor of 1.6 and m_b was increased by a factor of 3 (relative to the N setting). The stiffness of k_u was decreased to 4 Nm⁻¹; k_l remains at a stiffness between that of k_b and k_u. (Without a stiff enough lower cover, phonation with this parameter set is not possible as the upper-cover mass will never surpass the lower-cover mass to change the glottal profile from convergent to divergent – essential for
maintaining pressure conditions within the glottis that support oscillation rather than hinder it). Following Stevens (1998, pp. 82–85), vocal fold adduction was increased by shifting the body masses $m_b$ medially by 0.5 mm. Subglottal pressure was set low (400 Pa).

With these somewhat extreme settings, it was found that a pattern consistent with descriptions of creaky voice – i.e. one showing a substantial drop in F0, a lengthened closed phase, and “double pulsing” (Laver, 1980, pp. 122–126; Gerratt & Kreiman, 2001, p. 376) – was possible to achieve in the no VVFC condition. Double pulsing can be thought of as a single, complex glottal cycle: referring to the glottal flow plot in the VVFC condition (Figure 4.20), first there is a large pulse (arrow-1), which is quickly followed by a low amplitude pulse (arrow-2), and finally a phase of closure (arrow-3), then the pattern repeats.

The VVFC condition once again exhibits irregularity reinforcement, similar to that observed in case study #2. In this case, the irregular, double-pulsed cycle (bracket-1) is stabilized into a repeating pattern. There are moments in the no VVFC condition (e.g. between 0.16 and 0.24 s, bracket-2 in Figure 4.20) when the vocal folds exhibit this pattern, but without the reinforcement of the VVFC, their behaviour is more erratic.

Beside the stabilizing effect, VVFC lowers F0, this time by 9 Hz. The very high $\phi$ in both cases reflects the divergence in parameters for the upper- and lower-cover masses (the peak in lateral displacement for each mass occurring nearly 180 degrees out of phase). The spectral profile of each condition is similar, but, in the no VVFC condition, the noise amplitude is very high, while in the VVFC condition, the harmonics are much stronger overall; no obvious subharmonics are found because the small pulse (arrow-2) of
the double-pulse sequence is very low and its timing is not harmonically aligned with the large pulse (arrow-1). This distinguishes the double-pulsing behaviour from the period alternation seen in case studies #1 and #2.

Figure 4.20: Model state plots for Simulation #5 (Ps = 400 Pa; [a] context; for parameters, see Table 4.2). See Figure 4.17 details.

Case study #5: “enhanced glottal stop” (Figure 4.21)

As discussed in the introduction, glottal stop production is often observed with apparent ventricular incursion, suggesting VVFC occurs. The surprising thing about this is that conventional understanding predicts that vocal fold adduction alone should be
enough to produce glottal stop, yet this is not what is typically observed. Thus, glottal stop was simulated with and without VVFC to examine what VVFC might be doing in the production of this sound.

The approach to simulating glottal stop was to apply a ramp function that linearly increases the equilibrium length of the vocal fold body spring. This pushes the body and cover masses medially (from 100 and 200 ms). Then the reverse of this ramp function is applied (from 200 to 300 ms). Thus, the adduction of the vocal folds peaks at the 200 ms mark, and the duration of the simulated glottal-stop gesture is 200 ms. It is assumed that, as the vocal folds adduct, the body-cover stiffness becomes more stiff-sack because of TA contraction, which is supported by EMG observations of glottal closure gestures (Hirano & Ohala, 1969; Hillel, 2001). To simulate this effect, the stiffness parameters of the vocal folds were modified with the same ramp function used for simulating increased adductory force. These values were gradually adjusted from the normal configuration to the stiff-sack one and then back to normal again, according to the ramp function (see Table 4.2 for parameters used). Through experimentation, it was found that 1 mm of body-mass adduction (i.e. each m_b moves medially by 1 mm) was sufficient to stop vocal fold vibration entirely, with a 20 ms response time following the start of the ramp function. If the extent of medial displacement is reduced to 0.5 mm, however, vibration does not entirely halt. Thus, to illustrate the effect of VVFC more clearly, this value for body medialization was used in the simulations presented here.

To simulate glottal stop with VVFC applied, the same ramp-function based vocal fold parameter sweeping described above is used, but, in addition, the VVFC parameters using CO coupling are likewise increased and decreased by means of the ramp function.
The exception is that at the start and stop of the ramp function, the VVFC parameters are inactive. Unlike VVFC in the other simulations, the ventricular folds were medialized to the point of being in contact at the peak of the glottal stop gesture; this was timed such that contact occurred at 150 ms, was held constant for 100 ms, and then released at 250 ms, returning to its original configuration at 300 ms. The VVFC takes effect at approximately the point where the medial edges of the $m_v$ masses exceed the medial edges of the $m_b$ masses. Coupling parameters start at null and are then increased via the ramp function to a level commensurate with the coupling parameters used for the other simulations of VVFC (again, see Table 4.2).
Figure 4.21: Simulation of glottal stop without (top) and with (bottom) VVFC (Ps = 800 Pa; [a] context; for parameters, see Table 4.2). Thin black line: upper-cover mass $m_u$; light gray line: lower-cover mass $m_l$; dotted line: body mass $m_b$; thick black line: ventricular fold mass.

The results of these two simulations are shown in Figure 9. Without VVFC, the closure phase shows glottal leakage characterized by irregular pulses (bracket-1, Figure 4.21). At the peak of the gesture (bracket-2, Figure 4.21), the lower-cover is even slightly abducted and relatively stationary (although the upper-mass continues to vibrate). Technically (phonetically), this is not a glottal stop; however, it is still consistent with the phonetic description of glottal stop phonemes, which are often realized as a region of
creaky phonation in the transition between neighbouring vowels (Ladefoged & Maddieson, 1996, pp. 73–77). With VVFC applied, vocal fold oscillation is rapidly halted and there is no glottal leakage during the closure period (although the lower-cover masses once again slightly abduct).

Despite this effect, it was also found that, when adductory force of the vocal folds is increased (corresponding to a medial displacement of 1 mm for \( m_b \) in the model), vocal fold activity alone is sufficient to bring about glottal stop. However, the VVFC simulation demonstrates that ventricular incursion does have the effect of aiding in arresting vibration when the vocal folds do not adduct as strongly (0.5 vs. 1 mm of body mass adduction) and/or their biomechanical parameters (stiff-slack setting) do not favour oscillatory arrest.

4.3.5 Discussion

*Question A: Does the VVFC have the expected effects?*

The results of this low-dimensional, lumped-element simulation of VVFC reproduce the hypothesized effects: (1) effective mass increased, (2) damping increased, (3) new modes of oscillation and oscillatory irregularity emerged, and (4) mucosal wave interference occurred. Allen & Hollien (1973) and Laver (1975, 1980) speculate that (1) and (2) should be consequences of VVFC. In the model, increased effective mass (1) is indicated by a decrease in the fundamental frequency of the oscillating system (judged from the glottal flow); this happens despite the fact that VVFC actually introduces additional stiffness into the system. This additional stiffness, however, is accompanied by
(2) increased damping. In the model, the increased damping was probably responsible for the difficulty in achieving self-sustaining oscillation when VVFC was applied (and in some simulations it was necessary to turn off the VVFC damping to achieve vibration). When self-sustaining oscillation did occur, increased damping was manifest in reduced magnitude of displacement of the upper-cover mass and, consequently, the glottal flow. It is possible that concomitant narrowing of the airway immediately above the glottis at the ventricular level also plays a factor in reducing the glottal flow (van den Berg, 1955; van den Berg, Zantema, & Doornenbal, 1957).

Hypothesized effect (3), corresponding with Laver’s notion of VVFC constituting a “composite” oscillator, finds expression in the model in the form of irregular vocal fold vibration when VVFC is applied (e.g. case study #1). Although the effect can be described as irregular, it is not entirely aperiodic; rather, it can be generally described as producing “period-alternation” pattern, primarily in the form of 2:1 amplitude modulation of the glottal pulse (i.e. every second pulse is diminished in intensity). Surprisingly, in the case studies involving stiff-slack vocal fold parameters (#2, #3, #4, and #5), which corresponded with irregular vibrational behaviour even without VVFC (especially #2 and #4), the application of VVFC in such cases caused a reinforcement-type stabilization of the oscillation towards the semi-irregular, period-alternating or double-pulsing pattern (depending on the vocal fold parameters, compare case studies #2 and #4). This suggests that VVFC is intrinsically associated with such irregular patterns of vibration: it functions to stabilize them.

Some comment should be made on what distinguishes the two types of irregular patterns observed here and in natural speech (Gerratt & Kreiman, 2001). Although
similar in appearance to the period-alternation pattern, the double-pulsing pattern is
different because of the timing of the pulses: the two pulses are phased close together
(e.g., see arrow-1 and arrow-2 in Figure 4.20) and then followed by a longer closure
phase (arrow-3 in Figure 4.20). In this case, it could be said that each individual pulse
does not constitute a complete glottal cycle; rather, they form a single, complex cycle. In
the period-alternation pattern, the timing between the pulses is more equal (e.g. the pulses
identified by arrow-1 and arrow-2 in Figure 4.17). Each individual pulse constitutes a
complete glottal cycle although the movement of mass in each cycle is different. These
differences are manifest in the spectral profile: since period alternation is periodic in
nature (i.e. the pulses alternate in intensity but arrive at regular intervals), it contains
subharmonic structure – it is amplitude modulated; double-pulsing is less periodic in
nature and, thus, has no clear subharmonic structure but rather interharmonic noise
associated with the inconsistent timing and shape of each pulse.

The period-alternation effect is very similar to what is observed to happen to the
 glottal source in cases of vocal-ventricular (e.g. Lindestad, Södersten, Merker, &
Granqvist, 2001) and vocal-aryepiglottic phonation (Moisik, Esling, & Crevier-Buchman,
2010). In such cases of independent and simultaneous vibration of the vocal folds and
some portion of the epilaryngeal structures above, the effect is more likely to be aero-
acoustic in nature rather than mechanical. In the aero-acoustic case, the glottal airflow is
damped by the opening and closing of the epilaryngeal airway. This opening and closing
is driven by entrained or partially entrained oscillating tissue (i.e. the ventricular folds or
aryepiglottic folds) which have a lower natural frequency than that of the vocal folds and,
hence, gives rise to a amplitude modulation of the glottal source. Changes in the
pressures acting on the vocal folds arising from the interruptions to downstream flow cyclically perturb their vibration. In the specific case of vocal-ventricular vibration, during every second glottal period, when the ventricular folds come into complete contact, there is increased pressure recovery above the glottis, and this reduces the transglottal pressure drop, momentarily inhibiting self-sustained oscillation of the vocal folds (Bailly, Pelorson, Henrich, & Ruty, 2008; Bailly, Henrich, & Pelorson, 2010). The ventricular space must be open to allow for these effects to occur, but in the case of real VVFC, there is no ventricle (since vocal-ventricular fold contact mostly obliterates the ventricle).

In the mechanical case, the perturbation to vocal fold vibration is primarily attributable to an increase in the number of natural modes of vibration of the overall system resulting from coupling the ventricular fold mass to that of the vocal folds. It is well established that coupled oscillators will tend to lock into frequency regimes characterized by the superposition of these modes (Tokuda, Horáček, Švec, & Herzel, 2007), and this would appear to be true for the VVFC model. This does not rule out simultaneous and supporting aero-acoustic influence in triggering the period-alternating mode, but it is unlikely that, in the VVFC model presented here, such effects are solely responsible for the behaviour of the model in these cases. The damping of the system should also be considered, since it will likely cause vibrations associated with new modes of oscillation introduced by VVFC to decay rapidly – perhaps too quick to have any appreciable effect on the vocal folds.

There is evidence that VVFC causes (4) interference with the mucosal wave. For natural phonation, the phase difference ($\phi$) is typically 27 to 61° (Story & Titze, 1995, p.
In the VVFC model, the VVFC condition was almost always associated with an increase in $\phi$ towards values exceeding $61^\circ$ (although the extent varies with the specific parameters used). Large shifts in cover mass phasing can result in glottal leaks during the glottal closure phase if one mass starts abducting too early or one adducts too late. Examples of glottal leaks because of skewed cover mass phasing are in case studies #2 and #4. Such leaks might further destabilize vocal fold vibration if they grow with every glottal cycle. For example, see the no VVFC case in case study #4; ironically, in such cases, VVFC might help to stabilize the extent of irregularity by partially correcting the phasing of the upper-cover mass.

Overall, the focus here is on the **mechanical** role of VVFC. Aero-acoustic effects are not considered directly, although it is acknowledged that these likely play a significant role. Aero-acoustic considerations in the model follow the assumptions made in Birkholz (2005), one of which is that there is no pressure recovery at the exit of the glottis (or any cross-sectional expansion; pp. 60-61). In conventional models of glottal flow, a sudden expansion is assumed to occur at the glottal exit. In such cases, pressure recovery is assumed to follow the dynamic pressure loss (Titze, 2006, pp. 254, 278), but under conditions of a wide epilaryngeal duct area, the pressure recovery can be neglected and jet flow can be applied to sections above the minimum glottal diameter. If, however, the epilaryngeal duct area is very narrow, which is true for constricted phonation types, the pressure recovery is no longer negligible. The approach taken for the present model was to adhere closely to the aero-acoustic model in Birkholz (2005), which neglects pressure recovery throughout the entire vocal tract. Since this may have a significant impact on the behaviour of the model, it needs to be addressed in future research.
**Question B: What are the effects of the different types of VVFC?**

Although the original goal was to compare – on equal footing – the different types of VVFC (collision-and-mucous [CM] type and collision-only [CO] type and with or without coupling between m_v and m_b [VB coupling]), it was found that consistent parameter settings were very difficult to achieve as, depending on the specific conditions, the addition of VVFC very often resulted in failure of the vocal fold masses to engage in self-sustaining oscillation. It can be said, very generally, that self-sustaining oscillation was achieved with both types of coupling, and the effects were consistent with the VVFC hypothesis for both types of coupling. Based on the observation that mucous can lower the frequency response of normal vocal fold vibration (Ayache, Ouaknine, Dejonkere, Prindere, & Giovanni, 2004), it might be predicted that similar mucous adhesion forces caused by VVFC should also play a role in lowering F0 and possibly have an effect on the mucosal wave. Future modeling work will have to return to this question.

The presence or absence of VB coupling (represented by k_{vb} and d_{vb}) did not have a strong effect on F0 or \( \phi \), but it should be noted that values were necessarily kept low to allow for self-sustaining oscillation to occur (Table 1). It should also be noted that the geometry of the model is such that the ventricular mass m_v is centered closer to the body mass m_b than the upper-cover mass m_u, and since all masses in the model are constrained to just the lateromedial degree of freedom, the active, lateromedial component of VVFC force acting on m_b by means of k_{vb} and d_{vb} was relatively small in relation to other forces acting on m_b. The motivation for modeling VB coupling is that the ventricular fold
footprint on the vocal folds is not just over the medial part of the cover, but likely also
over its more lateral part (as depicted in Figure 1b). In the ST95 model, the vocal fold
cover is medial: there is no lateral section – the cover mass is lumped towards the medial
edge of the vocal folds. To better understand what effect the more lateral region of mass
distribution in VVFC might be having, it will be necessary to modify the model to
include the lateral portion of the vocal fold cover.

Question C: Can “constricted” sounds be simulated without VVFC?

It was demonstrated that, with somewhat extreme stiff-slack settings, it was
possible to simulate creaky-like phonation without VVFC. It was also possible to
simulate glottal stop without VVFC simply by increasing vocal fold adduction. In the
model, VVFC appears to enhance glottal stop by suppressing glottal leakage during the
stop phase when stiff-slack settings apply. In natural instances of constricted sounds, such
as creaky voice, harsh voice, and glottal stop, apparent ventricular-fold incursion upon
the vocal folds (“ventricular incursion”) is reported in laryngoscopic studies of such
sounds. Furthermore, we have good reason to believe that contact does indeed occur
based on the laminographic and laryngeal ultrasound evidence. What implications does
this contact have for the nature of constricted sounds like glottal stop and creaky and
harsh phonation?

One interpretation is that the essential component of “constricted” sounds is at the
vocal fold level (see e.g. Brunelle, Nguyễn, & Nguyễn, 2010): in this interpretation, the
epilaryngeal component is not essential; rather, it is merely an idiosyncratic feature, an
option, enhancement (*sensu*, e.g., Stevens & Keyser, 2010) or possibly a mechanical by-product of vocal fold tension and position adjustment used to produce such sounds but not functionally important. Another possibility is that laryngeal closure is intrinsically characterized by a cascade of gestures, each increasing in degree of stability, but each “costing” more time (and perhaps “effort”) to execute. At least three larynx internal components in this cascade can be identified: the vocal folds, the ventricular folds, and the aryepiglottic folds (working in concert with the epiglottis – the “AE level”). Stopping phonation with vocal fold adduction is quick but less stable than engaging vocal fold adduction in combination with ventricular incursion. Engagement of the AE level is the third stage in the cascade, and although it is phonetically classified as aryepiglottal stop, it can indeed occur in the production of even basic “glottal stop” as Lindqvist-Gauffin (1972) originally observed (and our own laryngoscopic observations corroborate this). This is presumably the most stable, oscillation-inhibiting degree of laryngeal closure, but also the most costly in terms of physiological effort and time required.

Support for the cascade view can be found in the literature. In examining glottalized stops in Fukienese/Taiwanese, Iwata et al. (1979) observed that apparent ventricular incursion always occurred in isolated forms which were relatively long in duration. However, at increased speech rates, laryngeal closure diminished: first, the ventricular fold involvement decreased, and then, at a relatively higher rate, vocal fold adduction diminished. Their interpretation was that ventricular incursion has “a

---

110 Additional support comes from observations made by the author that considerable posteroanterior motion of the epilarynx is evident even during very rapid vowel–glottal-stop–vowel sequences. In such cases, there is no obvious, visible adductory motion of the vocal folds and the ventricular folds only slightly medialize and do not make contact with each other.
reinforcing effect for the prevention of the vocal folds vibration [sic]” (Iwata, Sawashima, Hirose, & Niimi, 1979, p. 77). Possible factors making the vocal folds prone to continue vibration even during stop gestures include inertia, the impossibility of immediately nullifying subglottal pressure, and the concave downwards or “inlet valve” shape of the subglottal space (Fletcher, 1993; Edmondson, Chang, Hsieh, & Huang, 2011). As suggested by many in the literature and as demonstrated here, increased effective mass, damping, and stiffness from VVFC will rapidly diminish displacement of the folds, hinder the oscillation, and thereby limit glottal leakage in stop gestures. If ventricular fold adduction completely seals off the laryngeal airway, the (theoretically) infinitely high resistance to airflow will also greatly hinder phonation.

Yet another interpretation is that, as Lindqvist-Gauffin (1972) suggested, speech functions in general have been exapted from phylogenetically basic functions, one of which is laryngeal closure. In this view, glottal stop is inherently epilaryngeal in nature, being closely related to gestures such as effort closure and swallowing closure (Fink, 1974a), but refined for the purpose of speech under pressure to involve minimal effort and rapid production. This view is similar to the cascade view discussed above, but predicts that, given enough time, glottal stop will be realized with full epilaryngeal closure. The fact that TA contraction has as much relevance to vocal fold closure as it does to engagement of the epilaryngeal closure mechanism (first at the ventricular folds and then at the AE level) lends support to this. Furthermore, this view regards constricted phonation to be intrinsically associated with epilaryngeal function, not strictly vocal fold function. Furthermore, the fact that TA contraction generates the stiff-slack configuration (Hirano, 1975; Fujimura, 1981, pp. 279–280; Zemlin, 1998, p. 129; Esling, Zeroual, &
Crevier-Buchman, 2007; Deguchi, Kawahara, & Takahashi, 2011), as demonstrated by the VVFC model to be prone to irregular vibration, is consistent with this. A further point of interest is that, in parallel with the well-known fact that the aryepiglottic muscles are extensions of the oblique IA muscles, Reidenbach (1998a) has shown that the external branches of the TA muscles\textsuperscript{111} may, similarly, have extensions into the transverse IA muscles. Thus, in terms of musculature, glottal closure at the vocal fold level (via, minimally, the IA muscles) is intimately related to the mechanism driving epilaryngeal closure.

Determining which viewpoint is the most plausible cannot be decisively done here. While the model suggests that VVFC changes vocal fold behaviour in ways that are consistent with real observations of the production of constricted sounds, further physiological, imaging, and modeling studies are required which take into consideration the role of the epilarynx in constricted sound production in natural speaking and with various phonetic and phonological contexts. Understanding “normal” function should further shed light on abnormal behaviours of the larynx in disordered speech.

*Observations about model design and limitations*

Two different types of vocal-ventricular fold coupling were explored (one involving mucous adhesion forces and collision forces and one involving only collision forces). However, it proved difficult to simulate these two types on equal footing in terms of parameter settings, which made comparison difficult. It is believed that the collision-\textsuperscript{111} Reidenbach describes these extensions as part of the “posterolateral muscular system”, “a craniolateral extension of the thyroarytenoid muscle” (p. 365) found within the ventricular fold region.
and-mucous-type coupling is more physiologically realistic, but future research should refine how this is implemented. It may be that modeling adhesive coupling will continue to be relatively intractable in the lumped-element approach.

Increasing degrees of freedom or using a finite element approach are likely future directions to examine the effects of VVFC with greater degrees of physiological realism, including prediction of the effects of mucous. Increasing the resolution of the model in this way may also reveal other effects not reproducible with the present model. For example, in laryngoscopic observations, it is evident that apparent ventricular incursion influences the longitudinal dimension of the vocal folds, but these aspects cannot be studied in a one-dimensional, lumped-element model. Future modeling studies must also consider in more detail the concomitant aero-acoustic influences on vibration associated with very narrow epilaryngeal aperture and an obliterated ventricle. These effects were not considered in detail here, but they are likely highly important in determining the behaviour of the system.

4.3.6 Summary of the vocal-ventricular fold coupling model

This research has examined the hypothesis of vocal-ventricular fold coupling (VVFC) using a low-dimensional, lumped-element computational model of laryngeal biomechanics. We have shown that such a model can reproduce effects expected to occur with VVFC: increased effective mass and damping of the laryngeal system and the introduction of new modes of oscillation cause lowering of F0, inhibit oscillation, and can change vocal fold dynamics such that constricted phonation results. The products of simulation with coupling are classifiable as creaky-like or harsh-like. Other effects not
mentioned in previous literature include mucosal wave interference, reinforcement of irregular patterns of vibration ("irregularity reinforcement"), a form of (2:1) amplitude modulation ("period-alternation") consistent with previous observations of constricted forms of nonmodal phonation (Gerratt & Kreiman, 2001), and enhancement of laryngeal closure in glottal stop articulation.

The present model generally supports the view that vocal-ventricular fold coupling and, more generally, epilaryngeal stricture are highly important in the explanation of the nature of constricted phonation and laryngeal closure in speech and vocalization. This has significance for phonetic and phonological issues regarding the role of the epilaryngeal stricture mechanism in speech, particularly in relation to the physiological basis of sounds, such as glottal stop and creaky voice, both long considered to be primarily attributable to action of the vocal folds. The work also has significance for speech simulation, where it is desirable to use computationally inexpensive models (such as the lumped-element formulation presented here) to simulate natural sounding voices (Birkholz, 2005, 2011). The VVFC model may serve as a means to synthesize constricted phonation types and could improve the synthesis of glottal stop. While this work does not directly examine disordered speech, it is certainly relevant to this topic: for example, hyperfunctioning of the epilarynx in functional speech disorders might explain some of the characteristic constricted qualities of these voices, such as excessive glottal attack (Hess, Verdolini, Bierhals, Mansmann, & Gross, 1998). In general, this work constitutes a preliminary basis for future, more elaborate modeling studies which examine interaction between the epilarynx and the vocal folds.


4.4 Chapter summary: The epilarynx and the vocal folds

This chapter has dealt with the interaction between the vocal folds and the epilarynx. This concerns two types of phonetic laryngeal behaviour: stop production and phonation. Traditional models posit activity primarily in the glottic (i.e. vocal fold) plane. The view advocated by the research here, which corroborates and elaborates upon ideas put forth in previous research (e.g. Lindqvist, 1969; Lindqvist-Gauffin, 1972; Catford, 1977a, pp. 103–104, 163; Painter, 1986; Esling, 1996, 2005), is that, while the vocal fold level is important in speech, the role of the epilarynx cannot be overlooked in producing “constricted” sounds. The main effect is vocal-ventricular fold contact (see §4.1), which is facilitated by larynx raising (see §4.2) and which is considered to influence the biomechanical behaviour of the vocal folds, giving rise to creaky and (some forms of) harsh phonation (see §4.3) and an intrinsic part of glottal stop – but one that acts slower than the medial compression mechanism (see §4.2).

As hinted at in this chapter, the mechanism of vocal-ventricular fold contact is the first in a broader sequence of epilaryngeal constriction, which gradually progresses to full “collapse” of the laryngeal airway. At this more extreme end of constriction, more and more of the vocal tract configuration becomes compromised, leading to the intriguing possibilities of aryepiglotto-epiglottal and linguo-epiglottio-pharyngeal stricture, and the rather surprising effects characterizing raised larynx voice. These are the topics of Chapter 5.
Chapter 5
THE EPILARYNX AND THE SUPRALARYNGEAL VOCAL TRACT

In this chapter, several production studies are presented that explore the articulatory consequences of epilaryngeal constriction on the supralaryngeal vocal tract. These studies help illustrate the relationship between the larynx internal mechanism of epilaryngeal constriction and the primary external mechanisms, represented by tongue retraction and larynx raising. These three components characterize the role played by the epilarynx as a link between the vocal folds below and the supralaryngeal vocal tract above. A direct consequence is that manipulation of epilaryngeal state translates into a manipulation of voice quality, but it also means that the epilarynx has a complex relationship with vowel quality. The presence of epilaryngeal linkage between tongue and vocal folds exerts a subtle influence even within conventional “modal” vowels; on the other hand, more extreme epilaryngeal action is possible which induces a reconfiguration of the acoustic vocal tract.

This chapter thus covers a range of epilaryngeal behaviour. The first of the production studies (§5.1) illustrates the functional unity of the epilarynx in acting continuously from fully expanded to fully closed states. The productions allow us to revisit the subject epilaryngeal vibration (Chapter 3) and laryngeal closure (Chapter 4) but now contextualize these actions of the epilarynx within the larger picture of the vocal tract. The second study (§5.2) marks a return to the SLLUS technique, which was introduced in §4.2, but this time more extreme epilaryngeal stricture states are examined as they occur in pharyngeal consonants. The third study (§5.3) then addresses the peculiar
vowel quality effects associated with raised larynx voice (“pharyngealization”). Following this (§5.4) is a theoretical discussion of the subtle and extreme effects of extreme epilaryngeal stricture on vowel systems. The chapter is then summarized in (§5.5).

5.1 Visualizing epilaryngeal constriction continuum with videofluoroscopy

In §2.1, the relationship between the pharynx and the epilarynx was modeled as a tube-within-a-tube (see Figure 2.1). In swallowing, the epilarynx must maintain full closure while the pharynx temporarily dilates to allow for the passage of food, which suggests the independence of the structures. It can be further observed (as it was in §2.2) that the same is true in speech, leading to the generalization that pharyngeal stricture (largely under lingual control) does not entail epilaryngeal stricture¹¹². This can easily be demonstrated by considering the difference between [a], during which the pharynx becomes extremely narrow as Gauffin & Sundberg (1978) show, and [aʕ]¹¹³, which involves additional epilaryngeal stricture. In the literature discussing pharyngeal stricture in speech, the pharynx and larynx are assumed to act in unison, as if the larynx were merely an extension of the pharynx. It is especially common to encounter discussion of constriction in the laryngopharynx, but this conceptual collapsing of the pharynx together with the epilarynx leads to confusion in regard to how the epilarynx relates to vowel quality, since pharyngeal constriction does not imply epilaryngeal constriction.

¹¹² Shin et al. (1981, p. 177) demonstrate that the pharyngeal constrictors have a minor influence on laryngeal closure at the vocal fold level relative to the far more important intrinsic laryngeal musculature and the extrinsic musculature controlling larynx height as it is mediated through hyo-lingual-mandibular musculature.

¹¹³ As it might be spoken by Kermit the Frog (or by Mr. Bean).
This study provides a rare view of the independent action of the epilarynx with respect to the pharynx by examining videofluoroscopy data of three “pharyngeal” consonants. These data provide insight into the relative changes in pharyngeal and epilaryngeal tubes and provide some clues about the opportunistic nature of the epilaryngeal constriction.

5.1.1 Videofluoroscopy methodology

The videofluoroscopic data were obtained by John Esling with the assistance of Leonardo Fuks and Dr. Milton Melciades Barbosa Costa at the Instituto de Ciências Biomédicas, Departamento de Anatomia, Universidade Federal do Rio de Janeiro. Although a videofluoroscopic system was used, no barium-impregnated material was ingested by our participant because the objective was to image speech sound production rather than swallowing. The participant was screened radiographically in the lateral position (from the left). Beam intensity was set to provide optimal imaging of the epilarynx and pharynx. Audio was recorded using a handheld camcorder (as it was primarily intended to be a temporal guide in the illustrations given below). The participant produced three tokens of each sequence, and the best examples of each were used for illustration and qualitative analysis.

---

114 The data were obtained by John Esling with the assistance of Leonardo Fuks and Dr. Milton Melciades Barbosa Costa at the Instituto de Ciências Biomédicas, Departamento de Anatomia, Universidade Federal do Rio de Janeiro. These data were then analyzed by Scott Moisik (with some assistance from John Esling), as described in §5.1.1. The results and discussion represent the work of Scott Moisik. The data were originally presented (by John Esling) at the 2010 biennial colloquium of the British Association of Academic Phoneticians (BAAP) (March 31st, 2010; London England); and later published in Esling & Moisik (2011).
The data consist of canonical phonetic productions of intervocalic, [a]-context pharyngeal consonants: AE stop [ʔ], voiceless AE trill [n], and voiced AE trill [ʢ], all of which are produced by phonetically emulating the Iraqi Arabic trills described in §3.2. The participant is a 60-year-old male phonetician.

In the analysis, the relative midsagittal areas of the visible parts of the epilarynx and pharynx were compared, along with larynx height. The parameters were extracted using MATLAB (R2009a) with the help of two custom data-processing GUIs, as depicted in Figure 5.1.

Figure 5.1: Parameter extraction from videofluoroscopy data using custom MATLAB GUIs. The outline in (a) roughly corresponds to the epilarynx; the outline in (b) roughly corresponds to the pharynx; the star in (c) illustrates the approximate location manually selected each frame to estimate larynx height. The small white dot at the bottom of the videofluoroscopic window is the reference point for distance measurement.

The first GUI allows multiple area data to be collected simultaneously for a single video sequence by tracing regions of interest (ROIs) within a duplicated, time-sweepable frame and then luminance-thresholding the video to obtain a white-valued pixel region approximating the desired area. Two areas were defined: one over the epilarynx (Figure
5.1a) and one over the pharynx (Figure 5.1b) and care was taken to ensure that these areas did not overlap. The GUI converts the image data to black and white (from grayscale) by the user-selected luminance threshold values and superimposes these over the video frame to allow the user to visualize the measured area and judge whether it matches the ROI suitably. The user can sweep through the video to determine if the thresholding represents the sought-after area and adjust the value if necessary. This technique assumes that the sought-after areas, in this case, those of the epilarynx and pharynx in the videofluoroscopic video, are essentially uniform in grayscale value. This is admittedly an idealization, and both areas could not be perfectly extracted given grayscale variation in each region. This confound is mitigated, however, by the fact that both ROIs were subject to this problem. With the GUI method, it was judged that the extracted areas adequately represent changes occurring to each tube of interest.

Once a threshold value was obtained, the pixel area of each lumen over time was obtained using the MATLAB function `bwarea()`. The areas were made comparable by converting each into a percentage of their respective area maxima (for a given production).

The second GUI (Figure 5.1c) steps frame-by-frame through a video and allows the user to obtain the distance between a reference point and a manually located point. To increase the reliability of the results, the user is required to repeat the measuring process three times; the average of the three measurements is produced as output. Using this GUI, larynx height was tracked by locating the inferior most point of the cricoid arch (star in Figure 5.1c), which was the easiest reference point on the larynx to identify consistently on a frame-by-frame basis. As with the area data, the larynx height data were normalized
to a percentage wise measure relative to the global maximum in larynx height for each video.

5.1.2 Videofluoroscopy results and discussion

Selected, non-contiguous frames from the videofluoroscopic examination of epilaryngeal tube constriction are provided in Figure 5.2 to illustrate the various states of the epilaryngeal tube. Image (a) shows the laryngeal state associated with forceful inhalation. During this state the larynx is lowered and the tongue and epiglottis are advanced; the result is that the epilarynx is maximally expanded both in caliber and height to allow for unimpeded airflow into the lungs. Note the large separation between the hyoid bone and cricoid arch, which implies vertical expansion of the epilarynx and concomitant vocal-ventricular fold separation.

Image (b) shows the epilarynx in a relatively neutral configuration associated with quiet breathing. The state resembles (a) except the tongue is retracted such that the vallecula (v) no longer appears in the image and the epilarynx is not as widely patent, but there is still vocal-ventricular fold separation and a large posteroanterior separation between the epiglottis and the arytenoid cartilage complex.

115 It is more extreme than the inspiration that occurs during tidal breathing, hence it is labeled deep inspiration. This frame represents the most extreme inspiration observed in the data; generally, just prior to production a quick breath is drawn by the participant which does not approach the extremity seen in the illustrated case of deep inspiration (Figure 5.2a).

116 The somewhat retracted location of the hyoid bone here may be due to traction from the stylohyoid and stylopharyngeus muscles, which will lift the hyoid bone and dilate the pharynx. Note also that the cricoid is rotated anteriorly, which can be attributed to the larynx-lowering influence on cricoid orientation described by Honda et al. (1999).
Figure 5.2: Videofluoroscopic images of the epilarynx in its unconstricted and constricted states. Non-contiguous frames selected to show the epilaryngeal configuration continuum from fully unconstricted/open (a) to fully constricted/closed (d). Dotted line = approximate epilarynx sagittal area. Brace = hyoid-cricoid gap. Arc (a & b) = ventricular fold edge. States: deep inspiration (a); image neutral position (b); [ə] (c), and aryepiglottico-epiglottal stop [ʔ] (d). Below are traces that show the key details of frames (c) and (d). Grey region = epilaryngeal tube; a = arytenoid complex (arytenoid, cuneiform, and corniculate cartilages); c = cricoid cartilage; e = epiglottis (apex); et = tubercle of the epiglottis; h₁₂ = hyoid bone location in (a) and (b); h₃/d = hyoid bone location in (c) and (d); j = jaw/mandible; t =tongue; v = vallecula.
Images (c) and (d) illustrate two progressively more constricted states of the epilarynx. In (c) the epilarynx remains open, but it is clearly narrowed with respect to (a) and (b): the sound occurring at the time of this frame is the vowel [ɑ̃]117 which has been coloured by strong co-articulatory pharyngealization just prior to the [ʔ] (d). The tongue must further retract (i.e. move back and downwards to descend into the lower pharynx) for the production of [a]. To facilitate the retraction, it appears that the hyoid bone has been lowered (presumably by means of the thyrohyoid muscles), which we might further infer has benefits for the action of the primary agonists responsible for retracting the tongue – the hyoglossus muscles. The low hyoid position does not, however, mean that the larynx is low, since the gap between cricoid and hyoid has reduced considerably, which is made evident by comparing this distance as it appears in unconstricted-states (a) & (b) and constricted-states (c) & (d) (alternatively, one can compare the hyoid bone locations, \(h_{ab}\) and \(h_{cd}\) using the tracing).

During [ʔ] (d) the epilarynx completely closes (which is further illustrated in the videofluoroscopy tracings below), but, critically, the pharynx does not close completely118. While some pharyngeal narrowing does occur in [ʔ] (compared with [ɑ̃]), it cannot account for the epilaryngeal closure, which is assumed to be driven primarily by the intrinsic laryngeal musculature (see §2.1.2). Some additional tongue retraction (which accounts for the slight pharyngeal narrowing) appears to accompany the posterior displacement of the epiglottis, which aids in epilaryngeal closure; thus the mechanism in

---

117 Due to the somewhat spontaneous nature of the videofluoroscopy session, no non-pharyngealized [ɑ] was obtained. This oversight must be remedied by future research either using videofluoroscopy or MRI imaging to compare pharyngealized [ɑ̃] with non-pharyngealized [ɑ]. Comparison to non-pharyngealized [ɑ] produced by a different speaker and during an unrelated imaging session can be found in Figure 5.15b.

118 It would, however, close completely (in addition to complete epilarynx closure) during [ʔ], as in Amis (Edmondson, Esling, Harris, & Huang, 2005).
this case is not entirely a larynx internal one. (The mechanism may involve displacement and stressing of the pre-epiglottic body, a mass of fatty tissue just anterior to the epiglottis, by tongue retraction, which in turn could cause the displacement of the epiglottis itself towards the posterior border of the epilarynx and may also cause additional lowering of the ventricular folds onto the vocal folds. This is discussed further in §5.4.1.)

Figure 5.3, Figure 5.4, and Figure 5.5 respectively illustrate the relationship amongst estimated epilarynx area (solid line), estimated pharyngeal area (dashed line), and larynx height (dotted line) for [ʔ], [ʰ] and [ʝ]. (Note that all signals are relativized for temporal, rate-wise comparison; magnitude comparison is only possible within a measure – **not** across measures.)
Figure 5.3: VFS data for [ɑʔɑ]. Audio (top); Relative measures (bottom) expressed as percentages of the individual maxima of each time series; solid line = epilarynx area; dashed line = pharynx area; dotted line = larynx height. a = inspiratory breath preceding production; b = initial vowel; c = epilaryngeal closure; d = second vowel. Average F0: first vowel = 120 Hz; second vowel = 135 Hz.

In [ʔ] (Figure 5.3), after initial area expansion associated with a quick inspiratory breath (a), the epilarynx area diminishes at a much faster rate than pharynx area in anticipation of the short duration and heavily pharyngealized [ɑʕ] (b); then the epilarynx area closes entirely during [ʔ] (c), while pharyngeal area slightly narrows. Finally, the second vowel (d) is notably less pharyngealized towards its offset than the first vowel, and this corresponds with a relaxation of epilaryngeal stricture, and it becomes more patent than it was during (c). The pharynx also relaxes, but to roughly the same degree as
during (b). Larynx raising has two peaks during the production: one in the transition from (b) to (c), viz. the onset of [ʔ], and one during the second vowel (d), which probably reflects slightly increased F0 (~15 Hz higher than during the first vowel). Overall then, epilaryngeal area relates to “pharyngealization” and stop production, while pharyngeal area corresponds to vowel production; correlation between these two area signals reflects sharing of the lingual retraction.

Figure 5.4: VFS data for [alpha]. Audio (top); Relative measures (bottom) expressed as percentages of the individual maxima of each time series; solid line = epilarynx area; dashed line = pharynx area; dotted line = larynx height. a = inspiratory breath preceding production; b = initial vowel; c = epilaryngeal vibration; d = second vowel. Average F0: first vowel = 120 Hz; second vowel = 130 Hz.
In [h] (Figure 5.4), once again we observe initial area expansion of the pharynx and epilarynx associated with inspiration (a), but, unlike the initial breath in Figure 5.3, it is notably longer in duration and exhibits larynx lowering; a visual inspection reveals that we are looking at deep inspiration (as in Figure 5.2a). During the production itself, once again we observe a sharp drop in epilaryngeal area during the initial [ɑ] (b), which begins to exhibit some anticipatory epilaryngeal vibration. In the transition from (b) to (c) is a sudden, further drop in epilaryngeal area marking the onset of [h] (c). Epilaryngeal and pharyngeal area continually expand during (c), until (d), at which point epilaryngeal area is at its largest value (during the target sequence) while the pharynx starts to decrease in size, especially towards the end of the [ɑ], which becomes less and less “pharyngealized” and more like ordinary [ɑ]. As in the previous case, larynx height rises steadily throughout the production, showing a sudden jump during the [h] and continuing to rise throughout until its peak in (d), which also corresponds with a 10 Hz increase in F0 in comparison to the first vowel.
Figure 5.5: VFS data for [ɑʢa]. Audio (top); Relative measures (bottom) expressed as percentages of the individual maxima of each time series; solid line = epilarynx area; dashed line = pharynx area; dotted line = larynx height. a = inspiratory breath preceding production; b = initial vowel, epilaryngeal vibration, and onset of second vowel; c = offset of second vowel. Average F0: first vowel = 130 Hz; second vowel = 145 Hz.

The final case is [ʢ] (Figure 5.5), a voiced aryepiglottic trill. This production also illustrates inspiratory expansion of the epilarynx and pharynx with concomitant larynx lowering (a) – again, this is deep inspiration (as in Figure 5.2a).\textsuperscript{119} The first vowel is almost immediately subject to epilaryngeal vibration, and this is

\textsuperscript{119} It could be that the occurrence of deep inspiration prior to the trills (Figure 5.4a and Figure 5.5a) and not the stop (Figure 5.3a) reflects the participant preparing for the increased aerodynamic demands of engaging epilaryngeal vibration.
manifest in the immediate and extreme narrowing of the epilarynx (b), more so than what was observed for the voiceless AE trill in Figure 5.4. Pharynx area continues to diminish during (b) while epilarynx area expands somewhat, until (c) when the epilarynx suddenly springs open as the [ʕɑ] quickly takes on a normal quality ([ɑ]) and epilaryngeal vibration ceases. Larynx height once again rises throughout, and, at the peak, the average $F_0$ is 15 Hz higher than it was during the early part of the first vowel.

5.1.3 Videofluoroscopy summary

This study has demonstrated the sagittal appearance of epilaryngeal and pharyngeal actions during sounds traditionally described as pharyngeals, but the results of the area measurements suggest a different interpretation, as the pharyngeal cavity, which is narrowed intrinsically for [ɑ] “pharyngealized” quality of the vowels in the context of the “pharyngeal” consonants [ʔ], [ɦ], and [ʃ], which are actually aryepiglottal-epiglottals, as the data confirms. A suggestion for a better term for the secondary articulation of [ɑʕ] or [ɑʢ] would be epilaryngealization since changes in epilarynx area more strongly correlate with the auditory judgment, and the pharynx is certainly not responsible for the vibration occurring.

Any correspondence between the epilaryngeal and pharyngeal tubes (as judged by the area measures) indicates their shared influence by the tongue and larynx height. There is a third, and perhaps more fundamental mechanism to epilaryngeal control that the pharynx does not share – the intrinsic musculature of the larynx, which can close the epilarynx regardless of the status of the other two mechanisms. In the cases seen here, larynx height appeared to do less work on account of the intrinsically retracted tongue
position of the [a] vowel. It is apparent that tongue retraction does have an influence on the position of the epiglottis, and we might infer that this relationship is structurally mediated by the pre-epiglottic body (which will be discussed further in §5.4.1), which buckles during tongue retraction and no doubt transmits this buckling stress directly to the epiglottis in the region of the tubercle. A further consequence may be that these stresses are then transferred to the ventricular folds, which will further assist the intrinsic (eTA) mechanism in pushing the ventricular folds, particularly at their anterior extent, down into the vocal folds.

Such an epilarynx-mediated mechanism, helps us to understand the common observation that /ʕ/ often involves laryngealized phonation or is in fact a stop (Butcher & Ahmad, 1987, p. 166). Once the vocal folds become compromised by ventricular fold contact, their dynamics can be biased towards creakiness (or even harshness; in either case, see §4.3). The relationship is not so much between the pharynx and larynx, as suggested by some (Keyser & Stevens, 1994, p. 213), but rather between the tongue, the larynx height mechanism, and the vocal folds – with the epilarynx acting as the “conductor” of all of the activity. The next section (§5.2) explores these relationships as they play out in [ʔ], [ʰ], and [ʃ] using SLLUS.

5.2 SLLUS observations of upper epilaryngeal stricture

In the previous section, videofluoroscopic imaging of epilaryngeal stricture in sounds traditionally described as pharyngeal consonants revealed the relatively

---

120 Although, as Esling (1999, p. 354) suggests, laryngealization may in some cases be a mislabeling of epilaryngeal vibration.
independent nature of the epilaryngeal and pharyngeal tubes. Furthermore, in the [ɑ] we observed that larynx height seemed to change more with pitch than with epilaryngeal stricture, which can be explained by intrinsically retracted tongue position of [ɑ], which benefits from a relatively low hyoid bone position. It has often been observed, however, that pharyngeal consonants typically involve larynx raising. Some values are provided in Table 5.1; the values reported here indicate ~1 cm\textsuperscript{121} of larynx raising is typical of Arabic pharyngeals (also see Heselwood, 2007, p. 7). No study has shown how larynx height changes over time in the production of pharyngeals.

Table 5.1: Larynx raising in Arabic pharyngeal consonants (values in mm)

<table>
<thead>
<tr>
<th>Language</th>
<th>h</th>
<th>f</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tunisian Arabic</td>
<td>–</td>
<td>10</td>
<td>(Ghazeli, 1977, p. 36)</td>
</tr>
<tr>
<td>Iraqi Arabic</td>
<td>–</td>
<td>13</td>
<td>(El-Halees, 1985, p. 288)</td>
</tr>
<tr>
<td>Moroccan Arabic</td>
<td>4</td>
<td>7</td>
<td>(Boff-Dkhissi, 1983, p. 147)</td>
</tr>
<tr>
<td>Iraqi Arabic</td>
<td>7</td>
<td>7</td>
<td>(Butcher &amp; Ahmad, 1987)</td>
</tr>
</tbody>
</table>

This study\textsuperscript{122} returns to the SLLUS methodology (§4.2) to provide a different perspective of the same pharyngeal consonants. With SLLUS it is possible to correlate laryngeal state change with larynx height change, and the latter is quantifiable by means of optical flow analysis. This should make it possible to examine further how the external mechanisms influencing epilaryngeal constriction – tongue retraction and larynx raising –

\textsuperscript{121} Most of these measurements reflect male larynges; female larynges might be expected to exhibit less extreme larynx raising in terms of absolute value given the (somewhat) smaller overall scale of the structures.

\textsuperscript{122} The data shown in this section were originally presented (by John Esling) at the 2010 biennial colloquium of the British Association of Academic Phoneticians (BAAP) (March 31\textsuperscript{st}, 2010; London England) and then published in Esling & Moisik (2011); similar data with revised plotting appears in Esling & Moisik (2012). Analysis, interpretations, and text are attributed primarily to Scott R. Moisik with assistance from John Esling.
relate to the internal mechanism embodied by the laryngeal musculature driving anteroposterior narrowing of the epilarynx (§2.1.2).

5.2.1 SLLUS methodology

The methodology in this study was nearly the same as that used in §4.2, except the production targets were different: the sounds [ʔ], [ɯ], and [ʃ] were produced in [i] context. Another difference was the use of an Olympus ENF-P3 flexible fibreoptic nasal laryngoscope, as this allows for more natural [i] production than the orally-inserted rigid endoscope.

5.2.2 SLLUS results and discussion

At the start of the AE stop [ʔ] (Figure 5.6), only the vocal folds are adducted (frame 16), as necessary for modal vowel-phonation. Very quickly, the epilarynx begins to collapse: first at its lower border (frame 19), such that the ventricular folds medialize over (and presumably press into) the vocal folds, then at the upper border (frame 22). This two-staged closure is, in its first phase, basically the same mechanism observed during glottal stop (§4.2), but, in this case, the concomitant posteroanterior narrowing of the upper epilarynx continues to the point of contact between the aryepiglottic folds and the epiglottis (frame 22). Larynx height increases by 16.76 mm from frame 16 to 31, where it reaches its peak and then begins its descent during the following vowel. Comparing the laryngoscopy with the laryngeal ultrasound, it is evident that larynx raising is slower acting relative to epiglottis retraction and advancement of the aryepiglottic folds. Epiglottis (and presumably lingual) retraction appears to increase in
degree throughout the production and peaks towards the offset of the sound (frame 31). Anticipatory larynx ascent and descent are co-articulatory with the offset of the first vowel and onset of the second vowel, respectively. F0 is relatively level throughout, as is larynx height, but, at the periphery of the [ʔ], F0 is elevated\textsuperscript{123}, which is possibly attributable to the increase in larynx height, increased longitudinal vocal-fold stiffness (from TA contraction to drive anteroposterior narrowing), or a combination of both factors.

\textsuperscript{123} The sudden F0 jump before the [ʔ] (immediately after the frame 19 marker line), however, is probably due to F0-analysis error caused by creaky phonation.
Figure 5.6: SLLUS data for [i?i]. ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold.

For the voiceless AE trill [h] (Figure 5.7), upper epilaryngeal narrowing with simultaneous vocal fold abduction rapidly occurs in the time span from frame 15 to 17 (~66 ms). The resulting configuration suggests that tIA is relaxed while oIA-AE is strongly engaged. This causes contact and compression of the corniculate tubercles (as the arytenoids rotate about their anteroposterior axes inwards and towards each other) and a posterior gap in the cartilaginous glottis below. The ventricular folds (lower epilarynx) do not medialize as they do in [i?], and this further supports the interpretation that tIA, which has extensions into eTA (the agonists of ventricular adduction/descent), is
relatively relaxed. Vibration of the aryepiglottic folds starts occurring just after frame 17 and is evident in frame 24 (double-headed arrows). Larynx height increases by 9.48 mm from frame 15 to 24. It shows slight anticipatory increase prior to the engagement of the upper epilaryngeal stricture. Then it continues to increase throughout the vibration phase of the [h], and it finally peaks (just after frame 24) at the offset of the consonant. Intriguingly, larynx height and F0 seem to descend in parallel from frame 24 to 32, which reminds us of the co-functioning of larynx height in F0 control, particularly for F0 decrease (Honda, Hirai, Masaki, & Shimada, 1999).

Figure 5.7: SLLUS data for [iii]. ae = aryepiglottic fold; c = cuneiform tubercle; et = epiglottic tubercle; f = ventricular (false) fold; k = corniculate tubercle; m = mucosa of inner aryepiglottic wall; v = vocal fold.
The [ʕ] is illustrated in Figure 5.8. Like the other productions, full epilaryngeal stricture is rapidly achieved (between frame 14 and 18) at which point aryepiglottic vibration begins (indicated by double-headed arrows in frame 22). The extent of larynx raising is 11.22 mm (from frame 14 to 22). The raising is not obviously serving F0-change: despite the increase in larynx height prior to the [ʕ], pitch continuously declines (from frame 14 to 18)\textsuperscript{124}. F0 and larynx height correlation resumes after the offset of the [ʕ].

![Figure 5.8: SLLUS data for [iʕi].](image)

\textsuperscript{124} F0 analysis is inaccurate during and after the [ʕ] because STRAIGHT has difficulty with the epilaryngeal vibration.
The SLLUS data for the [i]-context pharyngeals evince the relatively more important role played by larynx height when tongue retraction is not facilitating epilaryngeal closure to the same extent. In [\(a\)] context (see §5.1), the tongue is in the ideal position for producing epilaryngeal stricture – all that is required is engagement of the internal mechanism (i.e. the intrinsic laryngeal musculature). For [i], the tongue is fronted. In this case, if the tongue is to play an assistive role in epilaryngeal stricture during a pharyngeal consonant, then the lingual-fronting mechanism (primarily the GGP) must disengage or diminish and the retraction mechanism must engage. It cannot be determined from the laryngoscopy videos whether the tongue actually retracts to produce the pharyngeal, and if so, by how much – or whether the epiglottis retraction occurs largely because of the action of the internal mechanism of epilaryngeal stricture acting alone – or some mix of the two of these. Further study is needed.

Larynx height change was in the neighbourhood of values reported for Arabic pharyngeals in previous studies (Table 5.1) and shows (visually) some correlation with F0 during the modal-voiced context vowels, especially for declining values of these two variables. However, we can interpret that larynx raising during these productions primarily assisted in epilaryngeal constriction during the pharyngeal consonants but not F0 increase (unlike what seemed to be the case for [\(a\)] in §5.1). This judgment is based on comparison of the temporal sequencing of the height change in relation to the consonantal stricture visible in the laryngoscopy videos and F0 change between the [\(a\)]-context (§5.1) and [i]-context productions. In the former case, larynx height peaked after the consonant in connection with a peak in F0, while in the latter case, the peak is delayed towards the offset phase of the consonant, while F0 typically did not increase in
the second syllable but rather tended to decline or level off (all attributed to specific performance rather than inherent dependencies between F0 and vowels following pharyngeal consonants).

The larynx height mechanism is relatively slow acting, especially in comparison to the internal mechanism of constriction. The fact that larynx height consistently lags behind intrinsic laryngeal articulations may be partly attributable to inertial factors, viz. the relative sizes of masses involved. We would expect the internal mechanism to accelerate/operate relatively faster than the height mechanism on the grounds that the ventricular folds, aryepiglottic folds, and epiglottis are all relatively less massive than the overall mass of the larynx (its velocity in the data ranged between 10 and 100 mm/s; no such measure is available for the fold structures, but the laryngoscopic evidence indicates they move much faster). Furthermore, if the tongue is involved, then the fact that epiglottis retraction appears to peak late into the pharyngeal consonant (as judged by the laryngoscopy video) may be associated with a late peak in lingual retraction. Early epiglottis retraction might be driven by the more quick acting intrinsic laryngeal muscles connected with the epiglottis, such as the external thyroarytenoids, thyroepiglotticus, and possibly the aryepiglottic muscles (Painter, 1986).

5.2.3 **SLLUS summary**

It seems to be the case that the internal mechanism (vis-à-vis the external mechanisms of tongue retraction and larynx height) is the primary agonist of epilaryngeal stricture, while tongue retraction and larynx raising play supportive roles which produce parallel, if not as extreme, effects as the internal mechanism but are relatively slow
It also appears to be the case that the relative importance of larynx height and tongue retraction is a function of the vowel context, and this suggests complementary functioning of the external mechanisms. In [a] context (as in §5.1), we can infer that tongue retraction dominates in assisting epilaryngeal stricture while in [i] context larynx height becomes relatively more essential since the tongue is pre-occupied with fronted-vowel production. (On a methodological note, we can observe that there is good visual correspondence between height changes seen in the laryngoscopy and the quantified larynx height data, which helps validate the overall analysis.)

With these basic relations established, we are now ready to move on to the subject which connects epilaryngeal stricture with the system of vowel quality and demonstrates the nature of raised larynx voice.

### 5.3 Vowel quality change as a function of larynx height

In this study\(^\text{125}\), laryngeal ultrasound is used to evaluate the effect of larynx height on vowel formant frequencies. Once again, this involves application of the SLLUS technique and optical flow (see §4.2) to quantify change in larynx height, and, from this basis, it is claimed that while larynx lowering (generally) yields expected lowering effects on formants, larynx raising has a lowering effect on F2 and F3 more characteristic of what has been labeled *pharyngealization*, and this effect occurs despite efforts to preserve vowel quality across height conditions.

---

\(^{125}\) This study was originally published as Moisik & Esling (2011b). The data were also presented (by John Esling) at the 2012 biennial colloquium of the British Association of Academic Phoneticians (BAAP) (March 28\(^\text{th}\), 2012; Leeds, England).
Larynx height is generally assumed to correlate positively with the resonant frequencies of the vocal tract. In the simple uniform tube model of the vocal tract, lengthening or shortening the tube to represent change to larynx height yields this effect, since resonant frequencies are inversely proportional to vocal tract length (Stevens, 1998, p. 139), as illustrated in Figure 5.9.

![Figure 5.9](image.jpg)

Figure 5.9: First three resonances of a hard-walled, uniform tube. Neutral vocal-tract length (at 0 displacement) assumed for calculation is 17.5 cm.

Sundberg & Nordström (1976; hereafter SN76) studied resonance effects of larynx height. They provide simulated results using an acoustic model with 1.5 cm changes to pharynx length (taken to represent larynx height change) and canonical results from phonetic productions made by two participants with “informal estimations” that laryngeal displacements were about 1.5 cm from the normal position with “normal speaking voice pitch” (p. 37). It is noteworthy that SN76 do not manipulate the cross-sectional area of the epilaryngeal tube (either by narrowing the ventricle, the vestibule, or
both) in their acoustic model. Overall their results indicate that formant frequency has a positive correlation with larynx height, but the strength of the effect differs by vowel: in absolute terms, open vowels show the greatest effect for F1, and close (front) vowels show the greatest effect for F2. F3 and F4 are claimed to positively correlate for all vowels. Their claim is that pharynx length drives the “first-order”\textsuperscript{126} effect of larynx height on vowel formant frequencies.

5.3.1 Methods

The objective here was to repeat the basic production-study protocol of SN76 but also to provide more reliable observations for larynx-height change. To accomplish this goal we collected laryngeal ultrasound data and processed these using optical flow analysis (following the methodology in §4.2.2) to calculate vertical change in larynx position.

Two phonetician participants (A and B) produced careful productions of [i æ a u] in normal (N), raised (R), and lowered (L) larynx states while attempting to maintain constant F0 and vowel quality. Productions followed two different larynx-height manipulation regimes: NRN and NLN, and NRNL and NLNR. Elicitations lasted ~6 seconds on average. Laryngeal ultrasound was manually administered using a 12L-RS probe with a 3.84-cm-wide FOV connected to a LOGIQe portable ultrasound machine set to 8 MHz with a 2.0 cm depth. The probe is applied to the participant’s neck, 1 cm

\textsuperscript{126} It is not clear what they mean by first-order in this context, and no explanation is offered in SN76. Order in the context of mathematical functions can have many interpretations, including the order of derivatives involved, the order of exponents, the order of approximation in the Taylor Series, the order of magnitude of a variable, the order of convergence of a numerical method in estimating a function, and so forth. It could simply mean that pharynx height explains most of the observed formant effects.
posterior to the thyroid notch. Audio and video data were routed through a Canopus TwinPact100 AD video converter and captured using Sony Vegas 8.0.

The first three formants were obtained by sliding-Gaussian-window LPC analysis (8th order) of the audio signal. Formant measurements are averages taken from ROIs defined for stable larynx height targets. F0 was measured using STRAIGHT (Kawahara, de Cheveigné, & Patterson, 1998) to evaluate consistency. All computation was performed in MATLAB R2009a.

5.3.2 Results

An example of the time series data illustrating an [i] vowel in NRNL regime produced by Participant B is in Figure 5.10. These data illustrate that (relatively) constant F0 (which SN76 claimed but never demonstrated) was maintained as much as possible while larynx height was manipulated. Although formant structure appears to change considerably, especially during the raised larynx phase, it must be emphasized that every effort was made to preserve vowel quality; thus, the sequence can be transcribed using the Voice Quality Symbols (VoQS) system of transcription (Ball, Esling, & Dickson, 1995) as [i::{L i:::} i::: {L i:::} i:], where {L} indicates raised larynx voice quality and {L} indicates lowered larynx voice quality. The point (and it is a crucial point) being that, within reason, vowel quality was preserved across the larynx height “transform” – what changes is voice quality.
Figure 5.10: Larynx height and formant frequency, NRNL regime, [i] vowel, Participant B. Plots: audio (top); larynx height (upper middle); formant frequency (lower middle); F0 (bottom).

Figure 5.11 shows box plots for larynx displacement by vowel and by larynx height condition. Larynx height targets are consistently achieved, but a tendency for elevated neutral position was noticed. The tendency was for the $N_2$ in $N_1RN_2L$ and $N_1LN_2R$ targets to be higher in elevation than $N_1$, indicating undershooting in return to neutral height, which negatively skews the data and pushes the mean value higher. For both raised and lowered, the change in height is usually on the order of 1 cm above or below neutral. Since we do not know absolute starting height, we cannot determine if a
particular vowel exhibits an absolute tendency towards being produced with a more raised or lowered larynx, but comparison of the patterns does provide some useful insights. Most obvious is the fact that the larynx height patterns for the vowels are non-uniform: the open vowels, [a] and especially [æ] (which was also observed by Perkell, 1969), have a more compact height distribution and tend towards raising in the neutral state. The close vowels [i] and [u] exhibit a larger range of displacement, and show less positive bias in the neutral larynx condition. This differentiation between the open and close vowels, the former tending towards larynx raising and the latter tending towards larynx lowering, is supported by similar observations made throughout the literature (Ewan, 1979, p. 40; Wood, 1979; Kröger, Hoole, Sader, et al., 2004) and discussed further in §5.4.1.

Figure 5.11: Average larynx height change by vowel, larynx height condition, and participant. L = lowered larynx; N = neutral larynx; R = raised larynx.
Average changes in formant frequency by vowel, larynx height condition, and participant are shown in Figure 5.12. Table 5.2 provides an cross-participant average of the changes in formant frequency relative to the neutral condition. Figure 5.13 shows non-linear regression of the larynx height data against formant frequency for both participants, which provides a view of the distribution of the data.

![Figure 5.12: Average formant change by larynx height condition and vowel Participant A = solid line; Participant B = dashed line; diamond = F1; circle = F2; square = F3.](image-url)
Figure 5.13: Actual larynx height data regressed on formant frequency using polynomial curve-fitting. Downwards-pointing triangle = lowered larynx; solid circle = neutral condition; upwards-pointing triangle = raised larynx.

Table 5.2: Formant change cross-participant averages relative to neutral height (%).

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>ã</th>
<th>a</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>+36</td>
<td>+22</td>
<td>+43</td>
<td>+53</td>
</tr>
<tr>
<td>F2</td>
<td>−20</td>
<td>−2</td>
<td>+8</td>
<td>+3</td>
</tr>
<tr>
<td>F3</td>
<td>−20</td>
<td>−20</td>
<td>−25</td>
<td>−16</td>
</tr>
<tr>
<td><strong>L</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>−4</td>
<td>−12</td>
<td>−13</td>
<td>+5</td>
</tr>
<tr>
<td>F2</td>
<td>−15</td>
<td>+2</td>
<td>+12</td>
<td>+6</td>
</tr>
<tr>
<td>F3</td>
<td>−10</td>
<td>+1</td>
<td>−7</td>
<td>−8</td>
</tr>
</tbody>
</table>

F1 uniformly raises in the raised-larynx condition regardless of the neutral F1 value (i.e. whether the vowel is relatively open or close). This is contrary to SN76 where close vowels, particularly the front ones, were claimed not to be sensitive to larynx height. The statement by SN76 that larynx height has no effect on the close vowels could be because they only consider absolute (vis-à-vis relative) changes. Since the influence of vocal tract length is non-linear (Figure 5.10), it is better to consider the relative values
when comparing across vowels, especially close compared with open vowels, which have
quite different intrinsic F1 values anyways. When considering larynx-raising changes
relative to neutral larynx-height value, as in Table 5.2, it becomes apparent that the
changes are on the same order of magnitude as those affecting other formants.

The larynx lowering condition is less impactful on F1, with a tendency towards
lower frequency, except for [u], where F1 actually increases for participant B despite
larynx lowering. A similar unexpected change for [u] is reported by SN76 for one of their
participants: larynx lowering increases F1 by 10 Hz (p. 38). Larynx lowering does
correspond to a smaller effect on [i] (or even a slight reverse effect [u] as noted), and this
is in agreement with SN76.

F2 also diverges from the expected pattern in a few cases: larynx lowering (and,
variably, pharyngeal expansion with concomitant anterior displacement of the
posterior bulk of the tongue) actually yields raised F2 frequency for [æ æ u], although [i]
shows the anticipated lowering. In the case of Participant A’ [æ], lowered-larynx F2 has
a particularly wide distribution (Figure 5.13), which has the effect of skewing the mean
(Figure 5.12) and regression line (Figure 5.13) upwards; this may indicate some error in
the LPC analysis. Larynx raising also presents unexpected results for F2; this time for [i
æ] there is a drop in frequency.

Finally, F3 was most well-behaved in regard to the expected effect of the larynx-
lowering condition: it always decreases with decreasing larynx height; however, larynx
raising once again surprises with an average F3 drop of −20.25% across the board, and in
almost all cases descends well below the F3 value associated with the lowered larynx
condition.
This unexpected pattern – larynx raising actually corresponding with a rise in F1 but a drop in F2 and F3 – is generally consistent across both speakers, although the exact details differ for particular formants of particular vowels. Some of the divergence from the expected pattern may be attributable to speaker-specific vowel production strategies. For example, the production of neutral larynx height was judged auditorily in several cases to be closer to raised-constricted than to the lowered posture, and this corresponds with the slight larynx raising bias in the neutral condition. However, it also may be an artifact of the production regime which involved switching back and forth between neutral and target-height conditions. This was thought to be more desirable than simply producing a target configuration in isolation as it allowed for more careful control of vowel quality and F0, but it may be that the neutral height became biased by contiguous conditions.

5.3.3 Discussion

These data represent an altogether different picture from what was reported by SN76. It is quantifiably clear from the laryngeal ultrasound that, in terms of production, the target conditions for larynx height were achieved by both participants. While we cannot rule out the possibility that the results partially reflect simultaneous change in lingual configuration (and, importantly, see §5.4.2), it seems probable that any lingual co-adjustment reflects what is entailed articulatorily in raising the larynx, i.e. “pharyngealization”. The occurrence of pharyngealization may be a consequence of strict control of F0: the pitch raising mechanism (CT muscles) counteracts constriction. As a corollary, since F0 change was actively prevented, larynx raising associated with airway
closure (as in swallowing) evidently was induced. This is consistent with the model discussed in §4.2.5: larynx height functions along two axes – to control pitch/F0 and to assist in epilaryngeal stricture; a prediction of the model is that maintaining steady F0 while raising the larynx will predispose constriction, and this is what was observed here. The implication is that SN76 did not control F0 well enough.

We can infer that the data in the present study represent a configuration involving strong retraction of the tongue in conjunction with larynx raising. Given that the larynx raising condition in this study is analyzed as “pharyngealization”, we might expect similar formant changes to be reported in association with actual instances of pharyngeals, pharyngealization, and related sounds occurring in real languages. This expectation is resoundingly confirmed.

Catford (1983, pp. 348–349) reports formant measurements of pharyngealized vowels in Tsakhur and Udi. In Tsakhur, for [i e] F3 decreases by ~200 Hz, for [o u] it decreases by ~1000 Hz, and for [a] it increases slightly. In Udi, for [i e] F3 decreases between ~300 to 500 Hz, for [o] it increases by ~100 Hz, for [a] it decreases by ~500 Hz, and for [u] F3 could not be measured. It is in this paper (also see Catford, 2002) that Catford draws the comparison between pharyngealization in Caucasian and rhotic quality in American English, stating that pharyngealized vowels sound “r-coloured” (quoting Kodsazov), and have a lingual configuration very similar to “double-bunched” American-English R (facts which will become very important in §5.4.2 and §7.3.1). Pharyngealization in Tsez (Maddieson, Rajabov, & Sonnenschein, 1996) is similar to the data reported by Catford and the changes associated with the raised larynx condition reported in the present study: F1 increases (for all vowels and particularly [i]) and F3
lowers. F2 is more complex: most important is the fact that it increases for back vowels. This is taken by Catford to be indicative of lingual fronting and motivates the use of the term “emphatic palatalization” (Trubetzkoy, 1939; Colarusso, 1985). Acoustic and auditory observations by Grawunder, Müller, and Abdulaev (2009) confirm the Kodzasov observation reported by Catford that release bursts of Tsez pharyngealized stops have this “‘r-coloured’ ‘centralization’” quality (concentration of F2 and F3 in the 1500 to 2000 Hz range).

Outside of the Caucasian languages, Ladefoged & Maddieson (1996, pp. 312–313) state that F2 and F3 show a tendency to converge in !Xóõ strident vowels, which are known to involve strong epilaryngeal stricture (see §3.1.2 and Figure 5.18). Data reported in Butcher & Ahmad (1987, p. 160) indicate that F3 lowers in Iraqi Arabic pharyngeals by ~200 to 400 Hz. Similar findings are reported for Saudi Arabic (Alkhairy, 1999).

In Tungusic language Even/Evenki (Aralova, Grawunder, & Winter, 2011), “tongue root retraction” harmony associated with “pharyngealization” (see Ladefoged & Maddieson, 1996, p. 307) has a lowering effect on F3 across the dialects in the study: Sebian “retracted” vowels involve F1 increase, F2 decrease, and F3 lowering (on the order of 200 Hz), although the effect is strongest on “retracted” [i] and [e], weaker for [o] and F3 increases very slightly for [u] (p. 242); in divergence with the general pattern, however, is the Bystraia dialect, which only exhibits F3 lowering for pharyngealized [i] while raising slightly for the other vowels. Amis pharyngeals (Maddieson & Wright, 1995, pp. 50–55) always cause an increase of F1, but the effects on other formants varies by vowel: F2 increases and F3 decreases extensively in /u_u/ context (giving the

---

127 Surprisingly, the authors do not make any comment on this change when they assert that increased F1 is primary in cueing pharyngeal place of articulation.
impression of [ɪ]; in /i_i/ context both F2 and F3 decrease; in /a_a/ context, F2 increases and F3 decreases but not as strongly as in the /u/ context. Finally, in Nuuchahnulth, Wilson (2007, p. 57) observes that uvulars do not cause an increase in F1 and a decrease in F3 to the same extent as pharyngeals (although the results are not statistically significant).

The results presented in the present study (and associated with the empirical evidence) must be taken as complementary to SN76 in demonstrating the other use of larynx raising – epilaryngeal constriction. This assumes that in SN76, pitch must have changed with larynx height condition (although they assert it was controlled, they do not provide evidence comparable to Figure 5.10). Nolan’s (1983, pp. 160–161, 182–187) examination of the effect of larynx height on vowel quality corroborates the present results and is likewise contrary to SN76 (also see Laver, 1980, pp. 27–28). Particularly noteworthy is the difficulty Nolan found with locating F3 in “laryngo-pharyngealized” voice (and less so in “pharyngealized voice”): it was not to be found in its typical 2.3 – 2.6 kHz band, and this forced the choice between a weak, sub-2.0 kHz resonance and a stronger one occurring in between 2.6 and 2.9 kHz (p. 161). What Nolan could discern is that, in raised larynx voice, F1 raises (except for [a]-like vowels), F2 lowers (except for [u]-like vowels), and F3 is always lowered. The conclusion Nolan draws from this is that larynx raising in raised larynx voice, necessarily induces pharyngealization. However,

128 This suggests that for Nolan (and Laver, whose data Nolan also considered) the “pharyngealization” is even more extreme than that observed for the two participants in the present study. More important than the absolute value, however, is the change relative to neutral, and in all cases, Nolan’s and those presented here, F3 is lowering relative to neutral – not raising as would be expected according to the uniform tube model.

129 An intriguing parallel here, is Heselwood and Plug’s (2010) assertion that rhoticization does not just involve a lowered F3: it involves sequestering of F3 into F2 and creating a large spectral gap in the region Nolan is describing. We should immediately think of the parallel to Kodzasov’s report that Caucasian pharyngealization sounds rhotic to American listeners (see Catford, 1983).
Nolan also observes that “whilst it is certainly not transparent how far the auditory quality abstraction of vowel quality from voice quality is possible under [conditions of raised larynx voice], to the extent that it is, vowel quality seemed to remain fairly constant” (1983, p. 185). What is meant here is that, unlike “pharyngealization”, which causes a change in vowel quality (often described as lowering, backing, or retraction), Nolan’s (and Laver’s) “laryngo-pharyngealization” is very similar to “pharyngealization” but allows for the preservation or re-mapping of vowel quality. This view diverges somewhat from Esling’s interpretation that raised larynx voice and pharyngealization use equivalent articulatory states but are distinguished by their different F0 biases: raised larynx voice quality is said to have higher F0 while pharyngealization is characterized by lower F0 (see, e.g., Esling, 1999; Edmondson & Esling, 2006).

The problem may simply be terminological: *pharyngealization* is too vague with respect to articulation to be useful in taxonomic classification of voice quality. On the one hand, it is used in some contexts to describe what would be better labeled as uvularization (see further discussion in §7.3.1); on the other hand, taxonomic classification needs to distinguish constriction in the lower vocal tract characterized by lingual retraction with and without additional epilaryngeal constriction, and *pharyngealization* is non-specific in this regard. Nolan’s/Laver’s distinction between *laryngo-pharyngealization* and *pharyngealization* is perhaps the most evocative of the articulatory what-and-where, but it is possible to be more phonetically precise if one considers the tube-in-a-tube model (see §2.1.1): there is the epilaryngeal tube and there is the pharyngeal tube. The pharyngeal tube can be roughly divided into two sections according to height: relatively high pharyngeal constrictions correspond with
“uvularization” and relatively low ones correspond with the pharyngeal component of Laver’s, Nolan’s, and Esling’s notion of pharyngeal constriction in “pharyngealization” – best call it lower pharyngealization or hypopharyngealization to be clear. However, neither of these cases of pharyngeal tube stricture necessarily entails epilaryngeal tube stricture – what might be called epilaryngealization. By making such a differentiation between pharyngeal and epilaryngeal strictures, it becomes possible to consider a hypopharyngeal-epilaryngeal stricture (or hypopharyngealization-epilaryngealization) which seems necessary to explain what is happening in the raised larynx condition when pitch is controlled.

5.3.4 Summary

The results here indicate that larynx raising has a more pronounced effect on formant structure than lowering, and the relation between larynx height and formant frequency is not a strictly positive correlation as the uniform tube model (and the SN76 study) would lead us to believe. Particularly interesting is the lowering behaviour of the third formant in the raised larynx case, which happens for all vowels that were examined – the “peripheral” vowels. From these results, it can be concluded that the larynx lowering data mostly conform with prior accounts, in which all formants decrease in correlation with expansion and lengthening of the vocal tract (actions which mainly affect the pharyngeal or back cavity). The non-linear effects on formant structure attested in the raised larynx case are attributed to engagement of epilaryngeal constriction, which

130 Laryngopharyngealization would also work, but unfortunately this term might lead to confusion as to what the epilarynx is doing. Furthermore, the constriction involved is not strictly bound to the laryngopharynx, but also involves the lower half of the oropharynx.
recruits lingual retraction into forming a very narrow closure within the hypopharynx. The results reported in this study show (mostly) strong correspondence with natural-language data.

The puzzle this work introduces is, as Nolan observed, how vowel quality could be maintained despite such extreme narrowing within the lower vocal tract (which might perhaps best be described as hypopharyngealization-epilaryngealization). One would expect vowels to be retracted across the board, but the situation seems more complex and involves voice quality change (i.e. “raised larynx voice quality”) more so than vowel quality change. The solution, which will be discussed in §5.4.2 has to do with the parallelism between pharyngealization and “double-bunched” R identified by Catford, and it reveals the plasticity of the vocal tract to match (remarkably) certain auditory-acoustic categories even under extremely perturbed conditions.

5.4 The epilarynx and vowels: A model of subtle and strong effects

One of the goals at the outset of this chapter was to integrate the consequences of epilaryngeal action into the vowel space and relate these consequences to the more broad category of voice quality. To accomplish this, the components of epilaryngeal constriction were examined in §5.1 §5.2, which demonstrated the flexible, but facilitating roles played by tongue retraction and larynx height as the external mechanisms of epilaryngeal constriction and which complement the primary internal mechanism (the intrinsic laryngeal musculature). Then, in §5.3, the effects of manipulating larynx height on vowel quality were examined. The result from this study is that, despite changes to vowel structure in the raised larynx condition which are suggestive of what is
traditionally thought of as pharyngealization, the fact that vowel quality is generally preserved or re-mapped into raised larynx voice quality merits explanation.

This section integrates all of the discussion in the preceding sections of the chapter into an account of the subtle (§5.4.1) and extreme (§5.4.2) effects associated with epilaryngeal constriction in the context of the vowel system. Fundamentally, these effects involve the degree of engagement of the external factors of epilaryngeal constriction. While the epilarynx can constrict under and lingual and larynx height setting – because of the internal mechanism – the more the external mechanisms engage, the more the internal mechanism is assisted in achieving the closure.

5.4.1 Subtle Effects: Open vowel susceptibility to epilaryngeal stricture

Esling (2005, pp. 40–41) states that “the quality of [a] is inherently susceptible to increasing degrees of laryngeal constriction”. [a] is the canonical retracted vowel and it involves canonical (lower) pharyngeal stricture, but it is not canonically characterized by epilaryngeal stricture: its intrinsic retraction biases (not predetermines) epilaryngeal stricture. The implication of Esling’s model, which is directly taken up here, is that vowels are not equal in regard to their laryngeal setting and that one cannot simply tease the larynx out of the vowel space: it is embedded within the vowel system via the epilarynx.

The typical pattern is for the open vowel (usually [a] or [æ]) to involve a slightly higher larynx height and the close vowels [i] and especially [u] to involve a slightly lowered larynx position (Griesman, 1943, p. 18; Lindau, 1975, p. 56; Ewan, 1979; Wood,
The acoustic-auditory benefits to vowel quality at this scale are manifest: lowering the larynx will drive F1 lower for “high”/close vowels and raising the larynx will drive F1 higher for “low”/open vowels, and we could imagine this will help maximize the perceptual difference between these vowels.

The changes do not simply involve consequences for F1, however; as we have seen, epilaryngeal constriction has a predisposition for a raised larynx setting, and the reason for this is reinforced when considering imaging data comparing close and open vowels. Van den Berg (1955) observed that the ventricle (the lower chamber of the epilarynx) is enlarged for [u] and reduced in size for [a].

131 It should be emphasized here that the relationship is not absolute. For example, based on a thyro-umbrometric examination of speakers from several different languages, Ewan (1979, p. 40) claims that either [i] or [a] have the highest larynx height. Furthermore, while [u] most often has the lowest larynx height, in some instances, [o] is even lower. Ohala (1987, pp. 208–209) survey of the literature turns up two patterns, from highest larynx height to lowest: [i a u] and [α], [i] ≈ [u]. Furthermore, Ohala points out that, regardless of the specific ranking of the vowels in terms of larynx height, by dint of the lingual-laryngeal relationship, the ventricle is expanded for relatively close vowels and narrowed for relatively open vowels. With the latter case, Ohala acknowledges the possibility for an increase in effective vibrating mass due to vocal-ventricular fold coupling. In support of this, Ohala cites data from Eijkman (1933) and Russell (1931) which demonstrate that relatively more open vowels have greater epiglottal tilting.
Figure 5.14: Vowel quality and passive epilarynx narrowing (retracings of data found in Sundberg, 1974, p. 839). e = epiglottis (apex); et = epiglottic tubercle; f = ventricular (false) fold; v = vocal fold.

Tracings obtained from Sundberg (1974, p. 839) of the epilarynx during operatic singing postures (Figure 5.14), where the larynx is held low (Sundberg, 1977), reveal this passive posteroanterior epilarynx narrowing still occurs for [a] (although the ventricle is prevented from narrowing due to vertical expansion associated with larynx lowering). We can infer that it is the retracted lingual configuration which causes rotational displacement and retraction of the epiglottis which is ultimately responsible for the passive epilaryngeal narrowing. Tongue retraction is coupled to larynx raising through the hyoid bone and the hyoglossus and thyrohyoid “chain-links” (e.g. Shin, Hirano,
Maeyama, Nozoe, & Ohkubo, 1981). So the larynx-raising bias in vowels with retraction and presumably any relatively open vowels – but especially [a] – is not a surprise.

More reflective of linguistic conditions are the tracings in Figure 5.15. This figure shows the three vowels, [i a u], as produced by a speaker of Standard German based on MRI data from Kröger et al. (2004). These traces show increased posteroanterior epilaryngeal narrowing and a marked decrease in the size of the ventricle for the open vowel in comparison to the close vowels. Furthermore, the [a] is associated with a diminished ventricular cavity in comparison to [i] and especially [u]. The exact cause of the diminished ventricle seems to involve both lingual retraction and larynx height. In the case of lingual retraction, there is retroversion of the apex of the epiglottis. The stress on the epiglottis from this epiglottic distortion is probably transmitted downwards to the base of the epiglottis where it is in proximity with the ventricular folds. Judging from the MRI traces, it cannot be determined if vocal-ventricular fold contact occurs at the anterior region of these structures, but, from this, we would predict [a] may occasionally involve such contact. Distortion of the pre-epiglottic body (dashed line) gives some indication of the relative degree of vertical compaction of the soft structures of the larynx, which is evidently more so for [a]. Like the epiglottis, this structure presumably also helps transmit deformation stress caused by tongue position downwards onto the epilaryngeal structures. X-ray data from Gauffin & Sundberg (1978) are comparable but also illustrate progressive anteroposterior narrowing of the epilarynx from [u] to [o] to [a].

The relationship between relatively open vowels and larynx raising is attributable to linguo-hyo-laryngeal linkage (Laver, 1980, p. 25; also see Zemlin, 1998). Muscles lowering the jaw (especially the anterior digastric, geniohyoid, and mylohyoid) and
tongue (especially the hyoglossus) exert an upwards and forwards pull on the hyoid bone; muscles raising the larynx act in part through the hyoid bone (especially the thyrohyoid muscles)\textsuperscript{132}.

![Figure 5.15: Midsagittal tracings of the vocal tract based on MRI data found in Kröger et al. (2004) (retraced for this dissertation). Dashed line indicates the pre-epiglottic body.](image)

The relationships involved in this epilaryngeal-stricture bias of retracted (and relatively open vowels) are schematized in Figure 5.16 (which shows a midsagittal view) and in Figure 5.17 (which shows an exaggeration of the consequences of this bias in the coronal section of the larynx). In Figure 5.16, tongue retraction (i) causes associated downwards displacement of laryngeal tissues by transmitting stress through the epiglottis (and possibly the pre-epiglottic body), which will induce anterior compaction (ii) of the epilarynx – purely passively, but this is a state that is certainly beneficial for more active

\textsuperscript{132} The hypoglossal nerve innervates the hyoglossus muscle, while a branch of the first cranial nerve, which travels alongside the hypoglossal nerve, innervates the thyrohyoid and geniohyoid muscles.
epilaryngeal constriction; possible larynx raising (iii) will only serve to increase the constriction in this configuration.

Figure 5.16: Midsagittal schematic of passive epilaryngeal stricture in [a]\(^{133}\) (the quintessential retracted vowel). Dashed line = location of coronal sections depicted in Figure 5.17.

In the coronal plane (Figure 5.17), the retraction and larynx raising create conditions that are favourable for contact and mutual compression of the soft tissues of the larynx (b). From above, tongue retraction causes the epiglottis to collide into the

\(^{133}\) Figure appears in Žygis, Brunner, & Moisik (under review). It was created by Scott Moisik.
ventricular folds (i), which will in turn collide into the vocal folds (ii). The vocal folds will be lifted upwards into this collision (iii) if there is concomitant larynx raising (which is typical for [a]) and, if the internal epilaryngeal stricture mechanism engages, then we would also expect concentric bulging of the vocal folds and ventricular folds, which will only serve to increase the likelihood of these structures coming into contact, and, as was demonstrated in Chapter 4, vocal-ventricular fold contact is a basis of constricted phonatory quality\textsuperscript{134}.

The significance is that it is through the epilarynx that the vocal folds become implicated in vowel articulation: \textbf{none} of these subtle interactions would occur if the epilaryngeal structures were not present to communicate effects happening at lingual level to the vocal fold level. Retracted and relatively open vowels are predicted to be more prone to triggering effects associated with epilaryngeal stricture. This could materialize as a tendency for creakiness or even glottal stop in the context of such vowels, but more extreme effects such as epilaryngeal vibration are possible, depending on the prevailing prosodic conditions and other factors\textsuperscript{135}. Conversely, the concomitant larynx-lowering tendency of relatively close vowels, such as [i] and [u], should be associated with the opposite bias towards vocal fold abduction, and thus we would expect a greater likelihood to encounter breathiness for such vowels.

\textsuperscript{134} Although he does not explicitly formulate the physiological mechanism, Ewan suggests a similar explanation for intrinsic F0 (Ewan, 1979, pp. 48–58): he posits that soft tissue compression occurs during relatively open vowels, resulting in increased effective mass of vibration and lowered F0.

\textsuperscript{135} Such as desire.
Figure 5.17: Coronal section of the larynx illustrating stricture bias of [α] corresponding with Figure 5.16.

It cannot be stressed enough that this relative narrowing does not constitute active epilarynx activity – this discussion certainly should not be interpreted to mean that [α] has the same epilaryngeal stricture as [α̃] or [ʔ], which represent cases of active epilaryngeal stricture. [α] is incomplete in terms of epilaryngeal stricture: even if there is a tendency for larynx height to be elevated, but it is not an intrinsic property of the vowel, since one can actively lower or raise the larynx further and actively engage the internal mechanism of epilaryngeal stricture. The effect is an important aspect of vowels to grasp since it will be argued to have reflexes in the phonological layer (§7.2.5 and §7.3).

---

136 Figure appears in Žygis, Brunner, & Moisik (under review). It was created by Scott Moisik.
137 The reader should be careful to note that the diagram in Figure 5.17b is an exaggeration of the expected bias effect and is not be interpreted as a mandatory configuration for vowels like [α]. The expectation is, rather, that, ceteris paribus, relatively open vowels are more likely to be subject to this form of laryngeal compaction than relatively close vowels.
138 As seems to be the case for some (e.g. Shahin, 2011a).
5.4.2 More dramatic effects: The “neo-tongue” & the “neo-pharynx”

We are now ready to discuss the more extreme possibility of the epilarynx-mediated interaction of vowel quality and voice quality in raised larynx voice, which was the subject of study in §5.3. Raised larynx voice quality is a voice quality, i.e. it can be superimposed over the segmental layer, which, as we know, includes vowels. Implicit in this is that there will be colouring or distortion of the vowel space, but vowel-quality relations will be preserved or transformed into an auditory space governed by the articulatory demands of the new voice quality.

As was noted in §5.3.3, following a provocative comment by Kodzasov, Catford (1983, also see 2002) made a connection between pharyngealized vowels and rhotic quality. The shared articulatory property of these sounds is the peculiar, simultaneous, palatal and (lower) pharyngeal stricture associated with what Catford called “double bunching” (1983, p. 349) of the tongue. We will return to this matter in the context of Caucasian languages in further detail §7.3.1, but it suffices to say here that what Catford could not demonstrate was how “double bunching” in [ɹ] variants differed from “double bunching” in Caucasian pharyngealization. Because of x-ray limitations (see §2.2.1), it is not possible to determine with certainty what the epilarynx is doing in these sounds and nor is it possible for us presently. For example, x-ray data of the sphincteric vowels and “pharyngealized” vowels of !Xóõ (Figure 5.18) reveal a strikingly similar “double bunched” configuration. Yet, it is difficult to say what the epilarynx is doing in these images (the question marks denote the uncertainty).
Figure 5.18: Palato-pharyngo-epiglottal configuration in !Xóõ. Tracings based on data found in Hess (1998); page numbers are provided for the source of each tracing. Arrows = double bunching (cause and effect). The question mark indicates possible epilaryngeal stricture, although the data are not perfectly clear.

The work of Esling (§2.2.3) demonstrates that pharyngeals and pharyngealization (or “raised larynx voice quality”) are produced with extreme epilaryngeal stricture. Assuming that this is true for the Caucasian languages (the ones which Catford examined) and taking into consideration the fact that !Xóõ “sphincteric” and pharyngealized vowels (as in Figure 5.18) are known to involve epilaryngeal stricture (see §3.1.2), then “double-bunched” pharyngealization can be interpreted as the “double-bunched” configuration (palatal and lower-pharyngeal stricture formed by the tongue) with the addition of epilaryngeal stricture. Furthermore, we can infer that the raised larynx vowels produced in Nolan’s study and those in the study in §5.3 all involve the configuration characterized by simultaneous “double-bunching” with epilaryngeal stricture.
Some proof of this is given in Figure 5.19. This figure shows tracings of five different mid-sagittal lingual ultrasound visualizations of the tongue as performed by the author. The “double bunching” in [ɹ] is illustrated in (a): the configuration gets its name from the bunched-up appearance of the tongue at two locations, one anterior (A) in the region of the hard palate, and one posterior (P) in the region of the lower pharynx. (The reduced anterior mass A denoted by the gray shaded region in Figure 5.19c, d, & e is called the “neo-tongue”; this concept will be discussed further below). Compare this with [i] shown in (b). Both sounds have a palatal stricture near (A), but differ in terms of the lower pharyngeal stricture at (P); crucially, neither of these sounds has epilaryngeal stricture. In (c) and (d), [ɹ] and [i] produced with extreme epilaryngeal stricture (or in “raised larynx voice”, symbolized as [ɾʰ] and [ɾ́] and comparable, in the case of [ɾ́], to the productions in §5.3). We can infer that the larynx has raised because of the dramatic upwards and forwards movement of the hyoid bone (H) (suggesting very strong geniohyoid contraction\textsuperscript{139}). Although (c) and (d) represent different vowel qualities, the lingual configuration is strikingly similar and resembles the ordinary [ɹ] in featuring “double bunching”. Example (e) is [ɾ́] – heavy whispery-voiced growling – and it remarkably has almost the same configuration as in (d).

\textsuperscript{139} The hyoid bone is the U-shaped bony scaffold found at the base of the tongue from which the larynx is suspended. In midsagittal lingual ultrasound (as in Figure 5.19) the body of the hyoid bone is orthogonal to the imaging plane. Compared to the surrounding soft tissues of the tongue, the hyoid bone has a very high acoustic impedance. Thus, it has low acoustic transmittance, which means that almost all ultrasound waves are reflected back to the probe. Few if any waves can penetrate the bone to enable us to see beyond it. Since acoustic waves radially propagate, this results in an arc-like non-imageable area on the ultrasound rendering (the area enclosed by dashed lines and marked “H” in Figure 5.19).

The hyoid bone shadow seen in midsagittal ultrasound is a quasi-useful landmark for making inferences about larynx activity. In Figure 5.19, the hyoid bone is drawn upwards and forwards. This indicates strong geniohyoid activity, which is known to be important in larynx raising (Shin, Hirano, Maeyama, Nozoe, & Ohkubo, 1981, p. 177). The hyoid-bone motion in Figure 5.19c, d, & e – as indicated by the hyoid shadow position – is thus an indicator that the larynx has raised.
Figure 5.19: “Neo-tongue” formation during “double-bunching” epilaryngealization (mid-sagittal ultrasound visualization of Moisik’s tongue). Gray mass = effective tongue: full sized in (c), “neo-tongue” sized in (a), (b), (d), and (e). A = anterior lingual mass; P = posterior lingual mass; Acoustic “shadows” (dashed-lines mark edges of shadows): M = mental symphysis (chin); H = hyoid bone (approximate location marked with ellipse). Arrows show direction of lingual motion visible in ultrasound video. The thin, solid lines indicate the ultrasound imaging area.

The question is – why use double bunching? The answer proposed here is that combining these two states triggers multiple effects: (1) it transposes the acoustic relations within the vocal tract which critically allow for the preservation (or re-mapping) of vowel quality subject to the raised larynx voice quality “transform”; (2) it benefits/enhances/synergizes with/facilitates epilaryngeal stricture; and (3) it leaves the anterior tongue still relatively free to manipulate vowel quality in cooperation with the jaw and the lips. Referring back to Figure 5.19, the subtle variation in the configuration
of the anterior (A) lingual mass seems to account for the different vowel qualities in (c) and (d): the surface just behind the anterior mass (chubby dark arrow in d) seems to dip more for [i̞] than for [i̞]. It is as if the tongue dorsum/body (at the chubby dark arrow in d) is behaving like the posterior tongue (P) in (b): advancing and concomitantly expanding the back resonating cavity. The difference between (b) and (d) is the location of this back cavity. In raised larynx voice, any vowel quality can be produced. The shift in back cavity position and concomitant reshaping of the anterior tongue seem to be key in understanding why this is possible. If raised larynx voice (or “pharyngealization”) simply caused a blunt, global retraction of the tongue, the goal of producing all vowel qualities would be far more difficult to achieve; instead, what we see is an unusual compromise between conflicting articulatory goals of simultaneously closing off the hypopharynx while still having flexibility in producing different vowel qualities.
The physiological mechanism causing “double bunching” is plausibly the middle genioglossus (GGM) muscle, and biomechanical modeling evidence exists which supports this analysis (Stavness, Gick, Derrick, & Fels, 2012). Figure 5.20 illustrates the effect of GGM contraction on tongue shape: (a) depicts the “neutral” tongue shape; (b) depicts the “double bunching” shape claimed to be associated with GGM contraction (small arrows on muscle line). Thanks to the hydrostatic nature of the tongue (e.g., see Smith & Kier, 1989), deformation stresses (large arrows) push the anterior tongue (A) towards the (anterior) hard palate and thrust the posterior tongue (P) or “tongue root” into the lower pharynx (hence the configuration is classifiable as a dual, palato-pharyngeal constriction). The tongue dorsum/body (D), being drawn by the GGM, advances and
descends, expanding the cavity in the region of the oropharyngeal isthmus (Gick, Anderson, Chen, et al., to appear).

Figure 5.20 is non-specific regarding the epilarynx (hence it is not shown). However, when these adjustments combine with larynx raising and extreme narrowing of the hypopharyngeal space (both the pharyngeal and epilaryngeal cavities) the acoustic effect is more than just linear vocal tract shortening as the term “raised larynx voice” would imply, and we know empirically that this is not the case either (based on §5.3).

It seems probable that the acoustic impedance of the hypopharyngeal space will be high enough during the “double-bunching” epilaryngeal configuration to significantly diminish the coupling between the upper resonating spaces and the hypopharyngeal cavities (i.e. the lower pharyngeal and epilaryngeal cavities). If the narrowing is extreme enough (but not fully closed) then the effective length of the vocal tract will no longer be primarily a function of larynx height, but rather it will be defined by the acoustic choke point located at the level of the upper border of the elevated epilarynx, roughly half way up the pharynx.
We can think of the vocal tract configuration in raised larynx voice (or “pharyngealization”) in terms of a “neo-tongue” and a “neo-pharynx”. These “neo” concepts are comparable to the term *neoglottis* in reference to when the upper esophageal sphincter is used in a phonatory capacity. All of these “neo” concepts convey that an alternate structure is substituted to fulfill the function typically performed by another structure (or space, in the case of the glottis). The “neo-tongue” (the anterior tongue bulge, A) stands in for the full tongue and the “neo-pharynx” (the cavity behind the neo-
tongue [the region of D in Figure 5.20] whose inferior boundary is the “choke point” of the hypopharynx), stands in for the pharynx proper. Figure 5.21 illustrates these concepts further with three tube-abstractions of the vocal tract: (a) is the “neutral” vocal tract configuration (as in [ə]); (b) is like (a), but with pitch-related larynx raising; and (c) shows the configuration responsible for “raised larynx voice quality”, which, as has been discussed, is articulatorily characterized as “double-bunching” epilaryngeal. The configuration in (c) entails more than just a shortening of the pharynx: it is characterized by a dramatically shortened pharynx – as if the pharynx had been cut in half – and a reduced-mobility but nonetheless still vocalically-versatile anterior tongue (A, the palatal bulge; compare with the gray region in Figure 5.19c, d, & e and A in Figure 5.20b). Thus, during raised larynx voice, it is as if the vocal tract had suddenly shrunk to new diminutive proportions.

“Double bunching” epilaryngeal and the corresponding “neo-tongue–neo-pharynx” configuration it creates are associated with pharyngeals and pharyngealization/raised larynx voice quality. The implication is that pharyngeals and pharyngealization will not necessarily have the “expected” (see §7.3.1) effect of lowering or retracting neighbouring vowels. It should be entirely possible to produce any vowel quality in the context of “pharyngeal” (epilaryngeal) articulation, but we might expect various effects associated with the “double bunching” epilaryngeal configuration to arise. For example, we might expect to find a palatal-stricture bias (“emphatic palatalization”). These possibilities do not negate the potential for pharyngeals to do the expected thing to

140 Comparable in some ways to the infant vocal tract with its ontogenetically-determined raised-larynx configuration and infant-sized tongue. Intriguingly, Sonya Bird reports, based on lingual ultrasound observations, that several infants employ a configuration very similar to that in Figure 5.19c & d.
vowels, such as retract them, but we should listen (and look) for signs, such as raised larynx voice quality, that suggest more is going on than just retraction of the tongue.

5.5 Chapter summary: The epilarynx and the supralaryngeal vocal tract

This chapter has examined how the epilarynx relates to the supralaryngeal vocal tract. Examination of the articulatory components involved in producing “pharyngeal” consonants (§5.1 and §5.2) shows that there is a primary mechanism responsible for constricting the epilarynx that is complemented by two external mechanisms, larynx height and tongue retraction. There is also an internal mechanism (the intrinsic laryngeal musculature), which is essential for constriction to occur, and allows the epilarynx to have a degree of independence from the external mechanisms. The complication introduced here is that the external mechanisms are complementary to epilaryngeal stricture, but their engagement does not take place without compromising the configuration of the supralaryngeal vocal tract, typically through lingual retraction and changes associated with raising the larynx (such as a shortened and narrowed pharynx); these mechanisms implicate the epilarynx in the manipulation of vowel quality, and, furthermore, it is via the epilarynx that vowel quality relates to changes in phonatory quality.

The third study of this chapter (§5.3) explored “raised larynx voice quality”. The results defied the predictions of the idealized acoustic model. Instead formant changes in raised larynx voice appeared to resemble pharyngealization, as observed earlier by several researchers (Laver, 1980, p. 27; Nolan, 1983, pp. 182–187; Esling, 1999). The idea that raised larynx voice quality is tantamount to pharyngealization is acceptable, as
long as it is understood that the constriction is in the lower pharynx and entails rather extreme epilaryngeal narrowing (with concomitant larynx raising). (If we are feeling innovative, we could use a term like hypopharyngealization-epilaryngealization, but that is admittedly cumbersome.)

These results point to several possibilities, explored in §5.4, for how sounds produced with epilaryngeal stricture will relate to vowel quality and how vowel quality may come to be associated with certain laryngeal properties. One predicted tendency is for relatively open vowels to be more prone to epilaryngeal stricture than relatively close ones, which may actually be associated with anti-constriction. Another possibility is for pharyngeals and pharyngealized sounds to be produced with the “double bunching” epilaryngealized configuration, a configuration characterized by the “neo-tongue” and “neo-pharynx”. In this case, there is a palatal bias, and while nearly any vowel quality can be produced, we should expect some limitations relative to the possibilities associated with a neutral vocal tract. Finally, we should also not be surprised to find epilaryngeal constriction simply corresponding with retracted vowels, as ordinary lingual retraction is a phonetically viable way to help the epilarynx narrow. These predictions, along with the predictions made in the previous two chapters, will be explored in the context of the “phonological” layer of the dissertation – Chapter 6.
Chapter 6

LOWER VOCAL TRACT PHONOLOGY: THEORY

Sounds of the lower vocal tract (LVT)\textsuperscript{141} – sounds with articulatory activity of the larynx and pharynx – have posed a problem to phonologists (e.g. see Clements & Hume, 1995; Rice, 2011): what are the phonetic and phonological properties of these sounds that explain their behaviour in phonological systems? Part of the problem is that, unlike sounds made in the upper vocal tract (UVT), many LVT sounds intertwine phonation and articulation together, and their production often requires broad changes to vocal-tract shape, not just localized articulatory movements. Thus, McCarthy (1991, 1994) proposed that, phonologically, the LVT is a broad region of articulation rather than one characterized by specific articularators, such as [labial], [coronal], or [dorsal].

The purpose of the present work is to bring phonological analysis of LVT sounds in line with physiological and phonetic understanding of epilaryngeal function. The proposal is that the epilarynx is critical to understanding the phonological behaviour of LVT sounds. The main theme of this proposal is that the epilarynx unifies phonatory and

\textsuperscript{141} Phonological discourse on sounds in the lower vocal tract (the vocal tract region encompassed by the pharynx) has traditionally used the labels guttural or post-velar. In the present work, the more neutral lower vocal tract sounds/phenomena is used in general contexts; guttural/post-velar is used when discussing the literature associated with these terms or when making reference to the traditional concept of a guttural/post-velar natural class (typically involving /h ? h ? x ? r q c/ etc.). The term guttural is strongly associated with work in Semitic and Cushitic phonology (e.g. see Hayward & Hayward, 1989) and does not convey information about anatomical structure. On the other hand, the term post-velar is defined relative to a specific place of articulation, i.e. the velar place. Physiologically, however, the velum is involved in the production of the uvular sounds. Thus, this term introduces possible confusion about the production of these sounds. The term lower vocal tract is very broad and somewhat vague as to its boundaries, but helps to convey the general anatomical region and does not connote an emphasis on any one particular part of the vocal tract. Other researchers have used the same terms, such as Davis (1995), but this is not to intimate that the views expressed here on “lower vocal tract phonology” are the same or even similar as those proposed by other researchers using similar terms.
articulatory activity: it bridges the gap between the vocal folds and the rest of the vocal tract – anatomically, physiologically, acoustically, aerodynamically.

The present approach contrasts with many competing phonological models that can be classified as part of the glotto-centric-linguocentric (GCLC) paradigm. GCLC models strongly ascribe the articulatory behaviour found in LVT (and related) phenomena to the actions of the “tongue root” and reduce the larynx to a uni-dimensional glottis. The abundant documentation of the phonetic nature of the epilarynx strongly suggests that GCLC approaches are not suitable from a phonetic perspective. The goal of this chapter is to demonstrate that the phonetic nature of the epilarynx also has reflexes in the phonological domain, reflexes that cannot be ignored.

First, §6.1 provides the theoretical context of the discussion and §6.2 contains a review of previous phonological models of LVT sounds. Then, in §6.3, a “phonetically-rich” formal model called the Model of Lower Vocal Tract Phonological Potentials (LPP) is proposed. The LPP is equipped to address the complex articulatory nature of the epilarynx in LVT sound production is proposed, outlined, and illustrated. Supporting argumentation for this model is then discussed in Chapter 7.

6.1 Introduction

A key problem phonologists face is the reconciliation of the physical reality of human communication with the categorical and systematic nature of the symbols forming the linguistic code. Generative Phonology (GP) asserts that the symbols are encoded with distinctive features (Chomsky & Halle, 1968; Clements, 2003). If these features exist, they need to be translatable into the properties of our physical systems that make
communication possible (Hale & Reiss, 2000a, 2008; Cohn, 2006; Hall, 2007; Odden, 2012). The features also need an origin: GP posits that features are innate and universal, but many phonologists have challenged this assumption, arguing that features emerge from sensory experience and cognitive interpretation of the physical properties of the code (e.g., recently, Mielke, 2004; Pulleyblank, Baumer, Montero, & Scanlon, 2006; Mohanan, Archangeli, & Pulleyblank, 2010). This view raises the question of what constrains feature emergence such that the features and their patterning are consistently and reliably reproduced within a given linguistic community and, furthermore, show uncanny cross-linguistic similarities? This is a complex question and the best answer will likely be one that considers multiple modalities of information since the linguistic code appears to take full advantage of our sensory and cognitive capacities (e.g. Fowler & Dekle, 1991; Gick & Derrick, 2009; Perkell, 2012).

As a contribution to this broader goal, the focus here is on a smaller slice of the nature of linguistic sound: how the organization of the LVT causes certain phonological patterns to emerge more readily than others\textsuperscript{142}. I propose a model of the relationships between the phonetics and phonology of the lower vocal tract that predicts the directions that LVT phonologies gravitate towards. The focus is on speech production in the LVT, especially concerning the epilarynx. The main claim is that the epilarynx is a key “pivot” point of the lower vocal tract system: it couples the vocal folds with the supralaryngeal

\textsuperscript{142} Various instantiations of this research have been published. A discussion of the “whole larynx” model of laryngeal features can be found in Moisik & Esling (2011a). A mixed conceptual–and–Feature-Geometric account using Revised Articulator Theory was presented (by Scott Moisik) at Phonology in the 21\textsuperscript{st} Century: In Honour of Glyne Piggott (Montreal-Ottawa-Toronto; MOT; May 7\textsuperscript{th}, 2011); see Moisik, Czaykowska-Higgins, & Esling (2011a). The research was later published as a proceedings paper in Moisik, Czaykowska-Higgins, & Esling (2012). Similar work was also presented (by Ewa Czaykowska-Higgins) at 50 years of Linguistics at MIT (December 10\textsuperscript{th}, 2011; MIT, Cambridge, Massachusetts); see Moisik, Czaykowska-Higgins, & Esling (2011b).

structures, particularly the tongue, and it is an articulatory-phonatory mechanism. Its hinge-like or pivot-like properties are mirrored in phonological systems as the basis for LVT phonological contrasts, with exact expression of correlated phonetic properties varying from one language to the next. (The pivot concept is discussed further in §6.3.2)

Since the epilarynx connects articulatory and phonatory behaviour (i.e. the supralaryngeal-laryngeal or place-placeless dichotomy), it applies to the behaviour of a broad range of LVT phenomena. Thus, the model I propose applies equally to consonants and vowels of the lower vocal tract and to phonation and tonal register systems: it serves to unify the account of these seemingly disparate phonological phenomena. The analysis put forth here illustrates how the proposed model provides a conceptual reinterpretation of old analyses of many so-called guttural or post-velars patterns. The goal is to identify the systematic organization of the physical system that serves as part of the universal substrate (the human body) for feature emergence in the phonology of lower vocal tract sounds.

6.1.1 Scope of work: Guttural / post-velar phonology and beyond

Contemporary views of the phonological status of lower vocal tract sounds reflect the arguments of Hayward and Hayward (1989) and McCarthy (1991, 1994) that laryngeal, pharyngeal, and uvular sounds constitute a natural class of sounds, traditionally called gutturals or post-velars. This is a surprising observation since the part of the vocal tract implicated in the production of these sounds is extensive: essentially the size of the pharynx. The definition McCarthy provides for the active articulators in this region is as follows (n.b.: pay close attention to his definition of laryngeals and pharyngeals):
The gutturals are produced by three distinct gestures: a purely glottal one in the laryngeals; retraction of the tongue root and epiglottis and advancement of the posterior wall of the laryngopharynx in the pharyngeals; and a superior-posterior movement of the tongue dorsum in the uvulars. (McCarthy, 1994, p. 196)

McCarthy reasons that, since the somatosensory endowment of the pharynx looks sparse in comparison to that of the oral region (McCarthy, 1994, pp. 198–202), one cannot define an articulator comparable to [labial], [coronal], or [dorsal] in order to define the guttural class. Rather, he claims that the guttural sounds share an entire region of articulation, designated [pharyngeal], and it is this property of their representation which underlies their unity as a natural class. Much of the evidence for justifying a natural class of gutturals comes from Semitic and Cushitic languages, although similar class-hood behaviour is observable in many languages of the Pacific Northwest (e.g. Bessell, 1992).

Table 6.1: Cross-linguistic variation in guttural/post-velar class membership

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Included</th>
<th>Excluded</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>Semitic</td>
<td>/ɣ χ ɣ ɣ ɣ ʔ/</td>
<td>/q/</td>
<td>(McCarthy, 1991, 1994; Rose, 1996, p. 75)</td>
</tr>
<tr>
<td>Somali</td>
<td>Cushitic</td>
<td>/q χ ɣ ɣ ʔ/</td>
<td>/h ʔ/</td>
<td>(Hayward &amp; Hayward, 1989, p. 184)</td>
</tr>
<tr>
<td>Interior Salish</td>
<td></td>
<td>/t q χ ɣ ɣ ʔ/</td>
<td>/h ʔ/</td>
<td>(Bessell &amp; Czaykowska-Higgins, 1992; Bessell, 1992)</td>
</tr>
<tr>
<td>Caucasian</td>
<td></td>
<td>/h q χ ɣ ɣ ʔ/</td>
<td></td>
<td>(Colarusso, 1975, pp. 298–299; Rose, 1996, p. 98; Bellem, 2005, 2009)</td>
</tr>
<tr>
<td>Chilcotin</td>
<td>N. Athapaskan</td>
<td>/t q χ ɣ/</td>
<td>/h ʔ/</td>
<td>(Cook, 1983)</td>
</tr>
<tr>
<td>Nuuchahnulth</td>
<td>S. Wakashan</td>
<td>/q h ɣ ɣ ʔ/</td>
<td></td>
<td>(Davidson, 2002; Werle, 2010)</td>
</tr>
</tbody>
</table>

143 Emphatics (e.g. Arabic) are represented with the symbol ‹ for convenience. Retracted coronals (in Interior Salish), and ‘flat’ coronals (in Chilcotin) are represented with ‹. In most cases and for the sake of brevity, only the plain members of each type are included in the list.
Based on this typology (Table 6.1), the following generalization is possible: pharyngeals, provided they exist in a language, pattern as gutturals/post-velars. Hayward and Hayward (1989, p. 188) assert that pharyngeals and [a] are the prototypes of the guttural/post-velar class. This strongly implicates the epilarynx in the guttural/post-velar class because it is responsible for the production of the class’ most constant members, the pharyngeals (Esling, 1996, 1999, 2005; Esling & Harris, 2003b). On the other hand, the laryngeals and uvulars are relatively peripheral to the class: laryngeals sometimes fail to pattern the same way as the other members (Rose, 1996); the uvulars do not have cohesion within the guttural class, since /q/ is often guttural-external in its phonological behaviour (Hayward & Hayward, 1989, p. 179; Trigo, 1991, pp. 122–126; McCarthy, 1994, pp. 202–204; Rose, 1996, pp. 98–101; Bin-Muqbil, 2006, pp. 243–247).

Several studies (e.g. Colarusso, 1985; Czaykowska-Higgins, 1987; Trigo, 1991) make a connection between the guttural sounds and other types of sounds which involve an association between the pharynx and larynx. These include “tense”, “head”, or “harsh” tonal registers in Tibeto-Burman and Mon-Khmer languages, in cross-height or so-called “tongue root” or ATR harmony languages of Niger-Congo and Nilo-Saharan families, pharyngealized vowels found in Caucasian and some Tungusic languages, and sphincteric/strident/epiglottalized vowels found in Khoisan-group languages.

Clearly the range of phenomena under examination is vast, and the present work will not address every conceivable issue associated with these data. Table 6.2 and Table 6.3 respectively provide an outline of what is and what is not addressed in this

<table>
<thead>
<tr>
<th>Oowekeyala</th>
<th>N. Wakashan</th>
<th>/q ʔχ hʔ/</th>
<th>(Howe, 2000, p. 74)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gitksan</td>
<td>Tsimshianic</td>
<td>/q ʔχ hʔ/</td>
<td>(Brown, 2008, p. 115; Yamane-Tanaka, 2006)</td>
</tr>
</tbody>
</table>
dissertation. The main focus in the material covered concerns the adequacy of conventional feature-based approaches (reviewed in §6.2) to the analysis of the relevant data and whether the issues would be better understood with consideration of the behaviour of the epilarynx.

Table 6.2: Lower vocal tract issues that are addressed in this dissertation (see Chapter 7)

<table>
<thead>
<tr>
<th>Behaviour/Tendency/Pattern</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>epilaryngeal vibration in language</td>
<td>7.1</td>
</tr>
<tr>
<td>tone and vocal register</td>
<td>7.1.2</td>
</tr>
<tr>
<td>“variable laryngeals”</td>
<td>7.2</td>
</tr>
<tr>
<td>glottal and pharyngeal stop relationship</td>
<td>7.2.1, 7.2.2, 7.2.2</td>
</tr>
<tr>
<td>ejectives and pharyngeals</td>
<td>7.2.2, 7.2.4</td>
</tr>
<tr>
<td>oral vs. pharyngeal uvulars</td>
<td>7.2.4</td>
</tr>
<tr>
<td>pharyngeal genesis</td>
<td>7.2.4</td>
</tr>
<tr>
<td>“glottal constriction” and relatively open vowels</td>
<td>7.2.5</td>
</tr>
<tr>
<td>glotturals and vowel lowering</td>
<td>7.2.5</td>
</tr>
<tr>
<td>pharyngeal and pharyngealized vowels and palatal articulation</td>
<td>7.3.1</td>
</tr>
<tr>
<td>the intersection of phonatory, vocal, and voice quality effects</td>
<td>7.2.5, 7.3</td>
</tr>
<tr>
<td>relationship among ATR, vocal-register, and pharyngealized vowels</td>
<td>7.3, 7.3.2, 7.3.3</td>
</tr>
</tbody>
</table>

Table 6.3: Lower vocal tract issues not addressed.

<table>
<thead>
<tr>
<th>Behaviour/Tendency/Pattern</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>emphasis spreading (direction, locality, blocking)</td>
<td>(Herzallah, 1990; Shahin, 2002)</td>
</tr>
<tr>
<td>guttural degemination</td>
<td>(McCarthy, 1991, 1994)</td>
</tr>
<tr>
<td>intrinsic F0</td>
<td>(Whalen &amp; Levitt, 1995; Whalen, Gick, Kumada, &amp; Honda, 1998)</td>
</tr>
<tr>
<td>morpheme structure constraints &amp; distributional restrictions</td>
<td>(McCarthy, 1991, 1994)</td>
</tr>
<tr>
<td>implosives and ejectives (distributional asymmetries)</td>
<td>(Clements &amp; Osu, 2002; Clements, 2003)</td>
</tr>
<tr>
<td>ATR harmony patterns (scope, morphophonemics, directionality)</td>
<td>(e.g. Archangeli &amp; Pulleyblank, 1994; Bakovic, 2003; Casali, 2008)</td>
</tr>
</tbody>
</table>
6.1.2 Broader theoretical concerns: The phonetics-phonology interface

In this Chapter, phonological issues of LVT sounds are addressed primarily from an articulatory viewpoint. On one hand, articulatory-based phonological analysis is a conventional practice, as segmental phonology is commonly discussed in terms of articulatory events in the vocal tract. This approach is the corner stone of traditional Feature Geometry\textsuperscript{144} (Clements, 1985; Sagey, 1986; Goad, 1993; Keyser & Stevens, 1994; Padgett, 1995; Clements & Hume, 1995; Halle, 1995; Halle, Vaux, & Wolfe, 2000; Uffmann, 2011) and of Articulatory/Gestural Phonology (Kelso & Tuller, 1983; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Browman & Goldstein, 1989; Borroff, 2007; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Tilsen, 2009; Pouplier, 2011)\textsuperscript{145}. Both of these approaches posit that the primitives of phonological representation directly relate to physical aspects of articulation, with the representational units interpreted (by

\textsuperscript{144} A different approach to Feature Geometry is the Modified Contrastive Specification theory (Dresher, 2003; Mackenzie & Dresher, 2004; Dresher & Zhang, 2005; Hall, 2007, 2011), which posits that the phonological activity of a particular feature is governed by its relationship to other features in the contrastive hierarchy; it may be phonologically inert (not part of the hierarchy), “prophylactic” (phonologically inactive but earmarking a segment for a specific phonetic realization; see Hall, 2007), or active over all or part of the system, depending on its position within the contrastive hierarchy. Thus, what seems like a [+high] vowel may or may not show [+high] behaviour, depending on the role of that feature in relation to other features in the overall system. Since specifying what exactly the contrastive hierarchy is for a given language relies on analysis of its phonological patters, questions of existence and uniqueness of a “solution” for what the hierarchy is remain to be fully addressed.

\textsuperscript{145} This work has arisen from more general theories of motion and action (e.g. Bernstein, 1967; Easton, 1972; Turvey, 1977). Such theories posit that movement is governed by functionally-coherent cooperation or synergization of biological structures (mainly muscles and joints). This idea of a synergistic, functionally-defined movement primitive has a long heritage, both in the speech literature and outside of it. Many terms exist varying in scope and intent: e.g. functional synergy (see Kelso & Schöner, 1988, pp. 27–28), muscle synergy (Bernstein, 1967), and coordinative structures (Easton, 1972; Turvey, 1977; Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984, p. 813). The key problem with (understanding) movement is the management of the vast number of degrees of freedom characterizing the possible motion of part of the body – say, the hand (with its many joints) or the vocal tract. Rather than controlling each degree of freedom to produce a specific task, the task itself provides internal cohesion in the coordination of muscle activity and joint motion – thereby obviating the degree of freedom problem. Articulatory Phonology extends the movement primitives to the phonological domain in the form of “articulators” functionally unified by the “gestures” they execute (Browman & Goldstein, 1989).
some researchers) as motor instructions sent to the articulators (e.g. Halle, Vaux, & Wolfe, 2000, p. 388). On the other hand, there are arguments that phonological representation is arbitrarily related to the aspects traditionally associated with the phonetic domain, and labels reflecting articulatory, acoustic, auditory, or other domains are at most a mere convenience, but have no real traction in phonological analysis (Hale & Reiss, 2000a, 2000b, 2008; Hall, 2007, pp. 13–20, cf. 2011, pp. 46–48; Odden, 2012).

The present work is articulatory in the simple sense that consideration is made primarily from the movement of physical structures in the vocal tract, as opposed to sound perception or other factors. This view does not in any way seek to mitigate the importance of non-articulatory factors in the organization of sound systems. Moreover, it is presumed here that there is considerable parallelism in the organization of the systems governing our speech abilities, as suggested by Wood (1979) and several others (Stevens, 1972, 1989; also see Clements & Hume, 1995, pp. 299–300; Stevens & Keyser, 2010; Perkell, 2012). However, the choice of an articulatory perspective is not made arbitrarily: this entire dissertation is concerned with a part of the body, the epilarynx, and its interaction with other body parts, such as the vocal folds, tongue, and pharynx. Consideration of the anatomo-physiological aspects of articulation is assumed to hold explanatory value in phonological theory insofar as it is accepted that phonological form is not arbitrarily related to such phonetic properties.

The nature of this relationship hinges upon one’s assumptions about the nature of the phonetics-phonology interface in general. In approaches that advocate a strict separation between phonetic substance and phonological form, such as Hale & Reiss (2008; also see Odden, 2012; Miller, 2012, pp. 204–205), the units of phonological
structure contain no information about phonetic content of speech (for example, a distinctive feature does not specifically encode information about formant frequencies or the auditory system’s response to such frequencies). Thus, the phonological objects are phonetically arbitrary: these units are abstract variables (e.g., Reiss, 2003) that are used in phonological “computation” to determine their organization and the relations they have to one another (such as various degrees of identity). A transducer mechanism, “independent of the cognitive system” (Hale & Reiss, 2000a, p. 8), is posited to convert back and forth between the phonetic properties of speech and their abstract phonological associations; Shahin (2011b, p. 326) points out that this independence, which allows for complete, modular separation of phonetics and phonology, is implausible since the transducer itself must function in terms of the phonological objects, which are cognitive objects.

There are other approaches to the phonetics-phonology interface which suggest that phonological constructs, such as distinctive features, are epiphenomenal, arguing instead that systematic phonological alternations originate from diachronic sound change driven by various factors (Ohala, 1983, 2005a, 2011; also see Blevins, 2004): thus, this view sees phonology and phonetics as being so closely interwoven that the notion of interface is of little use (“there is no interface”; Ohala, 1990). This view does not exclude the possibility of abstract organizational structure and categories, which Mielke (2004) argues can still arise from the residue of diachronic change.

Yet another way of looking at the interface is that there is a phonetics-phonology gradient, as argued for by Cohn (2006, 2010; cf. Pierrehumbert, 2003, p. 178). In this

---

146 Cohn (2010, p. 19) offers the insight of what might be called the computer paradigm of phonological computation (cf. Port & Leary, 2005). Phonological theory, particularly associated with the Generative tradition developed in a time when computer storage space (RAM, hard-drives, etc.) was limited; the idea is
view, phonological representations are simultaneously highly abstract (à la Generative Phonology) and highly detailed (à la Exemplar/Connectionist models), and various degrees of abstraction exist in between these extremes. Cohn (2006) presents arguments within several domains associated with phonology that show evidence for phonological gradedness\textsuperscript{147}, such as contrast, phonotactics, and morphophonemic and allophonic alternation; Hall (2009) similarly argues that gradedness as characterizes the phonological relation of predictability used in defining contrastive or allophonic relationships. Another form of gradience is advocated by Mielke (2004, pp. 143, 163), who argues that natural classhood lies along a sane-“crazy”, or (phonetically) natural-unnatural, continuum.

This view of a gradient phonetics-phonological interface is assumed here for several reasons: (1) phonetic and phonological components are scientific reifications used for communicating about the reality of speech (cf. Ohala, 1990, p. 161): we should not expect to find behaviour that is purely and unequivocally just phonetic or just

\textsuperscript{147} One might even conceive of Feature Geometry as an abstractness gradient: each level of the hierarchy arguably represents a different order of abstraction over phonetic properties. At the \textsc{Root} (the most abstract level of the geometry) the most abstract features are found [consonantal, sonorant] (for a suggestion that [sonorant] is more abstract than [consonantal] see Kaisse, 1992, p. 330; also see Hall, 2011, pp. 46–47); as Pouplier (2011) observes, these features cannot be strictly associated with articulatory gestures; they are manner of articulation features and vague with respect to place of articulation. Subordinate to this level are features such as [nasal], [lateral], [strident], and so forth (depending on which version of Feature Geometry is assumed; for a review of Feature Geometry see Uffmann, 2011; an implication of such abstraction is in the labeling of these features as “articulator-free”, see Halle, 1995, p. 6), and the \textsc{Laryngeal} and \textsc{Place} nodes. This level could be thought of as one degree less abstract than the \textsc{Root}: the features are more specific with regard to manner of articulation and the nodes provide vague indication of place of articulation. Moving to the next level (below \textsc{Laryngeal} and \textsc{Place} nodes) are features that are, yet again, one degree less abstract, as they correlate with specific places of articulation, the articulators involved, and the acoustic/auditory properties associated with sounds produced at these locations. Subordinate levels are arguably even less abstract, and so on. Thus, in this view, the individual tiers of the hierarchy do not all have the same status in terms of abstraction, but rather implicitly express a structure of graded abstractness. Similar observations can be made about suprasegmental organization (such as syllable and prosodic structure).
phonological. These categories merely reflect our best attempt at defining boundaries, and as the research of Cohn (2006, 2010), Hall (2009) and others shows, there is reason to doubt that the boundaries are well-defined; (2) a model which posits distinct modules for phonetic and phonological processing requires a “transducer” mechanism at the interface of these modules to explain the flow of information between them; a gradient interface allows for very flexible, dense, but organized connections within and across layers and no transducer needs to be posited since these connections are presumably a natural function of cognitive organization at the neural level (e.g. Joos, 1948, pp. 109–125; Meunier, Lambiotte, Fornito, Ersche, & Bullmore, 2009); finally, (3) focusing on the question of whether something is a matter of phonetics or phonology has the possible side effects of either duplicating explanation across these layers (Blevins, 2004, p. 5) or failing to make larger generalizations across the layers so as to provide better explanations for “phonological” phenomena; an example would be McCarthy’s (1994, p. 195) judgment that “laryngeal involvement” in the production of pharyngeals is merely a “superficial mechanical effect” and therefore “unsuitable as an explanation for a truly phonological property [McCarthy’s emphasis]” incapable of unifying the class of gutturals. In straining to find the one property abstract enough to suitably apply to his notion of guttural, he dismissed what is argued here to be the very essence of pharyngeal articulation, the one sound category always associated with the guttural class. The approach here is not to suggest that this “laryngeal involvement” is the entire story or the basis of the guttural/post-velar class, but rather to demonstrate that a structure that has been dismissed on “mechanical” grounds as being irrelevant to phonology actually is
extremely useful in explaining many of the phonological phenomena associated with lower vocal tract and is worthy of serious consideration in phonological modeling.

6.1.3 Broader theoretical concerns: Phonological features or potentials?

It was observed in the introduction (§6.1), that some researchers regard distinctive features as emergent components of phonological representation (Blevins, 2004; Mielke, 2004; Pulleyblank, Baumer, Montero, & Scanlon, 2006; Miller, 2012; for a discussion of self-organization in Phonology more generally, see Wedel, 2011) rather than phonological content that is innately determined by Universal Grammar (UG). Adoption of this view calls into question the pursuit of identifying and characterizing phonological features as possibly nothing more than applying “convenient labels” to phonological generalizations (Pulleyblank, Baumer, Montero, & Scanlon, 2006, p. 15): suitable for descriptive adequacy, but not explanatory adequacy (Cohn, 2010, p. 15). The perspective taken here is to remain relatively agnostic regarding the innateness debate. It seems probable that we have a drive as human beings to form abstractions and categories which facilitate relational thinking (e.g. Mareschal & Quinn, 2001), and this may be the very same mechanism used in forming phonological features. It could be that this drive to form abstractions originates from innately specified information, but the purpose here is not to speculate on this matter.

Whatever the resolution to this debate is, it is an incontrovertible fact that humans have vocal tracts: this is a human universal. The same applies to sign language: hands are a human universal. At the anatomo-physiological level there are innumerable, idiosyncratic variations of vocal tracts and hands. Remarkably, however, we get by with
the bodies we have, even if a finger is missing or the vocal folds have been resected. Our linguistic community, be it spoken or signed, imposes limitations on what counts as part of the language code. To borrow the expression from Laver (1994, p. 29; cf. Smolensky & Legendre, 2006, pp. 42–43, vol. 1), we, as speakers (and signers), must rely on a “notional vocal apparatus” (or notional signing apparatus in the case of sign language), to interface with our linguistic community: if we could not abstract sufficiently beyond the anatomo-physiological level, then the enterprise of communicating via language would simply not be possible. We would not be able to contend with the systematic but highly variable signal carrying the linguistic code in a noisy environment.

Leaving the issue of sign language aside for now, it is this notional vocal apparatus that we must consider in attempting to explain phonetic and phonological behaviour. Highly apropos to the topic of the present paper is Pulleyblank’s (2006, pp. 17–19) discussion of the so-called sphincteric vowels of the Khoisan languages, citing Traill (1986) and Ladefoged & Maddieson (1996, pp. 310–313). In considering the typological distribution of vowel features, Pulleyblank’s (rhetorical\textsuperscript{148}) puzzlement over why UG should require a sphincteric feature in all languages, when its occurrence is ostensibly rare, seems reasonable from the generative-nativist standpoint: what would the evolutionary advantage be? However, it is not the case that “sphincteric” voice is in any way rare when we consider the same property from the perspective of the notional vocal apparatus: given sufficient exposure, time, or training, anybody should be able to become

\textsuperscript{148} Pulleyblank (2006, p. 17) observes that it is discouraging that our main theoretical device for modeling the phonological behaviour of sounds – the distinctive feature – is rather weak at pinpointing particular phonological functions: for example, the fact that a vowel possesses a particular feature in a given language does not let us make strong predictions about how that vowel will behave in the phonological system. This is part of Pulleyblank’s main argument that distinctive features and phonological constraints are not “hard-wired” by UG, but rather emergent, in parallel with Blevin’s (2004) and Mielke’s (2004) arguments for emergent phonological organization.
a speaker of a Khoisan language and produce suitably accurate sphincteric vowels. When we approach from a phonological-systems perspective we conclude that sphincteric is rare, but there is nothing rare about it when we approach from a phonetic perspective: the mechanism for sphincteric vowels is our growing mechanism – epilaryngeal vibration – and it can be argued that it is quite commonly used in speech and vocalization (as the survey in §3.1 indicates). What is special, however, is the fact that “sphincteric” voice is a phonological possibility or potential that is realized in the Khoisan languages.

So with this discussion in mind, whether features are genetically endowed or emergent is tangential to the question of what the **potentials** of the notional vocal apparatus are and how they determine, in part, what can be phonologically encoded. Features may be innate or learned, but the notional/abstract vocal apparatus has the same potentials regardless. Individual vocal tracts may differ: maybe the person with supercricoid partial laryngectomy (Crevier-Buchman, Pillot-Loiseau, Rialland, et al., 2012) will have difficulty producing a difference between the plain and sphincteric vowels of !Xóõ; but assuming no deficiencies affect their auditory or cognitive systems, it is highly likely that they will perceive the difference and recognize its linguistic importance.

Emergent features are unique to each language learner: the individual members of a speech community must converge upon a similar set of emergent features, but nothing guarantees that individuals will acquire exactly the same set of abstractions. Thus emergent features at the community level are an unrestricted set: everyone’s are unique (Dalby, 2002, p. 5; Blevins, 2004, pt. III); however, phonological potentials are not unlimited: there are only so many unique things our notional vocal apparatus can do and
only a finite number of biases acting on these actions. This claim is based on evidence of categorical behaviour at the pre-cognitive, physiological level in the form of quantal biomechanical-articulatory (Fujimura, 1989; Buchaillard, Perrier, & Payan, 2009; Gick, Stavness, Chiu, & Fels, 2011; Nazari, Perrier, Chabanas, & Payan, 2011) and articulatory-acoustic relations (Stevens, 1972, 1989; Stevens & Keyser, 2010).

The assumption of a closed set of phonological potentials runs contrary to Port & Leary’s (2005, pp. 927–929) argument that phonetics is unlimited. It is true that individual speech production is highly idiosyncratic and phonetic objects are theoretically infinitely variable (as has been noted for growling). The point of abstraction to a notional vocal tract – which is a fundamental assumption in the field of phonetics that enables the concept of phonetic similarity to be defined – is that we can define a limited set of possibilities and principles. If the world was not characterized by discontinuities and non-linearities, it might be safer to default to the infinite set viewpoint, but the assumption here is that cross-domain non-linearities characterizing speech make defining a finite set of phonological potentials a tractable and productive activity (of course, this is not to suggest it is necessarily a small set either).

For our present purposes, one of the most important constraints on the set of possible speech-related actions required to produce speech sounds is the device (Gick, Stavness, Chiu, Flynn, & Fels, in preparation). Devices are functionally-defined neuromotor pathways underlying spatio-temporal patterns of muscle activation and the associated quantal biomechanical-articulatory relations responsible for the regulation of the many degrees of freedom characterizing complex movement/action (see §6.3.3). Devices are “constructed” around actions of body parts that are useful for a task, such as
producing a speech sounds. Key to the nature of devices as constraining the set of possible speech-related actions is the fact that speech biomechanics is characterized by non-linearities (e.g. Fujimura, 1989; Gick, Stavness, Chiu, & Fels, 2011; Nazari, Perrier, Chabanas, & Payan, 2011), much like articulatory-acoustic relations have non-linear relations or quantal properties (e.g. Stevens, 1972, 1989). We can assume that these biomechanical quantal properties are more or less shared by all humans, given that we are all subject to the same laws of physics, from which the non-linearities arise. Thus, it is reasonable to abstract away from individual variation in thinking about a device. By way of illustration, we might suppose that all speakers manage to construct similar devices for approximant-related protrusion-rounding of the lips as in [w], for compressing the lips as in [p] closure, and for labio-dental frication stricture, and so forth (Gick, Stavness, Chiu, & Fels, 2011) – but the list of possible devices is finite, given the non-linear biomechanical-articulatory relations. The devices view implies that articulation is not controlled piece-meal by continuously adjusting muscle-contraction parameters or by controlling the parameters of a finite set of articulators (such as the tongue tip, tongue root, and so forth); rather, articulation is controlled by engaging the right devices for the task at hand: devices are stricture specific.

Devices are one type of phonological potential, but they are a highly relevant to the present discussion given that the focus here is on articulation. To illustrate how devices fit into the theoretical model of phonological potentials, consider Figure 6.1. Following Gick et al. (in preparation), we will symbolize devices with white square brackets “⟦⟧”. We can posit that people construct “growling” devices when the need arises (say if you are a speaker of !Xóõ). To symbolize this abstract device, we will
borrow the IPA symbol for voiced epiglottal fricative – [ʢ]: this is at least “in the ball park” and, in the context of this dissertation, [ʢ] is understood to mean voiced aryepiglottic-epiglottal vibration (see Chapter 3). This device can be associated with a canonical phonetic product – [ʢ] – with coarsely definable properties in articulation and acoustics, and it can participate in systemic phonological behaviour, denoted /ʢ/ (or /Vʢ/ for vocalic use as in “sphincteric” phonation in !Xóô). It is important to remember that we are speaking abstractly here. Real devices in real people (represented by body₁, body₂, body₃ in Figure 6.1) vary idiosyncratically and produce infinitely variable phonetic output (see Figure 6.1); we might symbolize the real devices as [ʢ₁] and individual phonetic output tokens as [ʢ₁]₁, [ʢ₁]₂, [ʢ₁]₃, etc.; these real instances are “mirrored” or related through abstraction to [ʢ], which is associated with an abstract “universal body” and the canonical phonetic product [ʢ].

Figure 6.1 associates phonological potentials with diachronic phonology; this is because phonological potentials are assumed to exert their influence at the diachronic scale. They are trends that speakers in a speech community gravitate towards; they are susceptibilities of production to physical principles; they influence the trajectory that language change takes through time. Synchronic phonology is what real languages and real speakers actually do: it is realized potential.
The goal of this work is to understand the phonological potentials associated with the lower vocal tract and the role of the epilarynx in these potentials. To give them a formal casting, potentials are defined in (1):

(1) **PHONOLOGICAL POTENTIALS:** Physical properties of speech understood abstractly that form, bias, or influence the structure and apparent patterning of phonological systems and subsystems.

The concept of phonological potentials (or potential(s)) implicates that any sound producible by the human vocal tract could be useful as a speech sound. Phonetic research
has clearly established that some sounds are simply not attested in any of the known sound systems of the world (some examples are sublingual clicking, dental percussives, and ingressive velic trills). Thus, there is a likelihood-of-incorporation cline which predicts what sounds are likely to be incorporated into sound systems, and many of the lower-vocal-tract sounds explored in this work, such as “growling”, appear to be on the less-likely-to-be-incorporated end of this cline. The explanation for why a sound falls where it does on this cline is attributable to a confluence of factors which extend far beyond the scope of this work, which is mainly articulatory in focus. Suffice it to say, the explanation reduces to a multivariate function, and the articulatory factors are just one among many variables.

The notion of phonological potentials can be related to speech sounds and the symbols for these, but it is not to be confused with the idea of phonetic symbols (such as those of the IPA) and categories and the speech sounds these define, and it is also not to be confused with the idea of distinctive features or “gestures”.

Phonetic symbols and categories are representational devices for writing and communicating information about sound systems: they are metasymbols with varying degrees of specificity suitable for different communicative goals, and, hence, can be used to communicate about what is perceived of as more phonetic and more phonological information in the speech code (Port & Leary, 2005, pp. 927–929). On the other hand, distinctive features are thought of as cognitive abstractions formed over properties of speech sounds that create so-called natural classes in phonemic systems. Potentials are also not like “gestures”, the atomic representational units of Articulatory/Gestural Phonology (Browman & Goldstein, 1989). Gestures are instructions for the actions
performed by specific articulators (like the tongue tip and “tongue root”) and are thought to be phonologically coordinated to form articulatory events in speech.

Phonological potentials are the language-symbolic possibilities humans have for forming speech sounds and their realization is what is transcribable with written phonetic symbols. Potentials are an abstraction of the form and functioning of the vocal tract in linguistic sound production and its obeisance to the laws of physics. Unlike “gestures” or distinctive features, phonological potentials are not representational, cognitive objects\textsuperscript{149}. For example, consider /b/ (a voiced bilabial stop) and /ʕ/ (a voiced pharyngeal approximant). The voicing feature, say [+voice] (Keating, 1984), or glottal closing gesture (Goldstein & Browman, 1986) shared by both of these sounds in a given language are identical objects in each respective theory. The phonological potentials characterizing “voicing” in /b/ and /ʕ/, however, are quite different in nature. The difference is a matter of the proximity between the primary stricture (bilabial for /b/ and aryepiglotto-epiglottal for /ʕ/) and the mechanism causing voicing (i.e. vibration of the vocal folds). This difference is not just mechanical/phonetic detail to be ignored, it is an intrinsic part of the phonological potential of /ʕ/. Every time someone realizes a /ʕ/ there is an increased likelihood for some epilarynx-induced vocal fold perturbation to occur because of the nature of voicing in the /ʕ/. We could say that, by virtue of the “gross” physiological configuration or state, the voicing in /ʕ/ and /b/ will likely share more in common than with the voicing in /b/, but exact identity will not hold as if the gestures or

\textsuperscript{149} The difficulty here becomes where to draw the mind-body division. Devices relate to both gross physical structures of the body (such as the lips) but also to the neuromotor subsystem driving them. In some ways, these are much like the “neuremes” proposed by Joos (1948), invariant neuromotor organization associated with a particular speech segment.
features involved were identical\textsuperscript{150}. A speaker might conceivably develop some abstraction linking the voicing similarity between \textipa{ʕ} and \textipa{ɓ} at a higher level of cognitive organization associated with the phonemes, but that is a matter of idiosyncratic features, not potentials.

Interactions, such as that just described, between different parts of the vocal tract inherent to potentials and their interaction with other potentials is the object of study here. Defining what potentials exist will be a major task, and it is only possible to provide theoretical arguments as to what we might expect are the potentials associated with the lower vocal tract. A more rigorous defense will need to be made with biomechanical modeling, and appeal to a broader domain of phenomena than just articulatory ones. Following suggestions by Wood (1979) and Perkell (2012), it is likely that there is a great deal of parallelism across domains, such as auditory, somatosensory, somatomotor, and biomechanical domains characterizing the categorical and finite nature of the set of potentials.

On a final note, potentials should not be thought of as purely phonetic or purely phonological objects. They are abstractions that allow us to model relations between phonetic-like objects and phonological-like objects. This is a reasonable approach under the assumption of a gradient phonetics-phonological interface. The present research is not asking, “is this a phonetic or phonological object?”, but rather it is asking, “how do the potentials of speech production influence the organization (or perhaps more appropriately, self-organization; see Wedel, 2011) and behaviour of sound systems”.

\textsuperscript{150} These approaches would rule out such differences as irrelevant to phonological considerations. I am suggesting that, since the phonetics-phonology interface is not cleanly divisible, these types of effects are important to “phonology” and must be accounted for accordingly.
Before providing some answers to this question, we first need to start with what is known about how lower vocal tract sounds behave and previous models for analyzing this behaviour. This is the topic of the next section (§6.2).

6.2 Previous models of lower vocal tract phonology

The phonological nature and behaviour of lower vocal tract sounds is complicated by several paradoxes. Universalist (i.e. UG) approaches to phonology require the representational capacity for sphincteric/strident/“growled” vowels, yet they are (apparently) extremely rare in sound systems (as argued by Pulleyblank, Baumer, Montero, & Scanlon, 2006, p. 18). Laryngeals pattern as if they were “placeless” in some languages and “pharyngeal” in others (Bessell, 1992; Bessell & Czaykowska-Higgins, 1992; Rose, 1996; Paradis & LaCharité, 2001; Borroff, 2007; Rice, 2011, p. 532; Shahin, 2011a). Pharyngeal sounds are assumed to be specified for retraction of the tongue root in their phonological representation, yet in some languages they can pattern with palatals (coronals) and cause “emphatic palatalization” (Trubetzkoy, 1939, p. 124; cited in Colarusso, 1985, p. 366; Comrie, 2005; Bellem, 2005a, 2009).

This section briefly reviews the formalisms used in addressing guttural/post-velar/LVT phonology. The view here is that the “paradoxes” noted above stem from scientific myopia – failing to look outside of conventional analytical approaches to

---

151 Note that, given the articulation-oriented approach of this dissertation, the review of previous models is given largely from an articulatory perspective. Further justification of such an approach is that the models under review here are primarily based on articulatory considerations. John Ohala (personal communication) points out that a more comprehensive assessment could be made that considers aspects of aerodynamics and acoustics in addition to articulatory factors.

152 The point of identifying myopia in this research or any research is not to dismiss it as being uninformed or lacking in merit or value. On the contrary, the work reviewed here is incredibly valuable to
gain insight into a problem (see Ohala, 1990, p. 164): in the case of the strident vowels, the assumptions of UG makes these vowels seem “problematic”; in the “variable laryngeals” and “emphatic palatalization” cases, it is glottozentrisch-linguocentrism (GCLC) that caused myopic approaches to be taken. GCLC is so pervasive and entrenched in approaches to LVT phenomena, that it is paradigmatic in scale.

Since most of the LVT phenomena to be evaluated in this Chapter have long traditions of research in the GCLC paradigm, it is necessary to review details of the formalisms used in analyzing the sound patterns. Subsequent sections will return to these formalisms to demonstrate their limitations in predicting the phonetic and phonological behaviour associated with LVT sounds.

6.2.1 Glottozentrisch-linguocentrism in guttural/post-velar phonology

Many formal phonological models assume lingual-laryngeal independence: this is evident in the laryngeal-supralaryngeal dichotomization of distinctive features (Steriade, 1987) and their organization (Clements, 1985; Sagey, 1986; McCarthy, 1988, p. 89), and it has become the established, de facto view of the phonological function of the larynx. For example, as Montler (1998, p. 371) states: “In fact, it is well known that glottals are not supposed to affect vowels … since laryngeal articulation is physically independent of tongue articulations.” This separation is held so firmly to be true by phonologists that, as Uffmann (2011, p. 648) observes, “the general existence of a LARYNGEAL class node is

understanding LVT phonology and is exemplary in its descriptive adequacy; the point is that despite this value, it is research that is a product of a specific historical and paradigmatic context within the field. Understanding the grounding orientation (i.e. GCLC) of its many proposals, which still have currency in the field today, allows us to view these proposals more transparently and integrate their observations into more encompassing approaches to the phenomena at hand.
undisputed”. Such an entrenched position is surprising since numerous observations have been made, for example, by Czaykowska-Higgins (1987) and Trigo (1991), that phonological models must address lingual-laryngeal-pharyngeal interactions found in LVT sounds. Instead, in many phonological models (some of which are reviewed here), articulation in the lower vocal tract is phonologically essentialized in the reifications of glottal and radical articulators, which are marshaled into the analysis of the patterns and properties of LVT sounds under the banners of guttural or post-velar phonology.

6.2.2 Early approaches to features of lower vocal tract sounds

Many discussions on the representation of gutturals/post-velars (Czaykowska-Higgins, 1987, p. 9; Hayward & Hayward, 1989, pp. 184–185; McCarthy, 1991, p. 9; Bessell, 1992, p. 14; McCarthy, 1994, p. 197; Hess, 1998, p. 2; Bin-Muqbil, 2006, p. 63) begin with an evaluation of the adequacy of features proposed in Jakobson, Fant, and Halle’s (1952) “Preliminaries to Speech Analysis” (Preliminaries) and in Chomsky and Halle’s (1968) “Sound Patterns of English” (SPE) for defining the guttural/post-velar class. The general consensus is that early phonological modeling in Preliminaries or SPE cannot define the guttural/post-velar class, lack representational power for specific sounds, or they express the class in a phonetically unnatural way. In Preliminaries, [+flat] (which denotes lowering of some or all formants) applies to uvulars, emphatics, and pharyngeals, which lower F2, but not laryngeals, since these sounds are assumed to

---

153 John Ohala (personal communication) comments the “preliminary” nature of Preliminaries would suggest it should not be criticized too severely for failing to adequately extend to all aspects of sound structure. (This is interpretation is supported by the characterization of Preliminaries by the authors as “a provisional sketch”. Ohala further points out that Preliminaries succeeds in defining natural classes on purely acoustic (vis-à-vis articulatory) grounds.
not have an effect on F2 (McCarthy, 1994, pp. 196–197); additionally, Bin-Muqbil (2006, p. 6) argues that since [+flat] also applies to labials and rounded vowels, it fails to explain why the Arabic vowel /u/ can become pharyngealized (which would presumably involve assimilation of [+flat]). In the SPE system, McCarthy (1994, pp. 196–198; also see Hayward & Hayward, 1989) notes that gutturals/post-velars are [‒anterior] and [‒high], but this, McCarthy argues, implies anomalous articulation for uvulars, which he assumes have a “high tongue body” and therefore should be [+high] (p. 197). Furthermore, he assumes that [+low, +back] does not apply to pharyngeals since:

… the distinctive gesture in pharyngeals is with the tongue root/epiglottis and posterior pharyngeal wall, not the tongue body. In fact, the tongue body is not back but front with the Arabic pharyngeals, as we can see by the adjacent front allophone of the low vowel: compare pharyngeal [hææl] “condition” with uvular [χɑɑl] “maternal uncle” (McCarthy, 1994, p. 197)

SPE also contains a proposal for [covered], defined thus: “covered sounds are produced with a pharynx in which the walls are narrowed and tense and the larynx raised; uncovered sounds are produced without a special narrowing and tensing in the pharynx” (Chomsky & Halle, 1968, pp. 314–315). Painter (1973, pp. 97–98) evaluated [covered] in an x-ray study of Twi/Akan vowel harmony, which is based on what he called “tense” [i e æ o u] vs. “lax” [ɪ ɛ a ɔ ʊ] vowel sets. Painter concludes, in parallel with Lindau (1975, 1978), that, articulatorily, pharynx volume distinguishes the Twi/Akan vowel sets (n.b.: typical of early x-ray, see §2.2.1, the hypopharynx was not

154 Labials and alveolars are distinguished from the ‘guttural’ class with [+anterior] while palato-alveolars and velars are distinguished with [+high].
155 The adjustments in covered are said to correspond with “open and covered singing” (Chomsky & Halle, 1968, pp. 314–315), though this interpretation of the physiological mechanism in covered singing is opposite to that documented by Hertegård, Gauffin, and Sundberg (1990), who, based on laryngoscopic imaging, describe covering as involving a lowered larynx position and expanded pharynx.
imaged) but tense-lax, not an additional [covered] feature, characterizes the phonemic system. However, Painter also notes the entanglement of vowel, voice\(^{156}\), and phonatory quality intrinsic to the Twi/Akan contrast, which suggests the so-called tense vs. lax contrast, in auditory terms, is a matter of what he describes as breathy vs. non-breathy, or non-choked vs. choked (1973, p. 117)\(^{157}\).

The feature [covered] faded into obscurity, along with short-lived proposals by Lindau (1975, 1978) for [expanded] (cf. Keyser & Stevens, 1994; Davis, 1995), both being eclipsed by the linguocentric feature, [Advanced Tongue Root] or [ATR], originally proposed by Stewart\(^{158}\) (1967; cf. Halle & Stevens, 1969) for the Akan harmony system and extended to tonal register in Mon-Khmer by Gregerson (1976). By denotation, [ATR] gives primacy to movement of the so-called tongue root in the production of these contrasts which usually cross-cut entire vowel systems\(^{159}\).

Czaykowska-Higgins (1987) proposed a tongue root articulator with two different types of behaviour: Type-I is a “mutual cooperative relationship between tongue root position and laryngeal behaviour” (p. 7); Type-II is independent action of the tongue root

\(^{156}\) Although it is common place now to refer to vowel harmony in West African languages as “ATR harmony” (Casali, 2008), the more general “voice quality contrast” or “voice-quality harmony” used to be applied (Tucker, 1975; Jacobson, 1980).

\(^{157}\) Similar impressionistic labels are provided by Hall and Hall (Hall & Hall, 1980; quoting from citation by Czaykowska-Higgins, 1987, p. 5) tense vowels are “muffled, breathy, or hollow” and lax vowels are “creaky, bright, brassy, or non-hollow”. What Painter’s and these terms suggest is (possible) individual variation in execution of the contrast as a function of overall pharyngeal-laryngeal state: some speakers go for an expansion vs. neutral strategy involving breathy phonation associated with larynx lowering; others take the strangulation vs. neutral strategy, which correlates with creaky/brassy/bright qualities; yet others might take an expansion vs. strangulation strategy and use both non-modal qualities.

\(^{158}\) Earlier observations about vowel harmony in West African languages appear in Welmers (1946) and Berry (1957), among others (see Casali, 2008, p. 496); however, to my knowledge, the feature [ATR] first appeared in Stewart (1967).

\(^{159}\) Goad (1991, 1993, p. 22) argues that there are really two features: [ATR] and [RTR]. In her view, [ATR] is for vowels (it intersects with the vowel height), while [RTR] is the feature for active constriction in the pharynx (it is primarily a consonantal feature and associated with pharyngeals and pharyngealization).
in constricting the pharynx. These correspond, respectively, to the features [±lower pharynx] (or [±LP]) and [±upper pharynx] (or [±UP]). In justification of this cooperative-independent analysis of the tongue root, she argues that languages with Type-II tongue root behaviour may exhibit phonatory correlates in the retracted group but not in the non-retracted group (e.g. retracted vowels in Nxa’amxcin are creaky but non-retracted are not breathy); however, Type-I, she claims, always involve phonatory effects in both tongue root states (advanced and retracted).

Based on x-ray (Ghazeli, 1977) and acoustic evidence (Card, 1983), Czaykowska-Higgins suggests that Arabic emphasis is uvularization, or [+UP, −LP]; pharyngeals are represented with [−UP, +LP]. This gives the possibility of two types of secondary articulation: uvularized pharyngeals, attested in Palestinian Arabic (citing Card, 1983), and pharyngealized uvulars, attested in Northwest Caucasian languages (Ubykh, Abaza, and Abkhaz; citing Colarusso, 1975); she further suggests Interior Salish “retraction” (1987, pp. 14–16) involves a combination of both types of tongue root action. These representations are illustrated in Table 6.4:

<table>
<thead>
<tr>
<th></th>
<th>q</th>
<th>θ</th>
<th>q⁺</th>
<th>θ⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>[upper pharynx]</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>[lower pharynx]</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Czaykowska-Higgins’ approach is overall linguocentric in nature and also has a glottocentric view of the larynx\(^\text{160}\), but it critically established that correlated or

---

\[^{160}\text{Czaykowska-Higgins states that “the primary purpose of the tongue root is to form a constriction in the supra-laryngeal cavity” (1987, p. 12): this means that either the larynx is conceived of as being lower than}\]
cooperative lingual-laryngeal activity needs to be accommodated in phonological analysis. Trigo (1991) takes this notion one step further by exploring the hypothesis that tongue root position and larynx height are independent phonetic parameters, with corresponding phonological features [ATR/RTR] and [lowered larynx/raised larynx] (or [LL/RL]). Her survey spans “voice quality or register” contrast in African and Southeast Asian languages, laryngeal, pharyngeal, and uvular consonants, and the association between voiced stops and enlarged pharyngeal volume. In her conclusion, she concedes that the evidence for orthogonality of the features [ATR/RTR] and [LL/RL] is weak and that the “co-occurrence of pharyngeal and laryngeal effects is attributed to a mechanical link between pharynx and larynx” (p. 132). Thus, Trigo posits (1991, pp. 116–119) a mechanical relationship between larynx height and vocal fold functioning that operates independently of vocal fold abduction/adduction (phonologically expressed by [±spread glottis] and [±constricted glottis] features); this premise is used to explain the occurrence of breathy phonation with larynx lowering and “squeezed”/“pressed” (p. 117) phonation with larynx raising. Concomitant with this is a peripheralization-centralization\(^{161}\) of the vowels (moving towards or away from [ə] in vowel space, see Trigo, 1991, f.n. 5), which, in the case of Caucasian languages, she claims is caused by “a peculiar distortion of the tongue” (f.n. 7) similar to a bunched-r in American English.

---

\(^{161}\)Trigo (1991, p. 115) emphasizes that peripheralized/centralized distinction implied in the transcription of ATR vowel harmony associated with African languages is not to be confused with the tense/lax contrast of English, which also involves vowel length and lacks the other phonetic correlates of vowel harmony sets in these languages. However, she later applies the terms peripheralized and centralized to describe vowel quality changes corresponding with so-called chest and head registers in Akha (f.n. 10) and non-pharyngealized and pharyngealized vowels in Caucasian languages (f.n. 7).
While inheriting the orientation of the GCLC paradigm, Trigo’s analysis marks the most concerted effort to countenance the phonological relevance of lingual-laryngeal connection in forming register contrasts. Perhaps because of her doubtfulness about the necessity of encoding larynx height as a distinctive feature, her theoretical proposal did not result in progressing phonological analysis beyond the dominant glottocentric-linguocentric approach. Its lack of impact may also be attributed to its timing, as it also came at the time when Feature Geometry was in ascent as a popular formalism for phonological analysis. The requirement in Feature Geometry of strict dominance (a feature may be dominated by one, and only one, feature; see McCarthy, 1988; cf. Ohala, 2005a) may have caused reinforcement of the already well-established view that laryngeal and supralaryngeal activity were separate in phonological terms.

6.2.3 Feature geometric representations of gutturals/post-velars

Feature Geometry (FG) models are intended to represent classhood behaviour in phonological processes and patterns: the apparent systematic behaviour of groups of features spreading, deleting, or sharing distributional properties. The hierarchy represents dependencies amongst the features regarding these processes: for example, [consonantal] and [sonorant] are thought never to spread independent of any other features and hence they define the ROOT node. Feature Geometry represented a theoretical advancement beyond earlier models, such as SPE and Preliminaries, that posited unordered bundles of features. With FG it was possible to make principled statements about why certain features patterned together on a regular basis. While FG has fallen out of favour with
contemporary theoretical views\textsuperscript{162} (embodied in formalisms such as OT; cf. Uffmann, 2011), so much of the work on guttural/post-velar phonology is formulated in FG that it is essential to review the key proposals made under this framework. The number of different proposals for feature geometric representation in general is overwhelmingly large, and the “guttural” part of the tree is no exception. Bessell (1992, Chapter 2) provides a survey of some prior models; summaries of the “guttural” section of the tree are also provided by Clements and Hume (1995, p. 274) and more recently by Uffmann (2011, p. 648). For the present purposes, it suffices to examine five proposals depicted in Figure 6.2:

\begin{quotation}
\textsuperscript{162} Since OT is output oriented (the “goal” of OT is to choose the correct output form from a theoretically infinite set of inputs) organization of features in the input is not evaluated by constraints: all that matters is what features are present in the candidates (although, see Uffmann, 2011).
\end{quotation}
All models posit a representational unit (node or feature) associated with guttural/post-velar classhood: [pharyngeal], PHARYNGEAL, GUTTURAL, LOWER VT all of which suggest the idea of the pharynx or pharyngeal part of the vocal tract. For example, in Herzallah’s model (Figure 6.2a), which is framed in Unified Feature Theory\(^{163}\), gutturals/post-velars and the /a/ vowel are argued to be specified with the [pharyngeal] feature; in McCarthy’s model (Figure 6.2b), [pharyngeal] is tantamount to a region of articulation, rather than a specific articulator, hence its status as sister to the Oral node.

---

\(^{163}\) Unified Feature Theory (UFT; Clements, 1989; Clements & Hume, 1995; for a review see Halle, Vaux, & Wolfe, 2000, pp. 399–412) was an attempt to reconcile consonantal place of articulation features with High-Low-Front-Back vowel features. Primary consonantal articulations receive C-PLACE specification for articulator features; vowels and secondary articulations are V-PLACE.
(also a region of articulation). This regional status of [pharyngeal] is later emphasized in the proposals (Figure 6.2d-e) by Rose (1996) and Halle et al. (2000; cf. Halle, 1995) though it has nodal status rather than being a feature.

In all models except Davis’s (Figure 6.2c), pharyngeals proper (/h ʕ/) are associated with the idea of the tongue root, as denoted by [radical], [RTR], or TONGUE ROOT, which distinguishes them from laryngeals. Davis, who assumes a feature [Constricted Pharynx] or [CP] in addition to [RTR], proposes that uvulars are [RTR] and pharyngeals are [CP], stating that “pharyngeals do not just mainly involve a movement of the tongue root; … movements of the laryngopharynx are crucially involved” (1995, p. 471): provided “laryngopharynx” is not meant to entail activity of the larynx, the LARYNGEAL node he posits actually means “glottal”, or, “just the vocal folds”.

Attempts at accounting for guttural/post-velar patterning resulted in a shifting perspective on the nature of phonological place of articulation. Prior to these accounts, there were Place and Laryngeal categories of features (Clements, 1985; Sagey, 1986): the vocal tract was roughly split into laryngeal and supralaryngeal zones for phonological purposes, and laryngeals were viewed as being “placeless”. With the advent of guttural/post-velar phonology (e.g. Hayward & Hayward, 1989; Herzallah, 1990; McCarthy, 1991), phonologists were forced to reconcile the relationship of laryngeal features to place features. The introduction of a pharyngeal zone of articulation undermined the logical unity of the PLACE category. This is reflected in the change in nomenclature seen in Davis’, Rose’s and the Halle et al. models (Figure 6.2c, d & e): what was formerly PLACE becomes associated with the oral or upper vocal tract, and the pharyngeal part of the vocal tract is then given place status (although the Halle et al.
model resisted the renaming trend – but GUTTURAL-\textsc{-vs.-PLACE} is basically the same division). The status of the laryngeal node was also called into question, leading to the following split in approach: Rose’s model follows McCarthy’s proposal by keeping LARYNGEAL external to PLACE; the Davis’ and Halle et al.’s models admit LARYNGEAL/LARYNX into the domain of the other articulators.

Perhaps one of the most impactful proposals is that of Rose (Figure 6.2d), which owes its provenance to McCarthy. Rose’s representations of gutturals/post-velars are in Figure 6.3. For Rose, laryngeals do not automatically have PHARYNGEAL specification: they obtain it based on whether other gutturals/post-velars are found in the phonological inventory. Laryngeals with PHARYNGEAL node characteristically cause vowel lowering to /a/, which is a canonical phonological property of the guttural/post-velar class; laryngeals lacking Pharyngeal are placeless (they are transparent in transconsonantal harmony and fail to cause vowel lowering).

Rose’s proposal leads to the complication that uvulars need to be split into two types: ORAL and GUTTURAL (also see Trigo, 1991, pp. 122–126; Bessell, 1992, pp. 18–19; McCarthy, 1991, 1994); thus, the DORSAL place specification is associated with both of these nodes (see Figure 6.2d). If all uvulars were PHARYNGEAL, then, in languages with uvulars and laryngeals but no pharyngeals proper, the laryngeals would be expected to pattern as if they had PHARYNGEAL specification (e.g. they would cause vowel lowering\textsuperscript{164}). Rose (1996, pp. 100–101) observes, however, that such languages (Kashaya

\textsuperscript{164} According to Rose (1996, pp. 78–81), this is due to the Node Activation Condition. The Node Activation Condition causes a primary feature to be “activated” on both phonemes in a contrast whenever these two phonemes are solely distinguished by a single feature dependent on the primary feature. Thus, if [RTR] solely distinguishes, say, ′/r/′ from ′/h/′, then the Node Activation Condition will cause ′/h/′ to be PHARYNGEAL. If /h/ is PHARYNGEAL then it is not “placeless” and will cause vowel lowering. /q/ does not
being an example, as it has /q qʰ q’ ? h/) show no guttural/post-velar behaviour at all. Hence, Rose classifies uvular stops as “Oral Uvulars” and uvular fricatives and approximants as “Pharyngeal Dorsals”.

```
<table>
<thead>
<tr>
<th>Laryngeals</th>
<th>Pharyngeals</th>
<th>Oral Uvulars</th>
<th>Guttural Uvulars</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ʔ h/</td>
<td>/q qʰ q’</td>
<td>/χ b r</td>
<td></td>
</tr>
<tr>
<td>PLACE</td>
<td>PLACE</td>
<td>PLACE</td>
<td>PLACE</td>
</tr>
<tr>
<td>or ROOT</td>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharyngeal</td>
<td>Pharyngeal</td>
<td>Oral</td>
<td>Pharyngeal</td>
</tr>
<tr>
<td>[RTR]</td>
<td>Dorsal</td>
<td>[RTR]</td>
<td>Dorsal</td>
</tr>
</tbody>
</table>

Figure 6.3: Representations of post-velars following Rose (1996, p. 80).

This split in the treatment of laryngeal features reflects conflicting beliefs about how the organization of Feature Geometry should be defined and the deeper metatheoretical tension embodied in the debate about the phonetics-phonology interface: should definition proceed by patterning or by anatomical organization? Feature geometries are phonological, so they are assumed to be categorically organized on the basis of observed phonological patterns; however, at the time, there was a notion that FG is a mapping of the physical organization of the vocal tract (Keyser & Stevens, 1994; Davis, 1995; Halle, 1995; Halle, Vaux, & Wolfe, 2000). Thus, many of these models apparently cause Pharyngeal node activation on laryngeals, and this is attributed to it bearing an ORAL-DORSAL feature (thus the Node Activation Condition need not apply since /q/ will never be solely distinguished by [RTR] from the laryngeals).
constitute, more or less, a schematization of the physical layout and operation of the vocal tract\textsuperscript{165}.

It seems to be that the generalizations expressed by FG do reflect facts about the physical layout and interaction of the structures in the vocal tract used in producing speech sounds. Assuming the generalizations are valid observations, and assuming that phonological patterning is not arbitrarily related to the substance of speech, the work in FG represents an attempt at describing the relationship between the physical world and its abstraction and codification in spoken language. However, Feature Geometry is hamstrung by its simplicity and the strictness required of the dependencies it entails. Thus, while knowledge of lingual-laryngeal interaction in LVT sounds was available at the time that these models were being proposed, the limitations of FG forced maintenance of both glottocentrism and linguocentrism. In all of the models in Figure 6.2, what is labeled \textit{Laryngeal/Larynx} is a misnomer, since laryngeal action cannot be said to be independent from articulatory events occurring in pharyngeals. Furthermore, these nodes group features describing glottal states – the models are glottocentric. Linguocentrism is embodied by the tongue root retraction (encoded as [radical] or [RTR]) being selected as the representative articulator of gutturals/post-velars.

\textbf{6.2.4 \textit{A short excursus on OT and lower vocal tract sounds}}

With the advent of Optimality Theory (OT; Prince & Smolensky, 1997, 2004), the theoretical focus in phonological research has shifted away from questions about the

\textsuperscript{165} Another possibility is Dresher’s Contrastive Hierarchy (Dresher, 2003; Mackenzie & Dresher, 2004; Dresher & Zhang, 2005; Hall, 2007), which rests on the premise that the organization of phonological representation is solely defined by contrast and patterning.
representation of sounds and towards questions about the interaction of violable constraints operating on distinctive features. In this framework, variable cross-linguistic behaviour in sound patterning is a matter of the constraint hierarchy and not properties of the underlying representations of sounds per se, although the representational conventions established in previous research are often assumed in OT analysis.

OT research on gutturals/post-velars proceeds from conventions established in FG research. For example, Lombardi (2002, pp. 220–222) claims that the variable laryngeal behaviour identified by Rose (1996) is attributed to the markedness hierarchy for PLACE features: *DORS ⪰ *LAB ⪰ *COR ⪰ *PHAR (meaning that PHARYNGEAL is the least marked place of articulation); further, she claims that laryngeals have PHARYNGEAL specification and are distinguished from pharyngeals proper with a [+glottal] feature. With these assumptions, laryngeal patterning with the guttural/post-velar natural class is due to their PHARYNGEAL specification; their placelessness is due to the unmarked nature of *PHAR (although she only demonstrates unmarked nature of laryngeals as epenthetic segments, not their transparency in translaryngeal harmony).

To illustrate how translaryngeal harmony works, we can examine Yamane-Tanaka’s (2006) analysis of Gitksan transguttural harmony. This analysis uses a modified version of Lombardi’s place markedness hierarchy (given above): among other changes, *PHAR is split into several sub-constraints to account for the difference in behaviour between uvulars and laryngeals, and fricatives and stops of those categories\footnote{The *PHAR hierarchy she proposes for Gitksan is *PHAR(TR)OBS-VLINK ⪰ *PHAR(TR)FRIC-VLINK ⪰ *PHAROBS-VLINK ⪰ *PHARFRIC-VLINK (which roughly corresponds with /q/ ⪰ /h/ ⪰ /ʔ/ ⪰ /h/). The ranking of HARMONY with respect to this hierarchy determines which segments will be transparent. Thus, if Harmony is ranked in between *PHAROBS-VLINK and *PHARFRIC-VLINK, then /h/ but not /ʔ/ will be transparent in vowel harmony.}. The
analysis relies on a constraint HARMONY to drive vowel harmonization; the ranking of
this constraint relative to the place markedness hierarchy determines the class of
segments which are transparent and thus permit harmonization. The low ranking of the
*PHAR class of constraints is taken to explain the transparency of gutturals/post-velars in
the language, the typological variation in transguttural harmonies, and the graded,
diachronic development of transguttural harmony in Gitksan such that it applies with
fewer restrictions across laryngeals and only recently is applying across uvulars.

While this OT approach may satisfy descriptive adequacy for Gitksan and the
typological patterning of transguttural harmony, it does not explain the deeper questions
of why these harmonies occur or why there is cross-linguistic gradedness in transguttural
harmonies. Even if constraints are real cognitive entities, we still have the burden of
explaining what properties of the physical world give rise to them in the first place lest
the constraints arise ex nihilo, in which case we might wonder whether it is sheer
coincidence that they transparently reflect properties of the physical world. This must
involve, in part, understanding the nature of how these sounds are produced (for a more
acerbic criticism of OT, see Ohala, 2011).

6.2.5 Lingual-laryngeal encoding in Articulatory Phonology

Articulatory Phonology (AP) posits articulatory gestures as the atomic units of
phonological representation comprising internal temporal events of articulator
movements (ONSET, OFFSET, TARGET, RELEASE, etc.) and can specify varying degrees of gestural magnitude (Browman & Goldstein, 1989); gestures can overlap and form a “score” when multiple articulators are involved. There is a finite number of articulators defined for speech: LIPS, TT, TB, TR, VEL, GLO.

The work of Borroff (2005, 2007) attempts to address issues of guttural/post-velar phonology in the context of the AP framework (in combination with OT, see Gafos, 2002) Borroff’s suggests that laryngeals lack the ONSET and OFFSET components, which are used to explain some syllabic and sequential properties of these sounds (on the assumption that sounds linearize by aligning their gestural landmarks).

Borroff acknowledges (2007, pp. 77–81) that glottal stop can be phonetically characterized by supralaryngeal gestures, especially associated with the upper larynx. Her survey of glottal stop behaviour leads her to posit two types of glottal stop, a simplex one possessing only a pure glottal gesture in its representation, and a complex glottal stop with a synchronous Tongue Root gesture thought to entail the extra/secondary components of glottal stop in “guttural” languages. She then claims that languages with complex glottal stops are those that have other guttural/post-velar sounds in their

---

167 ONSET is the start of an articulatory gesture; OFFSET is its end point. TARGET is the moment of maximal displacement of the gesture, which may be sustained for a given length of time; RELEASE is the moment the maximal displacement is relaxed.

168 These correspond with physical structures in the vocal tract as follows (for detailed definitions and discussion, see Browman & Goldstein, 1989). LIPS involves the upper and lower lips and the jaw; its parameters are protrusion and aperture. TT involves the tongue tip, body, and jaw; its parameters are tongue tip constriction location and degree. TB involves the tongue body and jaw; its parameters are tongue body constriction location and degree. TR involves the tongue root; its parameters are tongue root constriction location and degree. VEL involves the velum; its parameter is degree. GLO involves the “glottis” (which is a space, not a structure, so presumably this articulator is the vocal folds?); its parameters are constriction degree and location.

169 Primarily meaning /ʔ/ since the majority of her analysis concerns this phoneme.

170 Borroff points out that relabeling this gesture to Epiglottis or something more reflective of the laryngoscopic data will not change her analysis (2007, p. 80).
inventory and that show evidence that glottal stop can align as an onset because of its extra Tongue Root gesture (which has ONSET and OFFSET events).

As Borroff recognizes (2007, p. 183), this model makes very specific predictions regarding the presence or absence of the extraglottal gestures of glottal stop. In languages with laryngeal transparency, these gestures should be absent; in languages with the so-called complex glottal stop, they should be present. The phonetic evidence does not support this interpretation. As Shahin (2011a) observes, Hebrew, a language which should have complex glottal stops because it has other gutturals, there is not necessarily an extraglottal gesture for these sounds, ventricular or likewise. Yet in a language like English or Swedish, which lack other “gutturals”, the extra gestures can occur (e.g. Roach, 1979). Furthermore, Borroff’s suggestion that additional gestural components in glottal stop are used to nullify the transglottal pressure difference are also suspect. It is possible to sustain phonation with extreme levels of extraglottal narrowing, whether ventricular or full epilaryngeal stricture is employed. Furthermore, the high epilaryngeal resistance to glottal flow (in addition to lower subglottal pressure) may actually help prolong the amount of time that phonation can be sustained during supralaryngeal closure (as in creaky voiced stops) because it will take much longer to fill the supralaryngeal plenum before the transglottal pressure difference is nullified. Her perspective suggests also that stopping phonation is the only function of the extra gestures, which misses the generalization that these gestures actually give rise to a wealth of constricted possibilities.

\[\text{\textsuperscript{171} A simple experiment is to compare how long phonation can be sustained for [b] with modal and with creaky phonation. My (informal) results suggest that creakiness enables phonation to be sustained for much longer than is possible with modal phonation.}\]
not limited to glottal stop, but also including various phonation types, such as creaky and pressed/harsh phonation.

6.2.6 Summary: General problems with previous approaches

This section has identified the overwhelming focus placed on the glottis and tongue root in the phonological characterization of LVT sounds (most often discussed under the rubric of guttural/post-velar phonology).

Glottocentrism is couched in discussion of the larynx, but the above review shows that generally, the larynx reduces to the “glottis”. With a “flat” view of the larynx, it is not surprising that only very recently (e.g. Borroff, 2007; Shahin, 2011a) has it become apparent to researchers that the vocal folds might interact with the structures above. The strength of glottocentrism to influence what counts as phonological explanation is embodied by the fact that such acknowledgements are just now emerging despite early, prominent observations of the lingual-pharyngeal-laryngeal interaction (Trigo, 1991; McCarthy, 1994; Keyser & Stevens, 1994; Halle, 1995) and awareness of the supraglottal structures of the larynx and their involvement in “glottal” sounds (Halle & Stevens, 1971).

The history of phonological thinking about the LVT suggests researchers do care about phonetic data and used it in formulating phonological representations for LVT sounds. As discussed in §2.2.1, the overwhelming majority of this research draws on x-ray imaging studies. Given the limitations of x-ray data for revealing LVT articulation, it is not surprising that the “tongue root” became a point of fixation for phonologists. This
is not to say that the tongue root (region) is irrelevant, but relegating focus to it does not help us to model the connection between the larynx and the rest of the vocal tract.

This rest of this chapter will argue that when anatomo-physiological and phonetic understanding of the epilarynx is admitted into the domain of “phonological” explanation, many issues raised in LVT phonological analyses can be attributed to the articulatory interaction between the epilarynx, vocal folds, and the rest of the vocal tract. This is not to say that the epilarynx explains everything in LVT phonology, but countenancing it more importantly frees analysis from adhering to GCLC ways of thinking about such sounds. The next section will provide a conceptual model of what the epilarynx does in speech production. This model is framed in terms of vocal potentials (§6.1.3), and these will be applied in the analysis of three topics of LVT phenomena: phonological uses of epilaryngeal vibration (§7.1), epilarynx interaction with the vocal folds (§7.2), and epilarynx interaction with the supralaryngeal vocal tract (§7.3).

6.3 LVT Phonology: A model of articulation-based phonological potentials

Thus far, three topics have been addressed. First, lower vocal tract phonology spans a disparate set of phonetic and phonological phenomena, much of which points to a connection between the larynx and the rest of the vocal tract (§6.1.1). Second, recent research supports the view of a gradient phonetics-phonology interface (§6.1.2), which is assumed here. Then, in §6.1.3, phonological potentials were introduced as the broad theoretical approach to analyzing phonological systems and patterns. Our present focus is on articulation-oriented phonological potentials related to lower vocal tract sounds. The existence of phonological features (cognitive abstractions over speech sounds) were
neither denied nor assumed; however, features are not the primary object of study here. §6.2 was dedicated to reviewing the previous, feature-based accounts related to LVT phonology. These models were classified as belonging to the glotto-centric–linguocentric (or GCLC) paradigm. Knowing about the GCLC approach is important in understanding how the proposed approach (articulation-oriented phonological potentials) can offer new insights into old problems that GCLC-type features handle poorly.

The current section develops the Model of Lower Vocal Tract Phonological Potentials (LPP) as it relates to LVT phonology and, in particular, to the role played by the epilarynx in LVT sound patterns and systems. The proposed model is presented as follows: §6.3.1 first outlines the core assumptions about the anatomo-physiological nature of the epilarynx which grounds many of the assumptions made in subsequent sections; §6.3.2 then outlines the phonological potentials of the epilarynx; §6.3.3 provides technical details and a summary of the formalism of the theoretical model. Finally, in §6.3.4, predictions based on the theoretical model are made concerning the pathways of LVT phonology defined by the epilaryngeal potentials. Chapter 7 then examines specific LVT phonological phenomena by regarding the classical feature-based approach and then illustrating how the phenomena are treated in the proposed model.

6.3.1 Physiological components of epilaryngeal operation

The epilarynx is a distributed, composite articulatory-system characterized by internal and external physiological control mechanisms which aid in its constriction. The three main physiological components of epilaryngeal operation are: (1) the intralaryngeal musculature; (2) the larynx height mechanism; (3) tongue retraction. (1) is an epilarynx
internal mechanism; (2) and (3) are epilarynx external mechanisms. Each component can be further broken down into subcomponents. This situation is illustrated in Figure 6.4:

Figure 6.4: Key components of epilarynx operation. Midsagittal schematic.

Detailed definitions of these components are provided in Table 6.5 (for a review of the anatomy and physiology of the epilarynx, see §2.1).

Table 6.5: Definitions of the components of epilaryngeal control

<table>
<thead>
<tr>
<th>Component</th>
<th>Subcomponents</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 intralaryngeal</td>
<td>a. vocal-ventricular</td>
<td>lower epilaryngeal</td>
</tr>
<tr>
<td></td>
<td>b. aryepiglottos-epiglottal</td>
<td>upper epilaryngeal</td>
</tr>
<tr>
<td>2 larynx height</td>
<td>a. thyro-hyoid approximation</td>
<td>“local” larynx raising</td>
</tr>
<tr>
<td></td>
<td>b. hyo-laryngeal elevation</td>
<td>“global” larynx raising</td>
</tr>
<tr>
<td>3 tongue retraction</td>
<td>a. hyoglossus</td>
<td>primary retraction</td>
</tr>
<tr>
<td></td>
<td>b. middle genioglossus</td>
<td>“double-bunching” retraction</td>
</tr>
</tbody>
</table>

(Dashed line separates epilarynx-internal [above] from epilarynx-external [below] components)
Component (1), the intralaryngeal musculature, is the most important in driving epilaryngeal constriction. The lower epilaryngeal subcomponent (1a) pertains to the musculature acting primarily in the ventricular plane to cause downward and medial motion of the ventricular folds; the relevant musculature comprises the external thyroarytenoid and its projections into the transverse interarytenoids. Vocal fold adduction (as in pre-phonatory or phonatory postures) enables constriction in this plane. The upper epilaryngeal subcomponent (1b) involves the musculature driving motion in the aryepiglottic plane, the thyroepiglottic muscles and the sphincteric chain formed by the lateral-cricoarytenoid, oblique interarytenoid, and aryepiglottic muscles. The vocal folds do not need to be adducted for constriction to occur in this plane.

Component (2) involves larynx height, which plays a key role in changing the vertical dimension of epilaryngeal constriction. The “local” subcomponent (2a) involves approximation of the thyroid cartilage and hyoid bone driven by the thyrohyoid muscles. This can occur regardless of the “global” position of the hyo-laryngeal complex – subcomponent (2b). Subcomponent (2b) involves the suprahypoid and pharyngo-laryngeal musculature to lift the larynx upwards. Epilaryngeal constriction is more efficiently accomplished when both subcomponents are engaged. Because of the muscular linkage through the hyoid bone, component (2) can cause the tongue to retract/lower, but this will depend on the activation level of the suprahypoid muscles and other forces acting on the tongue.

Component (3) is the tongue retraction component. The term tongue root is avoided because the action of this component influences the entire tongue, not just a localized motion of the tongue root – this is because the tongue is a muscular hydrostat
(see Smith & Kier, 1989). Primary retraction (3a) involves a hyoglossus-driven motion that is obliquely oriented downwards towards the larynx (as Figure 6.4 depicts). Secondary retraction (3b) involves displacement of the posterior body of the tongue under the influence of middle genioglossus action (the extreme contraction of which produces the “double bunching” state seen in a canonical variant of American-English R). An important effect of general tongue retraction (3) is to push the epiglottis posteriorly and to distort the pre-epiglottic body, the stresses of which are transferred to the level of the lower epilarynx and probably assist posterior displacement of the epiglottic tubercle.

Importantly, tongue retraction does not entail active epilaryngeal stricture, which depends on the contribution of the other components, (1) and (2). If retraction is strong, and the larynx is not lowered, then the epilaryngeal space will narrow, but this is passive – not active – epilarynx engagement. Thus, tongue retraction biases but is not sufficient for active epilarynx stricture.

Broadly speaking, the epilarynx has two basic states: constricted and unconstricted (see §2.2.3). Table 6.6 lists general physiological and auditory descriptors applied to these polar opposite configurations.

Table 6.6: Typical concomitant phonetic properties of epilaryngeal states

<table>
<thead>
<tr>
<th>property</th>
<th>anti-constriction</th>
<th>constriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1: intralaryngeal</td>
<td>relaxed</td>
<td>contracted</td>
</tr>
<tr>
<td>C2: larynx height</td>
<td>lowered</td>
<td>raised</td>
</tr>
<tr>
<td>C3: tongue position</td>
<td>advanced</td>
<td>retracted</td>
</tr>
<tr>
<td>phonatory quality</td>
<td>breathy or modal</td>
<td>creaky or harsh</td>
</tr>
<tr>
<td>voice quality</td>
<td>lowered larynx voice</td>
<td>raised larynx voice</td>
</tr>
<tr>
<td>vowel quality</td>
<td>peripheralized</td>
<td>condensed</td>
</tr>
</tbody>
</table>

(C# = component #; see Figure 6.4).
6.3.2 Phonological potentials of the epilarynx and lower vocal tract

The model proposed here claims that understanding the “phonological potentials” of the epilarynx can help us understand how the phonological LVT space is divided into phonemes and the patterning of these phonemes. The phonological potential of the epilarynx is related to a set of epilaryngeal (biomechanical) devices, [ʢ], [ʡ], [ʢ], and so forth, each of which relies on its own specific neuromotor organization governing spatio-temporal relations during device activation but which also share broad similarities in vocal tract posture/configuration. We cannot say these devices are exactly the same in any particular regard. For example, the exact execution of upper epilaryngeal (aryepiglottal-epiglottal) stricture in [ʡ] and [ʢ] will differ at the neuromotor and biomechanical levels – just as the device responsible for labial stricture in labial compression differs from that which is responsible for protrusion-rounding (Gick, Stavness, Chiu, Flynn, & Fels, in preparation). However, we can posit that they share a set of general configurations or postures of the vocal tract. Let us call these configurations/postures gross (physiological) states, and we will denote them with curly braces. For example, {↑lx} denotes the gross state of larynx raising (involving some combination of component 2 and its subcomponents, see Figure 6.4). These states may be static or involve dynamic behaviour, such as vibration; a list of gross states relevant to lower vocal tract phonology is provided in Table 6.7.

172 As discussed in §6.1.3, following Gick et al. (in preparation), white square brackets “[ ]” symbolize devices.
Table 6.7: Gross (physiological) states of the lower vocal tract. See Figure 6.5 for relationships.

<table>
<thead>
<tr>
<th>GS</th>
<th>Physiological Description</th>
<th>Musculature/Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>vfv</td>
<td>VF vibration</td>
<td>LCA, IA, sufficient airflow</td>
</tr>
<tr>
<td>vfo</td>
<td>VF open (abducted/&quot;spread glottis&quot;)</td>
<td>PCA and/or LCA relaxation</td>
</tr>
<tr>
<td>vfc</td>
<td>VF close (adducted/&quot;pre-phonation&quot;)</td>
<td>LCA, IA</td>
</tr>
<tr>
<td>epc</td>
<td>epilaryngeal constriction</td>
<td>iTA, cTA, TE, tIA, oIA, AE, LCA, CE (C1)</td>
</tr>
<tr>
<td>epv</td>
<td>epilaryngeal vibration</td>
<td>epilaryngeal constriction and high airflow</td>
</tr>
<tr>
<td>tfr</td>
<td>tongue fronting</td>
<td>GG and lingual floor muscles, or GGM</td>
</tr>
<tr>
<td>tre</td>
<td>tongue retraction</td>
<td>HG (C3a)</td>
</tr>
<tr>
<td>tra</td>
<td>tongue raising</td>
<td>SG, SL</td>
</tr>
<tr>
<td>tdb</td>
<td>(tongue) &quot;double bunching&quot;</td>
<td>strong GGm (C3b)</td>
</tr>
<tr>
<td>dbe</td>
<td>“double-bunching” epilaryngeal</td>
<td>strong GGm, epilaryngeal constriction</td>
</tr>
<tr>
<td>↑ lx</td>
<td>raised larynx</td>
<td>TH, PP, suprahyoid strap muscles (C2)</td>
</tr>
<tr>
<td>↓ lx</td>
<td>lowered larynx</td>
<td>infrahyoid strap muscles</td>
</tr>
<tr>
<td>vto</td>
<td>(upper) vocal tract rel. open</td>
<td>jaw lowering and/or lowered tongue position</td>
</tr>
<tr>
<td>vtc</td>
<td>(upper) vocal tract rel. close</td>
<td>jaw raising and/or elevated tongue position</td>
</tr>
<tr>
<td>H</td>
<td>high pitch</td>
<td>CT, iTA</td>
</tr>
<tr>
<td>L</td>
<td>lower pitch</td>
<td>infrahyoid strap muscles or epilar. constriction</td>
</tr>
</tbody>
</table>

(C#, e.g. C3a, refers to components of epilaryngeal constriction discussed in §6.3.1; GS = gross states; VF = vocal fold)

The gross states involved in epilaryngeal operation exhibit potential interaction. Many of the gross states of epilaryngeal constriction are diametrically opposed to other potential gross states, giving rise to articulatory conflicts (e.g. Gick & Wilson, 2006). These conflicts are a potential source of language and speaker variation, and they can also involve compromise-configurations in which the articulators involved meet each other “half-way”. On the other hand, many of the gross states of the lower vocal tract produce supportive configurations; thus, there is also articulatory-cooperation. These two aspects of articulation: cooperation and conflict might be thought of in terms of synergies and anti-synergies – the two potential “forces” of gross state interaction. Consider the

---

173 This state can also be characterized as palato-pharyngo-epilaryngeal or in conceptual terms of neo-tongue, neo-pharynx (NTNP). See discussion below and Figure 6.9 and §7.3.1 and Figure 7.12.
schematic in Figure 6.5 which relates the gross states (circles) defined in Table 6.7 through these “synergistic relations” and shows how these relate to different types of potential quality (i.e. vowel quality, phonatory quality, etc.):

Figure 6.5: Schematic of synergistic relations in the LPP: synergies and anti-synergies. Gross states are drawn with circles. Refer to Table 6.7 for definitions.

So synergistic relations encompass synergies and anti-synergies, which describe the relationships amongst the gross states of the (lower) vocal tract. Synergies, denoted by \{→\}, are complementary or mutually supportive relations between gross states\(^\text{174}\);

\(^\text{174}\) A complication that is deliberately neglected here is that only some synergies are mutual in nature: larynx raising \{↑lx\} complements tongue retraction \{tre\} via the linguo-hypo-laryngeal linkage (and the hyoglossus-thyrohyoid musculature). Other synergies seem to operate only in one “direction”: e.g., vocal
they do not imply a particular function, but may be useful towards achieving a functional goal, such as epilaryngeal stricture. An example is the synergy between tongue retraction and larynx raising, symbolized as \( \text{tre} \leftrightarrow \uparrow \text{lx} \); this synergy exists because of linguo-hyo-laryngeal linkage. One example of its relevance to speech is that it helps explain why “low” vowels, particularly “retracted” vowels like [a], tend to involve a higher larynx position (e.g. Wood, 1979). Anti-synergies, denoted by \{ \cdots \}, are “conflicts” between configurations: they do not preclude a particular state, but some compensation is needed within the system; for example, epilaryngeal vibration at higher glottal pitch \{ \text{epv} \cdots \text{H} \} is not impossible, but requires the antithetical combination of epilaryngeal constriction and cricothyroid contraction (Esling & Harris, 2005) needed for increasing glottal pitch (growling at higher glottal pitch will probably involve higher airflow as compensation).

The gross states (Table 6.7) and their synergistic relations (discussed below, see Figure 6.5) are based on previous research on epilaryngeal physiology and speech functioning (see Chapter 2) and are further grounded by additional observations made in the preceding three chapters of this dissertation (Chapter 3, Chapter 4, and Chapter 5). Most of these states are self-explanatory (e.g. \{ \text{vfo} \} is simply any kind of vocal fold abduction). Less obvious are the lingual gross states – tongue fronting \{ \text{tfr} \}, tongue raising (back and up) \{ \text{tra} \}, and tongue retraction (back and down) \{ \text{tre} \}: these are based directly on Esling’s (2005) Laryngeal Articulator Model discussed in §2.2.3\textsuperscript{175}. The

\[ \text{fold opening/abduction} \ \{ \text{vfo} \} \text{ does not have any appreciable influence on larynx lowering, but lowering the larynx} \ \{ \downarrow \text{lx} \} \text{ does predispose the vocal folds to (partial) abduction (see §4.2.5).} \]

\textsuperscript{175} \text{It is assumed (but not addressed in detail) that each of these gross states of the tongue is complemented by a partner component, such as the velic traverse articulator (Gick, Francis, Klenin, Mizrahi, & Tom, 2013) aiding in strictures associated with tongue raising \{ \text{tra} \} or the middle and inferior pharyngeal constrictor muscles aiding in tongue retraction \{ \text{tre} \}.}
(common) vocalic correspondences of these gross states, along with \( \{vtc\} \) and \( \{vto\} \), are depicted in Figure 6.6.

![Diagram showing vocalic correspondences of several gross states](image)

**Figure 6.6**: Vocalic correspondence of several gross states. Shading darkness indicates approximately how strongly a gross state characterizes a particular vowel (or its probability of association with a given vowel).

Before continuing on, let us stop momentarily to reflect on the model being proposed. The concept of phonological potentials is very broad. Phonemes are potential distinctive entities within sound systems. Devices are potential components of these phonemes underlying the physical (biomechanical) realities associated with the phoneme. Each device has phonological potentials that arise from the synergistic relations of gross
states associated with the device. The phonological potentials, taken together, help predict how certain phonemes might evolve or pattern based on articulatory considerations. (A complete model would need to take into account aerodynamics, perception, and so forth.) Thus, the concept of phonological potentials is assumed to primarily apply to broader tendencies at the diachronic or typological scale.

Since our focus is on LVT phonology let us now render how the epilarynx fits into this part of the vocal tract by surveying some relevant broad regions of activity or constriction and depicting of some of the key gross states involved. The rendering is done with a series of tube-abstraction cartoons of the vocal tract which depict abstract configurations (characterizable by gross states).

The first set of cartoons (Figure 6.7, Figure 6.8, and Figure 6.9) shows key regions of constriction of the lower vocal tract. The broad configurations of the pharyngeal tube are depicted in Figure 6.7 along with (a) “neutral” for comparison. The first observation is that there are two tubes – not one – shown in the tube-abstraction of the lower vocal tract: the pharyngeal tube and the epilaryngeal tube (see §2.1.1 and, in particular, Figure 2.1). This diagram deals exclusively with pharyngeal regions, not epilaryngeal ones.

The pharyngeal tube is controlled primarily by the position of the tongue, but undoubtedly the pharynx itself participates in pharyngeal narrowing by means of the pharyngeal constrictor and pharyngeal elevator muscles. There are two key regions: upper pharyngeal and lower pharyngeal. Upper pharyngeal (Figure 6.7b) corresponds with constriction at the tongue body/dorsum approximating with a variety of possible structures (velic traverse, palatoglossal arch, rear pharyngeal arch and upper rear
pharyngeal wall): highly relevant for our present purposes is the tongue raising (up and back) \textit{[tra]} gross state. Lower pharyngeal (Figure 6.7c) is the site of [a]-type stricture in the region of the so-called “tongue root”. It is possible that full linguo-pharyngeal contact can seal off this part of the pharynx without actively constricting the epilarynx, but, to date, this has not been linguistically attested. In any case, even during a plain [a] vowel, the lower pharynx is extremely narrow (Gauffin & Sundberg, 1978; Painter, 1986, p. 331), to the point that, if the epiglottis is curved enough, its lateral margins will likely contact the pharyngeal walls. While this retraction can passively narrow the epilarynx and synergizes with epilaryngeal constriction and larynx raising, it does not constitute active epilaryngeal constriction nor is it sufficient for epilaryngeal constriction: physiologically speaking, epilaryngeal-constriction components 1 (intralaryngeal) and 2 (larynx height) (see §6.3.1) are still free to engage or disengage (denoted by the question marks). For example, the larynx could be actively lowered, which is anti-synergistic with epilaryngeal stricture (see Figure 6.5), or the larynx could be further raised (which is synergistic with epilaryngeal stricture).
Figure 6.7: Broad pharyngeal regions of constriction and gross states (tube abstraction). Dashed line indicates boundary between oral/upper and pharyngeal/lower regions of the vocal tract. Gray double ogive indicates the laryngeal ventricle. Question mark indicates indeterminacy in configuration. Abbreviations: tre = tongue retraction (down and back); tra = tongue raising (up and back).

Two palatal-type regions of constriction and relevant gross states are depicted in Figure 6.8. Tongue fronting \{tfr\} pushes the anterior mass of the tongue forwards and upwards. The primary cause is genioglossus posterior contraction; the palatal vault limits the upward thrust of the tongue and jaw height and lingual muscles (such as the genioglossus anterior) provide means to modify the configuration for different (anterior/front) vowel qualities. The (tongue) “double-bunching” \{tdb\} gross state involves the mass of the tongue diverging in two directions: one palatal, the other lower pharyngeal (hence palato-pharyngeal). Thus, \{tdb\} is interpreted as simultaneously synergistic (see Figure 6.5) with tongue fronting \{tfr\} and tongue retraction \{tre\}. Furthermore, the dorsal-dip in the tongue (at the dashed line) makes \{tdb\} anti-synergistic with tongue raising \{tra\}. Such a manoeuvre results in the canonical tongue
shape for “double-bunched” American-English R (for more details on this tongue configuration see §5.4.2). Note that, like the regional pharyngeal constrictions depicted Figure 6.7, neither of the states in Figure 6.8 entails a particular epilaryngeal state; however, because tongue fronting \{tfr\} has the opposite effect of tongue retraction \{tre\}, it is posited here to be anti-synergistic with epilaryngeal constriction \{epc\}.

Figure 6.8: Palatal/palato-pharyngeal gross states (tube abstraction). Abbreviations: tdb = (tongue) “double-bunching”; tfr = tongue fronting (forward); See Figure 6.7 for additional details.

There are three regional epilaryngeal constrictions to consider (Figure 6.9). Like the pharyngeal constrictions, there are three regions of epilaryngeal constriction differentiated by the height of stricture: lower (a) and upper (b) epilaryngeal stricture and (c) “double bunching” epilaryngeal stricture. Lower epilaryngeal stricture (Figure 6.9a) entails moderate posteroanterior narrowing and, critically, vocal-ventricular fold contact. As far as we know, this does not occur without vocal fold adduction. Upper epilaryngeal (Figure 6.9b) stricture is primarily a matter of narrowing in the aryepiglottic plane. Vocal
fold adduction is an option but not required. Since, epilaryngeal constriction is externally controlled by components 2 (larynx height) and 3 (tongue retraction) (see §6.3.1), the gross state for epilaryngeal constriction \( \text{epc} \) should potentially synergize with \( \text{↑lx} \) and \( \text{tre} \). (Remember: these are potentials, not absolute requirements of epilaryngeal constriction.) It is possible to produce tight upper epilaryngeal stricture without compromising lingual mobility too much (as in whispering). Thus, the tongue is relatively free (b) when upper epilaryngeal stricture occurs (indicated by the crossed dotted arrows), although the synergizing of tongue retraction \( \text{tre} \) with epilaryngeal stricture \( \text{epc} \) (see Figure 6.5) will potentially be associated with retracted vowel quality (hence the asymmetry in the crossed arrows in Figure 6.9b). The implication is that it should be possible to produce any vowel quality with concomitant upper epilaryngeal stricture, as in whisper, but retracted quality will be favoured.
Figure 6.9: Epilaryngeal gross states (tube abstraction). Abbreviations: epc = epilaryngeal constriction; tdb = (tongue) “double-bunching”;dbe = “double-bunching” epilaryngeal; tre = tongue retraction (down and back); ↑lx = raised larynx; See Figure 6.7 for additional details.

“Double bunching” epilaryngeal stricture is depicted in Figure 6.9c: extreme engagement of all epilaryngeal stricture components. The \{epc\}, \{↑lx\}, and \{tre\} gross states are strongly implicated (the “cover” gross state is thus \{dbe\}). The result is very narrow laryngopharyngeal and epilaryngeal air spaces. From an aero-acoustic perspective, the hypopharynx is (nearly) effectively closed off: the impedance of this space will be very high, and, consequently, the resonances of the vocal tract will be drastically changed and the potentials of the tongue altered. To characterize these changes, the concepts of “neo-pharynx” and “neo-tongue” are introduced (defined in §5.4.2). The key point of these terms is that one can potentially reproduce any vowel quality in this configuration, even [i] – but with raised larynx voice quality (or “pharyngealization” [i̯]). This voice quality has a “condensing” effect on vowel quality –
vowels will sound less “peripheral” but retain their intrinsic quality ([i] is still “[i]” when [i\textsuperscript{\textdegree}], not [i] or some other quality).

The regions of constriction discussed above are very broad: they depict some of the key gross states, but they do not illustrate what a device “looks like”. There are many devices (not infinitely many) and showing all possibilities is impractical. However, it is worthwhile examining the outward appearance of the series of potential hypopharyngeal stop devices. Figure 6.10 depicts four “stopping” points interpreted as punctuated events along a continuum of hypopharyngeal constriction (also see Table 6.8). Stricture starts at the vocal fold level (a), rises up to the lower epilarynx (b), then upper epilarynx (c), and finally the lower pharynx is closed off (d). As one moves along the continuum, the articulatory complexity presumably increases, but the benefit is that the stability of the closure also increases. “Glottal stop”, i.e. forceful vocal fold adduction with no epilaryngeal support is the least stable but most articulatorily expedient. It is unstable because, as an inlet valve, the vocal folds are poor at maintaining high intrathoracic pressures without support from the ventricular folds above (see §2.1.2); with increasing subglottal pressure, the vocal folds will tend to abduct and possibly engage in self-sustaining oscillation. (We might assume that the stopping states are stable points in the biomechanical-articulatory relations governing the overall mechanical system – an important part of the attractiveness of a device for phonemic incorporation. As a further assumption, we might suspect that as the complexity increases, one pays the cost of “effort”.)
Table 6.8: Hypopharyngeal stop potentials

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>VF</th>
<th>VVF</th>
<th>AE</th>
<th>LEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>glottal stop</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>?</td>
<td>vocal-ventricular (VV) stop</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>?</td>
<td>aryepiglottal (AE) stop</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>?</td>
<td>linguo-epiglottal-pharyngeal (LEP) stop</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

(n.b.: VF = vocal fold contact; VVF = vocal-ventricular fold contact; AE = aryepiglottal contact; LEP = linguo-epiglottal-pharyngeal contact; dotted line = epilaryngeal pivot)

Figure 6.10: Schematic showing hypopharyngeal stop potentials along a continuum of complexity and stability. ae = aryepiglottic fold; c = cuneiform tubercle; e = epiglottis; f = ventricular (or false vocal) fold; k = corniculate tubercle; ppw = posterior pharyngeal wall; pt = posterior tongue; v = (true) vocal fold; dotted outline = cuneiform cartilage.

An important change happens as one progresses along the continuum from lower-epilaryngeal to upper-epilaryngeal closure: the directions of action change (as discussed in §2.1.2). The imaginary division between these steps along the continuum is called the “epilaryngeal pivot”. Throughout the discussion in Chapter 7, the concept of “pivot” is
invoked in connection with the organization of potential phonological systems and subsystems. These pivots (which could also be thought of as phonological fulcrums or hinges), can be identified as division lines which phonemic systems tend to organize around. Like real pivots, fulcrums, or hinges, the conceptual pivots have two directions of operation. (It could be presumed that phonological features can be defined along these pivots, but our focus here is not on defining features, it is on understanding the physical nature of LVT sound systems and the key role played by the epilarynx in shaping these systems.)

The pivots discussed in this work vary in their degree of abstractness, but the epilaryngeal pivot can be related directly to epilaryngeal physiology. The affiliation (or direction) of the constriction changes at the epilaryngeal pivot from being primarily affiliated with the vocal folds (in fact, the ventricular folds are dependent on the vocal folds) to being affiliated with the tongue, which participates more and more as one moves towards the extreme end of the continuum (d). Furthermore, as discussed in §2.1.2 and §6.3.1, these two levels of the epilarynx also appear to have two dedicated muscular subsystems.

We might wonder whether a similar continuum could be defined for hypopharyngeal fricatives. The answer is that, while vocal fold abduction (as in [h]) can be combined with upper epilaryngeal constriction (as in [h]), it is incompatible with lower epilaryngeal constriction because the ventricular folds (i.e. the lower epilarynx) cannot adduct when the vocal folds are abducted. (Hence, in Figure 6.5, \{vfo \cdots epc\}.) Thus, the continuum only applies for hypopharyngeal stops: the hypopharyngeal
fricatives are discontinuous because of the anti-synergy between opening the vocal folds and constricting the epilarynx (one must “jump” from [h] to [h]).

The final potential is epilaryngeal vibration, or growling (discussed in §6.1.3). It is highly synergistic with epilaryngeal constriction. Although simultaneous vocal fold and ventricular fold vibration (as in Kargyraa throat singing) is physically possible (see Chapter 3), this type of phonation has yet to be attested in linguistic contexts: it may be the anti-synergistic combination of larynx lowering with epilaryngeal constriction \( \text{\{lx} \cdots \text{epc}\) which makes vocal-ventricular phonation less “attractive” for incorporation into sound systems. However, vibration of the upper epilarynx, particularly the aryepiglottic folds, is attested. It should be combinable with upper epilaryngeal and “double bunching” epilaryngeal potentials.

6.3.3 Phonological potentials: Formalism

The LPP does not assume any traditional or current phonological formalism (such as FG or OT) – except the phoneme. The primary reason is because the LPP does not make reference to distinctive features (as discussed in §6.1.2, §6.1.3, and §6.2): the LPP is not a model of mental representations. Rather, all analysis in the LPP concerns potential phonemic organization at the physical/body level (possible divisions and patterning of the system), and compares this to “realized” systems – those that are attested in real life for real speakers. The idea of potentials\textsuperscript{176} is useful because it is not

\textsuperscript{176} Potentials can be thought of as analogous to potentials in physics, such as electric (field) potential, gravitational potential, potential energy, and so forth. All of these concepts describe a physical state that can potentially influence the behaviour of a system (such as a heavenly body in orbit or a charged particle moving in an electric field) but do not strictly determine this behaviour. Predicting the actual behaviour requires consideration of initial conditions, boundary conditions, and other applied forces. Knowing the
absolute or deterministic: it does not describe the way phonological systems ought to be, it describes the way they tend to be and the factors that go into determining these tendencies. Here we are focused on “articulatory” factors, but potentials are not limited in this way. Phonological potentials can also be aero-dynamic, acoustic, perceptual, and even cognitive in nature.

A potential phonological system or subsystem is characterized by the connection among phonemes, devices, and the gross states and their synergistic relations. While it is possible to outline these relationships with prose, schematics (such as that shown in Figure 6.11) are used when matters become more complicated and more and more factors are being related together. (In fact, Chapter 7 is organized such that the progression of the sections gradually introduces more and more complex schematics as more of the vocal tract becomes implicated in a given potential system or subsystem.)

potential of a system does go a long way in predicting how it might evolve (for example, it is easy to predict the likely behaviour of a compressed spring the moment it is released). Similarly, phonological potentials let us make a reasonable prediction (or at least ex post facto interpretation) of the behaviour of phonological systems.
Figure 6.11: Schematization of phonological potentials. Letters are used as illustrative stand-ins for actual speech-related symbols.

In Figure 6.11, the phonemic potential is denoted with the conventional notation (i.e. /A/) and maps directly to phonemic analysis in any conventional phonological formalism (not shown). The white square-bracket notation (e.g. [B]) denotes devices (“devices”), which often correspond in size or scope to gestures (sensu Browman & Goldstein, 1989). Following Gick et al. (in preparation), the devices are functional primitives of movement defined by neuromotor pathways (proportionally fixed spatio-temporal patterns of muscle activation); they can be tuned through repetition, lost through lack of usefulness, and “constructed” when the need exists for a new device for some new functional goal. A phoneme can comprise many devices (e.g., in Figure 6.11, /A/ comprises [B] and [C]) – a simple example (abstracting away some details) would be the
common devices used for /b/ realization (see Gick, Stavness, Chiu, Flynn, & Fels, in preparation): a lip closure device, symbolized $[^b]$, and a vocal fold vibration device, symbolized $[^\cdot]$\textsuperscript{177}.

It is assumed that the relation of devices to the phoneme is stochastic (or potential) – not absolute – and this is represented in Figure 6.11 using grayscale shading: darker phoneme-device association lines signify a greater likelihood of finding a particular device being used in connection with the phoneme. For example, in Figure 6.11, /A/ is very likely be realized with the engagement of $[^B]$; it is relatively less likely to be realized with the concomitant engagement $[^C]$. (Returning to /b/, we would suspect a very high probability of finding the lip closure device $[^b]$ when examining a speaker “realizing” the phoneme). The lighter the association line, the more the unlikely it is that a device will be used in the context of a particular phoneme. (No attempt is made to quantify these probabilities, but this is a goal for future development of the theory.) Furthermore, if, in an analysis using the formalism, it is necessary to indicate that a device will be diachronically abandoned (which becomes relevant in §7.2.4), this can be indicated with a dashed phoneme-device association line (as in the relationship between /A/ and $[^D]$ and its corresponding gross state $[^H]$ in Figure 6.11).

The devices should not be confused with allophones, even though there is some conceptual overlap. Like allophones, some devices may be engaged more often in correlation with specific phonological contexts. Unlike allophones, devices are not

\textsuperscript{177} The decomposition of a phoneme into corresponding devices is an empirical issue that will not be directly evaluated here. The exact decomposition is less important than the assertion that phonemes are associated with devices because they impose finiteness on phonological potentials. The main focus of the discussion is on the gross states (which are not specific to devices) and the synergistic relations relating these. Inclusion of the devices is done to allow for easy interfacing of the LPP with a larger theory of devices and to ground the gross states in a theoretical object that has well-defined articulatory properties.
directly interpretable as sounds – devices are neuromotor structures – but (non-abstractly speaking) their concerted functioning does produce sounds in the context of speech. In the present work, the devices are understood abstractly, they are also not language or speaker specific: they are potential. Languages and speakers will vary in which devices are “realized” (i.e. tend to be actually used). The goal of the present research is not to focus on the study of allophony but rather to focus on how the physical nature of the body (expressed in terms of devices and their associated gross states) influences the organization of phonemic systems.

Gross (physiological) states are general configurations or activities of the vocal tract and can be identified broadly with a certain auditory qualities. Gross states should not to be confused with devices. Gross states are the general, physical configurations or activities of speech divorced from the neurophysiological structure underlying these states (i.e. the devices). Gross states are introduced into the theory because they allow for statements to be made concerning the commonality of configurations across the devices. For example, sounds made at the lips may involve many different devices (one for compression, one for protrusion, one for closure, and so forth), but we can broadly group these as involving gross activity at “the lips” (and not, say, at “the toes”). The gross states taken alone are merely descriptive in nature (and nowhere near as descriptively precise as the devices), but they are conceptually useful because of the synergistic relations. The synergistic relations let us predict (or at least offer an *ex post facto* explanation of) how devices associated with particular gross states are likely to organize into phonological systems and patterns. In a way, the synergistic relations are the core of the potentials model: they are tendencies, biases, inclinations, susceptibilities – but not absolutes.
In Figure 6.11, gross states are represented with circles (in-text references to gross states are made with the \{E\} notation described earlier). Gross states are either directly or indirectly associated with the devices. Direct association (Figure 6.11b) means that the gross state is entailed by a particular device (for example, the lip closure device \[b\] entails gross activity of the lips, so this is a direct association). Indirect association (Figure 6.11c) means that a gross state \{I\} is associated with a device \[B\] through synergistic relations (in this case, a synergy) with some gross state \{E\} that is directly associated with \[B\].

Several times now, phonological potentials have been claimed to be “realizable”. Phonological potentials do not exist *per se* because they are really just abstractions and general principles about “phonological” behaviour. You cannot go out and measure a phonological potential since you would only ever be able to measure something real – what we call “phonetic” output or realization. Languages – and speakers of languages – realize only a subset of what is possible (what is potential) and these things are subject to variation – they are not human-universals – but it is claimed that phonological potentials are (see §6.1.3). All languages making use of a particular sound have a **chance** at realizing the potentials associated with that sound.
Figure 6.12: The “big picture”: phonological potentials and realities. Abbreviations: ↓lx/↑lx = lowered/raised larynx; blc = bilabial closure; vfc = vocal fold closure (or adduction/pre-phonatory configuration); vfv = vocal fold vibration. Devices: [b] bilabial closure; [] = vocal fold vibration; […] = vocal fold adduction for pre-phonation. “Inner” indices indicate uniqueness of all objects to (imaginary English native) speaker-1; “Outer” indices are to emphasize uniqueness of individual phonetic output tokens.

For a more concrete example of this potential-reality distinction, consider the diagram in Figure 6.12. This diagram introduces a “phonetic” layer between the phonemic and device layers; the purpose is to illustrate the “big picture” (to make things visually simpler in Chapter 7, the phonetic layer is omitted, as it is in Figure 6.11 and Figure 6.13; n.b.: Figure 6.12 is parallel to Figure 6.1). Potential /b/ and a realization of /b/ in the form of an English phoneme /b1/ that is somewhere inside speaker-1’s
brain/body are depicted. Everything in the real case must be indexed to emphasize that it is all unique to speaker-1. Everything is associated with a probability (represented by the darkness of association lines). In (simplified) potential /b/ there are two phonetic possibilities – equally likely to be found (as far as the present consideration is concerned). \([b]\) is the potential phonetic output associated with \([\text{[b]}]\) and \([\text{[.]}]\); \([\text{[b]}]\) is the potential phonetic output associated with \([\text{[b]}]\) and \([\text{[.]}]\) (\text{n.b.} timing is intrinsic to the nature of devices and we abstract away from it here). These devices are associated with several gross states, including a \{blc\} (not introduced previously) for basic bilabial closure. (There are some interesting laryngeal height synergies: voicing \{vfv\} is synergistic with larynx lowering \{↓lx\}; voicelessness \{vfc\} for the pre-phonation state – or vocal-fold adduction – is synergistic with larynx raising \{↑lx\}).

The right side of Figure 6.12 shows how these potentials are realized in one specific language, Canadian English, for one specific speaker, speaker-1. Imaginary field work reveals that this speaker has a very high probability of producing the voiceless unaspirated \([b]\) in association with /b\_1/ (all depicted with dark lines compared to the other possibility, which is depicted with very-light-gray lines). All phonetic output tokens are unique. Many tokens of \([b\_1]\) are found: \([b\_1]\)_1, \([b\_1]\)_2, \([b\_1]\)_3, and so forth. On the rare occasion, we observe \([b]\) (voiced unaspirated). We posit that speaker uses the devices \([\text{[b]}]\) and \([\text{[.]}]\) (close reflections of the more abstract, universally potential \([\text{[b]}]\) and \([\text{[.]}]\)) in realization of /b\_1/. We also suspect that the vocal tract conforms to the directly associated gross states (and may occasionally have some slight larynx raising). The present study is primarily concerned with the phonological potentials (of the epilarynx and lower vocal tract) rather than the realities (although these do come into play).
To further clarify how the formalism will be used in this dissertation (Chapter 7), consider Figure 6.13, which is a simplified schematic for a potential /ʔ/. In this example (we will see a much more complex one later on), the phonemic potential /ʔ/ has two potential devices: ⟦ʔ⟧ and ⟦ʔ⟧. ⟦ʔ⟧ is what is conventionally thought of as glottal stop: just vocal fold closure/adduction, but it also applies to the pre-phonatory state (as in unaspirated voiceless stops). ⟦ʔ⟧ is vocal-ventricular-fold stop stricture, which crucially uses vocal-ventricular fold contact (see Figure 6.10). ⟦ʔ⟧ is just associated with the vocal-fold-closure gross state, \{\textit{vfc}\}, which is synergistic with \{\textit{epc}\} and \{\textit{↑lx}\} (as depicted in the diagram). ⟦ʔ⟧ is primarily associated with \{\textit{vfc}\}, \{\textit{epc}\}, and \{\textit{↑lx}\} gross states, which are all synergistic (as depicted in Figure 6.5).

Figure 6.13: Simplified illustration of the potential glottal stop subsystem. Abbreviations: vfc = vocal fold closure/adduction; epc = epilaryngeal constriction; ↑lx = raised larynx. For description of these, see Figure 6.5 and Table 6.7. Devices: ⟦ʔ⟧ = vocal fold closure; ⟦ʔ⟧ = vocal-ventricular fold contact and adduction (lower epilaryngeal/LE) closure (see Figure 6.10).
In Figure 6.13, and elsewhere, general grayscale darkness is inherited from subordinate levels: for example, the darkness of device–gross-state association lines is “inherited” from phoneme-device association. (The synergistic relations also “inherit” in this way.) Note, however, that the gray tone is darker for $\hat{\text{ʔ}}$ to $\{\text{vfc}\}$ and $\{\text{epc}\}$ and lighter for $\hat{\text{ʔ}}$ to $\{\uparrow \text{lx}\}$. This indicates that $\{\text{vfc}\}$ and $\{\text{epc}\}$ are of relatively greater importance to $\hat{\text{ʔ}}$ than $\{\uparrow \text{lx}\}$ is. By “greater importance” it is meant that $\{\text{vfc}\}$ and $\{\text{epc}\}$ will almost always be found in the realization of the device (say if one were to look with a laryngoscope, one would see these states and infer that the participant’s $\hat{\text{ʔ}}$ device was being engaged).

In summary, this section and the previous one have examined the gross states and corresponding synergistic relations relevant to positioning the epilarynx within phonological theory and provided a formal framework for expressing these relations. Figure 6.5 provides the master key to all of the phonological potentials to be discussed in Chapter 7, and it is recommended that Figure 6.5 and Figure 6.11 be printed for easy reference.

6.3.4 Predictions about the pathways of lower vocal tract phonology

We have thus far made note of the key components of epilaryngeal function (§6.3.1), outlined the phonological potentials of the epilarynx by examining its potential devices and the gross states and synergistic relations characterizing these devices (§6.3.2), and demonstrated the formalism used to express the relations binding potential phonemes, (phonetics), devices, and the gross states and synergistic relations all together
The model of Lower Vocal Tract Phonological Potentials (LPP) makes the following predictions collected under three questions:

1) What are the potential roles and consequences of epilaryngeal vibration in phonology?

The LPP countenances the possibility for the epilarynx to interact with phonation, yielding distinctive harshness generally, and in the extreme case, distinctive epilaryngeal vibration. Where it occurs in tonal-register systems, growling shows a bias to low tones on account of its subharmonic nature (see Chapter 3), i.e. \{epv \leftrightarrow L\}, but because of the anti-synergy between epilaryngeal constriction and high glottal pitch \{epc \cdots H\} and epilaryngeal vibration and high glottal pitch \{epv \cdots H\}, constricted high tones should be uncommon. If they do occur they will likely be realized without epilaryngeal vibration. Finally, somewhat surprisingly, vocal fold abduction is predicted to be synergistic with epilaryngeal vibration \{vfo \leftrightarrow epv\}, yet is anti-synergistic with epilaryngeal stricture \{vfo \cdots epc\}, which leads to the potential for epilaryngeal vibration to be biased towards voiceless or whispery voiced phonatory qualities.

2) What is the phonological relationship between the epilarynx and the vocal folds?

The glottocentric account views glottal stop and glottalized or laryngealized sounds as non-interactive with the supralaryngeal vocal tract: in such models, these sounds are not predicted to interact with vowels or pharyngeal consonants. On the other
hand, the LPP predicts that glottal stricture is augmentable with (lower) epilaryngeal stricture (Figure 6.10). From this, we expect to find the potential closure mechanisms in the hypopharynx being realized within the sound systems of languages—extending the notion of “glottal” stop. Furthermore, we might expect to find glottal stops and glottalized sounds (the focus will be on ejectives) patterning with pharyngeals. Alternatively, glottal stops and glottalized sounds should exhibit non-placelessness, i.e. occasional interaction with vowels, most likely relatively open vowels, as expressed by \{tre ↔ epc\}. Conversely, since \{vfo ⋯ epc\}, /h/ is predicted to be less likely to interact with the pharyngeals and with relatively open vowels.

3) What is the relationship between the epilarynx and the supralaryngeal vocal tract?

In the linguocentric view, pharyngeals are [RTR]: they are therefore predicted to retract and lower vowels. In the LPP, however, pharyngeals are generally uncoupled from specific vowel quality effects, although several possibilities arise depending on the engagement of the three epilaryngeal stricture components. Pharyngeals may fail to retract vowels where other guttural/post-velar sounds, mainly uvulars, do (upper epilaryngeal potential); pharyngeals may indeed retract vowels across-the-board \{epc ↔ tre\}; or, pharyngeals may cause the extreme effect associated with the “double bunching” epilaryngeal potential, possibly becoming associated with fronting or palatalization \{dbe ↔ tfr\}; languages realizing this potential are likely to have phonological coupling between pharyngeals and palatal sounds. Moreover, languages can form entire register-type systems around the presence or absence of epilaryngeal constriction and will tend to
show vowel, phonatory, tonal, and voice quality correlates consistent with the gross states associated with epilaryngeal constriction and anti-constriction.

Before continuing with the study proper (Chapter 7), a note of the intention of the LPP is warranted. The LPP is intended to collect together the wealth of empirical insight stemming from phonetic research on laryngeal articulation in speech and marry it with what is known about lower vocal tract sound systems. It is not strictly comparable to traditional phonological formalisms that generally seek determine the cognitive organization behind speech sound systems. Rather, the LPP is a formalism grounded firmly in articulation (and concomitantly to acoustics, aerodynamics, and perception) that allows for generalization about the potential influences acting on phonological systems, but grounding (primary) causality outside of the representational realm or the realm of “phonological computation”. Nonetheless, as noted earlier, it is necessary to contextualize each topic of discussion in terms of what has come before: Generative-style analysis in the GCLC paradigm (§6.2). In most cases, the weakness of these models, i.e. their failure to predict particular sound patterns and organizations, is addressed and contrasted with the LPP.
Chapter 7

LOWER VOCAL TRACT PHONOLOGY: ARGUMENTS

This chapter explores the Lower Vocal Tract Phonological Potentials (LPP) model in a series of arguments which directly relate to the three chapters detailing epilaryngeal functions: epilaryngeal vibration as a source mechanism (introduced in Chapter 3, addressed in §7.1); epilaryngeal interaction with the vocal folds (introduced in Chapter 4, addressed in §7.2); and the relationship of the epilarynx to the rest of the vocal tract (introduced in Chapter 5, addressed in §7.3). §7.4 provides a summary of the three main topics addressed in this chapter and provides some theoretical discussion concerning the relation of the LPP model to other models of phonology. The chapter is summarized in §7.5 with a statement of the implications of the LPP model, and suggestions for future phonological research.

7.1 Expressions of epilaryngeal vibration in phonology

In mainstream phonology, laryngeal features describe the states of the vocal folds, such as their aperture, height, and tension (e.g. Halle & Stevens, 1971; Kingston, 1985; Goldstein & Browman, 1986; Avery & Idsardi, 2001). The phonological use of supraglottal laryngeal mechanisms of phonation is not predicted to occur according to such models.

Epilaryngeal vibration has never been countenanced in any model of phonology. Several factors could have resulted in this scenario. One factor is the apparent rarity of distinctive epilaryngeal vibration. Another factor could be the lack of visual evidence
showing epilaryngeal vibration is clearly an epilaryngeal mechanism and not a glottal one. Yet another factor is that, unlike glottal-level behaviour, phonetic and phonological epilaryngeal behaviour is relatively under-documented, and therefore it lacks consistent taxonomic description, which would allow for cases in diverse languages to be more easily compared. Finally, there is a stigma associated with epilaryngeal vibration that it is either pathological or that it can lead to pathology and is therefore not “normal” behaviour. This factor explains Gordon and Ladefoged’s remark that “[i]f the !Xóõ did not exist, and someone had suggested that [sphincteric/strident voice] could be used in a language, scholars would probably have said that this was a ridiculous notion” (2001, p. 401). Catford writes that the sounds with epilaryngeal vibration found in Agul “would be regarded as pathological if they were produced by a person speaking English” (1983, p. 347). Traill’s (1986) defense of sphincteric phonation as “phonetically normal” (p. 125) – despite the impression of “severe vocal fold pathology” that the acoustic signal gives (p. 126) – reflects the general perception of sounds produced by epilaryngeal vibration as abnormal or suggestive of voice disorder. Rose (1989, p. 242) also observes the prejudice that epilaryngeal vibration (or “growling” for Rose) is pathological, and counters with a quip from Ladefoged (1993, p. 351) “one person’s voice disorder is another person’s phoneme”.

In this section, it is asserted that not only is epilaryngeal vibration an entirely normal function of the human vocal mechanism, but it is also phonologically relevant: it can be phonologically distinctive in vowels and consonants, and it is relevant in phonological patterning, particularly involving tone, although vowel-system interactions also occur (but for reasons having more to do with epilaryngeal configuration than
This section demonstrates how epilaryngeal vibration is accommodated in the context of the LPP as a phonological potential, and how its patterning is an expression of the synergies which underlie its interactions with other states of the vocal tract.

First, the distinctive occurrence of epilaryngeal vibration as part of phonation-type in Khoisan languages is examined in §7.1.1, followed by a consideration of “trilling” in pharyngeal consonants in §7.1.2. Then, §7.1.3 presents a consideration of the dual nature of epilaryngeal vibration as phonation-like and trilling-like. Finally, §7.1.4 observes how epilaryngeal vibration functions in tonal systems. Some of the patterns noted in this section, especially those related to vowel quality, will be re-examined in subsequent sections (mainly §7.3.2).

7.1.1 Distinctive epilaryngeal vibration in Khoisan

This section argues that epilaryngeal vibration is a phonological behaviour that does not transparently reduce to traditional phonological features. This establishes the need for the LPP model which directly incorporates the epilarynx.

The strongest case for distinctive epilaryngeal vibration comes from languages of the Khoisan group (formerly thought to be a language family) where epilaryngeal vibration occurs as a distinctive phonation-type. Noteworthy examples are !Xóõ (Traill, 1985, 1986), Jul’hoansi (Miller, 2007), and N|uu (Miller, Brugman, Sands, et al., 2009).

Traill’s (1986) often discussed description of distinctive sphincteric phonation in !Xóõ provides some of the clearest evidence that it is possible for phonological systems to exploit epilaryngeal vibration for contrastive purposes. During sphincteric vowels,
... the tips of the arytenoid cartilages\textsuperscript{178} vibrate strongly against the tubercle of the epiglottis and with particularly forceful productions these vibrations may be imparted to the epiglottis whose tip can also be observed to vibrate (p. 125) ... [producing] a series of irregular, low frequency (around 50 Hz) pulses ... most likely due to the vibrations of the arytenoid cartilages (p. 127-128); (Traill, 1986)

In comparison to other “rare” sounds, like the [\textipa{t\textael}] in Wari’ and Oro Win (Ladefoged & Everett, 1996), sphincteric phonation may be typologically rare, but it is not rare within the !Xóô language: unlike [\textipa{t\textael}] (Ladefoged & Everett, 1996, p. 794), it is not relegated to a small subset of words or words limited to a particular semantic function (e.g. it is not simply onomatopoeic). Rather, “sphincteric” is, according to Traill, “fully integrated” (1986, p. 130) into !Xóô’s rich phonation-type system.

Traill provides the most concrete evidence in the form of laryngoscopic imaging of native speakers producing sphincteric vowels and x-ray imaging of his own attempts at phonetically reproducing the sound: in both cases epilaryngeal vibration is witnessed. Rose (1989, p. 240) corroborates this auditorily by suggesting “sphincteric” phonation in !Xóô is very similar to “growl” in Zhenhai Wu (see §7.1.4). Evidence for epilaryngeal vibration in the other example languages, Juǀ'hoansi and N|uu, is admittedly somewhat sparse. Juǀ'hoansi (Miller, 2007) exhibits a rich set of phonation-type contrasts: modal, breathy, glottalized, and “epiglottalized” \textsuperscript{179} (symbolized with a superscript [\textael]). Miller does not commit to the possibility of epilaryngeal vibration in epiglottalized vowels, but she suggest that they have aryepiglottic stricture and that “the use of [epiglottalized]...
follows articulatory investigations … by Esling (1996, 1999)” (p. 57) in which it was clearly demonstrated that [ʕ] implies epilaryngeal (aryepiglottic) vibration. More clear evidence comes from Nǀuu, which has a contrast between modal and epiglottalized phonation types; Miller (2009, p. 141) compares the former with harsh register in Bai (see §7.1.4) and Somali (see §7.1.2) citing Edmondson et al. (2007), which is strongly suggestive that these vowels involve epilaryngeal vibration since laryngoscopic evidence supports the occurrence of epilaryngeal vibration in both Bai and Somali. She also states that epiglottalized vowels in Nǀuu\(^{180}\) are similar to those in Juǀ’hoansi. It is thus assumed that all of these example cases, !Xôô, Juǀ’hoansi and Nǀuu, involve epilaryngeal vibration – an assumption that is subject to validation by further empirical studies.

Trigo (1991, p. 118) analyses the !Xôô system of phonation-type contrasts (see Table 7.1) with her tentative [raised larynx/RL] feature in conjunction with the familiar [constricted glottis/cg] and [spread glottis/sg] features. The full set of !Xôô phonation-type contrasts are: (1) modal, (2) breathy, (3) glottalized, (4) pharyngealized, (5) breathy-pharyngealized and (6) glottalized-pharyngealized (based on Trigo, 1991, p. 118).

---

\(^{180}\) Audio data of epiglottalized vowels in Nǀuu provide further evidence that her analysis is correct (see Figure 3.14 and Figure 3.15).
Table 7.1: Trigo’s representation !Xóõ phonation types

<table>
<thead>
<tr>
<th></th>
<th>[RL]</th>
<th>[sg]</th>
<th>[cg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>modal</td>
<td>V</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>breathy</td>
<td>h V</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>glottalized</td>
<td>? V</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>pharyngealized</td>
<td>ʕ V</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>breathy-pharyngealized (sphincteric)</td>
<td>hʕ V</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>glottalized-pharyngealized</td>
<td>ʔʕ V</td>
<td>+</td>
</tr>
</tbody>
</table>

Trigo’s [RL] is supposed to represent the pharyngealization component in the !Xóõ system and is intended to explain the fact that both pharyngealized and sphincteric vowels occur with [+back] vowels (the quality of sphincteric and epiglottalized vowels is discussed further in §7.3.2). More abstractly, Trigo supports her analysis by stating that sphincteric/breathy-pharyngealized vowels “fill the gap” in vowel typology (p. 118) by constituting the combination of [spread glottis] with the feature for pharyngealized vowels (Trigo’s [RL]).

Hess (1998, p. 274) provides a different interpretation based on her typological analysis of pharyngeal articulations using x-ray data; for !Xóõ, Hess concludes that pharyngealized vowels should be regrouped such that glottalized-pharyngealized vowels are excluded, there being no evidence that these involve any additional pharyngeal narrowing. Further, she proposes to represent pharyngealized vowels with PHARYNGEAL-[radical] (involving tongue retraction at the level of the suprahyoid portion of the

---

181 All symbols are those used by Trigo (1991, p. 118).
182 Trigo does not state whether this is also true for the glottalized-pharyngeal vowels, which her analysis predicts would be the case.
183 It is surprising that Trigo accepts this phonological argument connecting [RL] with [+back] as consistent with her proposal for an independent [RL] feature (vis-à-vis [RTR]) but then casts doubt on the independence of the [RL] feature based on Traill’s x-ray data that these sounds also involve retraction of the tongue.
epiglottis) and the sphincteric/breathy-pharyngealized vowels as PHARYNGEAL-[laryngopharynx] (involving narrowing of the laryngopharynx). Hess does not provide details on how her analysis provides more insight into the phonological system of !Xóõ; rather, she simply states that it provides a more phonetically-based model (p. 275) at the cost of phonological symmetry\textsuperscript{184}.

From the perspective of the LPP, the problem with Hess’s and Trigo’s analyses is that, by focusing solely on configuration of the vocal tract, they neglect the key fact that sphincteric vowels phonologically involve epilaryngeal vibration. Articulatory factors such as a raised-larynx position \{\(\uparrow\text{lx} \leftrightarrow \text{epc}\)\} and a retracted tongue \{\(\text{tre} \leftrightarrow \text{epc}\)\} are synergistic with narrowing the epilarynx, which facilitates its vibration \{\text{epc} \leftrightarrow \text{epv}\}. These relationships are intrinsic to the LPP: !Xóõ is completely predictable in this regard. The fact that sphincteric phonation is realized with an abducted vocal fold configuration, and with minor, if any, vocal fold vibration (see Traill, 1986), is probably attributable to diachronic factors more than phonological feature permutations, or may be a reflex of the \{\text{vfo} \leftrightarrow \text{epv}\} synergy (which implies that greater airflow is required to drive epilaryngeal vibration, so it makes sense to abduct the vocal folds partially). Traill may be correct in asserting (see 1986, p. 130) the linguistic specialization of this configuration. It is doubtful that upper epilaryngeal stricture with slightly abducted vocal folds and relatively open lower epilarynx occurs in any context other than non-linguistic vocalization and speech. In other words, it is not a configuration with obvious physiological, life-

\textsuperscript{184} Following Clements (2003, pp. 291–292), the symmetry argument would be framed in terms of feature economy. Feature Economy is the principle of representing as many sounds as possible with the fewest features as possible (and is expressed as \(E = S/F\), where \(E\) is the economy index, \(S\) is the number of segments in an inventory, and \(F\) is the number of features needed to represent those segments). According to Clements, feature economy and inventory symmetry are not equivalent principles; rather, Clements argues that feature economy is the cause for the symmetrical appearance and “gap” avoidance characteristic of many phonological systems.
supporting underpinnings. In the LPP, the combination of upper epilaryngeal stricture with vocal fold abduction is a phonological potential. Furthermore, it is in no way unusual: it constitutes the configuration used for whispering (Esling & Harris, 2005; Honda, Kitamura, Takemoto, et al., 2010), although whispering usually does not typically occur with epilaryngeal vibration.

7.1.2 Epilaryngeal vibration and pharyngeal consonants

This section connects epilaryngeal vibration with pharyngeal consonants. This motivates viewing pharyngeals as having an epilaryngeal rather than linguo-pharyngeal basis. Agul pharyngeals are discussed as a possible case of distinctive epilaryngeal trilling.

In the present model, pharyngeals are interpreted as being a phonological potential associated mainly with the upper epilaryngeal stricture region. As such, it is not surprising that epilaryngeal vibration might occur in the realization of these sounds. That it has phonemic potential in connection with pharyngeal sounds is found in Catford’s *Fundamental Problems in Phonetics* (1977a, p. 163):

… the breathy-voiced (or whispery-voiced) ventricular fricative trill [ɦ̃] … occurs in some of the Abkhazo-Adyghe languages of the north-west Caucasus, notably Abaza and in Arabic loan-words in Adyghe and Kabardian. According to Rogava *this sound formerly occurred as a regular phoneme* [Moisik emphasis] in all the Abkhazo-Adyghe languages, but has been lost in Adyghe and Kabardian … The Adyghe sound is produced much deeper in the throat with occasionally ‘bleat-like’ ventricular trill plus ventricular turbulence …

In this quote, Catford describes the possibility of distinctive ventricular vibration. Regardless of whether Catford’s assessment of the mechanism involved is correct or not,
the generalization is that epilaryngeal vibration (of which ventricular fold vibration is a subtype) apparently can be phonemic.

Another form of epilaryngeal vibration – aryepiglottic trilling – is visually attested in phonetic variants of the pharyngeal consonants for one speaker of Iraqi Arabic, and a combination of epiglottal and aryepiglottic vibration occurs for one speaker of Somali (Edmondson, Padayodi, Hassan, & Esling, 2007) (see §3.2; also see Table 7.2). The Iraqi Arabic case is intriguing because the vibration occurs even independent of contrastive consonantal length and in relatively quick pronunciation. It is difficult to determine just how common this phonetic variant is: auditorily, the voiced trill [ʕ] can give an impression of a stop [ʔ], similar to the “tight approximant” variant of ‘ayn reported on by Heselwood (2007); the voiceless trill [h] on the other hand is auditorily confusable with the voiceless fricative [h]. Since visual evidence is difficult to obtain, it is possible that the occurrence of this variant is under-reported in Arabic and in other languages with pharyngeal consonants. Esling (1999, p. 354) suggests that the use of labels such as “laryngealization” in reports on pharyngeals (Esling cites Butcher & Ahmad, 1987, p. 166) might have masked the occurrence of epilaryngeal vibration by conflating it with creaky phonation. Impressionistic labels used in describing pharyngeals, such as the “remarkably raucous” (Colarusso, 1985, p. 367), often are suggestive of trilling.
Table 7.2: Aryepiglottic trilling in Iraqi Arabic (see §3.2)

<table>
<thead>
<tr>
<th>/raħi:l/</th>
<th>[raʰɾi:l]</th>
<th>‘traveling’</th>
</tr>
</thead>
<tbody>
<tr>
<td>/raħ:i:l/</td>
<td>[ɾaɾi:l]</td>
<td>‘to travel a lot’</td>
</tr>
<tr>
<td>/saʕi:d/</td>
<td>[saɾi:d]</td>
<td>‘happy’</td>
</tr>
<tr>
<td>/saʕi:d/</td>
<td>[saɾi:d]</td>
<td>‘make people happy’</td>
</tr>
</tbody>
</table>

(n.b.: Following Esling (2010) [ɾ] and [ɾ] are used here to denote voiced and voiceless aryepiglottic trilling, respectively)

These cases all support the observation that epilaryngeal vibration is phonetically incorporated as part of the production of pharyngeals, possibly for phonetic enhancement, as Esling (1999, p. 364) suggests. There is also evidence, however, that it is phonologically instantiated as a manner contrast within the pharyngeal place category. Specifically, Catford (1983, p. 347) claims that there are five pharyngeal consonants in Burkikhan Agul, “pharyngeal” /ʔ/ and “deep pharyngeal” /ɾ/ and two laryngeals, /h/ and /ʔ/. He suggests that the “raucous” deep pharyngeals (singling out /ɾ/) could be considered epiglottal trills on account of low frequency, periodic vibration of the epiglottis characteristic of these sounds. In line with Catford’s fundamental analysis of a there being a contrast, Ladefoged & Maddieson (1996, pp. 167–169), interpret /h/ and /ʔ/ as fricatives produced at two different places of articulation – pharyngeal and epiglottal; Esling (1999, p. 364) counters by saying that these sounds are homorganic (aryepiglottal-epiglottal) and instead represent a manner contrast between fricative (noisy) and fricative-trill (noisier) associated with lowered and raised laryngeal settings, respectively. Neither Esling nor Ladefoged & Maddieson discuss whether /ʔ/ and /ɾ/ contrast in the same way, i.e. either by place or by manner.

Kodzasov’s recordings of two speakers (“Aghul,” n.d.), one representing the Burkikhan dialect (a female) and the other representing the Tpig dialect (a male), provide
numerous examples of both voiced and voiceless trilling (see §3.1.4). The transcription of this data found on the UCLA Phonetics Lab Archive website, where this data is made freely available, does not make the occurrence of epilaryngeal vibration clear. To remedy this, a narrow re-transcription is provided as a supplement (see Appendix C); this re-transcription carefully notes the occurrence of voiced and voiceless epilaryngeal vibration as denoted with the symbols [ʂ] and [ɬ]. Selected minimal pairs are illustrated in Table 7.3 using the UCLA and the supplementary transcription (“Moisik”); these data highlight evidence for contrasts found in the Burkikhan dialect and (likely) cognate contrasts found in the Tpig data. (Since the Tpig data are less complete than that for Burkikhan, the following discussion is framed in the context of the latter.)

There is little doubt that epilaryngeal vibration occurs in the recorded data (see §3.1.4). For the Burkikhan speaker, the presence of epilaryngeal vibration is more common in intervocalic and in coda contexts, while onset pharyngeals are most often realized as fricatives – even the [ɬ]\(^{185}\), which has a mildly whispery voiced quality when it is a fricative. The Tpig speaker exhibits trilling in onset, intervocalic, and coda contexts. In both dialects, epilaryngeal vibration also occurs in the context of /ɬ/. This sound is consistently realized as a stop, but, like /ɬ/\(^{186}\), it causes several different co-articulatory effects on neighbouring vowels, including harsh voice quality with and

\(^{185}\) The [ɬ] sound very similar to those in Morley Stoney, which are also realized as voiced pharyngeal fricatives. This seems to be true, despite there being arguments that such sounds do not tend to occur (Laufer, 1996).

\(^{186}\) In non-pharyngeal contexts, the Burkikhan speaker tended to realize voicing with some breathiness. It could be that this helps to enhance the constricted voice qualities associated with pharyngeal contexts. This is similar to the velarized quality of Russian which probably serves to enhance the distinctiveness of palatalized consonants in that language.
without epilaryngeal vibration\textsuperscript{187}, creakiness, and pharyngealization (usually raised larynx voice quality) (for the complete transcription, see Appendix C).

Table 7.3: Narrow phonetic transcriptions of Burkikhan and Tpig Agul\textsuperscript{188}

<table>
<thead>
<tr>
<th>Comparison</th>
<th>UCLA Burkikhan</th>
<th>Tpig</th>
<th>Moisik Burkikhan</th>
<th>Tpig</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>( \hat{\text{h}} )</td>
<td>( \text{had} )</td>
<td>( \text{\text{`ad}} )</td>
<td>( \text{`ad} )</td>
<td>'star'</td>
</tr>
<tr>
<td></td>
<td>( \text{`s} )</td>
<td>( \text{`ad} )</td>
<td>( \text{`ad} )</td>
<td>( \text{`ad} )</td>
<td>'partridge'</td>
</tr>
<tr>
<td></td>
<td>( \text{?} )</td>
<td>( \text{?ad} )</td>
<td>( \text{?ad} )</td>
<td>( \text{?ad} )</td>
<td>'two'</td>
</tr>
<tr>
<td>b.</td>
<td>( \text{`h} )</td>
<td>( \text{han} )</td>
<td>( \text{`an} )</td>
<td>( \text{`an} )</td>
<td>'courtyard'</td>
</tr>
<tr>
<td></td>
<td>( \text{`s} )</td>
<td>( \text{`an} )</td>
<td>( \text{`an} )</td>
<td>( \text{`an} )</td>
<td>'stomach'</td>
</tr>
<tr>
<td>c.</td>
<td>( \text{`h} )</td>
<td>( \text{har} )</td>
<td>( \text{`ur} )</td>
<td>( \text{`ur} )</td>
<td>'flour'</td>
</tr>
<tr>
<td></td>
<td>( \text{`s} )</td>
<td>( \text{`ar} )</td>
<td>( \text{`ur} )</td>
<td>( \text{`ur} )</td>
<td>'hare'</td>
</tr>
<tr>
<td>d.</td>
<td>( \text{`h} )</td>
<td>( \text{m`an} / \text{m`an`ar} )</td>
<td>( \text{m`eh} / \text{m`eh`r} )</td>
<td>( \text{m`eh} / \text{m`eh`r} )</td>
<td>'whey(s)'</td>
</tr>
<tr>
<td></td>
<td>( \text{`h} )</td>
<td>( \text{m`an} / \text{m`an`ar} )</td>
<td>( \text{m`eh} / \text{m`eh`r} )</td>
<td>( \text{m`eh} / \text{m`eh`r} )</td>
<td>'whey(s)'</td>
</tr>
<tr>
<td>e.</td>
<td>( \text{`s} )</td>
<td>( \text{`aj} )</td>
<td>( \text{`aj} )</td>
<td>( \text{`aj} )</td>
<td>'potato'</td>
</tr>
<tr>
<td></td>
<td>( \text{?} )</td>
<td>( \text{`aj} )</td>
<td>( \text{`aj} )</td>
<td>( \text{`aj} )</td>
<td>'plug'</td>
</tr>
<tr>
<td>f.</td>
<td>( \text{`s} )</td>
<td>( \text{`al} )</td>
<td>( \text{`ul} )</td>
<td>( \text{`al} )</td>
<td>'summer'</td>
</tr>
<tr>
<td></td>
<td>( \text{?} )</td>
<td>( \text{`al} )</td>
<td>( \text{`ul} )</td>
<td>( \text{`ul} )</td>
<td>'mouse'</td>
</tr>
<tr>
<td>g.</td>
<td>( \text{`a} )</td>
<td>( \text{har} )</td>
<td>( \text{`ur} )</td>
<td>( \text{`ur} )</td>
<td>'flour'</td>
</tr>
<tr>
<td></td>
<td>( \text{`e} )</td>
<td>( \text{h`ar`c`aj} )</td>
<td>( \text{h`e`r`c`aj} )</td>
<td>( \text{h`e`r`c`aj} )</td>
<td>'grasshopper'</td>
</tr>
<tr>
<td>h.</td>
<td>( \text{`a} )</td>
<td>( \text{``asuf} )</td>
<td>( \text{``asuf} )</td>
<td>( \text{``asuf} )</td>
<td>'wet'</td>
</tr>
<tr>
<td></td>
<td>( \text{`e} )</td>
<td>( \text{```as} )</td>
<td>( \text{````es} )</td>
<td>( \text{````es} )</td>
<td>'to cry'</td>
</tr>
</tbody>
</table>

\((n.b.:\text{ For more details and examples, see }\S 3.1.4 \text{ and Appendix C.})\)

\textsuperscript{187} Maddieson & Write (1995) note that in Amis (see \S 7.2.3), a language of Taiwan unrelated to Caucasian languages, the [?] is observed often to involve “impulse-like energy spikes” at onset and offset of the closure, which they interpret as “trill-like” vibrations of the epiglottis (p. 49).

\textsuperscript{188} In the UCLA transcription, \( \text{\text{\`s}} \) and \( \text{\`h} \) imply voiced and voiceless epiglottal fricatives, as recommended in Esling (2010). In the Moisik transcription, these symbols imply voiced and voiceless epilaryngeal (aryepiglottal-epiglottal) trills.

\textsuperscript{189} The Moisik transcriptions for several cases in Burkikhan show an initial \( \text{\`h} \) before \( \text{\`\`\`as} \) or \( \text{\`\`\`\`es} \). One could speculate that these vowels, which do not always occur for Burkikhan \( \text{\`\`\`as} \) and which were not heard in the Tpig data for \( \text{\`\`\`\`es} \), could phonetically support the perception and/or production of the pharyngeal sounds (just as some apical trills can be preceded by a schwa). However, their presence admittedly weakens the minimal pair argumentation.
Based on the previous accounts in the literature, we expect some form of contrast in Burkikhan Agul between the /h ʕ/ and /h ʢ/ sets. However, closer examination of these data (Appendix C) reveals that, in general, it is difficult to find solid evidence for this (regardless of the nature of the contrast). The data contain one minimal triplet (a), which establishes the contrastiveness of /h ʕ ʢ/. Case (b) provides further support that /h/ and /ʕ/ contrast in Burkikhan, and, in this example, these sounds correspond with [ʕ] and [ʢ] in Tpig, which differ solely by epilaryngeal vibration (although the UCLA transcription would imply they are the same). This would seem to support Esling’s suspicion that epilaryngeal vibration is distinctive in Agul. As (c) indicates, however, this pattern is not robust, since the Tpig speaker also neutralizes the contrast when it occurs in a different vowel context: the speaker produces identical-sounding [ʕ] approximants with no audible trace of epilaryngeal vibration in the form (likely) cognate with Burkikhan /ʕar/ ‘hare’.

Ladefoged & Maddieson (1996, pp. 167–169) and Esling (1999, p. 364) both form their arguments around the word for ‘wheys’ (d), but there are actually two pronunciations for this word in the data: one with [h] and the other with [ʢ]. Unless there is some other semantic factor accounting for this difference, this evidence undermines the claim that the /h ʢ/ contrast exists (regardless of its basis), and whether there is epilaryngeal vibration or not is a matter of free variation. Despite the problematic ‘wheys’ (d) case, the possibility that epilaryngeal vibration is distinctive in Agul should not be ruled out entirely: as we have seen, (b) shows that it might distinguish /ʕ/ and /ʢ/ in

\[\text{\footnotesize\textsuperscript{190}}\]

It must be stated that, while both recordings are poor quality, the Tpig data are especially poor (as a rough indication, judging from the Nyquist frequency, it seems that the sample rates are 22100 Hz for the Burkikhan data and 11050 Hz for the Tpig data). It is possible that epilaryngeal vibration may be occurring in other tokens, but a very conservative approach was taken to transcription of the data in Table 7.3 and Appendix C on account of the overall poor quality of this data.
Tpig, and (e) & (f) provide evidence that epilaryngeal vibration distinguishes /h/ and /ɾ/; the (f) case is particularly strong because a parallel pattern holds in both dialects.

Although the status of a contrast involving epilaryngeal vibration needs further empirical support, it seems plausible. In any case, the analysis presented here is in agreement with Esling (1999) that Agul does not represent a place contrast between pharyngeal and epiglottal consonants, viz. Ladefoged & Maddieson’s (1996, pp. 167–169) analysis. Most challenging for Ladefoged & Maddieson’s analysis is the copatterning of vowel quality with the supposed /h h/ contrast in Burkikhan Agul. In the UCLA transcription (g; also see Appendix C), 〈h〉 occurs with 〈a〉 and 〈o〉, 〈h〉 appears with 〈i〉 ([ɛ]) and 〈u〉 ([u] or [y]). These vowels contrast in other contexts, as (h) illustrates. In the present interpretation, any difference in the intensity or spectral profile of the noise associated with 〈h〉 and 〈h〉 is better interpreted as a coarticulatory effect of the different vowel contexts, rather than an indication of a place contrast between pharyngeals and epiglottals.

Based on Catford (1977b, pp. 294–295), related Lezgic languages Lezgi and Tabasaran, both contrast /æ/ and /a/, which appears to be the case in Agul. Catford points out that the somewhat “skewed” (p. 295) appearance of these vowel systems may be attributed to the fronting/palatalizing effect of pharyngeals and pharyngealized consonants (discussed further in §7.3). In the cases under consideration, however, both segments are essentially “pharyngeals”, so the more plausible analysis is that these vowels represent an independent contrast in Agul.

Further empirical study of the Burkikhan and Tpig dialects of Agul is clearly required to determine conclusively what the phonological status of epilaryngeal vibration
is. It may turn out to be a phonetic variant of pharyngeals, comparable to that found in Iraqi Arabic and Somali and parallel to uvular trilling during uvular fricatives. The ‘wheys’ (d) case in Table 7.3 supports this possibility at least for /h/ /h/. Yet, another possibility is that in both dialects the possible epilaryngeal contrasts are /h/ /h/ /h/. The fact that voiceless trills in general (particularly [ɾ]) tend to be less stable synchronically and diachronically than voiced ones (Solé, 2002, p. 680) might extend to epilaryngeals as well. As a general principle, voiceless trills have glottal friction, higher oral pressure, higher transconstriction flowrates, and a larger open quotient, all factors which make them phonetically similar to fricatives (Solé, 2002, p. 680). It seems reasonable that the same principles apply to epilaryngeal trills and we might predict that the /h/ should be more stable as a trill phoneme than /h/, which, like /ɾ/, will be more fricative-like, and, also like /ɾ/, might be predicted to be lost or merged with another fricative diachronically.

Regardless of the ultimate status of pharyngeals in Agul, the fact that it occurs at all and has significant influence on vowel quality (by overlapping with a large portion of neighbouring vowels) when it does occur may have consequences for sound change. These vowel effects may remain as a residue in the case of historical loss. Such might be a possible origin of distinctive harsh voice quality on vowels (with or without epilaryngeal vibration). The influence of epilaryngeal vibration of perceived pitch (being generally lower in the presence of epilaryngeal vibration) may have consequences for tone systems or the emergence of tone. The Agul data do not provide evidence for any of these speculations, but their role should be considered in a complete account of human speech sound systems.
7.1.3 The dual nature of epilaryngeal vibration

This section returns to the theoretical issue originally raised in §3.1.1 regarding the classification of epilaryngeal vibration as phonation-like or trilling-like. The argument here is that epilaryngeal vibration has a dual nature in phonological terms. Its expression in phonological systems as either phonation or trilling reflects this dual potential.

Broadly speaking, five structures in the vocal tract can produce linguistically relevant vibrations: the lips, anterior tongue, velum, epilarynx, and vocal folds. In phonetic classification we speak of two types of vibratory behaviour, phonation and trilling. In canonical theory, the lips, tongue tip, and uvula are said to produce bilabial, apical, and uvular trills; the vocal folds, on the other hand, are not said to trill. Instead, they are said to produce phonation when they vibrate. This linguistic dichotomization of vibratory behaviour reflects the phonetic properties that differentiate what is possible with each structure, and how phonological systems incorporate these mechanisms in forming distinction.

Generally, trilling implies far more restricted phonetic and phonological possibilities than phonation does. Trills are typically low in frequency, around 25 to 30 Hz (Ladefoged & Maddieson, 1996, pp. 217–230; Catford, 2002, p. 171), and no tone-like trilling-frequency contrasts have ever been attested and trilling quality is also non-contrastive (e.g. there are no contrasts between, say, “modal” trills with “breathy” trills).

---

191 Although, within limits, it is possible to modify the fundamental frequency at which these structures oscillate through adjustment to postural tension forming of the articulators forming the trill and by changing the rate of airflow. A number of authors have also commented on how (Ladefoged & Maddieson, 1996, pp. 217–230; Catford, 2002, p. 171), even though the structures differ in mass (e.g. the lips are much more massive than the uvula) the trills these structures produce tend to occur at roughly the same frequency. This of course does not take into account the stiffness characterizing these structures at the moment they are set into vibration (natural frequency is predicted by $f = \sqrt{k/m}/2\pi$, where $m$ and $k$ are the effective mass and stiffness, respectively).
More importantly, they are restricted in phonological distribution. Specifically, where they do occur distinctively, the vibrations of such structures are always consonantal in nature. It is true that tongue tip trills can be syllabic and receive stress (e.g. Czech /ɾ/; Dankovičová, 2009), but this is no different from other consonants, such as fricatives, nasals, or laterals taking on syllabic function. Furthermore, trills do not productively combine with other places of articulation as if they were a secondary articulation. Rather, they are manners of articulation unique to particular places of articulation (most often tongue-tip trills) or are allophonic or phonetic variants (most often the case for uvular trills). Bilabial trills are highly context-dependent within a phonological system. Most attested bilabial trills historically arise from prenasalized bilabial stops preceding relatively close back rounded vowels (like [u]; (Ladefoged & Maddieson, 1996, p. 130); the [t̠ʃ] found in Wari’ and Oro Win has a dental stop as a “necessary component” (Ladefoged & Everett, 1996, p. 798) but, similarly, always precedes [o] and [u] vowels.

Trilling has never been attested as a productive secondary feature of vowels comparable to, for instance, nasalization. The explanation likely has to do with the location of the vibration within the vocal tract. In each case, production of trilling significantly interferes with what else can be simultaneously accomplished and, consequently, does not permit a high degree of acoustic differentiability amongst the various possibilities (such as concomitant vowel quality, e.g. trilled-vowels [i̯ u̯ a̯] are highly similar sounding to each other).

In contrast to trilling, the phonatory behaviour of the vocal folds is much more unrestricted in its phonological nature. Vocal fold vibration is the basis of tonal contrasts and it is possible to modify the quality of the vibration in phonologically relevant ways
(giving, say, a contrast between modal and breathy phonation types). Laryngeal states can combine with either consonants or vowels, or both within a given language. Most languages have modal voiced consonants (usually stops and usually in contrast to voiceless stops; Gordon & Ladefoged, 2001, p. 384) and modal voiced vowels are universal (Ladefoged & Maddieson, 1996, p. 50). Although less common, some languages exhibit nonmodal phonatory qualities in combination with either consonants or vowels, and an even smaller number of languages exhibit nonmodal contrasts on both consonants and vowels (Miller, 2007, p. 79).

Trilling and phonation are, therefore, not simply labels applied by convention to identify the different vibrations found in the vocal tract: these categories reflect a significant differentiation in the systemic nature of these sounds. Thus, in characterizing what epilaryngeal vibration is in phonological terms, it is necessary to address whether this kind of vibration qualifies as trilling-like or phonatory-like in its phonetic and phonological nature. In this consideration, it is worth noting that form reflects function. The epilarynx occupies a location within the vocal tract that is between the lips, tongue tip, and uvula – which trill – and the vocal folds – which phonate. Perhaps it is not surprising then that epilaryngeal vibration could be regarded to have the phonologically-dual nature of being both trilling-like and phonation-like.

Like other trills, it tends to have fixed rate of vibration (around 50 Hz), no tonal contrasts involving voiceless epilarynx vibration at different rates are attested, and it is unlikely that languages could contrastively modify the quality of epilarynx vibration (without involving vocal fold vibration). Epilaryngeal vibration is further trilling-like where it occurs in the context of pharyngeal place of articulation. In such consonants, it is
restricted in its occurrence to consonantal contexts (although in this context, epilaryngeal vibration can bleed into vowels by means of co-articulation). Since the epilarynx is the primary mechanism involved in producing pharyngeals, homorganic vibration in such contexts resembles homorganic trilling of the uvula during uvular consonants: it is a phonetic option, possibly for enhancement (Esling, 1999). It is similar in nature to tongue tip trills (which often contrast with homorganic fricatives, such as /s/) in that it can be contrastive with other non-trilled manners of articulation homorganic to pharyngeal place of articulation, such as in Agul.

On the other hand, epilaryngeal vibration is phonologically phonation-like because in some languages, similar to vocal fold vibration, it can contrastively combine with either consonants or vowels, or, more rarely, both (such as Ju’hoansi; Miller, 2007). In such cases it distinctively ranges across consonant place of articulation and vowel quality; Ju’hoansi has /tʰ kʰ ᵁʰ tʰ kʰ ᵁʰ gʰ�ʰ gʰ�ʰ�ʰ / and /ɑʰ oʰ ɑʰ ɔʰ / (Miller-Ockhuizen, 2003, p. 20; Miller, 2007, pp. 58, 61). It participates with the vocal folds in expanding the (laryngeal) nonmodal phonatory possibilities, expanding the dimensions of available contrast.

Epilarynx vibration is a phonological rarity in comparison to other types of trills and other phonation types. Since epilaryngeal vibration is a function of the articulation involved in pharyngeal consonants, and since such sounds are also relatively uncommon, it seems unsurprising that epilaryngeal vibration would also be uncommon. In addition to the increased muscular effort and time required to engage epilaryngeal stricture, epilaryngeal vibration also probably requires increased airflow required to drive the vibration. Perceptual factors may also explain the infrequency of trilling contrast in
phonological systems. Consider the case of aryepiglottic trilling in Iraqi Arabic (see §3.2). It is possible that, because of a sudden drop in fundamental frequency, voiced aryepiglottic trills [ʕ] might be perceptually confusable with aryepiglottal-epiglottal stop. This is parallel to Heselwood’s (2007) argument that this is the case for the tight-approximant variant of ‘ayn found in several Arabic dialects – a sound that is articulatorily comparable to [ʕ]. With regard to voiceless aryepiglottic trilling, Solé (2002, pp. 680–681) argued that, although voiceless trills in general (i.e. apical, uvular, etc.) are articulatorily more robust, they tend to be synchronically and diachronically less stable than voiced ones because they lack acoustic/auditory distinctiveness from fricatives. The voiceless aryepiglottic trills [h] are very noisy. It therefore seems highly probable they have low discriminability from pure aryepiglottal-epiglottal fricatives [h].

### 7.1.4 Epilaryngeal vibration and tone systems: Zhenhai and Bai

This section examines epilaryngeal vibration in the context of tone systems. This examination illustrates the “realization” of phonological potentials associated with epilaryngeal vibration in Zhenhai and Bai.

A key acoustic property of epilaryngeal vibration is the subharmonic structure found in the frequency domain (see Chapter 3). Since the natural frequency of the epilarynx is relatively low, even when under increased tension (Moisik, Esling, & Crevier-Buchman, 2010), this subharmonic structure is low frequency (the first subharmonic is typically between 40 and 100 Hz). Furthermore, there is evidence that subharmonic structure is perceptually associated with low pitch (Gerratt & Kreiman, 2001, p. 371) and, consequently, gives the impression of large size (Ohala, 1996). The implication is that, if epilaryngeal vibration were to interact with a tonal system, we
might expect it to be associated with low tones (L), which is stated in the relationship \( \{ \text{epv} \leftrightarrow \text{L} \} \).

For physiological reasons, we also expect epilaryngeal vibration to be less likely to occur in conjunction with high tones, or \( \{ \text{epv} \cdots \text{H} \} \). As noted above in §6.3.2, combining the epilaryngeal stricture needed to enable vibration with increased cricothyroid activity to increase the fundamental frequency of the vocal folds (and thereby achieve a high tone target) results in a compromised configuration in which the vocal folds are longitudinally stretched and subjected to ventricular fold contact, but the epilarynx is relatively open (i.e. the anteroposterior distance is large relative to more fully constricted states) and the aryepiglottic folds are under increased tension and consequently thinned longitudinally. The prediction is that, in this state, vibration of the aryepiglottic folds (or other tissues of the epilarynx) will be less likely than in configurations with weaker cricothyroid activity.

The Zhenhai dialect of Wu represents an illustration of \( \{ \text{epv} \leftrightarrow \text{L} \} \) and \( \{ \text{epv} \cdots \text{H} \} \) and more broadly of \( \{ \text{epc} \leftrightarrow \text{L} \} \) and \( \{ \text{epc} \cdots \text{H} \} \). Rose (1989) observes that Yang tones in Zhenhai can exclusively exhibit whisper (voiceless), whispery voice, harsh voice, and “growl” (p. 230). Yang tones are distinguished by low F0 onsets (T3 is rising-falling; T4 is low-rising; T6 is rising). Table 7.4 provides examples extracted from Rose (1989) of the different environments conditioning the phonation types. Whisper tends to occur in syllables with voiceless obstruent onsets, as in (a) and (b), and it often briefly precedes whispery voice and growling tonal onsets (p. 233). Whispersy voice tends to be found with relatively close oral vowels, as in (c), (d), and (e), and in syllables with a sonorant onset, as in (d) and (e). Growling (and presumably harsh voice in general) is
never found on close oral vowels but rather tends to occur with relatively open vowels, as examples (g) to (l) indicate, nasalized vowels (l), and without any onset-segment correlation\(^{192}\).

### Table 7.4: Distribution of whisper, whispery voice, and growl in Zhenhai (Rose, 1989)

<table>
<thead>
<tr>
<th>example</th>
<th>meaning</th>
<th>tone</th>
<th>segmental</th>
<th>phonation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ปาฏ</td>
<td>‘sugar’</td>
<td>T3</td>
<td>p</td>
</tr>
<tr>
<td>b.</td>
<td>ฏง</td>
<td>‘stupid’</td>
<td>T4</td>
<td>p</td>
</tr>
<tr>
<td>c.</td>
<td>ฏิ</td>
<td>‘ground’</td>
<td>T4</td>
<td>t</td>
</tr>
<tr>
<td>d.</td>
<td>ฏิ</td>
<td>‘secret’</td>
<td>T6</td>
<td>m</td>
</tr>
<tr>
<td>e.</td>
<td>ฏร</td>
<td>‘rain’</td>
<td>T4</td>
<td>y</td>
</tr>
<tr>
<td>f.</td>
<td>ฏะ</td>
<td>‘rice’</td>
<td>T4</td>
<td>f</td>
</tr>
<tr>
<td>g.</td>
<td>ฏา</td>
<td>‘to spread out’</td>
<td>low-lvl?</td>
<td>a</td>
</tr>
<tr>
<td>h.</td>
<td>ฏเ ’ยำ</td>
<td>‘hair’</td>
<td>low-lvl?</td>
<td>æ</td>
</tr>
<tr>
<td>i.</td>
<td>ฏา</td>
<td>‘overcoat’</td>
<td>low-lvl?</td>
<td>a</td>
</tr>
<tr>
<td>j.</td>
<td>ฏง</td>
<td>‘straight’</td>
<td>T6</td>
<td>e</td>
</tr>
<tr>
<td>k.</td>
<td>ฏง</td>
<td>‘to complain’</td>
<td>T4</td>
<td>e</td>
</tr>
<tr>
<td>l.</td>
<td>ฏา</td>
<td>‘wall’</td>
<td>low-lvl?</td>
<td>à</td>
</tr>
</tbody>
</table>

\(n.b.: [V] = \text{whisper}; [V] = \text{whispery voice}; [V'] = \text{growl.}\)

In phonetic terms, Rose makes the following observations about growling:

sometimes perceivable as a second pitch, much lower than the pitch of the corresponding Yang tones with whispery voice … indicat[ing] a low frequency amplitude modulation of a higher frequency source – presumably the vocal cords – by some as yet unidentified structure coupled to the cords … [in] my imitations of Zhenhai growl … the whole visible laryngeal structure and the epiglottis were set into rather violent vibration\(^{193}\) (Rose, 1989, p. 238).

---

\(^{192}\) Although, judging from Rose’s examples, it could be that growling is predisposed to occur on relatively open vowels in syllables or words terminated by a glottal stop.

\(^{193}\) For T3, rising-falling, Rose comments that growling only occurs on the low F0 part of the initial rise, even though the falling part re-enters the F0 region where growling occurred in the rise. This suggests a phonetic “growling hysteresis” (cf. Tokuda, Horáček, Švec, & Herzel, 2007).
Growling – epilaryngeal vibration – in Zhenhai is of phonological interest because of its distribution relative to tonal and segmental properties. Although Rose is not confident about the phonological conditioning of the different phonation types (p. 243), he does assert that vowel height plays a key role, particularly in predicting whether growl will occur or not. Although the relationship between epilaryngeal constriction and vowel quality will be discussed further in §7.3, for now it is worth observing that, unlike the restriction of growling/epilaryngeal vibration to !Xóõ vowels /ɑ o u/, Zhenhai growling can occur on front vowels and on nasalized close vowels.

Rose recognizes the parallel laryngeal states involved in producing whisper/whispery voice and harsh voice (with or without growl/epilaryngeal vibration), which prefigures a very similar claim in Esling & Harris (2005) that the common factor shared by these qualities is constriction of the epilarynx. Moreover, in evaluating the Zhenhai system, Rose concludes that all three phonation types are reflexes of a more abstract register intrinsically characterized by epilaryngeal constriction (“epiglottalization” in Rose’s terms, p. 240), suggesting it is the phonetic underpinning of the phonological Yang tone class. In support of this interpretation is the fact that the Yin and Yang tone contours are essentially the same, neglecting the initial, low pitch component of the Yang tonal contours (p. 243), which Rose interprets as the realization of the register component. The low tone restriction of this “constricted” register supports the \{epc ↔ L\} generalization.

The restriction of growling to relatively low tones in Zhenhai Wu is mirrored in Jianchuan Bai, which has tonal-register contrasts, traditionally described as a tense-lax system. Using laryngoscopic evidence, Edmondson & Esling (2006, p. 173) show that the traditionally-labeled “tense” register, involves epilaryngeal constriction; moreover,
epilaryngeal vibration (growling) occurs when tone value is relatively low. In conformity with Esling & Harris (2005), Edmondson & Esling replace tense register with the term “harsh register”. Although modal, breathy, and harsh phonation types are phonetically attested, the system can be analyzed as based on two contrasting registers differing by whether there is epilaryngeal stricture (constricted) or not (unconstricted), as depicted in Table 7.5.

Table 7.5: Tonal-register contrasts in Bai based on Edmondson & Esling (2006, p. 173)

<table>
<thead>
<tr>
<th>tone</th>
<th>unconstricted</th>
<th>constricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>55 [tɕi^5:i]</td>
<td>‘much more’ [tɕi^5:i]</td>
</tr>
<tr>
<td>mid</td>
<td>33 [tɕi^3:i]</td>
<td>‘to pull’ [tɕi^3:i]</td>
</tr>
<tr>
<td>rising</td>
<td>35 –</td>
<td>[tɕi^3:i]</td>
</tr>
<tr>
<td>mid-falling</td>
<td>31 [tɕi^1:i]</td>
<td>‘nephew’ [tɕi^1:i]</td>
</tr>
<tr>
<td>low-falling</td>
<td>21 –</td>
<td>[tɕi^2:i]</td>
</tr>
</tbody>
</table>

(N.B.: All syllables are [tɕi:]. The use of [ii] allows for phonation type transitions during the syllable to be more easily represented. [V]= harsh voice; [V^i]= epilaryngeal vibration.)

Breathiness is restricted to the unconstricted register and occurs on the mid-falling tone (see ‘nephew’ in Table 7.5) and non-breathy realizations of this tone do not (apparently) occur (Edmondson, Esling, Harris, Li, & Lama, 2001; Edmondson & Esling, 2006). The implication under this analysis is that, more than just being a default, modal register, the unconstricted register is diametrically opposed to the constricted register via anti-constriction, i.e. adjustments to the laryngeal mechanism which oppose the engagement of epilaryngeal stricture, such as larynx lowering and vocal fold abduction or less adductory effort. When the tone level is high, modal results; when it is low or falling, breathiness results, congruent with the mutual compatibility of larynx lowering (which is supported by the laryngoscopic evidence), (slight) vocal fold abduction, and low pitch. In
Bai, then, we have instances of $\downarrow \text{l} \text{x} \cdots \text{epc}$ but $\downarrow \text{l} \text{x} \leftrightarrow \text{L}$ and $\downarrow \text{l} \text{x} \leftrightarrow \text{vfo}$ for the unconstricted mid-falling tone; the unconstricted mid- and high-level tones do not involve breathiness because this is not the essence of the register, and no synergies, other than $\uparrow \text{l} \text{x} \leftrightarrow \text{H}$ apply (meaning, we expect larynx raising will probably occur with the H tone): modal voice is fully compatible with this state.

Returning to the distribution of epilaryngeal vibration in Bai, we note that harsh voice quality is consistent with epilaryngeal constriction, which cross-cuts the entire harsh register (hence the name). At the highest tone levels, the $\{\text{epc} \cdots \text{H}\}$ anti-synergy must be overcome, and the result is a state where the lower epilarynx (i.e. the ventricular folds) impinges on the vocal folds, but the upper epilarynx is relatively patent and evidently under considerable tension due to the anteroposterior widening effect of strong cricothyroid contraction. As in Zhenhai, growling/epilaryngeal vibration in Bai is confined to relatively low tones. Unlike in Zhenhai, the Bai case provides further support for $\{\text{epv} \leftrightarrow \text{L}\}$, because whisper and whispery phonation do not occur as variants. It could be said that epilaryngeal vibration and breathiness parallel each other in their occurrence on the mid tone (Chao-digit 3) when it is followed by a low tone (Chao-digits 2 or 1). The tendency for these phonatory qualities to be paired with low tone suggests the broader generalization that non-modal phonation (whether constricted or not) is phonologically associated with low tone. The Bai case also provides an instance of epilaryngeal vibration paired with close front vowel [i] without any compromise to its quality (Edmondson, Esling, Harris, Li, & Lama, 2001). Rose’s (1989) laryngoscopic observations of his own imitations of the Zhenhai growl predicted this possibility when, through experimentation, he discovered that the “extrinsic nature of epiglottalization
must be emphasized: epiglottis was not being pushed back by a (low) back tongue position since it also occurred when I growled a high front vowel” (Rose, 1989, p. 239).

Comparing Zhenhai and !Xóõ, we encounter the possibility for epilaryngeal vibration to partner with vocal fold abduction, which produces a relatively whispery growl. This is a particularly interesting situation since it represents a “love triangle” of synergistic relations. Epilaryngeal constriction is assumed to be synergistic with epilaryngeal vibration \( \{epc \leftrightarrow epv\} \) while growling is presumably synergistic with vocal fold abduction \( \{vfo \leftrightarrow epv\} \). The latter appears to be an aerodynamically based synergy and can be considered parallel to Solé’s (2002, pp. 680–681) assessment that voiceless apical trilling has a broader tolerance for variation in pressure and lingual configuration and does not involve maintenance of simultaneous vocal fold vibration. The conflict arises from \( \{vfo \cdots epc\} \). We are reminded of how anti-synergies are not impossibilities. In Zhenhai, which exhibits whisper or whispery phonatory qualities in contexts not favourable for growling, we might infer that \( \{vfo \cdots epc\} \) is intrinsic to the low F0 region of Yang tones, a gross state which by itself biases low tone (as described above). At the moment, the origin of the partially abducted vocal fold configuration in !Xóõ remains unknown.

A more general comparison of the Zhenhai and Bai tonal-register systems reveals that epilaryngeal constriction can cross-cut tone systems in two ways: across the low tones, as in Zhenhai, or across all tones, as in Bai. Within the constricted set, epilaryngeal vibration is relegated to realization on low tone levels. It is not physiologically impossible for epilaryngeal vibration to occur with high pitch level, but, given the relatively low likelihood of epilaryngeal vibration becoming incorporated into speech
sound systems, the occurrence of epilaryngeal vibration with high tone is expected to be very unlikely.

### 7.2 The epilarynx and the vocal folds

At the core of the LPP is the idea that the epilarynx is the physiological coupling between the vocal folds and the rest of the supralaryngeal vocal tract, and this relationship is expressed in diverse ways in the phonological possibilities and patterning of lower vocal tract sounds. This section focuses on the interaction between the vocal folds and the epilarynx.

At stake here is the conceptualization of glottal stop and the “phonological glottis”. In the glotto-centric view, this “glottis” is fully independent of the rest of the vocal tract and corresponds with various vocal fold states. Glottal stop, glottalization, laryngealization, and related sounds are often framed in terms of a [constricted glottis] feature (Halle & Stevens, 1971), which dictates how these sounds will phonologically pattern.

The LPP shifts the perspective away from glotto-centricism: instead, it posits that patterns in lower vocal tract phonology must countenance the coupling role of the epilarynx, which underlies interactions that occur between vocal fold states and those of the rest of the supralaryngeal vocal tract. While the LPP is not a model of phonological features (which are interpreted as part of the system of cognitive abstractions a speaker forms over the phonological data in their language) the proposals made by the LPP conflict with previous feature-based accounts. Thus, the explanatory value of the feature-based approach is addressed in each case.
It has been recognized (Lloret, 1995; Clements, 2003) that the [constricted glottis] feature is problematic in expressing the phonological relationships amongst glottal stop, ejectives and implosive oral stops. Moreover, [constricted glottis] has also been tentatively extended to the representation of pharyngeal consonants by some researchers (Halle, 1995, p. 18; Paradis & LaCharité, 2001, pp. 285–286).

The proposal here is that phonological “glottal constriction” is more complex than the feature [constricted glottis] implies. The nature of the epilarynx gives rise to several phonological potentials exposing some unexpected properties of laryngeals and their relation to pharyngeals and other sounds: multiple “glottal stops” (§7.2.1); glottal stop interaction with /ʕ/ in Tigre (§7.2.2); complex hypopharyngeal allophony in Amis (§7.2.3); synergistic interaction with sounds involving varying degrees of lingual retraction and larynx raising, as illustrated in the classic phonological problem of pharyngeal genesis in Southern Wakashan languages §7.2.4; and susceptibility of relatively open vowels to “glottal constriction” (§7.2.5). In §7.3, the discussion will then turn to the second part of the phonological nature of the epilarynx as a coupling mechanism by examining vowel quality patterns in connection with epilaryngeal stricture.

### 7.2.1 Strong and weak glottal stops

This section describes systems which seem to have two glottal stops, a strong and a weak one. The phenomenon is connected to incremental engagement of epilaryngeal stricture. The implication is that the physiological proximity of “strong” glottal stops to the epilaryngeal pivot in the continuum of hypopharyngeal stops (see §6.3.2, Figure 6.10) makes them candidates for potential reinterpretation as pharyngeal sounds if a two-way contrast is to be maintained.
Phonetically, a glottal stop should be silent. The vocal folds adduct strongly enough to stabilize the vocal folds, preventing their oscillation or causing oscillation to cease entirely. This task is apparently not a simple matter since sounds phonologically analyzed as glottal stops are known to be phonetically realized with low intensity, irregular phonation, i.e. creakiness (Ladefoged & Maddieson, 1996, pp. 73–77).

Typically, if a language has a glottal stop phoneme, it only has one, but systems with two contrastive glottal stops have been attested. For convenience, I will apply the deliberately vague terms weak and strong, or /?w/ and /?/, to refer to the two varieties, following Catford (1977b, p. 289). Speaking generally, the strong glottal stop is usually silent, while the weak one tends to occur with creakiness or a lowering of F0.

There are several examples in the literature of languages with two glottal stop phonemes. Halle and Stevens (1971, pp. 57–58) suggest Jingpo possesses two types of glottal stop, which they analyze as combinations of their [+constricted glottis] feature with either [+stiff] (/ʔ/) or [+slack] (/ʔw/). When post-vocalic, the [+stiff] one causes a slight rise in F0, while the [+slack] one causes lowering of F0194. Ladefoged & Maddieson (1996, pp. 76–77) discuss Gimi, a language that “behave[s] as if [it] had contrasting voiced and voiceless glottal stops” (p. 76). Phonetically, the “voiceless” /ʔ/ is realized as silence, suggesting tight glottal closure, while the “voiced” /ʔ/195 is realized as creaky phonation on nearby vowels, suggesting weakened glottal closure. Evidence from neighboring languages indicates that these sounds originated from debuccalization of /k/ and /g/, respectively. Both sounds exhibit uniform phonological behaviour, each capable

194 Halle & Stevens (1971, pp. 57–58) do not make any statement regarding the effect these sounds have on phonatory quality (i.e. whether creakiness occurs or not).
195 Ladefoged & Maddieson (1996, pp. 76–77) symbolize this sound using an asterisk, ‹*›.
of triggering a de-nasalization process when they precede the nasals /m, n/.

Catford (1977b, p. 289) describes several Northeast Caucasian languages, namely the Nakh languages, and Tsez and Dargi, which contrast a weak glottal stop ؆, often realized simply as creak, with a strong glottal stop ؆ ؆, which involves tight closure of the ventricular bands (as well as the vocal cords) and some constriction of the pharynx. This is sometimes called a ‘pharyngeal stop’ in the literature, but is perhaps better described as a (pharyngeal) ventricular + glottal stop.

In terms of production, x-ray evidence for Dargi shows that neither of these sounds involve complete linguo-epiglottopharyngeal constriction (Gaprindašhvili, 1966; also see Hess, 1998, p. 14), but the “strong” glottal stop does appear to have more tongue retraction, which suggests it could be more than a “glottal” stop – it could possibly be an aryepiglottopharyngeal stop.

In parallel to the weak-strong glottal stop contrast, Quianviní Zapotec contrasts creaky vowels and interrupted vowels (see Table 7.6). Chávez-Peón (2010, p. 252) argues that the “interrupted” part of interrupted vowels cannot be attributed to a sequence of vowel and glottal stop: the interruption is regarded as intrinsic to the vowel.

<table>
<thead>
<tr>
<th>tone</th>
<th>creaky</th>
<th>interrupted</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>/tseɪɲ/ ‘fifteen’</td>
<td>/tseʔɪɲ/ ‘thirteen’</td>
</tr>
<tr>
<td>falling</td>
<td>/ɡa/   ‘nine’</td>
<td>/ɡaʔ/ ‘green’</td>
</tr>
<tr>
<td>falling</td>
<td>/ɡa/   ‘nine’</td>
<td>/ɡaʔ/ ‘green’</td>
</tr>
</tbody>
</table>

(based on Chávez-Peón, 2010, p. 236)

The analytical variability of glottal stop specification for the [±consonantal] feature ([+consonantal]; see Shaw, 1991; Davidson, 2002; cf. [−consonantal]: van Eijk, 1997; Stevens & Keyser, 1989, pp. 90, 93; Halle, 1995, p. 7) may very well reflect the
isolated occurrence of strong and weak glottal stops across languages. Halle & Stevens (1971, p. 59) also touch on this possibility at the very end of their note on laryngeal features, but they describe the difference in terms of [±syllabic]: for them, [−syllabic] entails that “the false vocal cords form a partial or complete constriction in addition to that formed by the true vocal cords, whereas this is not the case for the syllabic [segments]”.

Part of the variation in representation may arise from the difficulty in applying features like [±consonantal] and [±sonorant] to laryngeal segments. In SPE (Chomsky & Halle, 1968, p. 302), [+consonantal] entails “radical” (extreme) mid-sagittal constriction, which is not obviously applicable to the constriction involved in laryngeal sounds: thus these sounds are often considered “glottal glides”. In another definition, [+consonantal] requires supralaryngeal constriction greater than that found in vowels (Miller, 2012, p. 22). This implies that no epilaryngeal stop closure qualifies as [+consonantal], which has led some to outright dismiss the phonological possibility of pharyngeal stops (Halle, 1995, p. 7); from the perspective of this dissertation, this is highly dubious. Bessell (1992, p. 333) argues that treating laryngeals as [−consonantal] absurdly implies that glottal stops are more vowel-like than vowels in terms of oral-cavity constriction degree.

Since features are assumed in this work to be emergent, experience-based, speaker-specific abstractions, it cannot be guaranteed that they take the same shape from one speaker to the next. This means that the possibility of strong and weak glottal stop materializing as a contrast in a given language can be universally analyzed with any particular feature. The best analysis on a language-by-language, or speaker-by-speaker basis will probably reflect the diachronic residue that resulted in the contrast in the first
place. For example, perhaps [±voice] best explains Gimi’s so-called “voiced” and “voiceless” glottal stops which are reflexes of a velar stop voicing contrast. The argument here is that this will not necessarily apply to the behaviour of the strong-weak glottal stop contrast in, say, Dargi, Jingpho, or the parallel vocalic contrast in Quianviní Zapotec.

The fact remains, however, that the potential exists for phonologies to possess more than one hypopharyngeal stop: in the real minds of real speakers in real languages, the abstractions formed across phonemic categories will differ. So what is /ʔ/ in one language may not function the same way as /ʔ/ in another language. The goal is to understand – at the universal scale – what enables all of the specific languages described above, and other languages past and future, to possibly settle at a point in their evolution into the steady state of having multiple glottal stop phonemes. We know that it is possible for none, one, or two hypopharyngeal stops\textsuperscript{196} to be realized in the phonological system of a given language.

Prima facie, it seems reasonable to posit that systems with one hypopharyngeal stop will be more common than those with two or more. However, a deeper explanation must take into consideration the detailed biomechanical and perceptual aspects differentiating the instantiations of the different hypopharyngeal stop potentials.

\textsuperscript{196} No phonological system of three or more hypopharyngeal stops has ever been attested. In the LPP, a system of up to four hypopharyngeal stops is theoretically possible, but for perceptual reasons, such systems are highly unlikely as the cues distinguishing the different stops become much more subtle.
Figure 7.1: Potential hypopharyngeal stop systems in the LPP. Shading tone: lighter = weaker association (less likely to occur). Abbreviations: vfc = vocal fold closure/adduction; epc = epilaryngeal constriction; ↑lx = raised larynx; tre = tongue retraction (down and back). Devices: ⟦ʔ⟧ = vocal fold closure; ⟦ʔ⟧ = vocal-ventricular fold contact and adduction (lower epilaryngeal/LE) closure; ⟦ʔ⟧ = aryepiglottic-epiglottal (upper epilaryngeal/UE) closure; ⟦ʔ⟧ = linguo-epiglottic-pharyngeal closure (see Figure 6.10). Epilaryngeal pivot marks division between lower and upper epilaryngeal zones.

The LPP model predicts that a “basic” system with only one hypopharyngeal stop (Figure 7.1a) will favour the less stable, but simpler-production end of the hypopharyngeal stop series, i.e. ⟦ʔ⟧ or ⟦ʔ⟧ (represented with dark association lines). The potential phoneme /ʔ/ is neither strong nor weak and individual variation has the potential to freely range according to factors such as speech rate, individual physiology, sociophonetic factors, and so forth. As far as phonetic realization goes, some speakers may produce ⟦ʔ⟧ or – perhaps, when time permits – even ⟦ʔ⟧ but they may be found to produce ⟦ʔ⟧ at faster speaking rates. Some may creak when the sound is intervocalic. The
list could go on, but hopefully the point is made: actual speakers of actual languages producing these sounds in actual contexts will (predictably) vary. The phoneme is very unlikely to be associated with \[\tilde{\gamma}\] (light gray association lines) as this mechanism involves extreme closure of the hypopharynx.

The next phonological grade are cases which permit two hypopharyngeal stops, a strong/weak system (Figure 7.1b), such as those discussed above. Now there are several options for division of the potentials on offer. The /\tilde{\gamma}/ is more likely to be associated with \[\gamma\] but may sometimes involve \[\tilde{\gamma}\], while the /\tilde{\gamma}/ is more likely to be found associated with \[\gamma\] but may cross the epilaryngeal pivot by associating with \[\tilde{\gamma}\]. Thus, the LPP predicts that in strong/weak glottal stop systems, there is the potential for the strong member to “flirt” with “pharyngeal” (upper epilaryngeal) place more than those systems with a single glottal stop, which could have consequences for language change. As in the one-hypopharyngeal stop system (Figure 7.1a), \[\gamma\] is still less likely than \[\tilde{\gamma}\] to be associated with the phoneme, but there is a greater possibility (represented with darker association lines) of this in the two-hypopharyngeal stop system than in the one-hypopharyngeal stop system (and in §7.2.1, we will look at Amis, where this potential is actually realized).

Here we must be careful to distinguish between hypopharyngeal stop potential and actual realizations that involve creakiness. First, in any of the stop-like configurations there should also exist the potential to counteract the vibrational inhibition of tissue contact inherent to the stop potential with adjustments that enable phonation. In all the cases described above, there is a re-occurring tendency for the “weak” glottal stops to be realized with phonation, often creakiness or lowered F0, while the “strong” glottal stops
are more likely to be realized as a true stops – silence. Phonation is a physiological possibility for realization of any of the potentials\(^\text{197}\), but the stronger stops with greater reinforcement from supraglottal laryngeal closure are less likely to have vibration.

The goal here, however, is not to predict these phonetic details. It is to adumbrate, based on knowledge of what lies above the vocal folds, the universal possibilities within the lower vocal tract sound system associated with stop-like phonemic contrasts. One of the key properties of Figure 7.1b is that /ʔ/ is treading into the territory associated with “pharyngeal” or aryepiglottal-epiglottal stops [ʔ], but “balances” on the pivot. The implication is that these stops will not be phonologically stable, and will either graduate towards the “pharyngeal” side (the side with [ʔ], i.e. the side characterized by upper epilaryngeal stricture) of the epilaryngeal pivot or collapse back into a single hypopharyngeal stop system. In §7.2.2, the epilaryngeal pivot in the hypopharyngeal stop continuum is examined in the context of Tigre, where “pharyngeal” /ʔ/ and /ʕ/ interact according to the former possibility: the glottal stop becomes pharyngeal.

7.2.2 Laryngeal-pharyngeal interaction in Tigre

This section shows how through synergistic interaction, /ʔ/ has the potential to become a pharyngeal to the exclusion of /h/, which reflects the fact that /ʔ/ and potentially /ʕ/ lie along the continuum of hypopharyngeal stops.

Knowledge of how the epilarynx works in the LPP would lead us to expect two things about lower vocal tract sounds: first, there should be a special affinity between /ʔ/

\(^{197}\) With the [ʔ], realizations may involve tense sounding phonation; with [ʔ] we are more likely to encounter creakiness; with [ʔ] realizations, phonation may also be creaky but additional voice quality changes will start to occur, depending on the prevailing conditions, such as subglottal pressure, and the speaker. Thus with [ʔ], raised larynx voice quality or even harshness with growling increase in likelihood of occurrence in (phonetic) reality.
and /ʕ/; second, it should not be surprising if we encounter patterns involving interaction between ejectives and /ʔ/ and /ʕ/ and other lower vocal tract sounds.

Regarding the first expectation, it is fairly well-established and recognized now (Bin-Muqbil, 2006, p. 234; Hess, 1998, p. 19; Davidson, 2002, pp. 9–10), that /ʕ/, which is traditionally classified as a pharyngeal fricative, tends towards approximant or even stop realizations and that the main site of stricture is aryepiglottal (see §2.2.3 and Chapter 4). Furthermore, Heselwood (2007, pp. 18, 24–25) argues that even some approximant varieties (such as the “tight ’ayn” he describes in several dialects of Arabic) are actually perceived as stops. The same stop impression is given by the voiced aryepiglottic trill in Iraqi Arabic (Hassan, Esling, Moisik, & Crevier-Buchman, 2011) (see §3.2), which shares much in common with the sound Heselwood describes.

Given the often overlapping phonetic identity of /ʔ/ and /ʕ/ – a key assumption in the LPP which is substantiated by phonetic evidence – we might expect occasional, exclusive interaction between these sounds. Prunet (1996, p. 191) demonstrates that Proto-Ethiopian Semitic *ʕ and *ʔ merged as /ʕ/ in Inor, which is described as having “glottal closure” and /a/ characteristics (p. 192). Paradis & LaCharité (2001, pp. 285–286; cf. Halle, 1995, p. 18) recognize this possibility and attempt to account for structural similarity between /ʔ/ and /ʕ/ in their analysis of adaptations of post-velars in loan-word phonology. Their approach is to treat /ʕ/ as specified with a [+constricted glottis] feature, which allows them to account for conversions between glottal stop and pharyngeal stop in various adaptation processes. One example is found in Fula, which converts Arabic /h/ and /ʕ/ to /ʔ/ and /ɦ/ respectively: they argue that this conversion is caused by non-availability of the [RTR] feature in Fula. The reverse adaptation occurs in Afar, which
lacks /hʔ/ but has /ḥʕ/ and adapts Arabic /hʔ/ to the native /ḥʕ/; their analysis, thus, is that this is possible because [RTR] is an available primitive in Afar. Importantly, they posit that both /ʔ/ and /ʕ/ share [+constricted glottis] and minimally differ by [RTR], allowing them to map onto each other in cases of borrowing.

The second expectation also follows from the assumption that ejectives in some languages will exhibit epilaryngeal stricture in producing the ejection: ejectives (can) involve larynx raising and {↑lx ↔ epc}. Furthermore, given {tre ↔ epc}, we might also expect tongue retraction to become involved, and some phonologies may incorporate this synergistic relationship. Such an expectation is borne out in Tigre and Harsusi, where ejectives do in fact modulate vowel quality: for example, /k’oraː/ is realized as [k’orːa:] ‘frog’ in Tigre and /at’eːbex/ becomes [at’aːbex] in Harsusi (Rose, 1996, pp. 93, 97).

Rose notes that such data upsets her initial analysis that ejectives do not pattern as gutturals (1996, p. 81); thus, she suggests that, on the basis that ejectives in these languages are cognate with Arabic emphatics (1996, p. 94), ejectives in these languages are additionally specified as [+RTR] to account for the lowering. Other researchers have also made note of this connection between ejectives and emphatics (Fre Woldu, 1986; Bessell, 1992, pp. 322–323; Fallon, 1994; Lloret, 1995, p. 265; Watson & Bellem, 2009).

GCLC models do not incorporate a link between the vocal folds and the tongue, and consequently patterns involving connections, say, between pharyngeals and ejectives, and between ejectives and emphatics, are somewhat surprising from such a perspective: these models do not inherently predict that there should be interactions of this sort. The solution in such models is to apply remedies such as exceptional specification of [+RTR]

---

198 Ejectives more often tend not to interact with vowels, or other “glottalized” or “glottal” sounds for that matter; these issues will be addressed further in §7.2.5 and §7.3.
on ejectives to account for their lowering of vowels or a [+constricted glottis] feature on /ʕ/ to account for the phonological relationship of this with glottal stop. These remedies reflect the need for acknowledging the phonological reality of the lingual-laryngeal connection embodied in the LPP model, but they do not adequately capture all of the details of the relationship.

A good example comes from Tigre laryngeal-pharyngeal neutralization data, which has appeared in numerous places in the literature but which has never been given a full analysis\(^{199}\). These data are important in illustrating a potential asymmetry within laryngeal and pharyngeal sound systems: the different behaviour of /h/ and /ʔ/. Specifically, the epilarynx is the basis of an intimate relationship between /ʔ/ with other gutturals – but most importantly /ʕ/ – that does not exist as such between /h/ and the other gutturals. It will be argued that the LPP can readily explain this as somewhat unsurprising potential behaviour of the epilarynx.

Tigre, a language which contrasts /h h ? ʕ/ and ejective consonants (Rose, 1996, p. 92), has an optional process that neutralizes the contrast between /ʔ/ and /ʕ/ in the presence of pharyngeals and/or ejectives anywhere else in the word. For example, /ʔaddaha/ ‘noon’ is variably realized as [ʕaddaha] or [ʔaddaha] (Raz, 1983, p. 5; Hayward & Hayward, 1989, p. 181; McCarthy, 1994, p. 224). Critically, /h/ and /ʔ/ do not show neutralization under the same conditions. The significance of this is that /ʔ/ and /ʕ/ share a degree of phonological similarity not shared by their voiceless counterparts (i.e. /h/ and /h/).

\(^{199}\) Although an attempt has been made using Revised Articulator Theory (Moisik, Czykowska-Higgins, & Esling, 2012).
Even though /h/ and /ʔ/ are associated with their own laryngeal features, [+spread glottis] and [+constricted glottis] respectively, and even though one can analyze /ʕ/ and ejectives as bearing [+constricted glottis], the asymmetry cannot be accounted for in a GCLC model. In such models, a plausible analysis for the /ʔ ʕ/ neutralization would be that [RTR] optionally spreads onto /ʔ/, since /h/ is also a trigger. However, nothing prevents [RTR] from optionally spreading onto /h/ as well. An example framed using Rose’s (1996) model is in Figure 7.2.

Figure 7.2: Failed GCLC analysis of Tigre laryngeal-pharyngeal neutralization. [RTR] is incorrectly predicted to spread to /h/.

Perhaps one might propose that /h/ does not have a PHARYNGEAL node in the Tigre language, but /ʔ/ does. In such a case, /h/ could not receive [RTR]. Such an analysis, however, goes against Rose’s assumption of the Node Activation Condition (Rice & Avery, 1993): /h/ must bear PHARYNGEAL on account of its minimal contrast with /h/ in terms of [RTR], the dependent of the PHARYNGEAL node. In fact, if we accept the argument by Paradis & LaCharité (2001), then laryngeals always bear PHARYNGEAL. Furthermore, an appeal cannot be made to stricture features like [±continuant], [±consonantal], or [±sonorant]. In orthodox phonology (for an overview, see Bessell,
1992, Chapter 7), regardless of the specification for [±consonantal] and [±sonorant], /ʔ/ and /h/ are assumed to share that specification. Bessell (1992, pp. 348–367) reviews evidence that /h/ could be analyzed as [+continuant] and /ʔ/ as [−continuant] (which is assumed in SPE; Chomsky & Halle, 1968, p. 303). Despite this, none of these stricture features adequately characterize the generalization about the Tigre neutralization: namely, that guttural consonants (i.e. /ħ C’/) are exclusively targeting /ʔ/. The stricture features do not represent the guttural natural class, but PHARYNGEAL and [RTR] do. There is no available means of stating why /ʔ/ is more “guttural” than /h/ in GCLC models in these models. However, there is an explanation if we consider the role of the epilarynx.\footnote{In an Emergent Features framework, we could freely posit a representation which allows us to tease apart /h/ and /ʔ/ and which represents /ʔ/ as sufficiently more similar to gutturals than /h/ such that this pattern could be “explained”. The novel representations could be justified by saying that, as is to be expected with the framework of Emergent Features, these representations are language specific. Features arise from the collective similarity of speaker-specific emergent representations. However, this does not help us understand why these types of interactions occur in the first place, since features emerge in a post-hoc fashion: meaning that this pattern exists prior to the emergence of any representations in the minds of speakers. The problem becomes a matter of “chicken and the egg” causality, or, perhaps “Chicken Little” causality, if one adopts Ohala’s (2011) view of explanations in OT and conventional theoretical phonology.}
Figure 7.3: Potential Tigre sub-systems showing relations defining /ʔ ʕ/ similarity vis-à-vis /h h/. Only relevant associations are depicted, especially in (c). Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; ↑lx/↓lx = raised/lowered larynx; tre = tongue retraction (down and back). Devices: ⟦h⟧ = vocal fold abduction; ⟦h⟧ = aryepiglottic-epiglottal (upper epilaryngeal/UE) stricture with abducted vocal folds; ⟦ʔ⟧, ⟦ʔ⟧, and ⟦ʔ⟧ = see Figure 6.10. Epilaryngeal pivot marks division between lower and upper epilaryngeal zones.

In the LPP, we can state a potential scenario (Figure 7.3) that gives insight into why the Tigre pattern is happening: the LPP predicts that /ʔ/ is more likely to interact with the /ʔ h C'/ “gutturals” than /h/\(^{201}\). To understand what is going on, we must make two assumptions. First, we must accept that the /ʔ/ may range over ⟦ʔ⟧ and ⟦ʔ⟧, a prediction that could be empirically verified. We must also accept that /ʕ/ may involve ⟦ʔ⟧; no evidence for this exists in Tigre, but for the closely related language, Tigrinya, we

\(^{201}\) In §7.4 we will explain why /h/ might pattern with gutturals in other cases, but this goes beyond the scope of the role of the epilarynx in phonology.
have laryngoscopic evidence that /ʔ/ is produced (at least by one speaker) as an aryepiglottio-epiglottal stop [ʔ] (Esling, 2003; Esling & Harris, 2003b). Thus, the assumptions are reasonable, and the account rests on these assumptions.

Now onto the explanation: in the LPP, since glottal abduction effectively opens the lower half of the larynx and inhibits lower epilaryngeal closure (i.e. closure in the ventricular plane), we posit that \{vfo \cdots epc\} (see Figure 7.3a). Anti-synergistic relations do not mean that it is impossible for these gross states to combine: they do so in [h]; so anti-synergistic relations can be overcome, but some compromise in state will be required\(^{202}\). Furthermore, according to the LPP, if anything, we predict abducted vocal fold states synergize with larynx lowering (indirect association of \{vfo\} with [h]) or \{↓lx \leftrightarrow vfo\}, but, crucially, larynx lowering is anti-synergistic with epilaryngeal constriction, \{↓lx \cdots epc\}. The result is that in the hypopharyngeal fricative system, there is greater opposition between /h/ and other gutturals, like /ʔ/, for reasons completely unrelated to the mental representation of sounds.

Glottal stop /ʔ/ (see Figure 7.3b), on the other hand, is far more similar to the other “gutturals” in terms of its gross states and synergistic relations. Like the pharyngeals, it is synergistic with larynx raising \{↑lx \leftrightarrow vfc\}, and critically, in its “strong” form [ʔ] it engages lower epilaryngeal stricture. We would predict, based on this state of affairs, that in languages possessing a set of sounds like that in Tigre, it would not be surprising to find strong affinity between glottal stop and the other gutturals, stronger than the affinity between /h/ and the gutturals, and this affinity materializes in the Tigre neutralization. It is not surprising that glottal stop neutralizes

\(^{202}\) This very anti-synergy may in part explain why [h] (e.g. in Nuuchahnulth) is attested more than [ʔ], which has only been observed in Amis (Edmondson, Esling, Harris, & Huang, 2005).
specifically with /“ʕ”/ (i.e. [ʔ]): glottal stop in raised larynx configuration effectively becomes [ʔ] (Esling, 1999). We can go further by predicting that, when neutralization occurs in Tigre, an [ʔ] occurs instead of a glottal stop.

The fact that ejectives participate in this pattern is a little more surprising since, as will be discussed in §7.2.4, ejectives do not necessarily involve epilaryngeal stricture. As it has already been established above, however, Tigre ejectives are no ordinary ejectives: they share something in common with emphatics – something which interferes with vowel quality. It could be that ejectives in Tigre, like in other southern Semitic languages, have taken on epilaryngeal stricture. The use of /Cʔ/ in Figure 7.3c symbolizes this possibility. The influence on vowel quality would then follow from the mutual synergy between epilaryngeal stricture and tongue retraction {epc ↔ tre}. The participation of ejectives in this pattern would also be increased in likelihood. Tigre evidently instantiates this potential.

All of these considerations above demonstrate that the laryngeals, /h/ and /ʔ/, are not a cohesive set, and their differences, especially in regard to their relationship to other gutturals, is not adequately explained or even (correctly) predicted by traditional accounts using features like [+constricted glottis] or [RTR].

7.2.3 Hypopharyngeal allophony in Amis

This section demonstrates that the epilarynx is a key component of a phonological account of the complex set of hypopharyngeal stop and fricative allophones that are found in Amis. This section also illustrates the asymmetry between hypopharyngeal stops, which fall along a continuum, and fricatives, which do not because vocal fold abduction is anti-synergistic with epilaryngeal stricture.
In §7.2.1, a strong-weak contrast was identified and interpreted as instantiating two contrasts along the hypopharyngeal stop continuum. The possibility of languages exhibiting hypopharyngeal stops at the extreme end of the complexity-stability scale (see §6.3.2) was not addressed. Then, in §7.2.2, a special relationship was suggested to exist between /ʔ/ and /ʕ/ which involved the potentially stop-like nature of both sounds and their intimate dependence on the epilarynx. In this section, we now examine the upper epilaryngeal and lower pharyngeal potential regions of the hypopharyngeal stop set as they materialize in Amis.

At first blush, Amis instantiates a contrast resembling the strong-weak glottal stop contrast potentials discussed in §7.2.1. Earlier research even posited this to be true for Amis, suggesting a contrast of “heavy and light glottal stops” (Fey, 1986; see Maddieson & Wright, 1995, p. 49). Such contrasts, however, can exist entirely on the vocal-fold–lower-epilaryngeal (VF-LE) side of the epilaryngeal pivot (see §6.3.2, especially Figure 6.10). In some cases, they may “straddle” the pivot (i.e. one sound is variably realized using [ʔ], which is lower epilaryngeal/LE, or [ʔ], which is upper epilaryngeal/UE, biomechanical mechanisms), and in such cases the strong member will start to resemble a pharyngeal (which is very likely the case for Dargi; see §7.2.1). The LPP predicts the existence of languages that realize a hypopharyngeal contrast that straddles the epilaryngeal pivot in its behaviour: Amis is a case in point, but the straddling applies to both /ʔ/ and /h/. Furthermore, the Amis case illustrates very clearly the difference between stricture at the upper epilarynx level and stricture at the lower pharyngeal level. The difference is between epilaryngeal tube closure in [ʔ] and pharyngeal tube closure in [ʔ], which stacks on top of the [ʔ] (as far as we know).
This difference is essential because it is a difference that GCLC models cannot make. Consider the following: if the vowel /a/ is, by means of [−ATR] specification, distinguished from /æ/, an analysis suggested by Halle et al. (2000, p. 408 f.n. 9), and if we assume that [±ATR] is a vowel feature and [±RTR], a consonantal feature, as Rose does (1996, pp. 89–90), it is entirely reasonable to predict that the phonetic realization of the [+RTR] feature of pharyngeal consonants should correspond with narrower pharyngeal stricture than the stricture entailed by the [−ATR] feature associated with /a/. We would never expect /a/ to involve narrower pharyngeal stricture than pharyngeal consonants: this would imply that /a/ is more consonant-like than pharyngeal consonants themselves. Yet this is implicit in GCLC models because pharyngeals consonants are (typically) [+RTR] and thus should involve greater tongue retraction than the vocalic [−ATR] stricture associated with /a/.

The problem is that GCLC models assume that tongue root retraction is the key feature of pharyngeal consonants. The real difference is expressed in the two-tube model of the lower vocal tract: “pharyngeals” (which really ought to be called “epilaryngeals”) fundamentally involve constriction of the epilarynx (see Figure 6.9); vowels like /a/ involve constriction of the pharynx (see Figure 6.7). From this perspective, it is entirely possible for the pharyngeal constriction in [a] to be narrower than the pharyngeal constriction in [?]. In other words, GCLC models and the LPP model make predictions of a different nature. In GCLC models, pharyngeals should always be produced with tongue root retraction in excess of what occurs for /a/; in the LPP, pharyngeals can vary considerably in their degree of tongue retraction, which makes it possible for pharyngeals
to allophonically vary in regard to tongue retraction – this prediction is supported by data from Amis.

Amis provides a unique window into viewing this difference in the context of prosodically conditioned allophony of its pharyngeal consonants, /h ñ/ (Edmondson, Esling, Harris, & Huang, 2005): in syllable initial context, the sounds are produced as [h ñ] (aryepiglottal constriction – stricture of the upper epilarynx); in word final position\(^{203}\), however, these sounds show additional tongue retraction resulting in linguo-epiglottal-epiglottial constriction (lower pharyngeal constriction) in addition to the aryepiglottal constriction below, giving [h ñ]\(^{204}\). Thus, [h ñ] have strong pharyngeal tube stricture on top of their inherent epilaryngeal tube stricture.

The pattern is illustrated in Table 7.7. Examples (a) to (d) have onset pharyngeals that are aryepiglottal, while examples (e) to (h) have word final pharyngeals produced with additional linguo-epiglottal constriction (in addition to primary aryepiglottal constriction). Note that Amis has final stress (Maddieson & Wright, 1995, p. 47).

\(^{203}\) There are no examples of this sound occurring in non-final coda position. I follow the analysis in Edmondson, Esling, Harris, & Huang (2005) by assuming that intervocalic occurrences are analyzable as onsets and the pharyngeals are realized as aryepiglottal in this environment. Until further evidence can be adduced, I assume that the relevant distribution of these two productions is based on whether the sound occurs in word final position. The exact details are not as important as the fact that the productions are prosodically conditioned with different phonetic realizations.

\(^{204}\) Edmondson, Esling, Harris, & Huang (2005) represent the aryepiglottal-epiglottal with additional epiglottal-pharyngeal constriction as [ð] and [œ]. I choose to stray from their representation for simplicity and because the use of the retracted tongue root diacritic [ , ] helps to draw attention to the key phonetic difference between these sounds while emphasizing the primacy of the epilaryngeal constriction, which is represented by the symbols [?] and [h].
Table 7.7: Hypopharyngeal allophony in Amis (Edmondson, Esling, Harris, & Huang, 2005)

<table>
<thead>
<tr>
<th>Pharyngeals in Onsets</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) /tʰap/</td>
<td>→</td>
<td>[ʔʰapʰ]</td>
<td>‘betelnut’</td>
<td>p. 4</td>
</tr>
<tr>
<td>b) /ikoŋ/</td>
<td>→</td>
<td>[ʔi.‘koŋ]</td>
<td>‘to bend’</td>
<td>p. 6</td>
</tr>
<tr>
<td>c) /poʔot/</td>
<td>→</td>
<td>[po.ʔisʰ]</td>
<td>‘small knife’</td>
<td>p. 4</td>
</tr>
<tr>
<td>d) /ʔisʔis/</td>
<td>→</td>
<td>[ʔis.‘ʔis]</td>
<td>‘cut hair’</td>
<td>p. c.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pharyngeals in Word Final Position</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e) /rumaʔ/</td>
<td>→</td>
<td>[ru.‘mqʔ]</td>
<td>‘house’</td>
<td>p. 4</td>
</tr>
<tr>
<td>f) /rirʔ/</td>
<td>→</td>
<td>[ri.‘riʔ]</td>
<td>‘grasshopper’</td>
<td>p. 7</td>
</tr>
<tr>
<td>g) /tihiʔ/</td>
<td>→</td>
<td>[ti.‘hiʔ]</td>
<td>‘accompany’</td>
<td>p. 10</td>
</tr>
<tr>
<td>h) /felih/</td>
<td>→</td>
<td>[fe.‘lihh]</td>
<td>‘turn over’</td>
<td>p. 11</td>
</tr>
</tbody>
</table>

This pattern is independent of vowel quality: the coda pharyngeal becomes “retracted” regardless of whether the preceding vowel is /i/ or /a/. Thus, the data cannot be simply analyzed by positing that the additional retraction depends on the quality of the preceding vowel.

The challenge for GCLC models, which typically posit that pharyngeals are [RTR] by default, is to explain why there is extra tongue retraction on pharyngeals following vowels. We could assume this pattern stems from vocalic dorsality – vowels are inherently DORSAL/[dorsal] (Halle, 1995; Halle, Vaux, & Wolfe, 2000; Howe, 2004; Flynn, to appear). Support for vocalic dorsality comes from the cross-linguistic tendency for coda velarization: examples include coda consonants becoming dorsals, velar insertion, and coda velarization, and so forth (for details, see Howe, 2004). Such an analysis relies on the assumption of “intensified” phonotactics in syllable rhymes (e.g. Selkirk, 1982; Mohanan, 1993; Howe, 2004; Flynn, to appear) meaning that vowels and coda consonants are more restricted in their phonological content than onsets are with
respect to rhymes. The analysis would be that word-final pharyngeals receive \textit{DORSAL}/[dorsal] from the preceding, tautosyllabic vowel\textsuperscript{205}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure7_4.png}
\caption{Amis pharyngeal dorsalization in a GCLC analysis (a); the representation of guttural-uvulars (b) in Rose (1996).}
\end{figure}

Figure 7.4 illustrates this using Rose’s (1996) model as representative of the GCLC approach. An immediate problem lies in the fact that Rose’s model has two types of \textit{DORSAL}: \textit{PHARYNGEAL-DORSAL} and \textit{ORAL-DORSAL}. Presumably it is \textit{ORAL-DORSAL} that spreads in Amis (Figure 7.4a) since the pattern is non-vowel-quality specific. We must also assume that the \textit{ORAL-DORSAL} can be received by the \textit{PHARYNGEAL} node. A further issue is whether it is undesirable that the result is phonologically indistinguishable from the guttural-uvulars (i.e. /χʁ/ see §6.2.3 and Figure 7.4b)\textsuperscript{206}. Amis does not have

\textsuperscript{205}No non-word-final examples of such pharyngeals could be located in the data, so we cannot be certain the coda analysis is the best analysis.

\textsuperscript{206}It is often said that phonetics-to-phonology mapping is not “one-to-one”. While this may be true in a conventional sense (e.g. phonemes correspond with multiple allophones, which correspond to tremendous, if not infinite phonetic variation – depending on the level of granularity of the analysis), we should be careful to remember two things: first, “phonetics” and “phonology” are domains of scientific inquiry, which do not necessarily have a “one-to-one” mapping with the reality of language and speech, and second, we should be cautious with such a principle because it allows for non-accountability of phonological theory to new understanding arising from phonetic empiricism. A much better model of phonetics-phonology is one where they are “married”, as Ohala (forthcoming) suggests, and mutually inform each other. For example, “phonologists” have identified the guttural/post-velar class groupings and should welcome plausible arguments from “phoneticians” for why these groupings might occur, rather than fitting phonetically outdated phonological models to new data with procrustean determination.
uvulars, but /h/ and /ʔ/ correspond with uvulars in related languages (Maddieson & Wright, 1995, p. 49). We could infer that the Amis sounds were originally uvulars that debuccalized at some point; in such a case, we might wonder why pharyngeal dorsalization does not “undo” the debuccalization, by yielding uvulars instead of linguo-epiglottopharyngeals.

The problem with GCLC approaches – like that illustrated in Figure 7.4 – is that pharyngeals are, in phonological essence, stripped down uvulars; put another way, Halle (1995, p. 18) views uvulars as “pharyngeals with a secondary DORSAL articulation”. These models cannot tease apart uvular and pharyngeal constriction and thus encourage the view that uvulars are inherently pharyngeals but with extra dorsal component. The GCLC models lack the ability to distinguish between the pharyngeal tube and the epilaryngeal tube, which is key in distinguishing between uvulars and pharyngeals. In the LPP, uvulars are not pharyngeals with secondary dorsal articulation and treating them as such in the phonology mischaracterizes the nature of pharyngeals (think “epilaryngeals”) and fails to integrate the lingual-laryngeal connection into phonology.

In GCLC models, “guttural-uvulars” and pharyngeals are only distinguished by the presence of a DORSAL specification (as implied by Figure 7.4). Thus, these models run afoul when faced with the Caucasian languages 207. As illustrated Table 7.8, these languages are not only well endowed with plain guttural/post-velar consonants but also exhibit secondary pharyngealization on uvulars, including “guttural-uvulars” /χʕ/ (Colarusso, 1975, 1985; Catford, 1977a, p. 193, 1977b, 1983; Hess, 1998, pp. 21–31; Carlson & Esling, 2003, p. 187).

207 In her analysis, Rose (Rose, 1996, p. 98) does not mention the pharyngealized uvulars in these Caucasian languages, despite reporting data for Ubykh.
Since in GCLC models, uvulars are essentially already pharyngealized in their representation, pharyngealized uvulars cannot be represented: these models predict pharyngealized uvulars are impossible. Since they are possible, this indicates that something is seriously wrong with the understanding underlying the features used to represent these phonemes because such sounds do exist.

Models with an additional feature, such as [+constricted pharynx], can address this deficiency, by assuming that uvulars are [−constricted pharynx] and pharyngeals are [+constricted pharynx]. As discussed in §6.2, there are several proposals in the literature for an additional feature of this sort, although labels, assumptions, and details vary (Colarusso, 1975, p. 132; Czaykowska-Higgins, 1987; Trigo, 1991; Keyser & Stevens, 1994; Davis, 1995; Hess, 1998).

As discussed above (and in §2.1), in the LPP, the epilarynx is viewed as an independent tube nested within the pharyngeal tube. As depicted in Figure 7.5, the hypopharyngeal phonemic subsystem in Amis is interpreted as “straddling” the

---

208 Bzyb is a dialect of Abkhaz (Northwest Caucasian). The same system is found in the sub-dialect Lyxny (p. 442); Colarusso describes the Apsuy and Kuvin dialects of Ashkharwal (pp. 445-446) and the Tapanta, Kum-Lo, Lo-Kuban, and Dudaruko-Bibard dialects of Abaza (pp. 448-454). Note that Colarusso uses [Č] (superscript bar) diacritic in transcribing pharyngealized consonants (e.g., p. 407).
epilaryngeal pivot for both the fricatives and the stops. According to Maddieson & Wright\(^{209}\) (1995, p. 49) only one hypopharyngeal fricative exists phonemically (simply symbolized in Figure 7.5a as /h\(^{210}\)/), but according to Edmondson et al. (2005) this sound involves three-way allophony such that [h] occurs initially, [ħ] medially, and [ɭ] occurs word finally. The stop subsystem\(^{211}\) (Figure 7.5b) resembles the strong-weak glottal stop contrast (see §7.2.1), but the strong member lies on the other side of the epilaryngeal pivot, and is attested to be an aryepiglottito-epiglottal stop [ʔ] when initial.

\(^{209}\) Maddieson & Wright symbolize the hypopharyngeal fricative with /u/, which for them means an epiglottal fricative. Since, in this dissertation, this symbol is strictly interpreted as a voiceless aryepiglottio-epiglottal trill, the alternate symbol /h/ is used, which entails a voiceless aryepiglottio-epiglottal fricative.

\(^{210}\) Although this has not been formalized in the LPP, the fact that /h/ straddles the epilaryngeal pivot suggests an instability in Amis hypopharyngeal fricative subsystem and could be subject to phonological reorganization in future sound changes. We might predict that the more costly upper epilaryngeal components would be subject to diachronic attrition, leaving just /h/.

\(^{211}\) While Maddieson & Wright (1995, p. 49) assume a contrast between /ʔ/ and /ɭ/, Edmondson et al. (2005) assume that glottal stop is epenthetic on vowel initial and vowel final words.
Edmondson et al. (2005) suggest that the pattern is prosodically determined: final stress causes constriction enhancement on /h/ and /ʔ/ in word final position. The LPP invites a different type of explanation for why [h]/[h] and [ʔ] occur in onsets/medially while [ht] and [ʔt] occur word finally: the former do not compromise vowel quality as much as the latter do (in Figure 7.5a & b, this relative strength of retraction is indicated by the darkness of the association lines: it is gray between [h]/[ʔ] and {tre} and black between [h]/[ʔ] and {tre}). When in onset position, the allophones with less
constriction occur, which reduces the duration required to transition into the vowel. When word final, full linguo-epiglottal-pharyngeal closure can occur without compromising the onset of the vowel. The transitional cues in shifting out of the vowel into the full hypopharyngeal stricture may enhance the consonantal identity in this position.

Obviously such an account begins to enter the territory of perception, prosody, and temporal characteristics of articulation; developing a full account will not be pursued here. For the present, the Amis case is taken as evidence that it is necessary to have the ability to distinguish between pharyngeal and epilaryngeal stricture to make sense of potential allophonic differentiation of these constrictions. Consequently, pharyngeals can be more allophonically complex than a GCLC account would predict. Ignoring the intimate relationship this category of sound has to the larynx mischaracterizes the potentials associated with this part of the vocal tract.

7.2.4 Pharyngeal genesis in Southern Wakashan and Ejectives

In this section, the potential behaviour of the epilarynx is shown to directly predict the proneness gradient found in the Southern Wakashan pharyngeal genesis pattern.

In the LPP, larynx raising and tongue retraction are posited to be synergistic with epilaryngeal engagement. Accordingly, we expect to see sounds which share these states to induce epilaryngeal stricture and even become pharyngeals/“epilaryngeals”. In §7.2.2, some evidence from Tigre demonstrated that ejectives can bias epilaryngeal stricture, but these ejectives were viewed as atypical, being closely related to emphatics found in other Semitic languages. This section presents further evidence from the history of Southern Wakashan that larynx raising and tongue retraction understood as gross states are useful
in characterizing the nature of sound patterns of the lower vocal tract. Specifically, these states each synergize with epilaryngeal stricture and therefore bias it to occur, and when they combine, this bias becomes even stronger. The data also highlight the analytical limitations of GCLC features like [+constricted glottis] and [+RTR] in explaining sound patterns in lower vocal tract phonology.

Pharyngeal genesis is a well-known pattern (Jacobsen, 1969; Colarusso, 1985, p. 367; Trigo, 1991, p. 125; Davidson, 2002, p. 75) in the diachronic phonology of Southern Wakashan languages. The key facts are as follows: Proto Southern Wakashan (PSW) uvular ejectives, *q’ and *q’w’, merge into /ʕ/ in modern day Nuuchahnulth and Nitinaht; its uvular fricatives, *χ and *χw, become /h/ only in Nuuchahnulth; plain uvular stops do not participate in the change; and no changes occurred in the development of the southern-most Wakashan language, Makah. Thus, both changes occurred in Nuuchahnulth (Nootka), but the change only partially occurred in Nitinaht (and Ditidaht), which retains the historical uvular fricatives, while Makah retains both proto-phonemes; these details are summarized in Table 7.9.

Table 7.9: Pharyngeal genesis in the history of Southern Wakashan

<table>
<thead>
<tr>
<th>Proto-sounds</th>
<th>Makah</th>
<th>Nitinaht</th>
<th>Nuuchahnulth</th>
</tr>
</thead>
<tbody>
<tr>
<td>*q’, *q’w’</td>
<td>q’</td>
<td>ʕ</td>
<td>ʕ</td>
</tr>
<tr>
<td>*χ, *χw</td>
<td>χ</td>
<td>χw</td>
<td>h</td>
</tr>
</tbody>
</table>

(Cognates for all three languages can be found in Trigo, 1991, p. 125)

Carlson & Esling (2003) persuasively demonstrate using laryngoscopy that the Nuuchahnulth /ʕ/ is produced as an aryepiglottic-epiglottal stop [ʔ]; this agrees closely with Davidson (2002: 9), who classifies the sound as an “ejective pharyngeal stop”
because it shares phonotactic distribution and phonological patterning with the oral ejectives. Carlson & Esling (2003: 187) also demonstrate that Nuuchahnulth [h] is produced with strong lingual retraction, forming a secondary lingual-epiglottopharyngeal stricture, in addition to its primary epilaryngeal constriction. Finally, for comparison, Carlson and Esling also examine Nlaka’pamux (Thompson River Salish) uvular ejectives (2003, p. 186), which they report exhibit strong epilaryngeal stricture, exceeding what is observed for [t’] in Nuuchahnulth.

The process is also synchronically active in Southern Wakashan (Davidson, 2002; Wilson, 2007; Kim & Pulleyblank, 2009; Werle, 2010). As illustrated in Table 7.10, glottalizing suffixes and clitics trigger mutation of stem final plain uvulars into [ʕ] in Nuuchahnulth and Ditidaht (a dialect closely related to Nitinaht). In these two languages, as (a) and (d) indicate, glottalizing morphemes trigger the mutation of the underlying uvular stop; predictably, Makah (e) does not exhibit any change. Whether a change will occur is dependent on whether the morpheme causes glottalization, as shown in (b) and (c): Kim and Pulleyblank (2009) attribute this to a floating [constricted glottis] feature on the glottalizing morphemes; the non-glottalizing morphemes have this feature anchored to a ROOT node.

Table 7.10: Synchronic activity of Pharyngeal Genesis

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Nuuchahnulth</td>
<td>/ʔatɬiːqʷ-ʔiːs/ → ʔatɬiːʕis</td>
<td>‘consuming forty’</td>
<td>W</td>
</tr>
<tr>
<td>b</td>
<td>Nuuchahnulth</td>
<td>/yaq-yiːq-ʔaɬ/ → yaqyiːʕaɬ</td>
<td>‘which vessel now’</td>
<td>KP</td>
</tr>
<tr>
<td>c</td>
<td>Nuuchahnulth</td>
<td>/yaq-yiːq-ʔitq/ → yaqyiːʕitq</td>
<td>‘which vessel they’</td>
<td>KP</td>
</tr>
<tr>
<td>d</td>
<td>Ditidaht</td>
<td>/ʔajɬiːq=ʔa/ → ʔajɬeːʔa</td>
<td>‘it’s a lot’</td>
<td>W</td>
</tr>
<tr>
<td>e</td>
<td>Makah</td>
<td>/p’uqʷ-ʔɬt/ → p’uʔqʷʔɬt’</td>
<td>‘feather mattress’</td>
<td>W</td>
</tr>
</tbody>
</table>

The question is: can non-epilaryngeal models account for what has happened in the history of Southern Wakashan and continues to happen in the synchronic phonology of Nuuchahnulth and Nitinaht/Ditidaht? A successful account should minimally explain two facts: (1) only uvular ejectives and fricatives undergo change; (2) the change has a graded nature: plain uvular stops never changed, uvular fricatives changed in one language, and uvular ejectives changed in two languages.

Trigo (1991, p. 125) analyzes the sound change, which she refers to as “oral depletion”, by claiming that the dorsality (and labiality) of the sounds /q’ qʷ χ χʷ/ is lost, leaving behind a “pharyngeal residue”. Trigo’s analysis is based upon the assumption that uvulars are dorsals that possess a secondary pharyngeal component as part of their phonological specification (i.e. /q/ = /k^ʕ_212/). In feature geometric terms, oral depletion amounts to deletion of the place node, which results in the loss of primary dorsal and labial specifications and leaves behind the secondary pharyngeal specification inherent to uvulars. Trigo’s account fails because it predicts that all uvulars will change and because it does not provide insight into the gradient nature of the change.

Kim and Pulleyblank (2009, p. 572) offer the explanation, following a proposal by Howe (1996), that spreading of [constricted glottis] onto uvulars causes their representation to become too complex: hence, the change is motivated purely in terms of phonological computation. Their assumption is that uvulars are [dorsal, pharyngeal] and thus simplify to [pharyngeal] when they receive the [constricted glottis] feature. This account is unsatisfying for a number of reasons. First, this does not help to explain the 212 If the pharyngeal component were primary for uvulars, then we would expect ‘pure’ pharyngeals to appear in the inventory of proto-Nuuchahnulth, according to Trigo’s (Trigo, 1991, pp. 126–127) assumptions.
diachronic pattern: it does not account for the uvular fricatives which change despite not being glottalized\textsuperscript{213}. Second, the complexity argument requires an explanation of what constitutes too complex\textsuperscript{214} and an argument for why the grammar seeks to simplify representations, but they do not provide any explanation along these lines. Finally, resorting to a phonological complexity argument gives short shrift to the phonetic details outlined in Carlson and Esling (2003). Adopting the Kim and Pulleyblank explanation means that it is accidental that the phonetics and phonology are putting the same pressure on the sound system to change uvular ejectives, but the causes in each of these domains are totally independent: the assumption that the phonological system is driven to employ the simplest representations is completely unrelated to the nature of lower vocal tract articulation. Thus there is duplication of explanation which raise the question of whether one explanation is redundant (Ohala, 2005a; Blevins, 2004, p. 5; Odden, 2012, p. 6; Miller, 2012; Ohala, 2011).

Other GCLC models of pharyngeals also struggle with accounting for the data. For example, the models proposed by Rose (1996)\textsuperscript{215} and Halle et al. (2000) both predict

\footnotesize
\begin{enumerate}
\item It is doubtful that a recovery analysis could be made by positing that the *χ* *χ" changed because they were historically glottalized, i.e. *χ' *χ"*. First, glottalization is not contrastive for fricatives in any Southern Wakashan language; it only is contrastive for stops/affricates and sonorants (Werle, 2010, p. 3). Second, when subject to glottalizing suffixes, modern day Nuuchahnulth /χ χ"/ change into /w'/ (Davidson, 2002, p. 54). Furthermore, Werle (2010, pp. 4–5) observes that /h/ behaves like any other fricative in Nuuchahnulth, while /ɬ/ closely parallels /h/ in its phonotactic restrictions. Finally, Davidson (2002, pp. 10–11) observes that /ɬ/ continues to pattern with ejectives, which could be interpreted as a reflection of its glottalized heritage. Unless there is evidence to prove that these sounds were glottalized historically, it is assumed that this was not the case and not relevant to explaining the sound change.
\item In Optimality Theoretic terms, the answer would involve assumption of high ranking markedness of [constricted glottis] and [pharyngeal] (or an ad hoc constraint specifically targeting [dorsal, pharyngeal]). Given such markedness constraints, one might wonder why /q + [cg]/ does not become realized as [k'] instead. If it is homophony avoidance, one has to then explain the many other cases where glottalization results in identity between derived and underived glottalized segments: for example, /s + [cg]/ can yield [y'] (Kim & Pulleyblank, 2009, p. 587).
\item There are two things to note here. First, uvular fricatives receive different specification from uvular stops on account of her analysis that PHARYNGEAL specification on laryngeals depends on whether there
\end{enumerate}

\textsuperscript{213} 213 It is doubtful that a recovery analysis could be made by positing that the *χ* *χ" changed because they were historically glottalized, i.e. *χ' *χ"*. First, glottalization is not contrastive for fricatives in any Southern Wakashan language; it only is contrastive for stops/affricates and sonorants (Werle, 2010, p. 3). Second, when subject to glottalizing suffixes, modern day Nuuchahnulth /χ χ"/ change into /w'/ (Davidson, 2002, p. 54). Furthermore, Werle (2010, pp. 4–5) observes that /h/ behaves like any other fricative in Nuuchahnulth, while /ɬ/ closely parallels /h/ in its phonotactic restrictions. Finally, Davidson (2002, pp. 10–11) observes that /ɬ/ continues to pattern with ejectives, which could be interpreted as a reflection of its glottalized heritage. Unless there is evidence to prove that these sounds were glottalized historically, it is assumed that this was not the case and not relevant to explaining the sound change.

\textsuperscript{214} 214 In Optimality Theoretic terms, the answer would involve assumption of high ranking markedness of [constricted glottis] and [pharyngeal] (or an ad hoc constraint specifically targeting [dorsal, pharyngeal]). Given such markedness constraints, one might wonder why /q + [cg]/ does not become realized as [k'] instead. If it is homophony avoidance, one has to then explain the many other cases where glottalization results in identity between derived and underived glottalized segments: for example, /s + [cg]/ can yield [y'] (Kim & Pulleyblank, 2009, p. 587).

\textsuperscript{215} 215 There are two things to note here. First, uvular fricatives receive different specification from uvular stops on account of her analysis that PHARYNGEAL specification on laryngeals depends on whether there
that all uvulars will undergo the sound change, as depicted in Figure 7.6. There is an extra draw-back to Rose’s model in that two delinking operations are required to account for the loss of DORSAL, since Rose represents uvular fricatives with a PHARYNGEAL-linked DORSAL feature, but uvular stops and ejectives have an ORAL-linked DORSAL feature.

![Figure 7.6: Pharyngeal genesis in GCLC models: (a) Rose (1996); (b) Halle et al. (2000).](image)

Once again, as was noted in the previous section, the problem is that these models treat pharyngeals as a stripped down uvular segment. There is no distinction that can be

are other gutturals in the phonological inventory. In some languages, uvular fricatives and laryngeals behave as gutturals. So it is assumed that both sounds must be specified as PHARYNGEAL. The laryngeals receive PHARYNGEAL specification in these languages in accordance with the ‘node activation condition’, which entails the activation of class specifications on sounds differing solely in features dependent on the class nodes. So if uvular fricatives had an ORAL-[dorsal] feature, they would not trigger the PHARYNGEAL specification on laryngeals. Thus Rose decides they must have a PHARYNGEAL-[dorsal] feature, and she then concludes that this means [dorsal] is in some sense a ‘bridge’ between the oral and pharyngeal sections of the vocal tract.

Second, Rose (1996, p. 81) claims that ejectives do not pattern as gutturals. Thus apart from [constricted glottis] specification, there is no difference between the representations of the plain uvular stops and the ejectives. It does not make a large difference, however, because the oral depletion process is more general: it targets uvular fricatives too. To remedy the problem one might propose that the uvular ejectives should be regarded as post-velars and thus be given the same representation as the uvular fricatives (with a PHARYNGEAL-[dorsal]). This, however, runs up against Rose’s (1996, pp. 100–101) analysis of Kashaya (Pomo; California) that shows that uvular stops /q q' q''/ are opaque in translaryngeal harmony. Since she analyzes translaryngeal harmony as involving the spread of ORAL vowel features, she must claim that the Kashaya uvulars, including /q'/ possess an ORAL node so that they block the spreading. Thus, her model cannot support the analysis of both phenomena (Kashaya translaryngeal harmony and the Southern Wakashan sound change) without contradiction.
drawn to substantiate why uvular stops would fail to change under this linguocentric representation: uvular stops are as much inherently pharyngeals as the uvular fricatives and ejectives are. True, uvular ejectives must additionally bear a LARYNGEAL node along with the [constricted glottis] feature, but this does not account for the status of uvular ejectives as the most prone to change since there is no relationship between these features encoded in the geometry. The closest encoding of the relationship is in the Halle et al. (2000) model’s organization of the GUTTURAL node and its dependent articulator nodes, LARYNGEAL and TONGUE ROOT. This representation, however, only predicts that the GUTTURAL dependents should behave as a unit with regard to spreading and deletion effects (Clements & Hume, 1995, p. 274). Any interaction between the lingual and laryngeal articulators is merely insinuated to occur by the sisterhood of the LARYNGEAL and TONGUE ROOT. There is no formal way to express the actual pattern of change observed in Southern Wakashan diachrony.

There is evidence that uvular stops are not universally part of the guttural/post-velar class (as noted in §6.1.1). McCarthy (1994, pp. 204–205) draws this conclusion about non-guttural status of /q/ in Arabic from statistical data showing root-consonant co-occurrence patterns. Crucially, sounds from the same natural class tend not to occur in the same root. This is true for /ʔ h ʕ χ ρ/, which show very low or no co-occurrence with each other; however, /q/ does not pattern in this way, as it occurs with /ʔ h ʕ h/ in a number of words. When it came to phonologically representing these facts, McCarthy (1994, p. 221) did not have a solution: he proposed identical representations for /q/ and the “uvular gutturals” /χ ρ/. Furthermore, he proposes to differentiate these sounds by assuming /χ ρ h ʕ h ʔ/ are marked [approximant] (1994, p. 222) while emphatics and /q/
are not. Rose (1996) attempts to solve this by suggesting that there are two dorsal features, an **Oral-Dorsal** and a **Pharyngeal-Dorsal** (as noted above): she describes dual-natured **Dorsal** feature as the “bridge between the [Pharyngeal and Oral] regions” (p. 76).

Bin-Muqbil (2006, pp. 243–247) offers an explanation for the unique status of /q/ by appeal to “Overlapping Innervation Wave Theory” (citing Joos, 1948; Lindblom, Sussman, Modarresi, & Burlingame, 2002; Lindblom & Sussman, 2002), which holds that constrictions are driven by “neural waves” sent from the motor system to the muscles comprising the articulators. These neural waves are thought to be subject to the laws of superposition, and, thus, waves from contiguous segments can constructively or destructively combine in a linear fashion, which is thought to yield co-articulatory effects. Bin-Muqbil argues that the “guttural articulator” is the neuromotor unit defined by the vagus nerve and the muscles it innervates; likewise, the lingual articulator is defined by the hypoglossal nerve. He suggests uvulars involve both neural pathways, lingual and guttural, and, on the assumption that tighter constrictions require stronger neural wave signals, /q/ is expected have a strong lingual signal and thus behave differently from /χ/ and /ʁ/, which have weaker constrictions and thus weaker lingual neuromotor involvement. The implication is that /q/ is the least “guttural” of the gutturals.

Bin-Muqbil’s analysis might be used to justify the representational splitting of uvulars along the lines suggested by McCarthy (1994) and Rose (1996). However, several issues remain to be explained if the pharyngeal genesis pattern in Southern Wakashan is to be accounted for using a GCLC approach. Most important is how to
represent uvular ejectives. There is no independent reason in such models to represent them differently from plain uvular stops (see discussion in f.n. 215). Furthermore, remedying the representation by positing that uvular ejectives are like uvular fricatives (see Figure 7.6a), still does not solve the problem of requiring two delinking operations for what should be a single process (especially since it still occurs synchronically).

GCLC models fail to model the pharyngeal genesis pattern because they do not incorporate an understanding of the epilarynx: the lower vocal tract is treated as a single, pharyngeal tube, with a flat larynx at the bottom. Not only do GCLC models get the details wrong, they also provide a misleading interpretation for why the sound change would occur in the first place. For instance, the suggestion that the change occurred because the representation was too complex (see above)\textsuperscript{216} places the causal burden on the nature of representation, minimizing the causal import of physical factors.

The LPP is focused on the role of the epilarynx in phonological patterning. The pharyngeal genesis case is interpreted as evidence that some understanding of the epilarynx, its basic physiological nature, and its role in the production of pharyngeals provides insight into this Southern Wakashan sound pattern.

It is useful to start with some physiological observations about the larynx as a valve mechanism. Specifically, the morphology of the vocal folds makes them efficient (see Figure 2.7 in §2.1.2) at preventing air ingress into the trachea during pressure build-up in the supralaryngeal vocal tract (Negus, 1949; Fletcher, 1993). They are not efficient at resisting pressure build-up within the subglottal space, and, thus, vocal-ventricular fold contact becomes necessary to build up large sub-glottal pressure (as in effort closure or

\textsuperscript{216} Or by appealing to markedness.
coughing). There is, therefore, a physiological asymmetry in the nature of the “glottis” – something not hinted at by [+constricted glottis]. Moreover, we can infer that vocal-ventricular fold contact is most useful in inhibiting the passage of air, when the overpressure is in the subglottal plenum; when the overpressure is in the supralaryngeal plenum, the vocal folds suffice to resist air and prevent it from flowing into the trachea. Presumably this is true when the larynx is suddenly raised, as in ejectives.

A key idea being argued for here is that epilaryngeal stricture is synergistic with larynx raising, i.e. \{\uparrow l x \leftrightarrow \text{epc}\}. However, this does not mean that larynx raising entails epilaryngeal stricture or vice versa. For example, larynx raising occurs when pitch is increased, but this does not mean the epilarynx constricts to increase pitch. It only means that larynx raising creates conditions that are synergistic with epilaryngeal constriction.

The other assumption is that tongue retraction plays an important role in producing uvular stricture (e.g. Namdaran, 2006) and that this also synergizes with epilaryngeal stricture, i.e. \{\text{tre} \leftrightarrow \text{epc}\}.

The hypothesis is that the PSW uvulars form a gradient of proneness towards engaging epilaryngeal stricture (Table 7.11), and this gradient is paralleled in the geographical distribution of the sound change itself, occurring most pervasively in northern-most language Nuuchahnulth (where uvular ejectives and fricatives underwent

\footnote{The feature [+constricted glottis], in standard models (e.g. Lombardi, 1991), does not specify larynx height: in fact, it is necessarily larynx-height neutral because it serves in specifying both ejectives and implosives (cf. Clements, 2003, pp. 312–318). Despite Trigo’s (1991) arguments for its phonological status, larynx height is often relegated to phonetics (Avery & Idsardi, 2001).}

\footnote{The discussion about larynx raising in relation to ejectives ignores other means to compress the air within the vocal tract to build up sufficient pressure for ejective production. None of the argumentation here is dependent on the fact that such alternative actions can occur in producing ejectives, or that the larynx must raise considerably to produce an ejective. It is simply assumed that larynx raising is typical in ejective production.
the change), less so in Nitinaht (where only uvular ejectives underwent the change), and not at all in Makah, the southern-most language.

Table 7.11: Gradient of uvular proneness to induce epilaryngeal stricture

<table>
<thead>
<tr>
<th></th>
<th>least prone</th>
<th>more prone</th>
<th>most prone</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSW</td>
<td>*q *q&lt;sup&gt;w&lt;/sup&gt;</td>
<td>*χ *χ&lt;sup&gt;w&lt;/sup&gt;</td>
<td>*q’ *q’&lt;sup&gt;w&lt;/sup&gt;</td>
</tr>
<tr>
<td>Makah</td>
<td>q q&lt;sup&gt;w&lt;/sup&gt;</td>
<td>χ χ&lt;sup&gt;w&lt;/sup&gt;</td>
<td>q’ q’&lt;sup&gt;w&lt;/sup&gt;</td>
</tr>
<tr>
<td>Nitinaht</td>
<td>q q&lt;sup&gt;w&lt;/sup&gt;</td>
<td>χ χ&lt;sup&gt;w&lt;/sup&gt;</td>
<td>?</td>
</tr>
<tr>
<td>Nuuchahnulth</td>
<td>q q&lt;sup&gt;w&lt;/sup&gt;</td>
<td>h</td>
<td>?</td>
</tr>
</tbody>
</table>

The gradient nature of the change is a reflex of the degree of synergistic potential each uvular has in relation to epilaryngeal stricture. This is depicted in three parts spanning Figure 7.7 (uvular stops), Figure 7.8 (uvular fricatives), and Figure 7.9 (uvular ejectives).

Uvular stops (Figure 7.7), which require closure at the oropharyngeal isthmus (Gick, Anderson, Chen, et al., to appear), primarily have tongue raising \{\text{tra}\} and of secondary importance is retraction \{\text{tre}\} (indicated with gray association lines). Retraction does synergize with epilaryngeal stricture \{\text{tre} \leftrightarrow \text{epc}\}, but it is relatively weak because retraction is relatively less important than tongue raising is to uvular stop production. The \{\text{epc}\} and \{\text{1lx}\} gross states are depicted as only indirectly associated with \{q\} via \{\text{tre}\}, so the potential for laryngeal interaction exists, it is just relatively weak. Thus, according to the LPP, we interpret /q/ as being the least likely of the three to change into a pharyngeal. (If anything, it might lose the \{\text{tre}\} and become more velar in nature, but there are velar stops in Nuuchahnulth, so this would potentially create homophony.)
uvular stop: least prone to induce epilaryngeal stricture

<table>
<thead>
<tr>
<th>POTENTIAL phonemic devices (PSW)</th>
<th></th>
<th>SW (step-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/q/</td>
<td>/q/</td>
<td>/q/</td>
</tr>
<tr>
<td>✶[q]</td>
<td>✶[q]</td>
<td>✶[q]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gross states &amp; synergistic relations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| tre - tongue retraction (down and back); tra - tongue raising (up and back). Devices: ✶[q] = linguo-oropharyngeal isthmus stricture as in [q].

Figure 7.7: Potential pharyngeal genesis, part I: uvular stop – least prone to induce epilaryngeal stricture (i.e. change into a pharyngeal). PSW = Proto Southern Wakashan; SW = Southern Wakashan. Gray lines = relatively weaker gross state participation; 
Abbreviations: epc = epilaryngeal constriction; ↑lx = raised larynx; tre = tongue retraction (down and back); tra = tongue raising (up and back). Devices: ✶[q] = linguo-oropharyngeal isthmus stricture as in [q].

Uvular fricatives (Figure 7.8) are associated with less extreme tongue raising than /q/ since fricative stricture does not involve complete closure. This is depicted by stronger association of ✶[χ] with {tre} than with {tra} (by means of association-line darkness)\(^{219}\).

Thus, the lower, overall more retracted position of the tongue means there is stronger synergy with epilaryngeal stricture (and correspondingly the association lines with {epc} and {↑lx} are darker than they are in /q/). Thus, in the LPP we declare that these sounds have a greater likelihood of becoming pharyngeals, and, as we know, they did become pharyngeals in Nuuchahnulth. At some point (call it “step-1”), the uvular stricture was

\(^{219}\)This implies that {tre} is stronger (more likely to occur) than {tra} in the case of ✶[χ].
diminished (represented as dashed association lines in Figure 7.8) and then diachronically abandoned (in “step-2”) in favour of pharyngeal stricture\(^{220}\). (To explain the actuation of this change, we will need to consider the transition from favouring uvular cues to pharyngeal perceptual cues, but this is not our present goal.)

\[\text{Potential pharyngeal genesis, part II: uvular fricative – more prone to induce epilaryngeal stricture (i.e. change into a pharyngeal). PSW = Proto Southern Wakashan; Nuu. = Nuuchahnulth. Dashed lines = “oral depletion”/historical loss. Gray lines = relatively weaker associations; Abbreviations: epc = epilaryngeal constriction; ↑lx = raised larynx; tre = tongue retraction (down and back); tra = tongue raising (up and back). Devices: } \llbracket [h] \rrbracket = \text{linguo-epiglottio-pharyngeal stricture; } \llbracket [\chi] \rrbracket = \text{linguo-oropharyngeal isthmus stricture as in } [\chi].\]

\(^{220}\) A subtlety here that is being glossed over is the fact that \([\check{h}]\) is observed, not \([h]\) (Carlson & Esling, 2003), even though \([?]\) occurs, not \([?]?\). The LPP has an explanation: first, we assume that \([\chi]\) is associated with \{vfo\} (i.e. it is voiceless). Since vocal fold abduction is anti-synergistic with epilaryngeal stricture \{vfo \text{-} cet\}, we could propose that the extra retraction causes a shift to \([h]\) rather than \([h]\) (in parallel with the transition from \([q']\) to \([?]?\)) is a compensatory action in overcoming the \{vfo \text{-} cet\} anti-synergy. This aspect of the LPP may also be involved in explaining why /\gamma/ resists change more than /q'/. Of course, we would also be wise to consider other factors, such as perceptual ones, in influencing the change. (But the LPP is only concerned with the articulatory side of the story.)
Uvular ejectives (Figure 7.9) are the most synergistic with epilaryngeal stricture. Ejection [’] intrinsically relies on larynx raising {↑lx}221 and vocal fold adduction {vfc}. This reliance combined with the tongue retraction inherent to the [q] means that /q’/ carries the strongest potential to engage epilaryngeal stricture. Thus, /q’/ has the greatest chance of becoming a “pharyngeal” (i.e. [?]). In the diagram, step-1 shows a possible transitional state in the history of Southern Wakashan at which point the oral depletion was starting to take effect (again, represented with dashed association lines) with concomitant strengthening of the vocal fold component to an epilaryngeal stricture ([?]), which brings about direct association of the /q’/ with {epc}. This intermediate step may or may not have occurred, it is included in the diagram for the sake of exposition; it is a plausible phase, since modern day uvular ejectives in Nlaka’pamux have been shown to involve epilaryngeal stricture (Carlson & Esling, 2003, p. 186). Like in the case of /χ/ (Figure 7.9), at some point the oral depletion becomes complete and we are left with the pharyngeal residue. Thus, the gradient nature of the sound change is simply a matter of some uvulars showing more potential for synergizing with the substance that makes a “pharyngeal” a pharyngeal – i.e. epilaryngeal stricture.

221 Once again, the presence of the {↑lx} gross state is more likely to occur in the production of [’] (ejectives) than, say, {↓lx} or neutral larynx height for that matter.
Figure 7.9: Potential pharyngeal genesis, part III: uvular ejective – most prone to induce epilaryngeal stricture (i.e. change into a pharyngeal). PSW = Proto Southern Wakashan; Nit. = Nitinaht/Ditidaht; Nuu. = Nuuchahnulth. Dashed lines = “oral depletion”/historical loss. Gray lines = relatively weaker associations; Abbreviations: vfc = vocal fold closure/adduction; epc = epilaryngeal constriction; ↑lx = raised larynx; tre = tongue retraction (down and back); tra = tongue raising (up and back). Devices: [ʔ] = aryepiglottal-epiglottal (upper epilaryngeal/UE) stricture; [ʔ] = linguo-epiglottal-pharyngeal stricture; [q] = linguo-oropharyngeal isthmus stricture as in [q]; [ʔ] = ejection manoeuver.

In the LPP, uvulars are in no way inherently pharyngeals: they only share the gross state of tongue retraction, but this is not a sufficient condition for generating epilaryngeal stricture. Previous GCLC models fail to make this distinction by positing [RTR] as the distinctive essence of /ħʃ/. Therefore they incorrectly predict that all uvulars will undergo oral depletion. Importantly, these models do not predict interaction between laryngeal events and supralaryngeal ones. Without consideration of the
epilarynx, the laryngeal and supralaryngeal systems are phonologically uncoupled, and we have no basis of expectations to understand what is happening when interactions do occur.

A part of the failure of the GCLC models must also be attributed to [+constricted glottis], which is of questionable value in explaining phonological patterning involving sounds to which this feature is often applied. This has been known for some time, but GCLC accounts struggle to present an alternate account that gets at the heart of the issue. Several cases can be noted. In attempting to account for why ejectives pattern differently from glottal stop and implosives in various Oromo dialects, Lloret (1995, p. 263) proposes analyzing the former as PHARYNGEAL. Urbanczyk (1992) presents arguments citing data from Shuswap, Kwakwala, Klamath, Saanich, Spokane that glottalized resonants (sonorants) and glottal stop pattern differently from glottalized obstruents/occlusive (ejectives). Czaykowska-Higgins & Kinkade (1998, p. 14) observe that in Interior Salish in general the spread of [+constricted glottis] from glottalized resonants is not blocked by other segments thought to bear [+constricted glottis] (such as glottal stop and ejectives). Caldecott (1999) observes that, in Saanich and St’at’imcets, glottalized resonants and ejectives behave differently in morphophonological patterns involving glottalization. Miller (2007, pp. 80–81) cites a case in Cuzco Quechua (MacEachern, 1999, pp. 7–9) showing ejectives diverge in phonological behaviour from glottal stop: no roots contain two ejectives, but there are roots which contain two glottal stops or a glottal stop and an ejective222.

222 Clements (2003, pp. 312–314) argues that “feature economy” patterns support the analysis of glottal stop, glottalized consonants, and laryngealized sounds as sharing the feature [+constricted glottis]. Neglecting the fact that this type of data does not provide information on how these sounds interact within
In the LPP, ejectives are not necessarily produced with concomitant epilaryngeal stricture. Any larynx raising that occurs is interpreted as synergistic with epilaryngeal stricture, but, because of the valve-functional asymmetry of the vocal folds, ejectives are efficiently produced without any epilaryngeal stricture. Evidence for this is in Carlson & Esling (2003, p. 186), who note that [t’] in Nuuchahnulth lacks the same degree of epilaryngeal stricture as the uvular ejective in Nlaka’pamux. The interpretation is that epilaryngeal stricture may be useful for producing an ejective (it does reduce the available volume of air to be compressed), but it is not essential, since the vocal folds efficiently ensure air does not leak into the trachea when an ejective is produced. Sounds like glottal stop and glottalized resonants are different. In this case, the overpressure is in the subglottal space: epilaryngeal stricture in the form of vocal-ventricular fold contact will help in preventing vocal fold motion in glottal stop or help contribute to creaky phonation.

7.2.5 Relatively open vowels and glottal stops

This section examines how the relationship between vocal fold closure and relatively open vowel quality is a phonological potential arising from the

a given language, it is also not convincing evidence for features being the basis for the associations he cites since no comparison to any other means of coding the data is provided. Furthermore, Clements acknowledges that phonemic analysis of the languages in the UPSID database (used to perform the feature economy analysis) can have a researcher-bias and are subject to diverse theoretical assumptions, but he does not present language specific evidence to assuage any doubts that the patterns are based on reliable analyses. Clements also does not provide evidence that the feature economy effects are unique, i.e. he needs to show that not only do glottal stop and ejectives occur together, but that they also occur to the exclusion of other sounds, such as glottal fricative. Finally, he does not rule out historical factors that might independently explain such patterns. For example, it seems plausible that an ejective series might have arisen from co-articulation of plain oral stops with a phonemic glottal stop, perhaps by the mechanism described by Ohala (1997, pp. 5–6). Campell (1973, p. 45) suggests /p/ in Yucatecan and Cholan-Tzotzilan arose from *ʔb. Clements needs to show that feature economy and not (phonetically motivated) historical change underlies such associations.
synergistic relations of “glottal constriction” as understood in terms of epilaryngeal functioning.

It is useful to begin with a pertinent quote from Montler:

“In fact, it is well known that glottals are not supposed to affect vowels … since laryngeal articulation is physically independent of tongue articulations.” (Montler, 1998, p. 371)

Montler’s remark stems from his observations that in Klallam the assumption of lingual-laryngeal independence – a deeply seated assumption of GCLC models – does not appear to apply: glottal stop lowers vowels (see below). Furthermore, the titular question posed by Brunner & Żygis (2011), “why do glottal stops and low vowels like each other?”, reflects their puzzlement that there would be any connection between sounds which ought to behave independently.

In the LPP, one of the most fundamental claims is that the epilarynx serves to couple the vocal folds to the rest of the supralaryngeal vocal tract. This coupling is not expected to be general to laryngeals. Rather, glottal stops are expected to behave differently from glottal fricatives in this regard because of the affinity between glottal stop and epilaryngeal stricture.

This section presents evidence of differential behaviour between /ʔ/ and /h/, but it is important to keep in mind the following caveat. The pattern being discussed is not related to the broader tendency for gutturals/post-velars\(^{223}\), which can include /ʔ/ and /h/, to lower vowels to /a/\(^{224}\).

\(^{223}\) Another factor connecting gutturals to open vowels is the fact that constrictions downstream of upstream ones can obscure place of articulation. This is especially true for [h], which is well known to become co-articulated with neighbouring vowels as a homorganic oral fricative, as in the pronunciation of hue as [qju] (Ohala, 2005b). The implication is that, among other factors, the “guttural” class is unified by a pressure to
One of the well-known problems of lower vocal tract phonology concerns the analysis of laryngeals, /ʔ/ h\(^{225}\). Some evidence motivates treating them as placeless, i.e. lacking a PLACE node in the feature geometry: this includes phenomena such as translaryngeal harmony (Steriade, 1987) and debuccalization (Blevins, 1993). Other evidence shows that they pattern like other gutturals, and, in feature geometric terms, they should bear a PHARYNGEAL node. Much of the evidence for this perspective comes from Semitic and Cushitic languages (Hayward & Hayward, 1989; Herzallah, 1990; McCarthy, 1991, 1994; Rose, 1996) showing laryngeals patterning alongside pharyngeals and uvulars in morpheme structure constraints, degemination, phonotactic restrictions, morphophonemic vowel lowering patterns and so forth. Further evidence outside of Semitic and Cushitic is found in Oowekyala (Howe, 2000, pp. 74–75), Nisgha\(^{226}\), and Salish (Thompson, 1984, pp. 25–52; Shaw, 1991, 1994; Bessell, 1992, pp. 363–367; produce all of the sounds in the class with a relatively open vowel tract configuration so that their place of articulation does not become obscured.

\(^{224}\) It is also worth noting that, even though there is an abundance of positive evidence for a connection between glottal stop and relatively open vowels, many researchers implicitly sweep glottal fricative into the discussion by generalizing the pattern to laryngeals. For example, Borroff (2007), whose dissertation is supposed to be general to laryngeals notes that “certain languages make distinctions even among the laryngeal consonants themselves, and thus no generalization holding of glottal stop necessarily holds of [h] as well”. I have attempted to do my due diligence in checking everywhere for counter examples showing glottal fricative interacting with relatively open vowels, and cases do occur. All exceptions or grounds for doubt have been explicitly noted. Furthermore, as noted above, I am not stating here that [h] never patterns with open vowels while [ʔ] does. The focus here is on an extra layer of phonetic and phonological intimacy that holds between just glottal and relatively open vowels.

\(^{225}\) According to Ruhlen’s (1975; also see Bessell, 1992, p. 35) survey of 693 languages, 75% (520/693) have laryngeals in their phonemic inventory: glottal stop is found in 49% (520/693) and glottal fricative is found in 64% (442/693); 38% (264/693) have both. 11% (78/693) have glottal stop but not glottal fricative; 27% (178/693) have glottal fricative but not glottal stop. According to Mielke (1997), glottal stop occurs in 35% (195/548) of all languages in his survey, which is 14% lower than Ruhlen’s estimate.

\(^{226}\) The general pattern is that laryngeals lower reduplicative default vowel [i] to [a]. If the reduplicative base is glide initial, as in /jank\(^{w}\) / ‘mouldy’ or /wa/ ‘name’, then [hinjank\(^{w}\)] and [huwa] result. Glottal stop always surfaces with [a]. The glide pattern can be analyzed by assuming that there is a prohibition on homorganic glide + vowel sequences (i.e. */ji/ and */wu/) and a default or epenthetic /h/ segment is used instead. Rule ordering can capture the fact that the vowel fails to change, or it could be a matter of preserving the identity of the glide on the reduplicant vowel. It could also be taken as evidence that /h/ has less of a connection to the open vowels than /ʔ/.
Rose, 1996, pp. 99–100). There is evidence to implicate both /ʔ h/ as participating in these patterns.

Controversy still persists regarding the phonetic and phonological treatment of laryngeals. Some approaches (Lloret, 1995, p. 265; Rose, 1996) advocate language specific specification: laryngeals in languages with other gutturals are PHARYNGEAL, otherwise they are placeless. Paradis & LaCharité (2001) argue that laryngeals are always specified with PHARYNGEAL. Lombardi (2001, p. 29) observes that “it has been impossible to reconcile these two approaches”; ten years later, Rice (2011, p. 532) acknowledges that the problem has still not been resolved. Shahin (2011a, p. 614; also see Borroff, 2007) summarizes the phonological solutions proposed to date: they either assume two different representations for the sounds or attempt to account for the different behaviour through other means, such as through constraint ranking; Shahin suggests that the answer lies in natural language articulatory data, proposing that glottals lack the aryepiglottic constriction characterizing the “primary pharyngeals”\(^{227}\), which for Shahin\(^{228}\) means \([q ɢ ɴ r χ h ≈ h ʔ h ʕ h ʜ ʡ h ʕ]\) (2011a, p. 605), but bear the feature of all primary pharyngeals: she is not clear on what this feature is.

\(^{227}\) This label primary pharyngeals, which Shahin claims is “more transparent”\((2011a, p. 605)\) than a traditional label such as gutturals, is equally muddy in my opinion. Does it mean that the main stricture for all of these sounds (i.e. \([q ɢ ɴ r χ h ≈ h ʔ h ʕ]\)) is pharyngeal tube stricture? How is it equivalent to the term guttural (which includes the laryngeals)? It also leads to the same entanglement between pharyngeals and uvulars that plagues the GCLC view (that uvulars are pharyngeals with secondary dorsal articulation; Halle, 1995, p. 17).

\(^{228}\) The reader should, at this point, carefully note that this view does not in any way correspond with my own view. What Shahin implies in her article is that uvulars have aryepiglottic constriction, embodied in her remark that “glottal primary pharyngeals are simply glottal, and lack the aryepiglottic constriction of the other primary pharyngeals.” (2011a, p. 614). In my view, uvulars do NOT inherently have aryepiglottic stricture. They are simply uvular (linguo-velo-uvular stricture); the epilarynx may or may not be concomitantly constricted. If all we know is that a sound is ‘uvular’ we do not know anything about what the epilarynx is doing exactly. All we know is that the tongue will be retracted to an extent. Tongue retraction is a component of epilaryngeal stricture but it does not mean epilaryngeal stricture automatically occurs: it is up to the language to decide what to do with the retraction and the epilarynx.
One property of the discussion surrounding how to phonologically analyze the laryngeals is the fact that /ʔ h/ are often lumped together. It is true that there are numerous cases where these two sounds participate along with other members of the guttural class in patterns that provide evidence of the existence of a guttural natural class in the first place, as noted above. However, it is also true that glottal stop has a peculiar affinity for relatively open vowels, especially /a/, the most open of them all; the glottal fricative does not share this affinity outside of its contribution to guttural patterning. Several examples are available, which are discussed below.

First is the often discussed pattern in Salish languages in which glottal stop uniquely lowers vowels. The data is difficult to interpret because, although it suggests exclusion of /h/ in the patterning, this cannot be determined absolutely: positive evidence showing that /h/ fails to lower vowels is non-existent, often because of independent restrictions, such as those of phonotactics. For example, in St’át’imcets/Lillooet, epenthetic schwa, which shows various phonetic realizations depending on the segmental context, is realized as [a] when preceding a glottal stop (Shaw, 1994; Rose, 1996). Shaw does not provide evidence showing either way if /h/ behaves like /ʔ/ in this regard. Bessell and Czaykowska-Higgins (1992) present evidence that the realization of the epenthetic vowel next to laryngeals (i.e. /ʔ h/) is [a] in Nxa’amxcin/Moses-Columbian; Rose (1996, p. 88) takes exception to their argument that the default realization of the epenthetic vowel in this language is [a] and instead argues that the laryngeals cause

\[\text{Czaykowska-Higgins (personal communication) points out that, like /ʔ/, Nxa’amxcin /h/ can lower a preceding schwa to [a] (i.e. } → \text{ a } /_{-h}\text{), but there are no cases of /i u/ lowering in the context of either laryngeal (/ʔ h/). The implication is that Nxa’amxcin does not provide evidence either way for showing different vowel-interaction tendencies between /ʔ h/.}\]
lowering. In support of this argument, she cites Bessell’s (1992, p. 142)\textsuperscript{230} observation that contrast between /ə/ and /a/ is neutralized when /ʔ/ follows. Crucially, Rose points out that (once again) no data are cited for /h/ triggering a similar effect. Thompson, Thompson, and Efrat (1974) and Montler (1998, 2004) discuss vowel lowering in Central Salish language, Klallam; /i u ə/ become [ɛ o a] preceding glottal stop. Montler (2004, p. 204, f.n. 5) explicitly notes that, due to phonotactic restriction preventing /h/ from appearing in coda position, no evidence can be found that shows the effect is solely attributable to /ʔ/. Shahin and Blake (2004) and Shahin (2011a, pp. 613–614) interpret this effect to be strictly “low-level” phonetic\textsuperscript{231} (also see Bessell, 1992, p. 323)\textsuperscript{232}.

In Sliammon, evidence exists for differential behaviour of /ʔ/ and /h/ (Blake, 2000): /ʔ/ (but not /h/) alternates with a low vowel (e.g. /liʧʷʔm=min/ becomes [liʧʷəʔmin] ‘comb’; p. 37) in parallel with glide vocalization of /j/ and /w/, which become [i] and [u], respectively; schwa is lowered to [ɑ] preceding /ʔ/, but is variably realized as [ʌ] or [a] preceding /h/ (p. 42, 59); the archiphoneme /L’/ has several realizations by phonological context, and most relevant is the fact that it becomes [ʔ] when near /a/ vowels (e.g. /paL’agιɬ/ becomes [paʔagιɬ] ‘one canoe’; p. 50); finally, Blake claims that

\textsuperscript{230} Rose cites Bessell (1992, p. 91), but this appears to be a mistake: there is no mention of the neutralization on this page.
\textsuperscript{231} Another, often mentioned, “low-level” connection occurs in Jalapa Mazatec vowels, which have raised F1 in the context of laryngealized (creaky) vowels (Ladefoged, Maddieson, & Jackson, 1988). The suspicion is that articulatorily, the creaky Jalapa Mazatec vowels involve a raised larynx configuration in their production (which is substantiated in §4.2), which in turn ought to raise the frequency of F1 [but not necessarily F2 or F3, (Nolan, 1983, pp. 182–187)]. Since a higher F1 is associated with a lower vowel, this pattern conforms to the connection between glottal stop and relatively open vowels. We might also suspect epilaryngeal involvement in these Jalapa Mazatec vowels, possibly in the form of vocal-ventricular fold contact, and, as per the LPP, this is synergistic with larynx raising.
\textsuperscript{232} In Gitksan, /h/ /ʔ/ show some divergent behaviour (Brown, 2008, p. 20). Brown proposes that there are two /h/s in Gitksan: some behave as gutturals in triggering vowel lowering, as in Nisgha (Thompson, 1984; Shaw, 1991; Rose, 1996, p. 99), and some behave as if they were placeless. Brown suggests the latter are “historically derived”. Yamane-Tanaka (2006) presents Eastern Gitksan data showing transguttural harmony: curiously, only data for /ʔχ q’ q/ are presented, not /h/.
/a/ is retracted by laryngeals, but only provides clear evidence for this effect in the context of glottal stop (p. 73)\(^ {233} \).

Another case where glottal stop appears to differ from /h/ in connection with relatively open vowels comes from hiatus resolution patterns. In Malay (Onn, 1976; Durand, 1986; also see Bessell, 1992, pp. 338–341), epenthetic [ʔ] resolves hiatus between identical vowels, i.e. \( V_1V_1 \) becomes \( V_1ʔV_1 \). When the vowels differ, i.e. \( V_1V_2 \), the pattern is a little different: if \( V_1 \) is either of the high vowels /i u/, then homorganic glides ([j] and [w]) will break up the hiatus; if \( V_1 \) is one of /e a o ə/, then [ʔ] separates the vowels. Bessell (1992, p. 340 f.n. 17) rejects the possibility that [ʔ] is the ‘glide’ equivalent of the [−high] vowels, based on an interpretation of SPE-style specification of the vowels and glottal stop. Her analysis favours treating [ʔ] as a phonological default here because it avoids violating structure preservation (no laryngeal features are contrastive in Malay).

Another language with a very similar pattern is Shona. Mudzingwa (2010, pp. 161–177) reports the following facts for the language’s two principal dialects, Karanga and Zezuru: for \( V_1V_2 \) sequences, /j/ and /w/ break the hiatus when \( V_2 \) is /i e/ or /u o/, respectively; when \( V_2 \) is /a/, then [ʔ] is used. Only in the case that the sequence is preceded by /h\(^ {234} \) does an [hi] serve as the hiatus breaker. In both Malay and the Shona dialects, the glottal fricative is phonemic, but the glottal stop is not. Another example is Kiribati phonotactics: /w/ is restricted to C_{i, e} contexts and /ʔ/ occurs only in /C_a/ contexts (Groves, Groves, & Jacobs, 1985).

\(^ {233} \) The one case with /h/ also has /ʔ/ (p. 73): /ʔah/ becomes [ʔah] ‘sore’.
\(^ {234} \) Or in borrowings with /b/ preceding \( V_1 \), which is always realized as breathy [b] (Mudzingwa, 2010, p. 173): for example, Hebrew word /baal/ ‘Baal (a god)’ is borrowed as [báʕáří].
Lombardi (2002) analyzes glottal epenthesis as a matter of constraint ranking. She proposes the following Place markedness hierarchy: *DORS, *LAB ⪰ *COR ⪰ *PHAR; this is intended to explain the prevalence of /ʔ h/ epenthesis (she assumes glottals are always specified for PHARYNGEAL place) and to account for epenthetic coronals or other segments. Interestingly, the overwhelming majority of examples of laryngeal epenthesis in Lombardi’s paper involve glottal stop, and many of these examples show glottal stop in association with /a/. For example, in Tamil epenthesis (Christdas, 1988; Lombardi, 2002, p. 225), [w] occurs with round vowels /o u/, [j] with /i e/, and [ʔ] with /a/. A similar pattern appears in Ilokano, but there are instances where [w] and [ʔ] are apparently interchangeable (Hayes & Abad, 1989; Lombardi, 2002, p. 227). In intermorphemic contexts, Dutch vowel initial syllables are subject to epenthesis: [ʔ] is inserted after /a/ (i.e. a#V becomes a#ʔV); however, a schwa will get deleted (i.e. ə#V becomes Ø#V), and after all other vowels, even mid vowels, a corresponding, homorganic glide is inserted (i.e. i#V becomes i#jV, etc.) (Booij, 1995, pp. 65–66; Lombardi, 2002, p. 226 f.n. 6).

In summary, cases showing an extra layer of intimacy between glottal stop (rather than glottal fricative) and relatively open vowels are fairly abundant in the data. The LPP analysis proceeds from the assumption that the pattern is bona fide and merits explanation beyond “laryngeals are PHARYNGEAL and so is /a/, therefore these sounds interact”. Montler (2004, p. 205) “conjectures” that glottal stop in Klallam is actually epiglottal (citing Esling, Carlson, & Harris, 2002). From the perspective of the LPP, this conjecture is on the right track, but one does not need to posit an epiglottal stop in Klallam – after all, the conjecture could be wrong – to account for interaction between
glottal stop and relatively open vowels. One needs to consider relations between the epilarynx and the vocal folds.

In the LPP, /ʔ/ is predicted to interact with relatively open vowels independent of any interaction that might occur between laryngeals and such vowels more generally. The basis of the argument is that any sound engaging epilaryngeal stricture, whether just at the lower epilarynx (such as [ʔ]) or upper epilarynx (such as [ʔ]), favours lingual retraction: \{\text{tre} \leftrightarrow \text{epc}\}. Likewise, lingual retraction works to open the vocal tract \{\text{tre} \leftrightarrow \text{vto}\}, and also is synergistic with larynx raising \{\text{tre} \leftrightarrow \text{lx}\}. Basic glottal stop [ʔ] does not alone predispose retraction, for vocal fold adduction and lingual position do not have any synergies \textit{per se}, but the adduction in [ʔ] does synergize with epilaryngeal stricture \{\text{vfc} \leftrightarrow \text{epc}\} and, it is through the composition of these synergistic effects that we would come to predict that glottal stop might bias or be biased by relatively open vowels.

Conversely, close vowels, canonically represented by /i/ and /u/, show a tendency to involve larynx lowering rather than raising (possibly to enhance their low F1; see §5.4.1). They also act against the influence of lingual retraction. In the LPP, these two factors lead to the prediction that these vowels are less likely to be associated with glottal stop, which is substantiated by the phonological patterning discussed above. Larynx lowering and non-retracted lingual configuration are anti-synergistic with both vocal fold adduction and epilaryngeal stricture: \{\text{lx} \cdots \text{vfc}\} and \{\text{lx} \cdots \text{epc}\}. If anything, we might predict such vowels to tend towards slight breathiness because \{\text{lx} \leftrightarrow \text{vfo}\}.

These relationships are depicted below using canonical open and close vowels /ɑ/ and /i/ as examples. Figure 7.10 shows /ʔ/ and Figure 7.11 shows /h/ (for comparison). In
Figure 7.10a, /ʔ/ is associated with /a/ through its potential to be realized using [ʔ], which is synergistic with the gross states of /a/. /ʔ/ is less likely to pattern with /i/ on account of the anti-synergies between these sounds, as depicted in Figure 7.10b: {vfc ⋯ ↓lx}, {epc ⋯ tfr} and {epc ⋯ ↓lx}. Nothing stops [ʔ] from being produced in the context of an [i] (anti-synergies can be overcome), but the same tendency or potential does not exist for /ʔ/ and /i/ sounds to become associated and we therefore predict they will be less likely to pattern together (or “like each other”\textsuperscript{235} in the same way that glottal stop and relatively open vowels do).

\textsuperscript{235} This is a quote from Brunner & Żygis (2011). They provide perceptual evidence that laryngealization can bias perception of relatively lower vowels. It could be that the titular effect is entirely attributable to perceptual factors (one immediately thinks of intrinsic F0 of vowels as well). It undesirable to regard the effect as exclusively arising from one domain or the other. Much more interesting is that, for independent reasons, the perceptual and articulatory systems are congruent with respect to the relationship between glottal stop and relatively open vowels. It is not surprising if the two systems feed off of each other in producing the phonological patterns.
Figure 7.10: Relationship between glottal stop and open vs. close vowels. Gray = relatively weaker associations; Abbreviations: vfc = vocal fold closure (adduction); epc = epilaryngeal constriction; ↑lx/↓lx = raised/lowered larynx; tre = tongue retraction (down and back); tfr = tongue fronting; vto/vtc = relatively open/close (upper/anterior) vocal tract; Devices: ⟦ʔ⟧ and ⟦ʔ⟧ = see Figure 6.10; ⟦ɑ⟧ = maximally tongue-retracted vocalic stricture (as in [a]); ⟦i⟧ = maximally tongue-fronted vocalic stricture (as in [i]).

In Figure 7.11a, the synergistic relations shared between /h/ and relatively open vowels like /ɑ/ are illustrated. The anti-synergy between vocal fold abduction and larynx raising {vfo ⋅⋅⋅ ↑lx} is predicted to inhibit epilarynx-mediated interactions between /h/ and /ɑ/, which accounts for the asymmetric behaviour of /ʔ/ and /h/. The LPP also predicts that, just as /ʔ/ and relatively open vowels “like each other”, likewise /h/ (which is typically ⟦h⟧) may arise in the context of relatively close vowels – possibly as slight breathiness on such vowels.
Figure 7.11: Relationship between glottal fricative and open vs. close vowels. Gray = relatively weaker associations; Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; ↑lx/↓lx = raised/lowered larynx; tre = tongue retraction (down and back); tfr = tongue fronting (forward); vto/vtc = relatively open/close (anterior) vocal tract; Devices: [h] = vocal fold abduction; [a] = maximally tongue-retracted vocalic stricture (as in [a]); [i] = maximally tongue-fronted vocalic stricture (as in [i]).

Articulatory evidence (discussed in §5.4.1) supports the LPP interpretation of the relationship between vowel quality and gross configurational changes that bias (but do not entail) epilaryngeal stricture for relatively open vowels and bias anti-constriction in relatively close vowels. The phonological evidence gives us reason to believe that such relationships percolate into the patterning of these sounds, whether it is glottal stop that causes lowering of neighbouring vowels or whether low vowels show a preference for glottal stop. In any case, without special interpretation or reworking of the features, there is no principled reason why features like [+constricted glottis] would ever predict that
such special vowel interactions to occur, even if the feature does set the sound apart from glottal fricative.

7.3 The epilarynx and the supralaryngeal vocal tract

In §7.2.5, it was argued that there is evidence indicating a special relationship between glottal stop and (relatively) open vowels, especially /a/. In this section, the theme of vowel quality and the epilarynx is explored more deeply. The discussion is framed in relation to other lower vocal tract sounds, particularly pharyngeals. One of the most surprising facts about pharyngeal consonants and pharyngealization is that they do not always retract vowels, contrary to what GCLC accounts predict. This section explores why this is the case, how it is a limitation for standard (GCLC) models of phonological features, and how consideration of the epilarynx and its relationship to the tongue, pharynx, and vocal tract points to an answer to some of these long standing mysteries of pharyngeal sounds.

The premise of argumentation made in this section is that epilaryngeal constriction is not strictly dependent on tongue retraction but instead benefits or synergizes with it. Furthermore, assuming retraction can be driven by a global shift in tongue position and/or by means of “double bunching” (as in the bunched variety of American English R), three possibilities arise in combination with epilaryngeal stricture: first, epilaryngeal stricture may leave vowel quality unaltered; second, it may be associated with retracted/lowered vowel quality; or, third, it may be associated with palatalizing bias and (possibly) pseudo-rhotic quality. It is argued that the first and third possibilities are not predicted by GCLC accounts, and, more importantly, it is not vowel
quality alone that is at stake, but voice quality, which represents the confluence of vowel, phonatory, and “global” quality.

First, §7.3.1 presents evidence that indicates pharyngeal consonants and pharyngealized vowels behave in complex and unexpected ways with regard to vowel quality. In §7.3.2 is a discussion of vocal register systems which indicate that vowel quality is not sufficient for their characterization, and it will be argued that voice quality and phonatory quality effects point to epilaryngeal influence. Finally, in §7.3.3 the topic of so-called “ATR harmony” languages is addressed in relation to the LPP.

7.3.1 Pharyngeals, pharyngealization, and vowels

This section examines the connection between pharyngeals and unexpected palatal stricture. The argument is that the connection stems from the “double-bunching” epilaryngeal potential which has an intrinsic predisposition for concomitant palatal stricture. The discussion relates this state to “raised larynx voice quality” and provides an explanation for how vowel quality relations are re-mapped but preserved with this extreme stricture.

In GCLC accounts, pharyngeals are [RTR] (or [−ATR]) and/or PHARYNGEAL/[+low], this means that, if anything, we expect vowel quality to become retracted/lowered in the context of a pharyngeal: these features are phonetically turbid. However, it has been recognized for some time now that pharyngeals do not always do the expected thing to neighbouring vowels, as the following quotes reveal:

Neither Delattre nor Ghazeli made films of pharyngeals in different vocalic contexts. Thus, although we see some raising of the anterior portion of the tongue body [Moisik emphasis] during the pharyngeals, we cannot know whether this is the influence of the vowel or an additional requirement of the pharyngeal consonants. In the actual tokens that Ghazeli examined, the vowel
following the pharyngeal is [æ]. In Delattre’s data, the tongue-body position also looks fairly [æ]-like. (McCarthy, 1994, p. 194)

When I started working on Tsez in the mid-1990s, I was surprised to note that back vowels in pharyngeal environments sound distinctively fronted, this being quite the opposite of what I would have expected [Moisik emphasis]. When I broadened my base to consider another of the Tsezic languages, Bezhta, I was even more surprised to find umlauted vowels, with rather minimal pharyngealization, as the equivalents of these Tsez vowels. This leads first to a consideration of the precise phonetic nature of pharyngealization, which turns out often to be more accurately characterizable as epiglottalization, which then leaves most of the tongue free to do other things [Moisik emphasis], including the production of more fronted vowel qualities. (Comrie, 2005, p. 2)

Vowel alterations are a key phonological effect of post-velars, often taken as a hallmark of guttural/post-velar classhood (Hayward & Hayward, 1989; Bessell, 1992; McCarthy, 1994; Rose, 1996; Brown, 2008). Rose (1996, p. 81) identifies two basic types of vowel effect that gutturals/post-velars can have: [RTR]-spreading and PHARYNGEAL node spreading. While spread of [RTR] is phonetically variable in its realization, it necessarily yields a retracted vowel allophone; the spread of PHARYNGEAL results in lowering to [a]. Rose’s model (which is canonically GCLC) gives us the expectation that [RTR] gutturals/post-velars (i.e. uvulars and pharyngeals) are a uniform class in their vowel interaction effects. We thus have no basis to expect these sounds would ever behave differently in this regard.

One does not need to look far to find evidence that pharyngeals are quite different from uvulars in their vowel effects, as the two quotes above have foreshadowed. Whether we are talking about pharyngeals in Interior Salish, Semitic, or Caucasian, there have been reports of behaviour that does not match the GCLC account. In Hayward & Hayward (1989, p. 187), one of the early papers advocating for a guttural natural class,
diachronic evidence from Akkadian and Arbore shows that historical pharyngeals (*h *ʕ) paired with *a were lost (vanishing entirely in Akkadian and becoming laryngeals in Arbore) whilst the vowel raised and fronted to /e/\(^{236}\). Hayward & Hayward (1989, p. 183) also note that the gutturals in D’opasunte, /ʂχʕ h/\(^{237}\), which neutralize the /e/ ~ /a/ contrast, cause lowering to /a/, and this vowel is realized with “marked fronting (to [æ])” only when preceded by a pharyngeal (p. 183)\(^{238}\). McCarthy (1994, p. 197) notes that the Arabic low vowel /a/ is realized as more front [æ] near pharyngeals and as [a] in the context of uvulars\(^{239}\). Herzallah (1990, pp. 29, 59) and Rose (1996, p. 87) describe similar patterns in Palestinian and Iraqi Arabic respectively.

When discussing emphatic palatalization\(^{240}\) in Ubykh, Rose (1996, p. 98) shows awareness of the problem that pharyngeal fronting poses for her theory that uvulars and pharyngeals are [RTR]. However, Rose squirrels this fact away into Hayward & Hayward’s analysis of pharyngeal fronting:

It is odd that an articulation involving tongue backing and lowering should condition vowel fronting. We may note, however, that pharyngeal fricatives may also be produced by means of a lateral compression of the pharynx immediately behind the mouth; Catford (1977: 163) claims that this is the most common articulation of [h] and [ʕ], though tongue root retraction is usually characteristic of the pharyngeal component of pharyngealized sounds. Vowel fronting in the environment of pharyngeals is not surprising if the pharyngeals are of the

\(^{236}\) For example, in Akkadian, */ṣaprum/ ‘dust’ became /eprum/ while in Arbore */saʕ/ ‘cow’ and */laʔ/ ‘small stock animal’ have become /seʔ/ and /leʔ/.

\(^{237}\) /ʕ/ is a “pre-/post-glottalized voiced pharyngeal fricative” (Hayward & Hayward, 1989, p. 191, f.n. 10).

\(^{238}\) In fact, Hayward & Hayward (1989, p. 187) raise the evidence that pharyngeals do not cause “backing and lowering” of the tongue to argue that it makes no sense to classify these sounds a [±back, ±low]. Their paper predates the introduction of [RTR] into phonological analysis of these sounds, but the point is almost identical to the argument being made here: there is an inadequacy in the defacto “tongue retraction” treatment of pharyngeals.

\(^{239}\) For example, [æːɭ] ‘condition’ vs. [ɑːɭ] ‘maternal uncle’.

\(^{240}\) The fact that we have known about *emphatische-mouillierung* ‘emphatic softening (palatalization)’ in Caucasian languages since Trubetzkoy (1939, p. 124) and yet still have no reasonable account of this phenomena despite half a decade of intensive phonological research that largely treats pharyngeals as tongue root retracted sounds suggests something is wrong with how we are phonologically conceptualizing these sounds.
‘lateral compression’ type. The feature specification [+back, +low] would not seem to be fully adequate for these; indeed, their articulation could not be characterized in terms of features referring to the posture of the tongue only. (Hayward & Hayward, 1989, p. 187)

However, Hayward & Hayward (1989, p. 187) caution that the guttural class cannot be defined properly without solving the pharyngeal fronting problem. Relying on Catford’s (1977a, p. 163) model of “common” pharyngeals (in Arabic) as upper pharyngeal/faucal241 constrictions as the explanation for why tongue retraction is not occurring is extremely shaky in the face of recent phonetic evidence that establishes the core constriction in pharyngeals as being upper epilaryngeal (aryepiglottal-epiglottal) in nature (see §2.2 and §2.2.3).

If emphasis in Arabic languages is regarded as [+RTR] spread (McCarthy, 1997; Halle, Vaux, & Wolfe, 2000, pp. 425–426; Shahin, 2002)242 then Hoberman’s (1985, p. 223) observation that pharyngeals can become emphatic (compare [baʕn] ‘after’ with [bʕn] ‘some’) is puzzling if pharyngeals are underlyingly [+RTR]. Contrary to what Hayward & Hayward (1989, p. 187) and Rose (1996, p. 98) are assuming about the production of pharyngeals, Hoberman argues that /ʕ h/ have a “much lower” constriction than that which occurs in “pharyngealized”/emphatic segments (1985, p. 223) and states that the former involve larynx raising while the latter do not. For independent reasons, Czaykowska-Higgins (1987, pp. 8–13) reaches similar conclusions about “pharyngealization”/emphasis in Arabic, arguing instead that it is uvularization. She further cites Card’s (1983) observation that, in Palestinian Arabic, emphasis can spread

241 One immediate problem with this view is the fact that this is the site of uvular stricture: the oral pharyngeal isthmus (Gick, Francis, Klenin, Mizrahi, & Tom, 2013).

242 In a footnote, McCarthy (1997, p. 232, f.n. 1) states that he is merely assuming the correctness of [RTR] as the “harmonizing feature” in emphasis spreading, so as to focus on his OT argument.
onto pharyngeals and connects this possibility for uvularized pharyngeals to the converse case of pharyngealized uvulars in Northwest Caucasian languages Ubykh, Abaza, and Abkhaz (citing Colarusso, 1975) (also see §7.2.3 and Table 7.8). Bellem (2005a, 2009) also argues that Arabic emphasis is not the same as pharyngealization found in Caucasian languages.

Let us set the issue of Semitic emphasis aside for now and focus on the phonological nature of Caucasian pharyngeals and pharyngealization. The available evidence points towards two phonological generalizations: (1) they tend to pattern with sounds involving a fronted lingual configuration or “palatal” stricture, and (2) they have the effect of “centralizing” vowels, not just fronting them.

Evidence for the first generalization (1) primarily comes from Lak and Bezhta. Lak (Anderson, 1997, pp. 974–980) has the vowels /i a u iʕ aʕ uʕ/; pharyngealized vowels are realized as [ɛʕ], [æʕ], [œʕ] and, when present in a word, cause /k/ and /l/ to palatalize (p. 980); Furthermore, Anderson suggests that pharyngealization has autosegmental status, that it is blocked by dental sibilants, and that Russian-Lak bilinguals pharyngealize Russian [a] and [u] when these vowels follow a palatal(ized) sound (p. 975). Bezhta (Kibrik & Testelets, 2004, pp. 221–222) pharyngeals morphophonologically participate in a palatal harmony which groups /a o u i s z ts tsʼ/ and pharyngeal-palatal /aʕ oʕ uʕ i eʃ ž č čʼʕ ħ/. Catford (1977b, p. 291) observes that Abkhaz /u/ (the labial-palatal approximant) is cognate with Abaza /ʕw/, which suggests a palatal constriction was present in the proto Abkhaz-Abaza segment and probably persists in the Abaza /ʕw/.

---

243 These sets appear as /a o u i s z c cʼ/ and /ä ö ü i e š ž č čʼʕ ħ/ in Bellem (2005b) who points out that the second set contains pharyngealized vowels, pharyngeals, and palatals.
Evidence for the second generalization (2), vowel space “centralization” (or perhaps condensing would be more appropriate), comes from various sources (Comrie, 2005, p. 2; Bellem, 2005a, 2009) showing the acoustic consequences of pharyngealization, particularly involving the Dagestanian languages. Catford (1977b, pp. 294–295) describes pharyngealization in Tsakhur as involving “fronting” of back vowels, particularly /u o/ (the same is true for Udi) and suggests that pharyngealization explains the “skewed” vowel systems of languages like Lak, Lezgi, and Tabasaran (p. 295), which are imbalanced by an overabundance of “slightly pharyngealized” front and central vowels /e æ œ ɵ/. Catford (1983, pp. 348–350) recapitulates that pharyngealized vowels in Tsakhur and Udi are “centralized”, clarifying that “the front ones seem to be lowered and retracted, and the back appear fronted”. Furthermore, and crucially, Catford identifies a parallelism between the lingual state in Caucasian pharyngealization and the palatal-pharyngeal stricture in American English “double bunching” of certain /r/ variants based on a comment by Kodzasov that Americans perceive the vowels to be “r-coloured”. Double bunching is described by Catford as follows: “[t]he tongue root at about the level of the tip of the epiglottis bulges backwards into the pharynx, while a

---

244 Many names have been applied to the different types of American English /r/, but Catford’s “double bunching” specifically is in connection with those tongue-tip down variants with a palatal constriction, a longitudinal concavity in the velar surface of the tongue and a secondary bulge in the mid-to-low pharynx (comparable to the constriction location in [u]). There is an abundance of literature documenting this and related configurations for American English R (Uldall, 1958; Delattre & Freeman, 1968; Delattre, 1971; Lindau, 1985; Catford, 2002; Alwan, Narayanan, & Haker, 1997). Trigo (1991, pp. 132–133, f.n. 7) was also aware of Catford’s proposals and notes them in the context of vowel centralization in cross-height “ATR” contrasts but dismisses the connection in favour of a raised vs. lowered/retracted interpretation.

245 This insight was based on a comment by Kodzasov that pharyngealized vowels tend to be perceived by American listeners as “r-coloured”. One cannot help, at this point, but point out that [ɪ] shares much acoustically (Moisik, 2006) and articulatorily (Mielke, 2011) in common with open-mid front rounded vowels [o] and [ø], which are close in quality to possible realizations of pharyngealized vowels, especially /u/ and /o/. Another piece of evidence for the similarity is the fact that Frenchville French, spoken in Pennsylvania, has started to merge [œ] and [ɔ] allophones to /ɚ/ (e.g. French neufl ‘nine’ is [nœ-f]) (Bullock & Gerfen, 2004). Czaykowska-Higgins (personal communication) observes that, in French Immersion contexts, French [œ] is often pronounced as [ɚ].
depression is formed in the dorsal surface of the tongue approximately opposite the uvular, with a further upward bulge further forward on the tongue” (1983, p. 349). Catford provides acoustic and articulatory support for the connection: x-ray imaging of Tsakhur [o̞] and Udi [a̞] (based on Gaprindašhvili, 1966; also see Catford, 2002, pp. 176–177); Catford’s own acoustic data support the connection by demonstrating that, like American-English R, F3 lowering and F2-F3 approximation occur for pharyngealized vowels in both Tsakhur and Udi. Maddieson et al. (1996) provide acoustic evidence further corroborating the centralizing analysis for Tsez.

On the basis of these two generalizations, Bellem (2005a) asserts that a sufficient account of Caucasian guttural/post-velar phonology must predict the “correlation between epiglottopharyngealization and vowel centralising, whereby /a/ → [æ], /i/ → [ɛ], /u/ → [ø]). Bellem’s (2005a, 2009) approach to the problem is framed in terms of Element Theory (Kaye, Lowenstamm, & Vergnaud, 1985; Anderson, Ewen, & Staun, 1985), which posits the phonological primitives, A I U (called “elements”), which map onto certain phonetic-articulatory properties and which additively combine or “fuse” to yield various vowel quality distinctions. For example, the A element in isolation is realized as [ɑ]; likewise, I is realized as [i], but when fused, they yield [ɛ] or [æ] depending on which element is the “head” (Kaye, Lowenstamm, & Vergnaud, 1985, pp. 309–310). Bellem’s suggestion is to represent the intrinsic palatal component of Caucasian pharyngeals and pharyngealized vowels as the composition of A and I elements – or pharyngeal and palatal elements (alternatively, Bellem states that [pharyngeal] and [palatal] features are “in play”).
The view taken in the LPP is that, while Bellem’s assertion about needing to be able to account for the pharyngeal-palatal connection in Caucasian is correct, the Elements model she proposes (or its feature equivalent) is still a GCLC model: it lacks the epilarynx. The view that Caucasian pharyngeals and pharyngealization boils down phonologically to simultaneous palatal and pharyngeal stricture cannot differentiate these sounds from the similar, but distinct, constriction occurring in double-bunching American-English R. The difference is whether the epilarynx is constricted or not: in Caucasian pharyngeals and pharyngealization it is; in double bunching, as in American English R, it is not. The phonological significance is that pharyngeals can become associated with a lingual configuration that intrinsically has a palatal component – the double-bunching configuration. Thus, patterning with palatals or fronting vowels are potential effects associated with pharyngeals.

The effect Caucasian pharyngeals and pharyngealization have on vowel quality is interpreted as one of condensing the possibilities of the vowel space, a “centralizing” effect, but not one which ultimately reduces every vowel to [ə]. Rather, the vowel space effect is characteristic of raised larynx voice quality (RLVQ), which Esling (1999) argues is essentially “pharyngealization” (see §2.3 and §5.3.3). Crucially, it is theoretically possible to produce any vowel quality in this configuration, but while vowel quality relations are preserved, voice quality is altered.

The clue to understanding what is going on in the double-bunching configuration and why it would be related to pharyngealization and what the role of epilaryngeal stricture is, lies in the dip in the velar surface of the tongue, opposite the uvula. We might hypothesize that this concavity is a direct consequence of strong middle genioglossus
contraction (see Stavness, Gick, Derrick, & Fels, 2012). Contracting this portion of the genioglossus draws the velar surface of the tongue down and forwards, causing the sulation of this part; as a side effect of the hydrostatic nature of the tongue, it will cause displacement of the anterior mass of the tongue towards the palatal region and the posterior bulk of the tongue down into the lower pharynx (Figure 7.12b). If we stop at this point, we have the double-bunching configuration and the auditory quality would be comparable to [1].

Figure 7.12: Forming the neo-tongue and neo-pharynx. Shaded region indicates the effective back cavity (or “pharynx”). The question mark indicates an indeterminate state. Abbreviations: dbe = “double-bunching” epilaryngeal; epc = epilaryngeal constriction; ↑lx = raised larynx; tdb = tongue “double bunching”.

Constricting the epilarynx takes the configuration to the next stage: “double-bunching” epilaryngeal (or, alternatively, palato-pharyngeo-epilaryngeal, see Figure 7.12c). In §5.4.2, this state was characterized by the ideas of the neo-tongue and neo-
pharynx (or NT-NP). To reiterate what was said in that section, the “double-bunching” epilaryngeal configuration causes a sudden jump in the articulatory-acoustic configuration of the vocal tract: it becomes a “smaller” facsimile of its “neutral” state (compare Figure 7.12a & c). The pharynx (represented with gray shading) is now about half of its neutral-configuration size since the lower pharynx becomes acoustically closed (depending on the degree of epilaryngeal constriction and lingual retraction). The velar-surface concavity together with what volume remains in the oropharynx constitutes a new effective pharyngeal cavity (or lower resonating cavity) and the imaginary line separating front/“oral” cavity from the back/“pharyngeal” cavity must be redrawn as the front constriction is now mandatorily shifted forwards. Most compromised in quality in this state should be the vowel [u] since the velar concavity is in conflict with velar constriction made in this area for that vowel. However, even though the tongue is much more restricted in its maneuverability, it still can (with cooperation from the lips, jaw, and so forth) produce virtually any vowel quality, thanks to the re-scaling of the vocal tract. This new, reduced-mobility but relatively “free” (Comrie, 2005, p. 2) tongue – or neo-tongue – is a conceptual device to emphasize that the tongue in the “double-bunched” epilaryngeal state still has the same vowel-quality potential as its neutral tongue counter-part: the vowel space has “centralized” or, more appropriately, condensed, but it is theoretically possible to differentiate [i] and [ɪ] in this state. This, in essence, is raised larynx voice quality, but its acoustic/auditory quality is more than just a linear shortening of the length of the vocal tract would imply (see §5.3 and §5.4).

246 The relationship between the neutral vocal tract state and the neo-tongue-neo-pharynx state is parallel to the relationship between an adult vocal tract and child vocal tract: the same vowel quality produced in the two different vocal tracts will have considerably different absolute formant values, but we can still posit identity between these vowels on the basis of the location within its respective vowel space.
An important theme here is that pharyngeals and pharyngealization do not necessarily involve a change in vowel quality so much as they involve a change in voice quality. The benefits of this potential for pharyngealized vowels to combine as a register-like shift in the vowel space, preserving relations between vowels in the non-pharyngealized “register”, comes from the analysis of systems with pharyngealized vowels such as Udi, as depicted in Table 7.12. Diachronically, the pharyngealized vowels of Udi actually arose from vowel co-articulation with earlier pharyngeals, pharyngealized uvulars, and /r\(^247\).

Table 7.12: Vowels of Udi (Colarusso, 1975, p. 343)\(^248\)

<table>
<thead>
<tr>
<th>plain</th>
<th>pharyngealized</th>
</tr>
</thead>
<tbody>
<tr>
<td>i y u</td>
<td>i(^{\delta}) u(^{\delta})</td>
</tr>
<tr>
<td>e ø o</td>
<td>e(^{\delta}) o(^{\delta})</td>
</tr>
<tr>
<td>æ a</td>
<td>æ(^{\delta}) a(^{\delta})</td>
</tr>
</tbody>
</table>

Colarusso (1975, pp. 342–345) argues that the contrast between plain and pharyngealized vowels in Udi does not reduce to traditional features like [±back], [±low], and [±ATR]: he states “[t]here is no evidence that [ATR] plays any part in the vowel system of Udi” (p. 343). Furthermore, Colarusso (p. 342) takes Udi (along with related languages, Tsakhur and Rutul) as contradicting Kiparsky’s (1974) [±ATR]-induced dubiousness about the existence of contrastive high pharyngealized vowels /i u/ and /i\(^{\delta}\) u\(^{\delta}\)/. Now, certainly [±ATR] is phonetically inadequate: if it was applied to the Udi close

\(^{247}\) The connection to /r/ is not surprising given the discussion above about the relationship between the double-bunching epilaryngeal gross state and [\i].

\(^{248}\) Colarusso (1975, p. 343) uses /V\(^{\delta}\)/ to symbolize pharyngealized vowels. I follow the standard IPA practice by symbolizing these as /V\(^{\delta}\)/.
front vowel we would expect an [i ɪ] contrast. However, one might counter by saying that Udi vowels represent just a different phonetic implementation of [±ATR], but are (essentially) no different phonologically from the /i ɨ/ contrast in English.

From the perspective of the LPP, Colarusso’s arguments should be taken seriously: the analysis that this pharyngealized-vowel contrast is [±ATR] obfuscates the phonetic nature of the pharyngealized vowels, unnecessarily reduces the phonological possibilities of human language, and presents misleading interpretation about how such sounds might pattern in Udi and other Caucasian languages. In languages like Udi, Tsakhur, Rutul, Tsez, Lak, and Bezhta, we need to simultaneously account for the palatal nature of pharyngeals and pharyngealization while also accounting for the possibility of preservation of vowel quality relations in a condensed vowel space. The NT-NP interpretation allows us to do just that. The condensed vowel space is a reflex of the sudden, effective diminution of the vocal tract when the double-bunching epilaryngeal configuration is engaged. This is why it is possible to maintain an [i] quality with pharyngealization. Caucasian-style pharyngealization is a remapping or transformation of the vowel space into that of the space defined by the NT-NP. Thus, it is theoretically possible to distinguish [iʕ] from [iɬ]. Although vowel quality relations can be preserved, this does not mean that NT-NP does not have a significant impact on the overall quality of an individual vowel (for example, possibly making it rhotic sounding). It should not be surprising if subsequent generations reinterpret the quality change, leading to a loss of the “pharyngeal” component, as in the change in *ʕw modern day Abkhaz to /u/, or the genesis of front rounded vowels in other Dagestanian languages, like Lak, Lezgi, and Tabasaran (Catford, 1977b, p. 295).
What is being advocated here is that (true) pharyngealization (perhaps epilaryngealization?) should no longer be strictly framed in GCLC vowel-quality terms such as *fronting*, *lowering*, or *retracting*. These vowel quality changes may result from subsequent perceptual reinterpretation of the voice quality change induced by the NT-NP state, but the GCLC terms obscure the articulatory origin causing the change in overall quality in the first place.

Returning to Udi (Table 7.12), the vowel system includes pharyngealization across five of its six non-pharyngealized qualities, /i e a o u/, giving /iʕ eʕ aʕ oʕ uʕ/. The fact that /uʕ oʕ/ co-exist with /y ø/\(^{249}\) emphasizes that the change in quality is not simply fronting, although it would not be surprising if the vowels eventually merged.

The “gap” in the open-front part of the system, such that /æ/ lacks a pharyngealized counterpart\(^{250}\), might reflect the intrinsic compatibility of this vowel with epilaryngeal stricture complemented with some slight double-bunching. The /æ/ vowel appears in Lak, Lezgi, and Tabasaran, which Catford (1977b, p. 295) explains is often mistaken as arising from umlaut by Soviet scholars but more likely owes its provenance to pharyngealization. It is also similar to the unexpected allophone (most often transcribed [æ]) of the low vowel /a/ appearing in the context of pharyngeals in Arabic\(^{251}\).

---

\(^{249}\) The absence of /yʕ øʕ/ in Udi, however, may be due to the perceptual similarity of these vowels with /uʕ oʕ/, which are inherently fronted by the “pharyngealization” (or we might say that the NT-NP configuration makes raised vowels like /u o/ susceptible to fronting; in the LPP, this fronting bias is expressed as \{dbe ↔ tra\} and \{dbe ↔ tfr\}).

\(^{250}\) The lack of /a/ could be attributable to the relative inefficiency of rounding on open front vowels, as described in Esling’s model (2005, p. 24): “[j]aw opening is dominant over labial setting at the front as vowels open”.

\(^{251}\) The [æ] allophone is said to be the default realization of /a/ in Arabic, but the fact remains that pharyngeals do not retract this vowel, unlike uvulars; instead, the interpretation here is that they are fully compatible with it. Of course, pharyngeals should cause other voice quality effects like inducing constricted phonatory quality, e.g. Interior Salish pharyngeals induce creakiness on neighbouring vowels (Czaykowska-Higgins, 1990, p. 2), just as in Arabic (Butcher & Ahmad, 1987).
Thus, in Udi, it may be that there is no /æʕ/ simply because its quality is so similar to that of non-pharyngealized [æ]. The implication is that, although perhaps not as extreme as in cases of the NT-NP state, pharyngeals tend to involve some degree of engagement of the double-bunching epilaryngeal configuration, giving the slight “fronting” bias on the low vowel in all of these systems.

The Udi vowel system provides evidence that a vowel system can be transformed into the NT-NP/“pharyngealized” state without loss of phonological distinctions present in the neutral state. However, there are other pharyngealized-vowel-system potentials. The Tungusic language, Even/Evenki, likewise (nearly) completely exhibits a mapping of non-pharyngealized vowels into a contrastive pharyngealized set (see Table 7.13), i.e. non-pharyngealized [i u o ə] correspond with pharyngealized [iʕ uʕ əʕ ɑ] (following transcriptions in Ladefoged & Maddieson, 1996, p. 307) and not, say, [iʕ oʕ əʕ ɑ] or otherwise. Novikova’s (1960; reproduced in Ladefoged & Maddieson, 1996, p. 307) x-ray evidence shows a configuration that is quite different from double-bunching epilaryngeal and the corresponding NT-NP state. Rather, Ladefoged & Maddieson (1996, p. 306) suggest the pharyngealized vowels in Even appear to involve a narrowing of the

---

252 Kinkade (1967, p. 232) states that “the most notable [effect pharyngeals have on neighbouring vowels] is a marked fronting of /a/ in immediate proximity to /h/ and / ś/” and provides the Nxa’amxcin example, [ˈhæcəm] ‘tie’.
253 Ladefoged & Maddieson (1996, p. 306) point out that the original data indicate all vowels as nasalized, which they interpret as casting some doubt on the accuracy of the Novikova’s original tracings. Denning (1989, pp. 58–59) mentions that Novikova describes the pharyngealized vowels as having “noise”.
pharynx and larynx raising, although [i̞] and [u̞] appear to have slightly forward anterior
tongue position in comparison to its position for [i] and [u].

Table 7.13: Vowels of Even (Ladefoged & Maddieson, 1996, p. 307)

<table>
<thead>
<tr>
<th></th>
<th>plain</th>
<th>pharyngealized</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
<td>i̞</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>o̞</td>
</tr>
<tr>
<td>a</td>
<td>a̞</td>
<td>a̞</td>
</tr>
</tbody>
</table>

Although it is not possible to classify the Even pharyngealized vowels more
conclusively in articulatory terms, the system does show that, contrary to what might be
predicted by analyzing the system in terms of [±ATR], vowel height and vowel frontness
are not fully compromised in the pharyngealized set. Thus, the /i u/ and /i̞ u̞/ sets do not
(ostensibly) differ in vowel quality so much as voice quality. Unlike Udi, which
uniformly applies pharyngealization voice quality throughout the vowel system without
compromising vowel-quality relations, Even preserves vowel quality in the high vowels,
but compromises it in the low vowels, giving the correspondences [ə] ~ [a] and [ɔ] ~ [ɔ̞].
Thus, the pharyngealized versions of the low vowels phonetically blend retraction and
epilaryngeal tube constriction. Furthermore, instead of the theoretically possible [o̞] we
have [ɔ̞], and the low vowel is simply [a], which contrasts with the schwa: [akən]

Overall, we need an analytical tool that can manage the complexity of vowel
interactions entailed by epilaryngeal stricture. It causes more than just the often cited

---

254 Distortion to the lower epiglottal line suggests epilaryngeal stricture, but in pharyngealized and non-
pharyngealized, the hypopharynx is depicted as wide open, which is likely an artifact of choices made
regarding what to trace and poor imaging resolution (see §2.2.1). The tongue does not have the concavity
of the double-bunching configuration, but it may be that sulcalization occurred but was not traced.
255 This vowel may also be produced with larynx raising, judging from its appearance in the x-ray data, but
Ladefoged & Maddieson (1996, p. 307) transcribe the vowel simply as [a].
lowering and retraction of vowels. It is distinct from effects caused by uvulars and
uvularization-pharyngealization. It can condense the vowel space but preserve quality
relations by virtue of the potential NT-NP (sometimes described as raised larynx voice
quality) but it can non-uniformly apply across a vowel system, sharing the potential for
influence with basic lingual retraction.

The LPP can handle this complex picture of pharyngeal-vowel interaction
precisely because it recognizes that pharyngeals are, ironically, not strictly defined by
tongue retraction or narrowing of the pharynx: they are epilaryngeals with various
potentials for combing with lingual mechanisms to facilitate epilaryngeal narrowing.
Thus, we have (at least) the following potentials, which, as a matter of convenience, are
depicted for /ʕ/ in Figure 7.13. In (a) the upper epilaryngeal potential is depicted, which
synergizes with larynx raising and tongue retraction \{↑lx ↔ tre\}, but these are likely –
not necessary – gross states that will accompany such constriction (and this is depicted
with a dashed line quality). Furthermore, since \{tre ··· tfr\} are anti-synergistic, these
sounds are predicted to cause front vowels like /i/ and vowels like /u/ (which can involve
substantial contributions from the genioglossus posterior) to be compromised in quality,
an effect that can be characterized as lowering or retraction of vowel quality; however, it
should be remembered that the effect will be more holistically one of voice quality than
strictly vowel quality and we should look for other effects occurring simultaneously such
as a change in phonatory quality.
The next potential type is (b) double-bunching epilaryngeal, where the benefit to retraction gained by engaging the double-bunching mechanism now paradoxically enhances fronting of the tongue and puts the vocal tract into an approximation of the NT-NP state (depending on the degree of constriction involved). The effects are somewhat different now. Now fronting of the tongue is much more synergistic with this state, thanks to the double-bunching epilaryngeal configuration \( \{\text{dbe} \leftrightarrow \text{tfr}\} \) and we expect
that these sounds will be palatal-philic, and exhibit the possibility of patterning with palatals (as in Lak and Bezhta): this relies on the assumption that sounds like /ʊ/ and /i/ involve roughly parallel gross states characterized in part by lingual fronting \{tfr\}. Thus, the patterning of pharyngeals with palatal sounds is a reflex of the potential for pharyngeals to be produced with the somewhat unusual double-bunching epilaryngeal configuration. Furthermore, there is an anti-synergy now acting on lingual raising towards the oropharyngeal isthmus (Gick, Anderson, Chen, et al., to appear) \{dbe \ldots tra\} (not depicted in Figure 7.13) and we should not be surprised to find the quality of vowels like /u/ becoming strangely coloured and questionably qualifying as /u/-quality. This is likely the case for Udi; Catford reports that /u o/ are particularly compromised, having a “distinctly central quality” when pharyngealized (1977a, pp. 294–295). This is a normal part of the NT-NP mapping, given the concavity of the velar surface of the tongue, but we must be careful in realizing that this is still a mapping and further centralized qualities could be distinguished, /u̯/ vs. /u̯/, but are predicted to be highly unlikely.

In summary, we have redefined what pharyngealization means by drawing on knowledge of the epilarynx. Pharyngealization can be upper pharyngeal “uvularization” (as in Semitic emphatics) or it can be lower pharyngeal (as in American English R). When the epilarynx is introduced into the picture, we can posit the double-bunching epilaryngeal configuration as another phonological potential for pharyngeals, a different species from the upper epilaryngeal configuration with more or less ordinary tongue retraction. No GCLC model would ever predict these possibilities or the complex patterns in relation to vowel quality that occur. In the next section, we will extend this basic
model further to register-type systems where more extreme interactions occur and which are intrinsically associated with simultaneous changes in phonatory quality.

7.3.2 Register contrasts and vowel quality

This section examines register – a dimension of contrast that cross-cuts another, unrelated dimension of contrast. The focus is on vocal register, which potentially involves tone, phonatory, and vowel quality. The argument is that many of these systems are based on the opposition of gross states associated with epilaryngeal pro-constriction and anti-constriction synergistic relations.

In §7.1, growling was claimed to play a phonological role in phonatory and tonal subsystems of various languages, namely !Xóõ, Ju|’hoansi, and N|uu, Bai, and Zhenhai Wu. This section returns to the issue of vowel quality interaction. Some of these systems show vowel restrictions and some do not. The traditional features associated with GCLC accounts do not help us characterize these systems, since by using them not only do we fail to generalize shared properties of these systems, but we also treat voice quality as if it is not part of phonology: as Edmondson & Esling (2006) state “[t]he difficulty of the vowel-quality theory rests in its inability to connect vowel-quality change to vocal register changes” (p. 177).

If we look at the Khoisan-group languages for which pharyngealized, strident, or epiglottalized vowels have been attested, as listed in Table 7.14, then it becomes apparent that vowel quality in the pharyngealized set tends towards Esling’s (2005, p. 23) canonical “retracted” vowels [อำนาจ] (i.e. both relatively low and back in traditional terms), while /i/ does not appear in this set. There are some surprises, however, which indicate this is an attraction to the retracted vowels rather than an absolute requirement for retraction. Looking at N|uu first, we observe that while non-pharyngealized mid-
vowels in N|uu have “[−ATR]” allophones [ɛ ɔ], the pharyngealized mid-vowels are realized as “[+ATR]” [ɛ̅ ɔ̅]. N|uu also apparently maintains an [u̅] (i.e. not [o̅]), according to Miller et al. (2009, pp. 155–156). In !Xóó “pharyngealization” only occurs on /a o u/ (see §7.1.1), although the /u o/ contrast often is neutralized to [o], regardless of the set, giving, for example, [o o̅ o̅] (Traill, 1981; cited in Hess, 1998, p. 176). Trigo (1991, p. 118) suggests that pharyngealized and strident vowels are [+back] (following traditional, SPE specification256). Ju|’hoansi has the most restricted pharyngealized set [a̅] and [o̅]257, perfectly matching Esling’s retracted vowels, as noted above.

Table 7.14: Vowel subsets in !Xóó, Ju|’hoansi, and N|uu258

<table>
<thead>
<tr>
<th></th>
<th>“non-pharyngealized”</th>
<th>“pharyngealized”</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>uu</td>
<td>i  ũ e/ɛ o/ɔ a/ɑ</td>
</tr>
<tr>
<td>!Xóó</td>
<td>ĩ i̅  ũ u õ o̅ ã ã</td>
<td>u̅ u̅ o̅ o̅</td>
</tr>
<tr>
<td>Ju</td>
<td>’hoansi</td>
<td>ĩ i̅  ũ u õ o̅ ã ã</td>
</tr>
</tbody>
</table>

256 If we stick to the SPE analysis, the inclusion of /u/ in the set is unpredicted, given that pharyngeals are analyzed as [+back, +low]. Furthermore, it leaves us wanting for an explanation of why pharyngealization is acceptable on the front and high vowels of Udi, Rutul, Tsakhur, and Even (discussed in §7.3.1). Trigo’s (1991, p. 118) [raised larynx] feature, which is applied to pharyngealized, strident, and glottalized vowels in !Xóó, circumvents the problem of requiring [+low] pharyngealization to combine with [+high] vowels, but fails to explain why glottalized vowels do not pattern the same way as the pharyngealized and strident vowels.

257 The ‘epiglottalized’ consonants are not subject to distributional restrictions with regard to vowel quality, as they can combine with nuclear vowels other than /a o/ (e.g. /gû ût/ “to twist”).

258 Data sources: !Xóó (Hess, 1998, p. 176); Ju|’hoansi (Miller, 2007, p. 58); N|uu (Miller, Brugman, Sands, et al., 2009, pp. 155–156). Nasalized vowels occur in all three languages: in !Xóó and Ju|’hoansi nasalization is not restricted by vowel quality: in N|uu, /ĩ á ù a̅ o̅/ and /áî áû ʊ ʊ ʊ ʊ̅ á̅ o̅̅̅̅̅̅̅́ o̅̅̅̅̅̅̅́ o̅̅̅̅̅̅̅́/ are attested.
In §7.1.4, Bai and Zhenhai were presented as languages with tonal-register systems in which growling occurred in a subset of tonal and vocalic environments. Bai is particularly compelling since laryngoscopic evidence demonstrates that, even in [i] context, extreme epilaryngeal stricture and vibration can occur (Edmondson & Esling, 2006, p. 176). “Growling”/epilaryngeal vibration in Zhenhai Yang tone, as described by Rose (1989), is restricted to relatively open oral vowels (e.g. [œ ɛ ə]) but can occur on close nasalized vowels and tends to be followed by a glottal stop; whisper in Zhenhai, also a function of epilaryngeal stricture but lacking epilaryngeal vibration, has nearly complementary distribution, tending to occur on relatively close vowels (e.g. [i ŋ ɛ]).

Akh has been described as having two registers, “oral”/“low”/“head” and “laryngealized”/“high”/“chest” (Lewis, 1968; Wyss, 1976; Trigo, 1991, p. 117). Both vowel and phonatory quality show register-wise correlation: the head register gives a “tense, restrained and ‘choked’” auditory impression and is always terminated with [ʔ] when pre-pausal; the chest register gives the impression of “‘hollow’ or ‘soft’” auditory quality and “varying degrees of breathiness” (quotes from Wyss, 1976, p. 152; also see Trigo, 1991, p. 117). Trigo provides further details in a footnote (1991, p. 133, f.n. 10): head register vowels involve “narrowed pharyngeal walls, raised larynx and pressed voice” and are “centralised [i o a] etc.”; chest register vowels involve “dilated pharyngeal walls, lowered larynx and breathy voice” and are “peripheral [i u a] etc.” From these descriptions, we can infer that Akha is best described as involving a distinction between

259 This predisposition for nasalized close vowels to growling (epilaryngeal vibration) may stem from the palatopharyngeus-mediated physiological link between opening the velo-pharyngeal port and larynx raising. The other possibility is that, since nasalized vowels tend to be characterized by lowered F1 (Ohala, 2005b, p. 421), it may be that tongue position is free to lower without compromising the vowel quality significantly.
constricted (or “head”) and unconstricted (or “chest”) registers (e.g. see Table 6.6), comparable in phonetic properties to Edmondson & Esling’s (2006, p. 175) [±constricted] register distinction.

Gregerson (1976, pp. 329, 333; also see Czaykowska-Higgins, 1987, p. 6; Denning, 1989, p. 58) describes “lax” and “tense” registers in Mon-Khmer language, Rengao: the lax register is described as having pharyngeal expansion associated with “tongue root” advancement, larynx lowering, and the vowels [i e a o u]; tense register has a constricted pharynx associated with tongue root retraction, larynx raising, and the vowels [e i e a o u].

Various dialects of Dinka have vocal-register contrasts involving a suite of features including vowel and phonatory quality, length, and tone. In Agar Dinka (Andersen, 1993, p. 4; also see Denning, 1989, p. 28) modal [i e a o u], breathy [i ø e ø a ø o u], and creaky [i ø e ø a ø o] registers occur with the only vowel restriction being that /u/ is never modal or creaky. Luanyjang Dinka (Remijsen & Ladd, 2008) has breathy and modal registers: these both combine with its seven vowel phonemes /i e a o u/, with the exception that, as in Agar, /u/ is always breathy. Bor Dinka has the most elaborate system, exhibiting four registers: modal, harsh, breathy, and faucalized (Denning, 1989, pp. 68, 131; Edmondson, Esling, Harris, et al., 2003; Edmondson & Esling, 2006, pp. 182–185). Table 7.15 provides Denning’s (1989, p. 68) summary of the general phonetic properties of the entire Bor Dinka register system.
Table 7.15: Phonetics of register in Bor Dinka (based on Denning, 1989, p. 68, table 3.27)

<table>
<thead>
<tr>
<th></th>
<th>hollow/faucalized</th>
<th>breathy</th>
<th>normal</th>
<th>harsh</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>glottal aperture</strong></td>
<td>greatest</td>
<td>←</td>
<td>→</td>
<td>least</td>
</tr>
<tr>
<td><strong>tongue root</strong></td>
<td>advanced</td>
<td>advanced</td>
<td>normal</td>
<td>retracted</td>
</tr>
<tr>
<td><strong>mandible</strong></td>
<td>advanced, lowered</td>
<td>advanced, lowered</td>
<td>normal</td>
<td>retracted, lowered</td>
</tr>
<tr>
<td><strong>mandible</strong></td>
<td>advanced, lowered</td>
<td>advanced, lowered</td>
<td>normal</td>
<td>retracted, lowered</td>
</tr>
<tr>
<td><strong>larynx</strong></td>
<td>lowered</td>
<td>normal or lowered</td>
<td>normal</td>
<td>raised</td>
</tr>
<tr>
<td><strong>rounding</strong></td>
<td>rounder if round</td>
<td>rounder if round</td>
<td>normal</td>
<td>normal</td>
</tr>
<tr>
<td><strong>fauces</strong></td>
<td>expanded</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
</tr>
<tr>
<td><strong>vowel height</strong></td>
<td>closer</td>
<td>←</td>
<td>→</td>
<td>opener</td>
</tr>
</tbody>
</table>

According to Edmondson et al. (2003) the basic vowel system of Bor Dinka is /i e a ɔ o/, but the vowels change depending on the register, as depicted in Table 7.16 (ignoring faucalized register) in connection with morphological alternation:

Table 7.16: Vowel alternations in Bor Dinka (see Edmondson et al. 2003)

<table>
<thead>
<tr>
<th></th>
<th>Modal</th>
<th>Breathy</th>
<th>Harsh</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>i</td>
<td>iɛ</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>eɛ</td>
<td>eɛ</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>aɛ</td>
<td>aɛ</td>
<td></td>
</tr>
<tr>
<td>ɔ</td>
<td>ɔ ɐ</td>
<td>ɔ ɐ</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>ŋ ɐ</td>
<td>ŋ ɐ</td>
<td></td>
</tr>
</tbody>
</table>

Somali (Edmondson & Esling, 2006, p. 175) also has a register-like system contrasting modal and harsh sets with correlated vowel quality (and associated harmony processes): the modal vowel set contains /i e æ ɔ œ u/ (transcription based on Saeed, 1999, pp. 11–16), while the harsh set exhibits /i ɛ a ɔ u/ and accompanying creakiness and raised larynx voice quality.

Overwhelmingly, and typical of the GCLC paradigm, [±ATR] is applied in the analysis of many of the systems described above. For example, Saeed (1999, pp. 11–16)
proposes the feature $[\pm \text{ATR}]$ as the basis for the Somali system such that the modal set is $[+\text{ATR}]$ and harsh set is $[-\text{ATR}]$. Likewise, Gregerson (1976) applies $[\pm \text{ATR}]$ to Mon Khmer.

Another approach would be to apply Lindau’s $[\pm \text{expanded pharynx}]$ feature (1975, 1978) to characterize these patterns, but this still does not satisfyingly account for lingual-laryngeal linkage. Halle (1995, p. 18) and Halle et al. (2000), claim that lingual-laryngeal interaction is a corollary of the structure of their vision of feature geometry, in which the GUTTURAL node exclusively dominates the TONGUE ROOT and LARYNGEAL articulator nodes, implying a phonological lingual-laryngeal affinity for segments specified with TONGUE ROOT features like $[\pm \text{ATR}]$. The problem with this approach is, again, one of causality: are we to assume that it is the representation that causes the lingual-laryngeal interaction? The problem is that the feature geometry formulated by Halle does not intrinsically express the direction of correlation amongst the phonetic properties characterizing all of the registers. There is no a priori (or even a posteriori) reason for the articulatory configuration entailed by $[\pm \text{ATR}]$ (an advanced/non-advanced tongue root) to correlate with any particular laryngeal state. Furthermore, if we are to assume that the patterning is a matter of mental representation, then the features should be coherent with phonetic observations regarding the pattern: presumably this is part of the usefulness of the feature, that it expresses a generalization about the basis of the kind of registers being discussed here. However, several of the cases described in this section do not nicely match the pattern predicted by $[\pm \text{ATR}]$: for example, the occurrence of very retracted, very pharyngeally-narrow $[\alpha]$ with the chest register (typically analyzed as $[+\text{ATR}][+\text{expanded pharynx}]$) and less retracted, possibly-even-somewhat-fronted $[\alpha]$. 
with the head register (typically analyzed as $[-\text{ATR}]/[\text{expanded pharynx}]$) should cause us to ask whether $[\pm\text{ATR}]$ or $[\pm\text{expanded pharynx}]$ are really the best analytical tools for languages with “chest”/“head” registers, like Akha. Furthermore (returning languages discussed in §7.1), it is desirable to relate Bai “harsh” register to “growling” in Zhenhai Wu to “sphincteric”/”epiglottalized” in Khoisan because these are effectively produced with one and the same mechanism – epilaryngeal vibration (ignoring vocal fold state). However, neither $[\pm\text{ATR}]$ or $[\pm\text{expanded pharynx}]$ capture the essence of all of these patterns since it is not just the pharynx that is implicated in the contrasts and framing it in terms of pharyngeal stricture unnecessarily locks us into a particular view of how vowels ought to behave in the different registers.

The point is that $[\pm\text{ATR}]$ (and GCLC analysis in general) does not properly relate the changes occurring at the laryngeal level with those at the lingual level. $[\pm\text{ATR}]$ does not convey voice quality: if anything, it conveys so-called tense-lax vowel quality. Attempting to circumscribe all of these different patterns with different vowel features such as $[\pm\text{low}]$, $[\pm\text{back}]$, and so forth, may be an appropriate task for an Emergent Features (Mielke, 2004) approach, and it may be that this enterprise would succeed at the scale of individual languages. Such an approach, however, does little to provide broader generalizations about what all of these patterns have in common, and it would not help us understand why they differ where they do because there would be no overarching theme connecting the patterns together.

In the LPP, there is a unified analysis: all of these patterns reflect various instantiations of register fundamentally characterized by phonetically contrasting states of the epilarynx. As noted above, Edmondson & Esling (2006, p. 175; also see Edmondson,
Padayodi, Hassan, & Esling, 2007) suggest the feature [±constricted], which Miller et al. (2009, p. 141) regard as being applicable to epiglottalized vowels in N|uu. The approach in the LPP, which deals in potentials not features, is to mirror and extend this analysis to every case discussed in this section. Thus, the interpretation here is that the different languages discussed above represent different realizations of epilarynx-related phonological potential. Figure 7.14 depicts these unconstricted (a) and constricted (b) epilaryngeal potentials: absence of epilaryngeal stricture and association with all gross states that act against (or are anti-synergistic) with epilaryngeal stricture (i.e. anti-constriction synergies) defines the unconstricted registers (which could be realized in a variety of related ways in terms of vowel and phonatory quality); conversely, the constricted register is defined by the presence of epilaryngeal stricture and is associated with a constellation of gross states that synergize with epilaryngeal stricture (i.e. pro-constriction synergies). Moreover, Figure 7.14 depicts these two basic potential registers as opposites because many of the gross states differentiating the two registers are anti-synergistic (e.g. \{↓lx \cdots ↑lx\}, \{vto \cdots vtc\}, and \{vfo \cdots vfc\}).
Figure 7.14: Potential unconstricted-constricted registers. Diagram simplifications:

Brackets = broad possibilities for association (not specific in this abstract case); Gray boxes = synergistic relations (see Figure 6.5). Devices (simplified/not fully decomposed):

$\llbracket V \rrbracket =$ modal; $\llbracket Y \rrbracket =$ breathy; $\llbracket Y \rrbracket =$ faucalized register; $\llbracket Y \rrbracket =$ creaky; $\llbracket Y \rrbracket =$ harsh (without epilaryngeal vibration); $\llbracket Y^t \rrbracket =$ “raised larynx voice”. Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; vto/vtc = relatively open/close (anterior) vocal tract; dbe = “double-bunching” epilaryngeal configuration; tfr = tongue fronting (forward); tre = tongue retraction (down and back); tra = tongue raising (up and back); ↑lx/↓lx = raised/lowered larynx. (n.b.: many synergistic relations have been omitted from this diagram; see Appendix F, Figure 8.1 for a more comprehensive version of the diagram.)
In the Khoisan-group languages, the retracted vowels [a ɔ] appear to be the vocalic “anchor” of the constricted registers, but [eɨ] and [uɨ]/[oɨ] are also attested. X-ray evidence in §5.4.2 of pharyngealized and sphincteric vowels in !Xóõ and auditory impression of epiglottalized vowels in Nǀuũ suggest that these languages may involve engagement of the neo-tongue–neo-pharynx (or raised larynx voice quality) configuration similar to pharyngealization in Caucasian (hence the inclusion of {db} in Figure 7.14). Bai represents one of the clearest cases of a language preserving voice quality despite engagement of epilaryngeal stricture, which emphasizes the need to divorce our conceptualization of these patterns from the linguocentric perspective. The potential registers are associated with retracted, open vowels by synergistic relation but nothing precludes these vowels from being paired with vowels like /i/ in any form of the constricted register, although we would predict this would be less typical than restriction or vowel-quality compromise. Zhenhai Wu is very similar to Bai if we accept the analysis that Yang tone in general involves epilaryngeal stricture (see §7.1.4). The LPP can predict the tendency for growl to occur in relatively open vowel contexts: growl is synergistic with epilaryngeal stricture, epilaryngeal stricture is synergistic with retraction and retraction is synergistic with an open vocal tract, or \{epc ↔ epv\}, \{epc ↔ tre\} and \{tre ↔ vto\}. The unusual tendency for /u/ to be associated with the unconstricted register in Dinka, especially since it is a breathy register, substantiates \{tra ↔ vtc\} in connection with \{↓lx ↔ vtc\} and \{↓lx ↔ vfo\}.

Languages like Akha, which have a head register that correlates with condensation or centralization of the vowel space and a chest register with expansion or peripheralization of vowel space, strongly suggest that the head register vowel space is
characterized by “double bunching” epilaryngealization \{d\text{be}\} (implying the NT-NP configuration and likely constricted phonatory quality, such as harsh or creaky). The appearance of [a] in the head/constricted register is no longer a puzzle but rather realization of a potential type of epilaryngeal-based register contrast (and predicted by the LPP), and echoes similar changes described for Caucasian languages and for pharyngeals in general being associated with a more fronted low vowel (see §7.3.1). The occurrence of [a] in the chest register, on the other hand, illustrates the need to differentiate between pharyngeal and epilaryngeal constriction (which is emphasized in the LPP). Neither [+ATR] nor [+expanded pharynx] can sensibly be used to predict that this vowel would likely occur in Akha’s chest register: [a] is the vowel with the greatest pharyngeal tube constriction and those features denote pharyngeal expansion (if they have any meaning at all). The difference between the registers is a matter of epilaryngeal constriction – not just the pharyngeal constriction. Finally, the Dinka dialects and Somali are cases further supporting the need for a model, such as the LPP, which associates vowel quality changes to particular registers but which is not fixated around such changes. Agar and Luanyjang Dinka registers are phonatory-quality rather than vowel-quality dominant, allowing the combination of most vowel qualities – both [+ATR] (e.g. [e o]) and [−ATR] (e.g. [ɛ ɔ]) – with the distinctive phonatory qualities; Somali and Bor Dinka on the other hand present a greater degree of vowel-quality–phonatory-quality correlation. However, Bor Dinka is an extreme case where the register system has subdivided the continuum of (epilaryngeal) constriction into four registers rather than two: faucalized and harsh represent extreme opposite ends of this spectrum.
All of the register possibilities discussed here are potential pathways for phonologies to develop. In the LPP, it is a matter of which synergy becomes dominant in the evolution of a given language (as far as articulatory forces of language change are concerned). The LPP provides a means to predict the pathways.

In summary, although the exact phonetic details differ, these languages all exhibit register-like systems involving a constellation of correlated phonetic properties which can be traced to whether or not there is engagement of the epilarynx. Traditional accounts apply labels such as head/chest, tense/lax, expanded/constricted pharynx or +ATR/–ATR; a better interpretation is that these are (epilaryngeally) unconstricted/constricted registers. It is not possible to generalize across all cases with regard to a particular tongue-pharynx feature like [±ATR] or [±expanded pharynx] because the state of the pharynx does not intrinsically characterize these systems: one needs to consider the epilarynx. There is a bias towards relatively open, relatively retracted vowels, but vowel quality does not strictly define the basis of these contrasts. In the next section, this analysis will be taken one final step further: reinterpreting “ATR harmony” languages in terms of epilaryngeal potentials.

7.3.3 “ATR languages” and their relationship to register

This section examines the nature of ATR harmony in the context of epilaryngeal operation. The argument is that ATR-type systems are vowel oriented but, on many occasions, are similar to register-like patterns in that they involve more than just vowel quality modulation. The implication is that ATR systems are the vocalic extension of the phonological potential of the epilarynx. An LPP account of ATR is provided and then extended further by a demonstration of how ATR fits into a broader set of potential contrasts that includes consonantal laryngeal contrasts, register systems, and ATR under one framework. At the center of this broad picture is the lingual-laryngeal coupling inherent in epilaryngeal operation and the associated synergistic relations.
Many languages of Africa possess a cross-height vowel harmony system that is commonly thought to be a function of tongue root advancement, encoded as the feature [±ATR] (Welmers, 1946; Halle & Stevens, 1969; Stewart, 1967; Painter, 1973; Lindau, 1975, 1978; Denning, 1989; Archangeli & Pulleyblank, 1994; Casali, 2008). Unlike many of the languages discussed thus far, which have received comparatively little attention, “ATR languages” and “ATR harmony” have been the subjects of intense debate in the phonological literature for many years. (To constrain what could otherwise be an intractable subject for this work, our attention will be confined to a very small number of cases; issues such as morphophonological scope, spreading direction, and dominance will be neglected here.)

To start, it will be useful to consider a typological sketch of ATR systems of the sub-Saharan languages of African. Casali’s (2008) comprehensive review of ATR-harmony languages provides just such a sketch and it has been summarized in Table 7.17:
Table 7.17: “ATR” Systems based on Casali (2008)\textsuperscript{260,261}

<table>
<thead>
<tr>
<th>Type</th>
<th>+ATR</th>
<th>–ATR</th>
<th>Location / Families</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-V</td>
<td>i ə o u</td>
<td>i ə a o</td>
<td>Saharan</td>
<td>Akposso, Bongo, Diola-Fogny</td>
</tr>
<tr>
<td>9-V</td>
<td>i e o u</td>
<td>i e a o</td>
<td>NC, NS</td>
<td>Akan, Maasai; very common</td>
</tr>
<tr>
<td>7-V: ‘1IU’</td>
<td>i e o u</td>
<td>e a o</td>
<td>West &amp; Central Africa</td>
<td>Yoruba; Wolof; common</td>
</tr>
<tr>
<td>7-V: ‘2IU’</td>
<td>i u</td>
<td>i e a o</td>
<td>East Africa</td>
<td>Kinande; relatively rare</td>
</tr>
<tr>
<td>5-V</td>
<td>i u</td>
<td>e a</td>
<td>West Atlantic, Bantu</td>
<td>Pulaar, Tsonga, Zulu</td>
</tr>
</tbody>
</table>

(NC, NS = Niger-Congo, Nilo-Saharan)

It is not particularly common for ATR languages to have a phonemic [±ATR] contrast for the “lowest” vowel in the system, usually transcribed [a]. This vowel is, in relation to phonetic and phonological properties, regarded as [–ATR], although in [+ATR] contexts, it can exhibit phonetic properties of a [+ATR] vowel, such as tongue root advancement (Gick, Pulleyblank, Campbell, & Mutaka, 2006; Casali, 2008, p. 532). However, the /a/ vowel is not always realized phonetically in a way consistent with that predicted by the feature [±ATR]. For example, Tugen Kalenjin has a 10-V system with [+ATR] low vowel contrast. Lodge & Local (2004, pp. 13–14) note that, although rarely commented on in the literature, the quality of these vowels is “counter-intuitive” in light of [±ATR] analysis: the [+ATR] vowel is [a] and the [–ATR] vowel is [a]. They emphatically conclude\textsuperscript{262} that “in Tugen the tongue body position is clearly not DETERMINED by the size of the pharynx” and thus the low vowels require a “contrary

\textsuperscript{260} Although Casali restricts his attention to “underlying” inventories, he includes the 5-V systems, which are purely allophonic ATR systems (Casali, 2008, pp. 503–504).

\textsuperscript{261} Also of interest are the neutrality patterns of the different vowels. The high vowels /i u/ can be neutral in 1IU systems, such as Wolof and Lokaa. The vowel /a/ is most often regarded to be neutral, which is attested in several 9-V and 2IU languages. Despite this, /a/ does not have ambiguous phonetic properties (for example, Casali notes that it still has the characteristic [–ATR] voice quality), and its phonological status as “[–ATR]” is not in doubt (p. 529): when it appears alone in a root, it causes affixes to become [–ATR]. Furthermore, Gick et al. (2006) have shown that, even when /a/ behaves transparently, as in Kinande, it is still phonetically affected by harmony; for example, it is produced with greater tongue advancement in [+ATR] contexts.

\textsuperscript{262} Lodge & Local (2004, pp. 13–14) consider this to be evidence against the “Intrinsic Phonetic Interpretation” hypothesis, which holds that features intrinsic or intuitive interpretation.
interpretation of [±ATR] to [the] interpretation for the non-low vowels – not a happy conclusion for universals of phonetic implementation” (p. 13). We should recognize that this pattern is similar to that identified in Akha (see §7.3.2): rather than [±ATR] vowel sets, the vowel sets are peripheralized/condensed. The implication of the Kalenjin case is that pharyngeal volume does not cleanly characterize cross-vowel register-like patterns in either language, but these vowel sets are consistent with the changes associated with the neo-tongue–neo-pharynx configuration.

Lindau (1975, 1978) surveys and provides an articulatory analysis (based on x-ray imaging) of the [±ATR] vowel sets in Akan, Igbo, and Ijo (Niger-Congo), Ateso and Dho-lo (Nilo-Saharan), and Shilluck and Dinka (Nilotic). Figure 7.15 provides Lindau’s tracings for Akan (a) and Ateso (b). Lindau observes that, in Akan, a combination of tongue root advancement and larynx lowering is used for the [+ATR] vowels, while the [−ATR] vowels have a somewhat neutral or retracted tongue position and larynx raising. The Ateso speaker appears to use neutral larynx height in both sets, which differ largely by the antero-posterior position of the tongue. Note that, judging from the midsagittal profile, it is evident that it is the whole tongue which displaces not just the “tongue root”.

These x-rays also show some signs of epilaryngeal stricture, particularly for the front Ateso vowels, /ɛ/ and /i/. In the Akan traces, it appears that the speaker is simply raising the larynx, although a shift in epiglottis orientation and position suggests epilaryngeal engagement. Laryngoscopic evidence for Akan (Edmondson, Padayodi, Hassan, & Esling, 2007, p. 2067; also see Edmondson, 2009) and Kabiye (Edmondson & Esling, 2006, pp. 180–181; Edmondson, Padayodi, Hassan, & Esling, 2007, p. 2067; also see Edmondson, 2009) confirms epilaryngeal constriction occurs.
Based on these observations, Lindau argues that [±ATR] systems are not a function of tongue root activity (the traditional analysis); rather, they appear to integrate pharyngeal cavity expansion, which she claims is controlled both by the tongue root and the larynx. (Similar conclusions are reached by Painter (1973) for Twi/Akan.) Consequently, Lindau proposes the feature [±expanded pharynx], which is phonetically implemented as a combination of tongue retraction and larynx height, accounting for phonetic variation. Tiede’s (1996) MRI study of Akan and English [±ATR] vowel sets supports Lindau’s observation that it is the pharyngeal volume that distinguishes the vowel sets for Akan, while the difference in the English tense/lax vowel contrast is primarily one of tongue height, with generally less extreme pharyngeal volume changes than for the Akan vowels. Tiede notes (1996, p. 419), however, that the English and Akan patterns diverge at the level below the epiglottis: while Akan consistently expands or
constricts the hypopharynx in correlation with [+ATR] or [−ATR] vowels, English only shows marginal and inconsistent variation of this region. Although Tiede collapses measurement of epilaryngeal area with piriform fossae cross-sectional area, such observations strongly support the view that epilaryngeal involvement is important in the ATR contrast, at least in Akan, as Edmondson et al. (2007, p. 2067) attest.

As the terms *ATR language* or *ATR harmony* and Casali’s typological description (Table 7.17) imply, the *de facto* analysis of ATR phenomena is that it is a vowel quality contrast. However, some earlier researchers framed the phenomenon in terms of “voice quality contrast” or “voice-quality harmony” (Tucker, 1975; Jacobson, 1980). Furthermore, it is frequently reported that [+ATR] vowels correspond (approximately) with breathiness and [−ATR] vowels correspond (approximately) with creakiness (Stewart, 1967; Painter, 1973, p. 117; Lindau, 1975; Hall & Hall, 1980; also see Czaykowska-Higgins, 1987, pp. 3–5; Denning, 1989; Trigo, 1991, pp. 119–120; Keyser & Stevens, 1994, p. 213; Guion, Post, & Payne, 2004; Casali, 2008, p. 510). Although Casali (2008) suggests that phonatory quality is “more subtle than some of the impressionistic labels might imply” (p. 510), he also suggests that phonatory quality is probably more common than descriptions of ATR languages let on, especially in West African languages and in East African Nilo-Saharan languages. Lodge & Local (2004, p. 6) point out, however, that the pattern is not universal. Tugen Kalenjin (and possibly Sabaot Kalenjin) atypically implements voice quality in correlation with ATR vowel sets: [−ATR] vowels actually involve somewhat more breathy phonation than the [+ATR] vowels, which is the opposite of the conventional pattern.
Trigo (1991, p. 119) argues that larynx raising (LR)/larynx lowering (LL) mechanically influences the vocal fold aperture, such that lowering opens the glottis and raising closes it\(^{263}\) and consequently explains “why [LL] vowels tend to be breathy voiced while [RL] vowels tend to be pressed voice” (assertions which closely agree with the LPP). Trigo then discusses Turkana, which has a [–ATR] “head” register with [ɪ e ɔʊ] and a [+ATR] “chest” register with [i e o u]; quoting Dimmendaal (1983, p. 18; from Trigo, 1991, pp. 133–134, f.n. 15), “vowels with the feature [–ATR] have a hard voice phonetically … [+ATR] vowels normally sound somewhat breathy”; in Noske’s (1996) interpretation “[–ATR] vowels of Turkana sound ‘choked’, back vowels more so than front vowels” (p. 64), but he also claims that the [+ATR] vowels lack breathy quality. Following an observation by Dimmendaal, Trigo points out that, in certain contexts, the head and chest registers coalesce on [–high, –low] vowels, yielding “[ɛ ʊ]” which are produced with harsh phonatory quality according to Dimmendaal (1983; also see Trigo, 1991, p. 120), as the underdot implies in Dimmendaal’s transcription. Noske (1996, p. 92) dismisses both Dimmendaal’s observation about these vowels being harsh and Trigo’s suggestion that they represent a mixed register\(^{264}\).

---

\(^{263}\) Trigo’s (1991, p. 119) evidence for this argument comes from experiments conducted on the larynges of mongrel dogs by Shin et al. (1981), which confirms that closure can occur even when intrinsic laryngeal musculature is “deleted” through dissection of the recurrent laryngeal nerves. Oddly, Trigo suggests that the Shin et al. study involves closure at the aryepiglottic level, yet Shin et al. do not discuss this level in relation to laryngeal closure. The parameters Shin et al. measure for laryngeal closure are vocal fold length, interligamental and interarytenoid distance, and glottal area. It is possible that Trigo mistakenly inferred that Shin et al. actually removed the intrinsic laryngeal musculature, as suggested by her comment “it is possible to open/close the glottis after removal of the intrinsic laryngeal musculature” (p. 119). Rather, Shin et al. merely disabled these muscles in their experiment by severing the nerve that innervates them; they did not remove the muscles.

\(^{264}\) Trigo (1991, p. 120) suggests that the mixed register vowels in Turkana plausibly represent a combination of [+ATR] with her [raised larynx] feature. This could be reframed, however, in terms of the debate in the literature regarding the relationship between [±ATR] and [±RTR]. Goad (1991, 1993, pp. 2–3) argues that these features are both monovalent and have different locations in feature geometry to reflect their different roles: [ATR] is a vowel height feature (e.g. /i/ vs. /u/) and [RTR] is dependent of [pharyngeal]
Edmondson & Esling (2006, pp. 178–181) point out, however, that voice quality or register is the key correlate of [±ATR] vowels, which does not necessarily implicate strict phonatory quality correlates. In support of this, they point out that the Kabiye speaker with whom they worked produces [−ATR] vowels with all the hallmarks of epilaryngeal constriction. The exception, however, is that the ventricular folds do not medialize in the “constricted” set (although they may still come into contact with the vocal folds below by descending), and, consistent with this visual observation, is the auditory impression that, while phonatory quality of the [−ATR] set does not seem notably different, it “unmistakably involves ‘raised larynx voice’” (p. 180). Harsh or pressed voice and raised larynx voice are both functions of epilaryngeal stricture.

place but can specify both vowels and consonants. Thus, for Goad, [+ATR] dominant harmony is simply [ATR] harmony: [−ATR] dominant harmony (e.g. in Yoruba) is reinterpreted in terms of height (Goad, 1993, pp. 150–183). Rose (1996, pp. 89–90) takes this a step further and asserts that [ATR] is strictly for vowel contrasts and [RTR] is strictly for consonant contrasts, but it can spread to vowels. So Turkana’s mixed register, if it exists at all (cf. Noske, 1996, pp. 91–92) could be reinterpreted as a combination of [ATR] and [RTR] (Goad, 1993, pp. 22–25, Ch. 2 f.n. 9, 121–124, Ch. 4 f.n. 13).

Critical to this debate is another aspect of Turkana phonology and a parallel pattern in Akha (1991, pp. 122–123), involving velar retraction in the context of the vowels /a o/ (Dimmendaal, 1983; Trigo, 1991, pp. 122–123; Noske, 1996, pp. 64–66). In Turkana, retraction occurs when /k/ is tautosyllabic with one of these vowels (e.g. /na-kima-k/ [ɲa.kɪ.mɑq] ‘old women’ and /ni-kodjo/ [ɲi.qod.jo] ‘taxes’), and if the stop is flanked by these vowels, it spirantizes to [x], [ɣ], or [χ] (e.g. /a-kɔk/ [ɑχɔχ] ‘to stay’). In Akha (Trigo, 1991, p. 123), /a o/ equivalents in either head or chest register cause retraction of /x/ to [χ] (using Trigo’s transcription, e.g. [xhɛ] ‘classifier for doors’ and [xɛʕ] ‘to break’ but [χhɔl] ‘trough’ and [χɔʔ] ‘to draw water’. The set /a o/ is traditionally assigned the conjunction of feature [+back, –high]: issues concerning the underlying representation of uvulars using these features aside, the same “natural class” is provided by referring to vowels involving tongue retraction or {rtd} (backwards and downwards motion of the tongue).

Both of these cases provide further evidence that vowels in the neighbourhood of [a o] and uvulars both involve lingual retraction and that this is a different matter from what is involved in Akha register or Turkana cross-height harmony vowel sets. Furthermore, these cases are also consistent with the reinterpretation of Arabic emphasis as uvularization, wholly different from pharyngealization-proper (epilaryngealization).

Bifurcating tongue root behaviour into [ATR] and [RTR] can express that velar retraction and ATR-harmony/register are orthogonal processes, but forces us into assuming, as Goad (1993, pp. 23–24) does, that ATR-type harmony only ever involves tongue root advancement. This is the essence of a GCLC account, which has no means of relating lingual activity to laryngeal activity because it is confined to phonetically inadequate features. The “choked” quality of Turkana’s so-called [−ATR] vowels is /ɪ ɛ ɔʊ/, the raised larynx voice quality in Kabiye, and the visually evident epilaryngeal constriction in Akan are just as valid phonologically as the facts relating to vowel quality. We need a model that can contend with both sets of correlates.
As far as the LPP is concerned, the challenge of ATR is to relate the vowel quality correlates with phonatory correlates and to provide some insight into the role of particular vowels within an ATR system. No previous phonological theory has been able to adequately “ground” ATR around a physiological mechanism that can coherently relate these properties, although Trigo (1991) came very close, and Pulleyblank & Archangeli (1994) present a reasonable model for how [±ATR] interacts with vowel height and backness. These models do not adequately model the lingual-laryngeal relationship because they do not include the epilarynx, which is the physiological mechanism that links vocal fold state, larynx height, and lingual state together.

First, it is useful to abstract away from language-specific and speaker-specific variation, the overarching pattern involves the intersection of vowel quality, phonatory quality, and overall voice quality. ATR resembles many of the register phenomena discussed in the previous section (§7.3.2) and echoes the behaviour of (some) pharyngeals too (§7.3.1). All of these can be related by voice quality types that correlate with epilaryngeal constriction, but each has its own orientation: register is more oriented around phonatory quality and somewhat orthogonal to vowel quality; pharyngeals involve extreme epilaryngeal stricture with phonatory and vowel quality effects serving a subsidiary role; ATR represents the most vowel-quality oriented of the three.

With this in mind, consider the schematic in Figure 7.16, which depicts the relationships involved in ATR for a handful of illustrative (unrounded) vowels (a more complex diagram is provided in Appendix F, Figure 8.2). The typical vowels associated with [±ATR] fall on the appropriate side of the “ATR pivot” (dashed gray line). This line also divides gross states according to their synergistic relation with epilaryngeal stricture.
(to keep the diagram simple, not all associations are shown, but they would be consistent with those in Figure 6.5). Thus, [+ATR] vowels (Figure 7.16a) are associated with the gross states that synergize with epilaryngeal anti-constriction (including \{tfr\}), and [−ATR] (Figure 7.16b) vowels are associated with gross states that synergize with epilaryngeal constriction (including \{tre\}). This is the key articulatory basis for potential association of ATR-like systems with correlated phonatory and voice quality effects: it is a relationship mediated by the state of the epilarynx – not the pharynx. In the LPP, gross states relevant to phonation, such as \{vfo\}, are not directly associated with gross states that are highly significant in vowel production (such as \{vto\}, \{vtc\}, \{tre\}, \{tfr\}, and \{tra\}; see Figure 6.5). This signifies that ATR, unlike phonatory register, is vowel-oriented, not phonation-oriented.
Figure 7.16: Cross-height harmony ("ATR") in the LPP. Grayness of the association lines represents gross-state association strength (e.g. $[\varepsilon]$ involves greater tongue fronting \{tfr\} than retraction \{tre\}); Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; vto/vtc = relatively open/close (anterior) vocal tract; tfr = tongue fronting (forward); tre = tongue retraction (down and back); ↑lx/↓lx = raised/lowered larynx. (See Appendix F, Figure 8.2 for a more comprehensive diagram.)

As Casali’s typology (Table 7.17) indicates, “ATR” is not often realized as a simple cross-height system (as in 10-V systems), but rather manifests in asymmetric or “imbalanced” ways (which is true for all of the other, non-10-V, ATR-type systems listed in Table 7.17). The claim here is that the “ATR” vowel system is “imbalanced” across
“the axis of [eo]”\textsuperscript{265} pivot (gray dash-lined box in Figure 7.16; \textit{n.b.:} there should also be an [o]/[ɔ] but rounded vowels have been omitted from this diagram to maintain simplicity; see Appendix F, Figure 8.2 for a more comprehensive diagram), which roughly separates vowels with \{v tc\} from those with \{v to\} gross states. The axis of [eo] is yet another potential pivot, but it is one that exerts its main influence in the vowel domain.

To the left (referring to Figure 7.16) of the axis of [eo] are the “high” vowels (such as /i/) and to the right are the “low” vowels (e.g. /a/). From Casali’s typology (Table 7.17), we know that if any [+ATR] vowels exist in the system, they will most likely be /i u/: these vowels are found to the left of the axis of [eo]; likewise, the canonical [–ATR] vowel is /a/, which lies to the right of the axis of [eo] (in Figure 7.16). Thus, “high” vowels – those that are strongly \{v tc\} – are attracted or biased towards the [+ATR] set; “low” vowels – those that are strongly \{v to\} – are attracted or biased towards the [–ATR] set.

Returning to the epilarynx: in the broadest sense, epilaryngeal stricture can apply to any vowel height: thus, in Figure 7.16, the “ATR” pivot is depicted as cross-cutting the entire vowel system (keep in mind that not all vowels are shown in Figure 7.16; see Appendix F, Figure 8.2). However, the relevance of the axis of [eo] to epilaryngeal stricture is the following: through synergistic relations, the epilarynx is aligned with vowels on the right side (in Figure 7.16) of the axis of [eo] – the relatively open vowels. As discussed in previous sections (e.g., see §7.2.5), relatively open vowels, especially retracted vowels like /a/, have a particular affinity for epilaryngeal stricture, which is

\textsuperscript{265} The idea of “the axis of [eo]” is credited to Esling (personal communication).
expressed through combinations of synergistic relations, such as \{\uparrow \text{lx} \leftrightarrow \text{vto}\}, \{\text{tre} \leftrightarrow \text{vto}\}, \{\uparrow \text{lx} \leftrightarrow \text{epc}\}, \{\text{tre} \leftrightarrow \text{epc}\} (not all of which are depicted in Figure 7.16). The vowels on the left side of the axis of [eo] – the relatively close ones – are anti-constricting in their synergistic relations: these vowels synergize with larynx lowering \{\downarrow \text{lx} \leftrightarrow \text{vtc}\} – an anti-constricting gross state that “undoes” or opposes epilaryngeal stricture \{\downarrow \text{lx} \cdots \text{epc}\} and synergizes with vocal fold abduction \{\downarrow \text{lx} \leftrightarrow \text{vfo}\} (also anti-constricting in nature \{\text{vfo} \cdots \text{epc}\}).

If we believe that ATR can represent the intersection of vowel quality and phonatory quality, then we need a model that explains the concomitancy of both of these properties and accounts for the imbalance of vowel quality across the axis of [eo]. In Generative Phonology, one must posit complex cognitively-bound rules or constraints for explaining the relationship between [±ATR], [±low], and [±high] and then connect these mental properties to phonatory quality – which is typically neglected in such accounts. The LPP account is oriented around the two basic states of the epilarynx and the synergistic relations associated with these states. The same system applied to all other cases discussed thus far extends to making the lingual-laryngeal connection in ATR systems – at least potential ATR systems.

Voice quality can involve more or less subtle changes in ATR phonetics. Raised larynx voice quality, harshness, and tenseness are consistent with gross states associated with (and including) epilaryngeal stricture and will most likely occur with [−ATR] vowels; slackness, breathiness, lowered larynx voice: all of these are predicted to normally occur with the [+ATR] vowels. Do ATR languages necessarily implement these properties? No – languages evolve and change. The strong prediction is that [+ATR] will
be associated with voice quality properties that are consistent with an unconstricted epilaryngeal configuration, while [–ATR] will show the opposite association. In the majority of cases, this appears to be true.

We might wonder about Tugen Kalenjin, which defies some of the usual patterns. The fact that [ɑ] occurs in the [+ATR] set in the language is similar to Akha (see §7.3.2), and the explanation in the LPP is that peripheralization/condensation is a potential correlate of an unconstricted/constricted epilarynx (not to imply that epilaryngeal stricture occurs in Kalenjin, which may plausibly be organized around the unconstricted state). However, if the [–ATR] set indeed is breathy (and not whispery), then Kalenjin represents a reversal of the pattern and an exception to the LPP (we would expect breathy with the [+ATR] set), one that is rare and merits further investigation (although, once again, the data are from a single speaker). It should be remembered that, by talking about potentials, we are free from framing phonology in absolute terms. Individual languages will gravitate towards certain phonological potentials defined by the epilarynx and its operation. However, not every case will end up the same because many other factors operate in shaping the sound system (as will be elaborated upon below).

An extension of the LPP account of ATR is to follow up Ohala’s266 (1994, p. 492) not entirely uncontroversial, but nonetheless very intriguing suggestion that cross-height/ATR systems originated as (and, as we have seen, often persist in being) a phonation-type contrast. Ohala’s proposal is especially interesting in light of Denning’s (1989, p. 9) vocal-register implicational universal, which states that, if there is a relationship between vowel height and voicing or phonation, the “laxer” the phonation

266 Ohala refers to ATR harmony as cross-height harmony, which, in light of the present work, strikes a note of greater appeal than the “tongue root” feature.
associated with the vowel, the greater the vowel’s height. Denning’s universal is based on a survey that includes numerous languages from Sino-Tibetan, Austroasiatic, Afro-Asiatic, Niger-Congo, Nilo-Saharan, Khoisan, Indo-European, Central Amerind and Na-Dene language groups. (In the LPP, Denning’s universal immediately relates to the posited synergistic relationships connecting (partially) abducted phonation (“lax” or “breathy”) via larynx lowering to relatively close vowels: \{\downarrow lx \leftrightarrow vfo} and \{\downarrow lx \leftrightarrow vtc\}267.)

Perhaps the most tantalizing aspect of Denning’s proposal is that it implies a possible pathway of evolution for ATR-like systems through not just phonation type but also consonantal laryngeal contrasts. Denning (1989, pp. 50–55) presents several cases, including Javanese, Lungtu, Lhasa Tibetan, Mnong, Murle, Middle Khmer, Mon, English and Buchan Aberdeenshire Scots (attributable to Stewart, 1967), for which consonantal voicing correlates with relatively higher/closer vowels, lower tone level268, and “laxer”/breathier phonation (with phonetic details varying by language; cf. Huffman, 1976). Much of this evidence reflects Haudricourt’s (1946; also see Gregerson, 1976, pp. 328, 342, 333; Huffman, 1976; Trigo, 1991, pp. 128–129) theory that Mon-Khmer register originated as voicing contrast, such that voiceless (“tense”) stops are associated with [-ATR] “head” register and voiced (“lax”) ones with [+ATR] “chest”/“breathy” register. Cohn (1993) describes the same relationship in Madurese, noting that voicing or

267 The larynx lowering adjustment has the corresponding acoustic-auditory “synergy” with high/close vowels by causing lowering of F1, which we might presume will increase the height/closeness of the perceived vowel (and the opposite is true for \{\uparrow lx \leftrightarrow vto\}).

268 Not directly addressed in this section (but see Figure 8.3, Appendix F) is the role of consonant voicing in tonogenesis: high and low tones emerge distinctively on vowels following syllable-initial voiceless and voiced stops, respectively. Ewan (1979, pp. 60–82) provides evidence that larynx height is typically lower for voiced consonants, and attributes the tonogenesis patterns to this larynx-height differentiation of consonantal voicing.
aspiration conditions [+ATR] vowels and unaspirated stops condition [–ATR] vowels: for example, \([p\text{ɛl}]\) ‘choose’, \([\text{birn}]\) ‘shy’, and \([p^h\text{ip}^h\text{it}]\) ‘seed’ (p. 107). In the same vein, Vaux (1996), buttressing earlier arguments made by Trigo (1987, also see 1991, pp. 128–131) and citing further evidence from Babine, Jingpho, Kirzan Armenian, and southern dialects of Akan, argues that voiced obstruents are [+ATR] (but voiceless ones not necessarily [–ATR]).

Thus, there are considerable grounds in the literature for drawing a connection across consonant voicing, phonatory quality, tonal quality, vowel quality, and overall voice quality. However, the feature \([±ATR]\), being GCLC, does not gracefully capture these facts. Even though \([±ATR]\) can be grafted on to the representation of consonants, as Vaux attempts to do, it does not clearly relate to phonation in its phonetic implications leaving us with an ad hoc analysis, that voiced obstruents are [+ATR]. This immediately clashes with proposals that [ATR] is strictly a vowel feature (Goad, 1991, 1993; Rose, 1996, pp. 89–90), and leaves us with the unfortunate situation of representing voiced uvulars and emphatics as simultaneously [+ATR] and [+RTR]. A proper model of the connection between lingual and laryngeal activity is required and that model centers around epilaryngeal state.

It could be contended, as it is in the LPP, that the connection amongst all of these properties – consonant voicing, phonation type, vowel quality, tonal bias, and voice quality – is larynx height, as understood in relation to epilaryngeal stricture. Larynx

---

\(^{269}\) Acoustic/auditory explanations for a relationship between phonatory and vowel quality also can be noted here. Breathier phonation vis-à-vis tense or creaky phonation will result in a sharper decline in harmonic intensity as a function of frequency such that relatively more acoustic energy is concentrated in the low frequency region. This may possibly predispose the perception of a higher vowel quality. The reverse causality may also hold, such that relatively close vowel constrictions cause increased spectral damping in the upper frequency range and thus are somewhat more likely to be associated with a breathy phonatory quality.
lowering can be viewed as facilitating consonant (mainly obstruent) voicing, via the robustly substantiated aerodynamic voicing constraint (Trigo, 1991, p. 130; Ohala, 2011). It can also be viewed as favourable for vocal fold abduction (see §2.1.2 and §4.2.5). Larynx raising has the opposite effects of inhibiting phonation and enhancing vocal fold adduction. The LPP does not directly address matters of sustaining or inhibiting phonation, but such relationships are not inconsistent with the proposals of the LPP.

Rather than attributing the relationships strictly to mental representation (as in Vaux, 1996), the LPP externalizes the relationships by attributing them to the body. A schematic showing these relationships is in Figure 7.17 (a more comprehensive diagram can be found in Appendix F, Figure 8.3).

Assuming a consonantal starting point, as in proto Mon-Khmer (e.g. Huffman, 1976), we have various possibilities for contrast, but they tend to fall on either side of the contrast pivot line (gray horizontal dashed line). The “lax”/“tense” distinction (e.g. Denning, 1989) is often a voiced/voiceless distinction, but aspirated/unaspirated voicelessness can also be split across the contrast pivot, which is how we will proceed with our illustration. Figure 7.17a depicts a potential consonantal contrast between aspirated (to represent “lax” stops and symbolized as [h]) and unaspirated stops (to represent “tense” stops and symbolized as [l]) that we can take as a starting point. These two types of stops are associated with the anti-constriction and pro-constriction gross states respectively, as depicted in the diagram.

It is typically claimed (e.g. Huffman, 1976) that co-articulation with adjacent vowels begins the process of register formation. The states associated with each incipient register gradually become phonologized on the vowels when the register-inducing
properties bleed-off of neighbouring “lax” (typically voiced or aspirated) and “tense” (typically voiceless or voiced unaspirated) consonants. This typically yields the so-called “chest” and “head” registers. Thus, in the Figure 7.17 example, the consonantal contrast (a) develops into a register contrast (b). The result is a phonatory-quality oriented contrast that may also have correlated vowel-quality changes. In the example, the registers resulting from “lax” and “tense” consonantal contrasts correspond to breathy (⟦Y⟧) and creaky (⟦Y⟧) vowels. The prediction of the LPP is that whatever the exact realization of the system, it will tend to settle into the familiar pattern formed by the opposition between epilaryngeal anti-constriction and pro-constriction.

270 Note that in actual languages, the contrast does not always need to involve extreme settings in terms of epilaryngeal anti-constriction/constriction. It is possible for some systems to have one state that is relatively more neutral and one with a more extreme setting (in either direction).
271 Appendix F, Figure 8.3 presents some extended possibilities. The “chest” register may be associated with low tone, lax/slack or breathy phonation, and vowel quality that occupies the full vowel space (i.e. peripheral, e.g. [i a u]) or there may be a bias towards close vowels. All of these properties can be traced to epilaryngeal anti-constriction largely involving a lowered larynx and/or some vocal fold abduction.

The “head” register will tend be associated with the hallmark properties of epilaryngeal stricture, which owes its provenance to synergizing with larynx raising and vocal fold adduction (see Figure 7.17a & b). If epilaryngeal constriction is not heavily involved, there may be a bias towards high tone (↑lx → H) (not depicted in Figure 7.17, but see). If there is epilaryngeal stricture, phonatory quality in head register will become “constricted”, either involving creakiness or harshness (of the pressed/tense variety), or in more extreme cases, epilaryngeal vibration may result (growling); varying degrees of raised larynx voice are also predicted to be a potential voice quality associated with this register. Vowel quality, which will be entangled with raised larynx voice quality, will tend to be condensed (n.b., not centralized, i.e. not [a]-like) or biased towards being relatively open or retracted, e.g. [i a o].
Only a few potentials are shown (e.g. “ATR” contrast is illustrated with canonical [+ATR] /i/ and [-ATR] /a/). Devices: [h] = vocal fold abduction (voiceless, as in aspirated stops); [], = pre-phonation (voiceless, as in unaspirated stops); [V] = breathy; [V] = creaky. Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; vto/vtc = relatively open/close (anterior) vocal tract; tfr = tongue fronting (forward); tre = tongue retraction (down and back); ↑lx/↓lx = raised/lowered larynx. (n.b.: not all anti-synergies are represented.)

Finally, Figure 7.17c represents the potential for register-like systems to develop into a cross-height contrast (such as “ATR”). The [+ATR]/[-ATR] contrast is illustrated with canonical /i/ and /a/ (purely for illustration; see Figure 7.16). To highlight the tendency for phonatory-quality association, /i/ is depicted as breathy ([V]) and /a/ is depicted as creaky ([V]).
At the point of (c), the system has become vowel-oriented, but there may be concomitant phonatory, tonal, and voice quality correlates. Some languages may start to abandon the non-vocalic properties and shift towards a basic, Germanic-style tense/lax vowel contrast. Others may maintain the non-vocalic properties. The LPP fits with the typical case of ATR-vowel correlates such that [+ATR] vowels /i e ə o u/ tend to be relatively more unconstricted in phonatory quality and voice quality while [–ATR] vowels /ɪ ɛ a ɔ u/ tend to be relatively more constricted in regard to these parameters (Figure 7.17c).

Ohala’s claim is that ATR languages started as a phonatory-register contrast, as in (b), and subsequently evolved into a cross-height contrast, as in (b). His claim is consistent with the LPP. We do not know whether there was an earlier consonantal stage. The only possible case of such a connection is Stewart’s (1967; also see Vaux, 1996) suggestion that there is a relationship between consonant voicing and vowel quality in eastern dialects of Akan. Such a connection is not essential since one might consider the potentials as evolutionary ruts that languages slip into for a time only to be pushed out of later by other forces that shape the sound system.

In short, the LPP predicts that the systemic re-alignment with particular articulatory (and resulting auditory) properties will happen according to the synergistic relations associated with epilaryngeal functioning. In actual cases, languages will merge into and branch off of the development path outlined in Figure 7.17 – not all languages must go from (a) to (b) to (c). Factors outside of the articulatory domain, such as perceptual factors, are likely to play a significant role in the developmental pathway of
individual languages. (Keep in mind that Figure 7.17 represents only a very limited number of potentials and more can be seen in Appendix F, Figure 8.3).

7.4 Discussion: The phonological epilarynx and beyond

We have now completed our tour of those aspects of lower vocal tract phonology that can benefit from articulatory considerations of epilaryngeal function. This section steps back and summarizes what can be accounted for with the Lower Vocal Tract Phonological Potentials (LPP). The three questions posed in §6.3.4 relating to three functions of the epilarynx will be answered and related together in §7.4.1. Then, in §7.4.2, a summary is given of the problems in lower vocal tract phonology that lie beyond consideration of the epilarynx. In §7.4.3, the LPP is compared to other models of phonology and, finally, in §7.4.4, we return to the matter of distinctive features in relation to the LPP.

7.4.1 Summary of the LPP account

The LPP posits three conceptual objects to help us think about lower vocal tract phonology, assuming a more organic, less modular view of “phonetics” and “phonology”: devices, gross states, and synergistic relations. To avoid reification of our conceptual tools, everything is framed in terms of potentials, rather than realities. All of these objects of the LPP theory are abstractions, but they provide a map of the articulatory biases that form a basis for predicting trends in realized phonologies in actual languages.
Potential devices are an abstraction or averaging of activities that the human vocal tract is capable of performing, like epilaryngeal vibration [ʢ]. These devices might share gross (physiological) states such as larynx raising {↑lx}, but it is understood that the neuromotor implementation of larynx raising in one device such as strong ejection [ʡ] may not be the same as in another, such as a raised larynx [ʢ] (for which no unique symbol exists): the former may make greater use of the palatopharyngeus and suprahyoid muscles, while the latter might (tend) to involve more thyrohyoid contraction (although real speakers with real mechanisms will find their own idiosyncratic devices for these actions).

At the core of the LPP are the synergistic relations, which can be used to associate all of the conceptual objects of the theory via the gross states. This is done by means of a schematic, introduced in §6.3.2 (Figure 6.5), which for convenience is repeated in Figure 7.18. The depicted relations are based on physiological knowledge and empirical data, but their usefulness is echoed in the various systems that were studied in this chapter.
In §6.3.4, three questions were posed regarding epilaryngeal function in phonological capacity. What follows is a review of these patterns, which are taken as support for the overall usefulness of the LPP as a model of lower vocal tract phonology:

1) What are the phonological roles and consequences of epilaryngeal vibration? (

Epilaryngeal vibration has a dual nature: it is “quasi-phonation, quasi-trill”. No previous model of phonology can predict that growling would occur distinctively in speech or
have allophonic distribution that is dictated in part by tonal quality and vowel quality: this is because no previous “phonological” model considers the epilarynx.

Understanding growling, as encapsulated in the LPP, provides in-roads to understanding potential interactions between tone systems and vowel quality systems. Growling is (strongly) synergistic with epilaryngeal constriction and the latter has synergistic relationships with larynx raising, vocal fold adduction, relatively open vowels, and a retracted lingual configuration.

The fact that growling tends to be whispery in !Xôô and Zhenhai Wu presents the possibility for the trumping\(^\text{272}\) of one synergy over another. Normally epilaryngeal stricture should favour vocal fold adduction, \(\{\text{vfc} \leftrightarrow \text{epc}\}\), but epilaryngeal vibration, which is synergistic with epilaryngeal stricture \(\{\text{epc} \leftrightarrow \text{epv}\}\), seems to profit from abducted vocal folds, as expressed by \(\{\text{vfo} \leftrightarrow \text{epv}\}\). Certainly epilaryngeal vibration with relatively more vocal fold adduction (e.g. voiced aryepiglottic trilling, found in Bai, N|uu, and Agul) and that with relatively more vocal fold abduction (e.g. whispery voiced or voiceless aryepiglottic trilling, found in !Xôô, Zhenhai Wu, and Agul) are both potentials. (See Appendix F, Table 8.5 for a summary of the claims of the LPP regarding epilaryngeal vibration in phonology.)

\(^{272}\) If this sounds reminiscent of how Optimality Theory works, please see the comment in §7.4.3 which discusses how the LPP and OT are different. In the LPP, this “trumping” would be stochastically modeled unlike conventional OT, which posits variably-ranked and cognitively-bound universal constraints.
2) What is the phonological relationship between the epilarynx and the vocal folds? 

(§7.2)

In glottocentric-linguocentric phonological models, the vocal folds should never interact with anything up above them. These models treat the larynx as a flat entity in a different plane of existence from the rest of the vocal tract: there is a glottis and then there is the supralaryngeal vocal tract. In reality, there is indeed a great deal of physiological independence separating the vocal fold level from the epilaryngeal levels, but more important are the dependencies and synergistic relations, which are not considered in traditional models. Features such as [+constricted glottis] do not divide the phonological space up effectively, and they certainly lack articulatory clarity.

The remarkable fact about the epilarynx is that it can come into contact with the vocal folds via vocal-ventricular fold contact, and this can influence vocal fold behaviour. This simple potential is the basis for lingual-laryngeal interaction and it provides a means to relate “glottal” activity, laryngeal activity, and pharyngeal activity.

The LPP posits a continuum of hypopharyngeal stop potentials, which increase in complexity and stability as the number of laryngeal closures increases. Along this continuum there is a “pivot” in the configuration of the epilarynx from lateromedial action to posteroanterior action which divides the ventricular plane or lower epilarynx from the aryepiglottic plane or upper epilarynx. Phonological systems will tend to straddle the epilaryngeal pivot if possible, or stay on one side or the other. Multiple possibilities are available on either side. A case like Amis shows that the “pharynx” must be viewed as two tubes, an epilaryngeal tube and a pharyngeal tube, in order to
understand the complex behaviour of pharyngeals in that language which prosodically alternate between upper epilaryngeal and lower pharyngeal stricture.

The progression along the continuum is synergistic with larynx raising and tongue retraction. One implication is that the gross states of pharyngeals might bias glottal stop towards being pharyngeal too, as in Tigre, or that sounds strongly associated with gross states of pharyngeals can diachronically mutate into pharyngeals, as in Southern Wakashan. Finally, the epilarynx provides a means to relate laryngeal closure to relatively open vowel qualities which involve gross states synergistic with epilaryngeal stricture: the most likely candidate being the retracted vowel [ə]. (See Appendix F, Table 8.6 for a summary of the claims of the LPP regarding the epilarynx and the vocal folds.)

3) What is the relationship between the epilarynx and the supralaryngeal vocal tract?

§7.3

Perhaps one of the most counter-intuitive aspects of pharyngeals – which has been known for a long time, but never previously resolved – is the potential for these sounds to become affiliated with palatals or exhibit a relatively fronted lingual position. In GCLC accounts, there is no explanation for this: pharyngeals are [RTR] in such approaches and should retract vowels. In the LPP, one potential configuration of the vocal tract is double-bunching epilaryngeal, which engages strong middle genioglossus activity to help retract the tongue. Without an epilarynx, however, this state is merely the state for “double-bunched” American-English R. When epilaryngeal constriction is added into the picture, we note that all vowel qualities are still possible although voice quality has changed (to
“pharyngealized” or raised larynx quality). To characterize these new potentials, the terms neo-tongue and neo-pharynx are applied: the vocal tract, especially the pharynx, has suddenly “shrunk” and the movement of the tongue is suddenly diminished in this state. Despite the overall condensing effect this has on vowel quality, the vocal apparatus is flexible enough that, even in this highly perturbed state, it is still possible to produce any vowel quality.

In phonologies, this means that some languages may preserve vowel quality in the context of pharyngeals (although voice quality will not be preserved). Other languages may show a tendency for restriction of vowel quality predictable by anti-synergistic relations. The synergies and anti-synergies associated with epilaryngeal constriction and anti-constriction form a strong pivot that many phonologies realize in the form of register-like phenomena, including vocal-register, tonal-register, and cross-height harmony systems (i.e. ATR). In the LPP, these are all regarded as related phenomena, and one can say very generally that there is a constricted register and an unconstricted register. These oppositions have associations across a constellation of phonetic properties, and languages generally follow the same pattern, even though differing by exact location within phonetic space as to where the phonological divisions are drawn.

A model that does not explicitly connect the vocal folds to the rest of the vocal tract via the epilarynx cannot explain why these register tendencies exist in language or why they are so similar from one case to the next. The LPP points to the vocal tract, and to the operation of the epilarynx, specifically, as a significant factor shaping the nature and trajectory of register-like systems, and brings together tone, phonatory, vowel, and voice quality, and laryngeal contrasts on obstruents. This model can be gracefully
connected into a model of aerodynamic factors influencing phonology. (See Appendix F, Table 8.7 for a summary of the claims of the LPP regarding the epilarynx and the supralaryngeal vocal tract.)

7.4.2 Summary of what is outside the scope of the model

The LPP holds that many of the patterns, contrasts, and diachronic tendencies associated with lower vocal tract sounds involves the articulatory nature of the epilarynx. The LPP is not a model of perception and it does not make any claims about the nature of phonological abstraction at the cognitive level.

Perception might be invoked as part of the explanation for why pharyngeals tend to be associated with relatively fronted vowels or even palatals, especially in contrast to uvular consonants, which tend to be associated with retracted vowels: for example, one might posit that [ʔʰæ] conveys pharyngeal identity better than [ʔʰɑ], especially if [qɑ] is in perceptual competition with pharyngeals. (One cannot help but think of how this is similar to the generally velarized/uvularized quality of Russian, which putatively is this way because of the presence of distinctive palatalization.) The idea of articulatory differentiation subserving perceptual differentiability is familiar in the literature (Cohn, 1995), and it is also often said that these perceptually-driven articulatory differentiations are offloaded onto vowels (Wilson, 2007, p. 56). In parallel with this perceptual organization, pharyngeals also involve articulatory mechanisms (such as double-bunching epilaryngeal) that bias fronted configuration of the tongue, which is a less likely option for the articulation of uvulars (since, unlike pharyngeals, these sounds depend on a greater degree of tongue retraction). Similar articulatory-perceptual parallelism may hold
for the relationship between glottal stop and relatively open vowels (Brunner & Żygis, 2011; Żygis, Brunner, & Moisik, 2012). The LPP does not explore these relationships in detail, but it seems there may be a great deal of compatibility between the perceptual-oriented explanations and the articulatory-oriented account of the LPP.

The LPP does not explicitly make any predictions about how epilaryngeal activity in phonology is organized in time, although the coordination of its many components may be important in explaining aspects of its distribution in language and its behaviour within phonological systems. For example, *a priori*, we might suspect that the relatively slow engagement of upper epilaryngeal stricture makes it less than ideal as a basis for producing speech sounds, and we might thus take this as part of the explanation for why pharyngeals are relatively uncommon in the world’s languages. Borroff (2005, 2007) provides some insight into temporal matters from an Articulatory Phonology perspective, but the work will likely be expanded upon by more sophisticated computational modeling of the vocal tract.

The LPP also does not consider higher-level organization, such as syllable structure, prosody, or morphology, which have been shown to be relevant in guttural/post-velar phonology (McCarthy, 1994; Rose, 1996). The epilarynx has little role in explaining why gutturals/post-velars as a class tend to degeminate, dissimilate, are subject to certain phonotactic restrictions, or, perhaps most importantly, why the guttural/post-velar class exists in the first place. Much ado has been made about the problem of finding a unifying correlate for this seemingly disparate class of sounds, and it should be emphasized that, although the epilarynx is important in understanding this class, it alone does not unify the class.
7.4.3 Comparison of the LPP to mainstream models of phonology

Much of the discussion about how the LPP relates to patterns of lower vocal tract phonology was contextualized in discussion about the standard features we, Phonologists, have conceptually inherited from what has been referred to here as GCLC models, or glottocentric-linguocentric models: features like [constricted glottis], [pharyngeal], [ATR], and [RTR].

All theoretical models are cartoon-like in their abstraction of reality, when we start to consider how complex reality really is. The question is, how useful are our cartoons? We can do a sort-of composite sketch of the vocal tract based on the information contained in features in GCLC models and compare it to a similar sketch representing the basic understanding at the heart of the LPP. If we do, we will see that, unlike the LPP, the GCLC model has a flat larynx lacking an epilarynx and views the pharynx as a single tube controlled by tongue root position and larynx height. Furthermore, GCLC models have an articulatory heritage, and this heritage shapes the way we think about what is possible in sound systems and what is “merely phonetic”. It also influences how we hear sounds: in GCLC models we are deaf to voice quality because only glottal phonatory quality and vowel quality can be expressed. Slackness in phonetically-accountable phonology has resulted in a serious slippage in the usefulness of the features proposed in phonological analysis under stringent pre-conceived notions of what the brain is and is not capable of doing. It has led to myopia in the face of facts that have been with us for quite a while: as far as the LPP is concerned, in 1969, Jan Gauffin told the basic LPP story only a year after SPE was published.
The LPP is not a theory of phonological representation. It is a theory of how the epilarynx is the articulatory nexus of the lower vocal tract. In some ways, it parallels some of the original concepts in FG, such as strong articulatory grounding. However, FG, especially (Revised) Articulator Theory, is too rigid to be able to express some of the complex physiological relationships spanning the vocal tract. The LPP by comparison is organic, but not freewheeling: the synergistic relations and gross states are invariant. {[^lx ↔ epc]: Larynx raising predisposes but does not entail epilaryngeal stricture. The average individual will vary according to their own devices, but applied to a population of speakers, this generalization should hold true. One can imagine that FG was an attempt to express similar structural relations of the vocal tract, such as the relation between tongue root behaviour and phonation being expressed by the dominance of LARYNGEAL and TONGUE ROOT by GUTTURAL in Articulator Theory (Halle, 1995) or Rose’s (1996) notion that DORSAL “bridges” the ORAL and PHARYNGEAL vocal tracts. The limiting factor on FG’s usefulness in this capacity is that it was designed under the pressure to be “phonology”: a model of the mental representation of sound and our competence to use such representations in the process of forming words. Thus, mechanical entailments and interactions, so important to understanding how LVT sounds work, could not be modeled.

The LPP might be said to be a covert expression of Optimality Theory since it has a non-absolutist approach to phonological prediction: this is very similar to the idea of constraint violability, a core postulate of OT. In the LPP, pharyngeals are capable of biasing vowel retraction, but some languages might instead involve a palatal bias. One might, in OT, posit a constraint *[+RTR]/[–back] and thereby account for the palatal
affiliation of pharyngeals in many Caucasian languages by saying this universal markedness constraint becomes violated to satisfy a higher-ranking faithfulness constraint demanding the realization of input [+RTR, –back] pharyngeals as output [+RTR, –back] pharyngeals and then be done with it. Doing so takes causality away from the physical nature of sound production and places it firmly in the hands of constraints – objects in the mind: we might start to wonder why the features are named the way they are in the first place.

The LPP is also not a model of phonetic grounding of phonology. Synergistic relations are not a categorical force acting like a switch in the grammar; rather, they exert a constant, steady pressure on sound production attracting sound systems one way or the other. The scale is thus diachronic in nature, although its effects can be detected at the synchronic scale in the form of co-articulatory tendencies, such as the biasing of glottal stop towards “pharyngealhood” in Tigre. The objects of the LPP, potentials, gross states, and synergies are also fictions – they are articulatory abstractions: reality and individual behaviour is much more complex. However, it is believed that they are useful fictions, ones that let us organize information about lower vocal tract sound systems, let us predict the types of phonetic correlates involved in such contrasts, and help us explain, in conjunction with other factors beyond articulation, why sound patterns of the lower vocal tract are the way they are and give us a hint as to where they might be going.

7.4.4 Features revisited: [±constricted epilaryngeal tube]

It has been assumed that distinctive phonological features are emergent, idiosyncratic abstractions pertaining to sound classes and that they are formed by
speakers of a language through experience of that language in its ecological context. It seems plausible that these abstractions could relate, via neural connectivity, to innumerable dimensions correlated with phonemic identity, such as properties associated with the somatomotor and somatosensory systems – including tactile, proprioceptive, and auditory properties of sound and its production, visual properties of articulation – lexical associations, orthographic representation – maybe even emotions. It is not clear how one can posit a feature such that it corresponds to a discrete, well-defined entity in reality, rather than a distributed, nebulously defined one.

Features are post-hoc entities in such a model, and as Mielke (2004) points out, they may influence sound change non-linearly by regrouping the phonemes into sets that pattern together in language change. This requires the leap of faith that a community of language users forms similar abstractions – behaviour that might resemble flocking.

Flocking is a form of “consensus without central direction” (Cucker & Smale, 2007, p. 852): a group of individuals lacking in leadership or extrinsic organization imposed on the group, nonetheless exhibit cohesion and order. The classic example is birds in flight, but there are many other examples, such as insect swarming, fish shoaling and schooling, and herding behaviour of land-dwelling animals. Flocking is thought to be a form of emergent behaviour, and, as such, arises from the combination of a simple set of rules. For example, mathematically stateable rules associated with the principles in 1-4 (Flierl, Grünbaum, Levins, & Olson, 1999, p. 400) can be used to model the behaviour of aquatic animals (such as krill) in turbulent flows:

1) locomotory forces, such as those resulting in preferred swimming speeds;
2) social forces, such as attraction or repulsion between individuals;
3) arrayal forces, equalizing speeds and directions of neighboring animals; and
4) environmental effects, such as chemical gradients, which could lead to
directional biases in movement.

These principles can be mapped into linguistic terms with relative ease. Principle
(1) is a statement about biomechanics, which relates directly to devices responsible for
constriction of the vocal tract and govern vocal tract behaviour in time. Principle (2) is
fundamental to language: it is a social behaviour and we should expect sociolinguistic
principles apply, such as convergence to and divergence away from phonetic speech
norms for the purposes of establishing social identity. Principle (3) can be considered
speech normalization to other (local) members of the speech community for reasons other
than social identity (such as controlling tempo of speech to maintain intelligibility).
Finally, principle (4), is a statement about directional bias, or “articulatory anisotropy”; this is strikingly similar to the concept of synergistic relations. These relations exert an
influence that does not directly intervene in the behaviour of an individual, but
presumably exerts an attraction at the scale of the speech community, just as a scent trail
might influence the direction of shoaling fish but not strictly determine the exact path of
an individual fish. Features can also be associated with the idea of directional bias in
principle (4), but one that operates at the level of cognitive abstractions, rather than at the
level of physical biases, where the synergistic relations are assumed to operate.

So, sounds produced with epilaryngeal involvement are subject to biases at the
synergistic relation level, and we might suppose biases also operate at the abstract level.
In close adherence to Edmondson & Esling’s (2006) [+constricted], we might be tempted
to posit a feature [+constricted epilaryngeal tube] or [+cet], which would have the
conceptual benefit of implying a phonetic conjunction of voice quality effects, localizing
these factors not just to larynx height, or tongue position, but the epilarynx. We might also speculate that this feature is strongly rooted in phylogenetically deep functions of swallowing ([+cet]) and inspiration([-cet]), as Lindqvist/Lindqvist-Gauffin/Gauffin (1969; 1972; 1972; Lindblom, 2009) had originally advocated. After all, when it comes to swallowing or breathing, it is not just larynx raising or lowering or tongue retraction or advancement that matters, what matters is whether closure or opening of the epilarynx occurs: all other activity supports these two actions fundamental to human life.

7.5 Summary and directions for phonology

The main theme of Chapter 7 has been that traditional approaches to phonological representation have a gap where the epilarynx should be: these are the glottocentric-linguocentric or GCLC approaches. In such models, there is no intrinsic means to express the connectivity between the two systems (lingual and glottal), which has consequences for phonological understanding of the possibilities of voice quality (the intersection of phonatory, tonal, vowel, and “global” quality of the voice). Since these models were designed with the vocal tract in mind, these models must be held accountable to physical reality, and, given what we now know (and, to a certain extent, knew a long time ago) they simply are not. The consequence is that GCLC models have given us a relatively inorganic picture of lower vocal tract phonology. GCLC models can (weakly) describe phonatory quality and (weakly) describe vowel quality, but they are myopic to voice quality. They have established a model where interaction between the glottal level and the lingual level are unexpected. They have given us the view that the larynx is a flat,
one-dimensional “glottis” and the pharynx is a single tube controlled by glottal height and “tongue root” position.

At the outset, the hypothesis of the model proposed here, the Lower Vocal Tract Phonological Potentials (LPP) model, was that phonological patterns of the lower vocal tract require consideration of the epilarynx. If phonology is not entirely substance free, if phonology can be determined by the form and functioning of the vocal tract, if phonology in any way lies in the body and not just in the mind, then it makes a good deal of sense to incorporate the epilarynx into a model of phonology. The epilarynx is a structure that physically links the vocal folds to the rest of the vocal tract, and this encourages a disposal of the old view of strict laryngeal-supralaryngeal separation. When admitted into consideration of lower vocal tract phonology, knowledge of the epilarynx helps clarify many of the problems raised in the literature. We can readily identify what “growling” is and make sense of how it interacts with tone and vowel subsystems. We can understand why ejectives sometimes behave differently from laryngealized/glottalized sounds but why they might bias pharyngeal genesis. The notion that we can have pharyngeal stops, multiple different allophonically related hypopharyngeal stops, or even multiple glottal stops can be attributed to epilaryngeal potentials. The very idea of glottal stop is reinvented in the LPP, and we now know to expect this sound to occur in the context of relatively open vowels and we now know why this might happen. We can understand what is involved in “emphatic palatalization” and the wholly unexpected patterning of pharyngeals with palatal sounds or association with non-retracted vowels. Finally, the LPP grants some insight into how phenomena as seemingly diverse as consonant voicing contrasts and “ATR harmony” are related.
The LPP is anchored in an understanding of the body – a vocal tract that includes an epilarynx – and what it is capable of doing, what it tends to do, and how this capacity might direct phonological systems, particularly those with lower vocal tract sounds. The LPP focuses on three principle functions of the epilarynx relevant to linguistic sound systems: epilaryngeal vibration, epilaryngeal interaction with the vocal folds, and epilaryngeal interaction with the supralaryngeal vocal tract. The interpretation of epilaryngeal behaviour in speech can be reduced to a statement about the physical location of the epilarynx in the vocal tract: it sits between the vocal folds and the rest of the vocal tract. It is not surprising then that its behaviour should parallel the vocal folds somewhat and also mirror other moving structures in the vocal tract or that it should act to couple phonatory behaviour to vocalic behaviour. Adopting the epilarynx into a model of speech sounds seriously conflicts with the traditional understanding of vowels as confined to a lingual space: the LPP reaffirms that it makes a great deal of sense to consider Esling’s “laryngeal logic” (2005, p. 24) of vowels, particularly the affinity canonically retracted vowels, \([\alpha \ ɒ \ ʌ \ ɔ]\) have for epilaryngeal stricture. However, note that the operative word is **affinity** – such vowels are not intrinsically constricted\(^{273}\) – but they do have a greater potential for engaging epilaryngeal stricture.

---

\(^{273}\) As Shahin (2011a) does when she misinterprets Esling’s model as meaning that “retracted” vowels inherently involve aryepiglottic stricture. This is **not** what Esling intends. A careful reading reveals that Esling’s view is perfectly in line with (and in fact, the inspiration of) the LPP. What Esling means is embodied in two quotes from the *There are no back vowels* paper (Esling, 2005): “[\(\alpha\)] is not just a low back vowel. It is related to the laryngeal constrictor mechanism in a complex chain of events that, ultimately, lead to the complete closure of the airway. … the quality of [\(\alpha\)] is inherently susceptible to increasing degrees of laryngeal constriction” (p. 40) and “retracted vowels are inherently the most susceptible to the effects of [the laryngeal constrictor mechanism], but the oral (front and raised vowels can also be strongly affect by laryngeal constriction” (p. 41). The susceptibility Esling speaks of is exactly parallel to the synergistic relations between retraction and epilaryngeal constriction in the LPP (although the terminology is somewhat different). Vowels like [\(\alpha\)] should **not** be considered equivalent to sounds like [\(\text{S}\)] in terms of the degree of epilaryngeal constriction: [\(\alpha\)] is susceptible or biases epilaryngeal stricture, possibly involving
Looking forward, there are numerous new questions that arise because of the LPP that we can pursue in future research. First, the LPP is a theoretical tool that promises some new insight into many old problems in phonology. One of the immediate extensions would be to examine individual languages and diachronic patterns in more detail using the LPP to see if its predictions hold true or if it is necessary to extend it. The positive result is that it is very explicit about the types of phonetic behaviour predicted to occur. More empirical work on epilaryngeal behaviour in sound systems is also required, especially where epilaryngeal vibration is concerned.

Second, implicit in all of the accounts is that phonemes can be decomposed into devices. This is a new way of evaluating phonemic identity and introduces phonological theory to a new methodology for evaluation – biomechanical modeling. With biomechanical modeling, we can begin to identify the biomechanical-articulatory factors that make some mechanisms more reliable in terms of production than others and quantify what it means for a certain production to be more “costly" or “effortful" in nature (which is presumed to be true for ⟦ʔ⟧ compared to ⟦ʕ⟧). The implication of such data would be a more accurate means to predict the rate of occurrence of particular sounds based, in part, on their biomechanical expense, which we might suppose plays a role in their rate of adoption as phonemes. Such modeling could thus help explain the relatively uncommon phonemic status of epilaryngeal vibration.

Alongside the question of the exact characterization of these devices is the question of whether phonemes can be decomposed this way, and, if so, how many devices characterize any given phoneme. We might wonder, for instance, whether ⟦ˈ⟧

passive narrowing of the epilarynx, but it is indeterminate in comparison to [ʕ] which has active and extreme epilaryngeal constriction.
contains [ʔ] as part of the mechanism of ejection, or whether an ejective like /q’/ is not something like [q’] (a holistic device just for the uvular ejective). For example, in §7.2.4 on pharyngeal genesis, /q’/ was decomposed into [’] and [q] based on a priori considerations: regardless of the decomposition, the gross states are the same. However, decomposition does matter in terms of whether we predict that, for example, the [’] is exactly the same in all ejectives or varies from one ejective to the next, possibly as a function of the accompanying oral stricture mechanism. Understanding the biomechanical mechanisms will impact our understanding of phonemes and features. The mechanisms have a neural component (they are neuromuscular organizations) and thus plausibly are subsumed into phonemic or even featural identity at the level of the neural substrate of phonological organization.

At certain points in the discussion, the LPP (a physiological, articulation-oriented model) crossed into perceptual and aerodynamic considerations. The second question is thus whether a unifying model can be created that spans these domains. There is no reason at the outset to believe that these domains would necessarily overlap or parallel each other given the different physical principles governing each system. The hypothesis tentatively offered here is that there is considerable parallelism across these domains.

As an example, according to the LPP, larynx lowering {↓lx} is anti-constricting, i.e. {↓lx ··· epc}. In aerodynamic terms, larynx lowering is beneficial for promoting phonation during oral obstruction as in [d] (by prolonging the time before pressure equalization occurs between the sub- and supraglottal spaces during phonation): voiced stops are also associated with unconstricted registers (see §7.3.2 and §7.3.3), which, in the LPP, is because of this predisposition towards larynx lowering. In acoustic-auditory
terms (and presumably perceptual terms), larynx lowering is associated with close vowels because it enhances F1 lowering: in the LPP, larynx lowering is associated with the close vowels, which are less favourable for epilaryngeal constriction and opposite in nature to the relatively open vowels, which are more favourable for epilaryngeal constriction (see §7.2.5; another example is the perceptual connection between relatively open vowels and glottal stop – entirely independent, but parallel to this is the articulatory affiliation between these sounds, which was also the topic of §7.2.5).

A deeper account will minimally bring together these three factors, articulation, perception, and aerodynamics (which in some ways is the bridge between articulation and acoustics) in conjunction with other factors, such as sociophonetic influences. It is an open question just how much feature abstraction will also tend to parallel these other domains, but it is unlikely to be entirely orthogonal to the other factors.
Chapter 8

CONCLUSION

The goal of this dissertation has been to establish a theory of the epilarynx in speech. To accomplish this goal, a multi-layered approach was taken. The individual layers of consideration are the anatomo-physiological, phonetic, and phonological layers, (although the degree of isolability of any one layer is taken to reflect the exigencies of scientific exposition, rather than the structuring of reality).

The substrate of the theory is the anatomo-physiological layer (see §2.1), which has fairly firm grounding in the laryngological literature, with the work of Fink (1956, 1962, 1974a, 1974b, 1975) and Painter (1986, 1991) being of great value in this regard. Layered upon this fundamental knowledge is the phonetic literature (see §2.2) which demonstrates how laryngeal physiology related to the epilarynx is expressed in speech production. Epilaryngeal functioning in speech has been less thoroughly investigated than other, easier to reach parts of the vocal tract or structures that have more obvious functional relevance to speech (such as the vocal folds). Nonetheless, a strong current of thought originating in the late 1960s and finally flourishing in the mid 1990s can be traced, and much of this work is taken as the backbone of the theory of the epilarynx in speech. Particularly notable are the ideas stemming from the research of Lindqvist/Lindqvist-Gauffin/Gauffin (1969; 1972; 1972), Esling (1996, 1999, 2005) and colleagues (e.g. Carlson & Esling, 2003; Esling & Harris, 2005; Edmondson & Esling, 2006; Edmondson, Padayodi, Hassan, & Esling, 2007; Esling, Zeroual, & Crevier-Buchman, 2007; inter alia); also this dissertation also heavily relies upon the work of

The layer of greatest physical abstraction, the phonological layer, is likewise dependent on the ideas originally developed by Czaykowska-Higgins (1987), Denning (1989), Hayward & Hayward (1989), Trigo (1987, 1991), and Hess (1998) which advocated for the phonological significance of a link between the laryngeal and supralaryngeal domains (see §6.2). The phonological considerations in this dissertation also draw heavily upon the work of McCarthy (1991, 1994) and, in particular, Rose (1996), for their elucidation of the theoretical issues surrounding “guttural” phonology.

8.1 The theory of the epilarynx in speech: A summary

The main premise of the theory of the epilarynx in speech is that the epilarynx needs to be considered as a functionally-cohesive unit rather than as decomposed collection of independent parts. The unity of the epilarynx is reflected in its most primitive physiological function, which is effective closure of the laryngeal airway, something the vocal folds acting alone are not capable of doing in all respects (since they are an inlet, vis-à-vis outlet, valve). The theory posits that the epilarynx functions intrinsically and extrinsically as a distributed mechanism rather than a single, isolable,
“active” articulator such as the aryepiglottic folds or “tongue root”, or otherwise. The theory also does not abstract beyond the epilarynx to a general laryngeal constriction mechanism, since the vocal folds asymmetrically participate in epilaryngeal function: adductory assistance is required for full closure, but the epilarynx can still constrict with the vocal folds abducted.

This is a conceptually different approach from that advocated in the Valves of the Throat model developed by Edmondson & Esling (2006) (also see §2.2.3). In the “Valves” model, the emphasis is on the individual valves – not their interaction; in the model proposed here, the emphasis is on the interactions of the structures comprised by the valves. The epilarynx model owes much of its provenance to the Valves model, but diverges by taking a more holistic view of the mechanisms of the lower vocal tract and regards the epilarynx as a key locus of the actions of speech-related activity within the lower vocal tract and emphasizes the interrelatedness of this activity through structural and aero-acoustic coupling.

Every aspect of the speech-functioning of the epilarynx, and, in fact, beyond speech too, can be understood as a form of coupling, or basis of interaction, between the vocal folds and the supralaryngeal structures. The organization of this dissertation has been framed around three core functions: epilaryngeal vibration (Chapter 3; Chapter 6, §7.1), the relationship between the epilarynx and the vocal folds (Chapter 4; Chapter 6, §7.2), and the relationship between the epilarynx and the supralaryngeal vocal tract (Chapter 5; Chapter 6, §7.3). While this organization was used for expository convenience, the reality of the epilarynx is best understood in terms of coupling/interaction in each case. The structure the epilarynx, with its lower border (the
ventricular folds) in close proximity to the vocal folds and its upper border oriented optimally for lingually-assisted closure, defines its status as a vocal-folds–tongue coupling mechanism. Furthermore, its position within the vocal tract, in between tongue and vocal folds, translates directly into much of the speech-related effects of the epilarynx.

8.1.1 Epilaryngeal vibration

Epilarynx vibration (Chapter 3 and §7.1), responsible for harsh, “growly” phonatory quality, acts not so much as an independent source, but more as a source modulator, with amplitude modulation being its main effect on pulsatile glottal flow. This is an aero-acoustic effect associated with time-varying supraglottal laryngeal aperture. A direct mechanical interaction is also plausible through vocal-ventricular fold coupling (Chapter 4 and §4.3 in particular), which has similar effects on the voice source – it corresponds with constricted vibratory modes represented by phonetic categories such as creaky or harsh. However, whether the effect is aero-acoustic, direct-mechanical, or a combination of both, coupling inherently biases the overall laryngeal-oscillating system towards lower vibratory rates because the effective mass of this system increases when the epilarynx is “added in”.

Theoretically, the epilarynx ambivalently classifies in phonetic terms as a phonatory-like and trilling-like mechanism, which was argued to be a consequence of its location within the vocal tract between the canonical phonation mechanism (the vocal folds) and the canonical trilling mechanisms (such as the tongue tip, uvula, and lips). This duality of phonetic classification is mirrored in its phonologically-oriented functioning:
epilaryngeal vibration behaves like a contrastive phonatory quality in some languages, but in other languages it is associated just with pharyngeal consonants and may even be a distinctive manner in Agul.

The low frequency bias of epilarynx vibration has consequences for its occurrence in tone systems as it tends to be confined to relatively low tone. Although epilaryngeal vibration at higher F0 is certainly possible, it would seem that when it occurs in tone systems it is restricted towards the low tonemes, as in Bai and Zhenhai. Another possibility is for epilaryngeal vibration to co-occur with vocal fold abduction, which generates a voiceless/whispered or whispery voiced growling and is attested in !Xóõ “sphincteric phonation” and in Zhenhai Wu “growl”. The likely explanation for this pairing of opposed laryngeal actions is the benefit gained in aerodynamic power required for epilaryngeal vibration; abducting the vocal folds lowers the resistance of the glottal airway and hence provides greater driving-pressure input to the epilaryngeal structures.

8.1.2 The epilarynx and lingual-laryngeal linkage

Vocal fold and vocal tract interaction were discussed separately in this dissertation (Chapter 4 & §7.2 and Chapter 5 & §7.3 respectively), but in reality, these two effects are intimately related. The lingual-laryngeal linkage is characterized by structural relationships on two levels, which were taken up separately in dedicated chapters.

The first relationship is embodied by the vocal-ventricular fold coupling hypothesis (Chapter 4 and §4.3), which posits that contact and mechanical coupling
occurs during epilaryngeal stricture and plays an important role in the production of constricted-larynx sounds such as glottal stop and creaky and harsh phonation.

The second relationship concerns the role of the epilarynx-external components which facilitate or synergize with epilaryngeal stricture: tongue retraction and larynx raising. While neither of these components is required for epilaryngeal closure, their presence is favourable for epilaryngeal constriction, which, even if not active in a given sound, can still be passively induced when one of these components is present (provided other mechanisms are not engaged which directly oppose epilaryngeal stricture, such as pitch raising). A corollary is that at least two levels of interaction should occur, a subtle interaction of relatively-open-vowel susceptibility to epilaryngeal stricture and a more extreme recruitment of the lingual and laryngeal components to the point that the acoustic configuration of the vocal tract becomes redefined. To describe what happens in the extreme case, the concepts of “neo-tongue” and “neo-pharynx” were introduced. Extreme epilaryngeal and hypopharyngeal narrowing effectively cuts the pharynx in half, which, in articulatory-acoustic terms, causes the back cavity to be regulated by the extent of concavity of the dorsal surface of the tongue rather than the degree of occupation of the pharynx proper by the posterior tongue. Vowels (particularly close ones like [i]) in this configuration take on the “double bunching” appearance of some varieties of American English R (§5.4.2), but the difference is that the configuration also involves the addition of epilaryngeal closure.

In order to understand sound patterns associated with the lower vocal tract, it was argued that consideration of the epilarynx is essential (Chapter 6 and Chapter 7). Many sound patterns that appear mysterious in a glotto-centric–linguo-centric account – which
lacks consideration and formalization of the epilarynx-mediated lingual-laryngeal linkage – become tractable when the epilarynx is countenanced. Traditional phonological models, characterized by the use of features like [ATR], [RTR], and [constricted glottis], obscure the phonological potentials associated with lower vocal tract sounds. These models, which were originally designed with articulation in mind, have not maintained pace with developments in phonetic understanding concerning the lower vocal tract and the epilarynx in particular. The core failing of these models is the notion that the vocal folds and tongue are phonologically independent systems. The argument made in this dissertation is that the laryngeal-supralaryngeal dichotomy is too rigid to adequately capture the behaviour of sounds which rely on the linkage between these systems.

Ample evidence exists that shows the phonological reality of this linkage, and as Trigo discovered, the link does not just boil down to larynx raising and tongue retraction: neither of these components directly controls whether epilaryngeal stricture will occur or not, they simply bias epilaryngeal stricture. The polar opposites of these actions, tongue advancement and larynx lowering, have the opposite effect on the epilarynx. These two core states have their use in register-like systems which involve correlated modification of vowel, phonatory, tonal, and overall voice quality. The correlation is not an accident, but rather reflects the action of constricting or unconstricting the epilarynx and the changes to the epilarynx-external mechanisms which facilitate the state change.

Acknowledgement of the epilarynx in phonological patterning allows two of the most perplexing patterns associated with lower vocal tract phonology to be resolved quite readily. Interaction between glottal stop and relatively open vowels, especially /a/, is a direct consequence of the lingual-laryngeal linkage; the tendency for epilaryngeal
vibration to occur with these relatively open vowels is a parallel case. On the other hand, the seemingly anomalous palatal or fronting bias of pharyngeals and pharyngealization, most famously embodied by the “emphatic palatalization” of Caucasian languages, is explained by the extreme “double-bunching” epilaryngealization configuration, which has a palatal bias because the lingual mass associated with the “neo-tongue” is pushed forward by the action of muscles used simultaneously to augment tongue retraction in service of epilaryngeal narrowing. This configuration also helps explain why vowel quality can be (unexpectedly) preserved during pharyngealized vowels or similar configurations found in many register languages. This configuration makes it possible to achieve the paradoxical goal of articulating a very anterior vowel quality (as in [i]) with simultaneous epilaryngeal stricture. Of course, the more ordinary effect of vowel retraction associated with many pharyngeals is also a possible and natural articulatory consequence according to the theory of the epilarynx in speech.

The theory of the epilarynx in speech allows for the unified explanation of patterning spanning glottal stop, ejectives, and primary and secondary (bona fide) pharyngeal articulation, “growled” and “constricted” phonation types and their tonal distribution, and vocal-register and tonal-register systems and cross-height harmony systems involving phonatory and voice quality correlations. No previous model has been able to demonstrate how and why all of these seemingly unrelated aspects of lower vocal tract phonology are related. None of these things can be plausibly united without considering the epilarynx. In physiological terms, the epilarynx is the definitive lingual-laryngeal link, and this status is strongly reflected in phonology. Without the epilarynx,
we have no principled reason to relate what happens at the lingual level to what happens at the vocal fold level.

### 8.2 Directions for future epilarynx-related research

This dissertation leads immediately to three primary and related directions of future research: the first is empirical research (§8.2.1); the second is computational modeling (§8.2.2); and the third is phonological research (§8.2.3). In §8.2.4, extended domains of research relating to the epilarynx are discussed.

#### 8.2.1 Future empirical studies of the epilarynx

While much empirical work on epilaryngeal function in speech has been done, some of which now represents the contributions of this dissertation, it is necessary to continue to bring evidence to bear on the hypotheses developed in this body of work. Perhaps the single greatest weakness of epilaryngeal theory is its overreliance on data contributed by trained phoneticians. Much of what we know is grounded in such data: the tradition goes back to Lindqvist/Lindqvist-Gauffin/Gauffin (1969; 1972; 1972), who laryngoscopically examined his own productions of glottal stop, but Traill (1986), Rose (1989), and Esling (1996, 1999) all have conducted similar self-observations. Among all of these, there is agreement on the primacy of epilaryngeal activity, and studies which collected native speaker language data (see §2.2) corroborate what is observed in the canonical phonetic case (ignoring individual variation). Because of the general rarity of natural language data, and the difficulty in obtaining such data, phonetic observations are a valuable tool to examine the mechanism. Moving forward, it is hoped that as
researchers become increasingly aware of the epilarynx, more evidence from native speakers will be obtained using more advanced methodologies. Such data, as exemplified by Brunelle et al. (2010), will be essential in coming to understand the details of epilaryngeal function and its distribution in speech.

Innovation in imaging technology will play a critical role in developing a deeper understanding of the epilarynx. Real-time MRI (e.g. Lammert, Proctor, & Narayanan, 2010; Narayanan, Bresch, Ghosh, et al., 2011) will eventually be able to resolve a number of difficult articulatory problems. Based on this dissertation, several research questions immediately arise:

Table 8.1: Future empirical (imaging) research: physiological-articulatory simulation

<table>
<thead>
<tr>
<th>#</th>
<th>Research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Does vocal-ventricular fold contact actually occur and is it truly as significant an effect as proposed in this dissertation (see Chapter 4 and §4.3 in particular)?</td>
</tr>
<tr>
<td>A2</td>
<td>Can vocal-ventricular fold contact occur even without ventricular fold medialization?</td>
</tr>
<tr>
<td>A3</td>
<td>What is the role of epilaryngeal stricture in ejective production?</td>
</tr>
<tr>
<td>A4</td>
<td>What exactly happens to the tongue in raised larynx voice (and similar states)? How are different vowels articulated?</td>
</tr>
<tr>
<td>A5</td>
<td>How does epilaryngeal stricture vary as a function of vowel quality and prosody in various speaking situations, such as calm or excited speech?</td>
</tr>
<tr>
<td>A6</td>
<td>What is the relationship between the vocal and ventricular folds during vocal fold adduction with and without simultaneous upper epilaryngeal stricture?</td>
</tr>
<tr>
<td>A7</td>
<td>How do we characterize the physiological mechanism of growl in various languages (Khoisan-group, Bai, Ningpo-region Wu dialects, Agul)? Is it always (primarily) aryepiglottic in these languages?</td>
</tr>
</tbody>
</table>

Of course, MRI is still prohibitively expensive and in some situations it is entirely impractical, such as in the context of field research. This dissertation has shown that laryngeal ultrasound can serve the researcher interested in examining the vertical components of laryngeal function. It is both cheaper and (far) more portable than MRI,
which makes it an attractive alternative to consider – especially if one wishes to collect laryngeal articulatory data in remote locals or when circumstances do not permit the use of MRI.

8.2.2 Future computational studies of the epilarynx

As technology matures, more sophisticated computational modeling of laryngeal articulation becomes tractable. The modeling presented in this dissertation is exclusively in the lumped-element formulation, which entails a high order of abstraction in the spatial distribution of mass. More detailed modeling of the larynx is becoming a reality with tools such as ArtiSynth (Stavness, Lloyd, Payan, & Fels, 2011) (www.artisynth.org), which combines Finite Element Modeling (FEM) of deformable bodies, simulation of rigid bodies coupled to FEM components, simulation of collisions and contacts amongst all of these structures, and computational fluid dynamics to model airflow and resulting fluid-structure interactions. Such a tool will be instrumental in advancing our understanding of how the epilarynx works, especially in relation to other structures in the vocal tract. It would be possible to conduct parallel exploration of the questions discussed above (A1 – A6) using a tool like ArtiSynth to complement imaging data. Much of what we think we know about epilaryngeal constriction, even on the most basic physiological level, could be tested using such computational modeling. Some of the key questions arising from this dissertation are:
Table 8.2: Future computational modeling research I: physiological-articulatory simulation

<table>
<thead>
<tr>
<th>#</th>
<th>Research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Lower epilaryngeal border: What controls ventricular fold positioning? How is it affected by vocal fold positioning?</td>
</tr>
<tr>
<td>B2</td>
<td>Upper epilaryngeal border: How is the epiglottis controlled? What is the role of the aryepiglottic muscles?</td>
</tr>
<tr>
<td>B3</td>
<td>How does the intrinsic mechanism of epilaryngeal stricture operate? What is the role of crico-thyroid rotation in epilaryngeal stricture?</td>
</tr>
<tr>
<td>B4</td>
<td>How does epilaryngeal constriction differ as a function of tongue position and/or larynx height? Is it possible to define articulatory (stricture) efficiencies of the epilarynx associated with combinations of these states?</td>
</tr>
<tr>
<td>B5</td>
<td>What do the pharyngeal constrictor muscles contribute to epilaryngeal stricture?</td>
</tr>
<tr>
<td>B6</td>
<td>What is responsible for the “double bunching” epilaryngeal configuration? How is the “neo-tongue” controlled in this configuration?</td>
</tr>
<tr>
<td>B7</td>
<td>What is the role of the jaw in epilaryngeal stricture?</td>
</tr>
</tbody>
</table>

As the technology improves, simultaneous aero-acoustic and mechanical modeling will become feasible. A simulation featuring fluid-structure interaction will enable a number of additional questions to be explored:
Table 8.3: Future computational modeling research II: fluid-structure interaction simulation

<table>
<thead>
<tr>
<th>#</th>
<th>Research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>How do the different core types of epilaryngeal vibration differ? What are the enabling mechanisms? Is voiceless vibration of the upper epilarynx easier to engage than voiced vibration?</td>
</tr>
<tr>
<td>C2</td>
<td>How do different types of epilaryngeal vibration differ in terms of their effects on vocal fold vibration and the resulting glottal airflow?</td>
</tr>
<tr>
<td>C3</td>
<td>Is vocal-ventricular fold phonation truly more regular than aryepiglottic vibration?</td>
</tr>
<tr>
<td>C4</td>
<td>What are the restrictions on what co-oscillations are possible (e.g. can ventricular vibration occur simultaneously with aryepiglottic vibration)?</td>
</tr>
<tr>
<td>C5</td>
<td>Is vocal-ventricular vibration, such as that which occurs in certain throat-singing styles, possible without lowering the larynx?</td>
</tr>
<tr>
<td>C6</td>
<td>What is the role of the epilarynx in turbulence generation during fricative-like sounds?</td>
</tr>
<tr>
<td>C7</td>
<td>What is the role of the ventricular folds in glottal stop and constricted phonation production?</td>
</tr>
<tr>
<td>C8</td>
<td>Is there a functional asymmetry in vocal fold adduction during ejective production and during glottal stop production as predicted by the valve abstraction of the vocal folds?</td>
</tr>
</tbody>
</table>

8.2.3 Future phonological research related to the epilarynx

The phonological model proposed in the dissertation assumes a finite set of abstract, devices underlie the act of “articulation”. A key hypothesis is that these devices rely upon quantal biomechanical-articulatory relations which make the devices well-suited to reliably producing a given speech event. This dissertation predicts that several such quantal effects should be identifiable in epilaryngeal physiology that correspond to speech sound production. With computational modeling using tools like ArtiSynth, it should be possible to test various configurations predicted to occur and characterize their quantal articulatory properties. Similar work has already been conducted for the lips (Gick, Stavness, Chiu, & Fels, 2011), tongue (Stavness, Gick, Derrick, & Fels, 2012),
and soft palate (Gick, personal communication), so the proposed future research would follow the methodological approach pursued in these projects. A list of possible candidates for quantal behaviour in the biomechanical-articulatory domain is given in Table 8.4.

Table 8.4: Predicted quantal biomechanical-articulatory effects of the epilarynx.

<table>
<thead>
<tr>
<th>#</th>
<th>Effect</th>
<th>Description/Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>“epilaryngeal vibration”</td>
<td>biomechanical properties of epilarynx (e.g. greater effective mass than vocal folds), geometry, and topology bias epilarynx towards growling quality, regardless of what exactly is vibrating</td>
</tr>
<tr>
<td>Q2</td>
<td>vocal-ventricular fold contact</td>
<td>saturation of muscular effort once contact occurs; reliably produces constricted sounds (either glottal stop, creaky or harsh phonation)</td>
</tr>
<tr>
<td>Q3</td>
<td>aryepiglotto-epiglottal contact</td>
<td>stable point in epilaryngeal stricture continuum that underlies pharyngeal articulation</td>
</tr>
<tr>
<td>Q4</td>
<td>neo-tongue–neo-pharynx</td>
<td>extreme hypopharyngeal compaction constitutes the most extreme resting point in the epilaryngeal stricture continuum and simultaneous causes a sudden, register-like change in the control of vocal tract acoustics</td>
</tr>
</tbody>
</table>

Non-linearities in biomechanical-articulatory relations will help to explain the categorical status of certain epilaryngeal states, but these do not alone enable a complete description of articulatory influence on phonological systems and processes. The second consideration to be made is the nature of interactions between various quantal states and the gross physiological configurations (i.e. gross states). What is required is testing of the synergistic relations outlined in the Lower Vocal Tract Phonological Potentials (LPP) model. Such testing could be performed using computational modeling, which could serve as the basis for a model that can begin to make quantitative predictions about the behaviour of sound systems involving a rich complement of lower vocal tract
phenomena. For example, if the epilaryngeal susceptibility of relatively open vowels could be quantified with a biomechanical model, it might then be possible to develop a corresponding stochastic model that predicts the relative frequency of occurrence of epilaryngeal stricture by vowel. For example, a measure of the articulatory cost or efficiency of particular epilaryngeal states as a function of vowel or consonantal context could help predict the occurrence of epilaryngeal stricture in glottal stop.

8.2.4 Extended domains of epilaryngeal research

This dissertation has been focused on the articulatory nature of the epilarynx. There are many possible extensions of speech-related epilaryngeal research, but three are particularly important: language acquisition, prosody, and perception.

Early infant phonetic acquisition research at the University of Victoria has been ongoing for over a decade (e.g. Esling, Benner, Bettany, & Zeroual, 2004; Bettany, 2004; Benner, Grenon, & Esling, 2007; Benner, 2009; Benner & Grenon, 2011). This work demonstrates the importance of laryngeal constriction as a key “starting location” for early phonetic behaviour involving the modulation of phonatory quality, much of which likely involves epilaryngeal activity, most significantly in the production of harsh wailing or “cry”, which are among our first attempts at communicating with our caregivers. The epilarynx may through constriction-moderated source-filter coupling (Titze, 2008) increase the acoustic efficiency of cry while concomitantly minimizing stress on the vocal folds. The accompanying harshness may serve a role in increasing the perceptual salience of cry. Both of these have fairly straightforward advantages to the infant (to get the attention of care givers without damaging the vocal folds).
Speech prosody probably plays a very significant role in predicting when epilaryngeal constriction will occur. With large-scale data analysis techniques, and awareness of the acoustic indices of epilaryngeal constriction, particularly growl, it should be possible to conduct studies which identify the amount of epilaryngeal constriction or epilaryngeal vibration that occurs in various speech contexts. The prediction would be that acoustic markers of epilaryngeal activity will increase in the context of constriction-susceptible vowels such as [ɑ] and/or during moments of increased intensity, such as emphatically stressed syllables. It is the opinion of the author that this kind of epilaryngeal activity is quite common, even in English.

Finally, while perceptual explanations were hinted at several times throughout the dissertation, these were not given a full treatment. Epilaryngeal activity is associated with some very dramatic auditory effects, such as growling or raised larynx voice quality. The fact that this alone does not translate into widespread incorporation into phonological systems would seem to suggest that epilaryngeal vibration is either “too costly” to produce or “too extreme” an auditory effect for incorporation into language or perhaps “too associated” with emphatic speech or emotionally inflected speech. Yet several languages have incorporated these effects with more or less phonological function. Thus, the epilarynx is an interesting basis for asking questions about the “usefulness” vs. “costliness” relationship governing the shape and content of phonetic and phonemic systems.
8.3 Epilaryngeal coda

The epilarynx is quite possibly the most neglected part of the vocal tract and, as discussed in this dissertation, there is good reason as to why it was overlooked. The fundamental message laid out in the present document is that epilaryngeal influence is far more pervasive in speech than the relatively sparse treatment which it has been given in the literature would suggest. Painter's summary (see §2.2.2) of the linguistic applications of the epilarynx could not have been more insightful, and this dissertation stands as a testament to the validity of his intuition. The epilarynx exposes an entirely new dimension to sounds conventionally ascribed to lingual or vocal fold action. Whether one looks at vowel quality, tone, intonation, phonatory quality, laryngeal and/or pharyngeal consonants, ejectives, emphatics and/or retracted consonants, register and ATR harmony patterns, or whether one is looking at speech-peripheral cases such as singing, theatrical performance, imitations/impersonations or vocal caricaturization, proceeding without considering potential epilaryngeal effects will need justification: if one takes this dissertation seriously, then one cannot view the vocal folds as independent from the rest of the vocal tract any longer.
Bibliography


Bgaţba, X. S. (1964). *Bzybskij dialekt Abxazskogo jazyka (the Bzyb dialect of the Abkhaz language).* Tbilisi, USSR.


Mielke, J. (2004). *The emergence of distinctive features* (Doctoral dissertation). The Ohio State University, Columbus, Ohio.


ultrasound. Presented at the The 9th Phonetics Conference of China, Tianjin, China.


## Appendix A

### LIST OF LANGUAGES

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abaza/Ashkharwa/Ashkharua</td>
<td>Northwest Caucasian</td>
<td>Russia (Karachay-Cherkess Republic)</td>
</tr>
<tr>
<td>Abkhaz</td>
<td>Northwest Caucasian</td>
<td>Republic of Abkhazia</td>
</tr>
<tr>
<td>Agar Dinka</td>
<td>Nilo-Saharan (Nilotic)</td>
<td>South Sudan</td>
</tr>
<tr>
<td>Agul/Aghul</td>
<td>Northeast Caucasian</td>
<td>Russia (Southeastern Dagestan)</td>
</tr>
<tr>
<td>Akan</td>
<td>Niger-Congo (Kwa)</td>
<td>Ghana, Ivory Coast, Benin</td>
</tr>
<tr>
<td>Akkadian (extinct)</td>
<td>Semitic</td>
<td>Assyria, Babylonia</td>
</tr>
<tr>
<td>Akposso/Kposo/Ikposo</td>
<td>Niger-Congo (Kwa)</td>
<td>Ghana</td>
</tr>
<tr>
<td>Amis</td>
<td>Austronesian (East Formosan)</td>
<td>Taiwan</td>
</tr>
<tr>
<td>Arbore</td>
<td>Cushitic</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>Ateso</td>
<td>Nilo-Saharan (Nilotic)</td>
<td>Kenya, Uganda</td>
</tr>
<tr>
<td>Babine(-Witsuwit'en)</td>
<td>Na-Dené (North Athabaskan)</td>
<td>Canada (British Columbia)</td>
</tr>
<tr>
<td>Besleney/Kabardian</td>
<td>Northwest Caucasian</td>
<td>Circassia, Turkey, Jordan, Syria, Iraq</td>
</tr>
<tr>
<td>Bezhta</td>
<td>Northeast Caucasian</td>
<td>Russia (Southern Dagestan)</td>
</tr>
<tr>
<td>Bongo/Bonggo/Armopa</td>
<td>Austronesian (Sarmii)</td>
<td>Papua New Guinea, Indonesia</td>
</tr>
<tr>
<td>Bor Dinka</td>
<td>Nilo-Saharan (Nilotic)</td>
<td>South Sudan</td>
</tr>
<tr>
<td>Bruu</td>
<td>Mon-Khmer (Katuic)</td>
<td>Laos, Vietnam, Thailand</td>
</tr>
<tr>
<td>Buchan Aberdeenshire Scots</td>
<td>Indo-European (Germanic)</td>
<td>United Kingdom (Scotland)</td>
</tr>
<tr>
<td>Bzyb (dialect of Abkhaz)</td>
<td>Northwest Caucasian</td>
<td>Turkey, Abkhazia</td>
</tr>
<tr>
<td>Chilcotin/Tsilhqot’in</td>
<td>Na-Dené (North Athabaskan)</td>
<td>Canada (Chilcotin Country, BC)</td>
</tr>
<tr>
<td>D’opasunte</td>
<td>Each Cushitic</td>
<td>Ethiopia (East of the Wät’t o River)</td>
</tr>
<tr>
<td>Dargi/Dargwa/Dargin</td>
<td>Northeast Caucasian</td>
<td>Russia (Dagestan)</td>
</tr>
<tr>
<td>Dho-Luo</td>
<td>Nilo-Saharan (Nilotic)</td>
<td>Kenya, Tanzania</td>
</tr>
<tr>
<td>Dinghai</td>
<td>Sino-Tibetan</td>
<td>China (Ningbo, Zhejiang)</td>
</tr>
<tr>
<td>Dinka</td>
<td>Nilo-Saharan (Nilotic)</td>
<td>South Sudan</td>
</tr>
<tr>
<td>Diola-Fogny/Jola/Diola</td>
<td>Niger-Congo</td>
<td>Gambia, Senegal</td>
</tr>
<tr>
<td>Efik</td>
<td>Niger-Congo</td>
<td>Southern Nigeria</td>
</tr>
<tr>
<td>Even</td>
<td>Tungusic</td>
<td>Russia (Russia Far East)</td>
</tr>
<tr>
<td>Fula/Fulani/Pulaar</td>
<td>Niger-Congo</td>
<td>Mauritania, Senegal, Mali, Guinea…</td>
</tr>
<tr>
<td>Ge’ez (extinct)</td>
<td>Semitic (Ethiopian)</td>
<td>Ethiopia, Eritrea, Israel</td>
</tr>
<tr>
<td>Gimi</td>
<td>Kainantu-Goroka (Fore)</td>
<td>Papua New Guinea (Eastern Highlands)</td>
</tr>
<tr>
<td>Gitksan/Gitxsan</td>
<td>Tsimshian</td>
<td>Canada (Skeena region, BC)</td>
</tr>
<tr>
<td>Harsusi</td>
<td>Semitic (South Semitic)</td>
<td>Oman</td>
</tr>
<tr>
<td>Igbo</td>
<td>Niger-Congo</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Ijo/Ijaw</td>
<td>Niger-Congo</td>
<td>Southern Nigeria</td>
</tr>
<tr>
<td>Ilokano</td>
<td>Austronesian (Philippine)</td>
<td>Philippines (Northern Luzon)</td>
</tr>
<tr>
<td>Iraqw</td>
<td>Cushitic</td>
<td>Tanzania (Arusha Region)</td>
</tr>
<tr>
<td>Japanese/Nihongo</td>
<td>Japonic</td>
<td>Japan</td>
</tr>
<tr>
<td>Javanese</td>
<td>Austronesian (Sunda-Sulawesi)</td>
<td>Indonesia (Java), Suriname</td>
</tr>
<tr>
<td>Jianchuan Bai</td>
<td>Tibeto-Burman</td>
<td>China (Yunnan Province)</td>
</tr>
<tr>
<td>Jingpo</td>
<td>Sino-Tibetan</td>
<td>Burma, China (Yunnan Province)</td>
</tr>
<tr>
<td>Ju’hoansi/Southeastern</td>
<td>Kx’a (Khoisan-Group)</td>
<td>Namibia, Botswana</td>
</tr>
<tr>
<td>’Xuun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kabiyé</td>
<td>Gur (Gurunsi)</td>
<td>Togo, Benin, Ghana</td>
</tr>
<tr>
<td>Kasem</td>
<td>Niger-Congo (Gur)</td>
<td>Ghana, Burkina Faso</td>
</tr>
<tr>
<td>Kashaya</td>
<td>Fomo</td>
<td>United States (California)</td>
</tr>
<tr>
<td>Language/Region</td>
<td>Language Family</td>
<td>Location</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Kinande/Nande</td>
<td>Niger-Congo (Bantu)</td>
<td>Democratic Republic of the Congo</td>
</tr>
<tr>
<td>Kiribati/Gilbertese</td>
<td>Austronesian, Micronesian</td>
<td>Republic of Kiribati</td>
</tr>
<tr>
<td>Kirzan Armenian</td>
<td>Indo-European (Armenian)</td>
<td>Armenia</td>
</tr>
<tr>
<td>Lak</td>
<td>Northeast Caucasian</td>
<td>Russia (Southern Dagestan)</td>
</tr>
<tr>
<td>Lhasa Tibetan (Standard)</td>
<td>Sino-Tibetan</td>
<td>China (Tibet), Nepal, India</td>
</tr>
<tr>
<td>Lungtu/Lungu</td>
<td>Niger-Congo (Bantu)</td>
<td>Tanzania, Zambia</td>
</tr>
<tr>
<td>Maasai</td>
<td>Niger-Congo (Eastern Nilotic)</td>
<td>Kenya, Tanzania</td>
</tr>
<tr>
<td>Makah (extinct)</td>
<td>Southern Wakashan</td>
<td>US (Olympic Peninsula, Washington)</td>
</tr>
<tr>
<td>Malay</td>
<td>Austronesian (Malayic)</td>
<td>Indonesian, Malaysia, Brunei</td>
</tr>
<tr>
<td>Mehri</td>
<td>Semitic (South Semitic)</td>
<td>Yemen, Oman, Somalia</td>
</tr>
<tr>
<td>Monong</td>
<td>Mon-Khmer (Bahnaric)</td>
<td>Vietnam, Cambodia</td>
</tr>
<tr>
<td>Mon</td>
<td>Mon-Khmer (Monic)</td>
<td>Burma, Thailand</td>
</tr>
<tr>
<td>Morley Stoney (Nakoda)</td>
<td>Siouan</td>
<td>Southern Alberta, Canada</td>
</tr>
<tr>
<td>Mpi</td>
<td>Tibeto-Burman</td>
<td>Thailand</td>
</tr>
<tr>
<td>Murle</td>
<td>Nilo-Saharan (Surmic)</td>
<td>South Sudan, Ethiopia</td>
</tr>
<tr>
<td>Najdi Bedouin Arabic</td>
<td>Semitic (Central Semitic)</td>
<td>Saudi Arabia, Jordan, Iraq, Syria</td>
</tr>
<tr>
<td>Nahk</td>
<td>Northeast Caucasian</td>
<td>Russia (Chechnya and Ingushetia)</td>
</tr>
<tr>
<td>Nitinaht/Dicti dah</td>
<td>Southern Wakashan</td>
<td>Canada (Southern Vancouver Island, BC)</td>
</tr>
<tr>
<td>Naka’pamux/Thompson</td>
<td>Salishan (Interior Salish)</td>
<td>Thompson River Valley, BC</td>
</tr>
<tr>
<td>Nuuchahnulth</td>
<td>Southern Wakashan</td>
<td>Canada (Southern Vancouver Island, BC)</td>
</tr>
<tr>
<td>Nxa’amxcin/Moses-Columbian</td>
<td>Salishan (Interior Salish)</td>
<td>United States (Northern Idaho)</td>
</tr>
<tr>
<td>Oowekeyala</td>
<td>Northern Wakashan</td>
<td>Canada (North-central Coastal BC)</td>
</tr>
<tr>
<td>Oro Win</td>
<td>Chapakuran</td>
<td>Brazil, Bolivia</td>
</tr>
<tr>
<td>Quianvini Zapotec</td>
<td>Oto-manguan</td>
<td>Mexico (Oaxaca)</td>
</tr>
<tr>
<td>Rengao</td>
<td>North Bahnaric</td>
<td>Vietnam</td>
</tr>
<tr>
<td>Rutul</td>
<td>(related to Udi)</td>
<td>Russia (Southern Dagestan)</td>
</tr>
<tr>
<td>Shilluck/Shilluk</td>
<td>Nilo-Saharan (Nilotic)</td>
<td>South Sudan (Upper Nile State)</td>
</tr>
<tr>
<td>Shona (Karanga, Zezuru)</td>
<td>Niger-Congo (Bantu)</td>
<td>Zimbabwe, Mozambique, South Africa</td>
</tr>
<tr>
<td>Somali</td>
<td>Cushitic</td>
<td>Ethiopia, Djibouti, Kenya</td>
</tr>
<tr>
<td>St’át’imcets/Lillooet</td>
<td>Salishan (Interior Salish)</td>
<td>Canada (Middle Fraser, BC)</td>
</tr>
<tr>
<td>Takelma (extinct)</td>
<td>Penutian (or an isolate)</td>
<td>United States (Rogue Valley, Oregon)</td>
</tr>
<tr>
<td>Tamil</td>
<td>Dravidian</td>
<td>India, Sri Lanka, Malaysia</td>
</tr>
<tr>
<td>Tiberian Hebrew</td>
<td>Semitic (Central Semitic)</td>
<td>Israel</td>
</tr>
<tr>
<td>Tigre</td>
<td>Semitic (South Semitic)</td>
<td>Eritrea, Sudan</td>
</tr>
<tr>
<td>Toba</td>
<td>Guaiacuruan</td>
<td>Argentina, Paraguay, Bolivia</td>
</tr>
<tr>
<td>Tsez</td>
<td>Northeast Caucasian</td>
<td>Russia (Southern Dagestan)</td>
</tr>
<tr>
<td>Tsonga</td>
<td>Niger-Congo (Bantu)</td>
<td>Mozambique, South Africa, Swaziland</td>
</tr>
<tr>
<td>Tugen Kalenjin</td>
<td>Nilo-Saharan (Southern Nilotic)</td>
<td>Kenya (West Central Kenya)</td>
</tr>
<tr>
<td>Turkana</td>
<td>Nilo-Saharan (Eastern Nilotic)</td>
<td>Kenya (west of Lake Turkana)</td>
</tr>
<tr>
<td>Ubykh</td>
<td>Northwest Caucasian</td>
<td>Turkey</td>
</tr>
<tr>
<td>Udi</td>
<td>Northeast Caucasian</td>
<td>Azerbaijan, Russia, Georgia</td>
</tr>
<tr>
<td>Wari†</td>
<td>Chapakuran</td>
<td>Brazil, Bolivia</td>
</tr>
<tr>
<td>Yoruba</td>
<td>Niger-Congo (Volta-Niger)</td>
<td>Nigeria, Togo, Benin</td>
</tr>
<tr>
<td>Zhenhai (Wu)</td>
<td>Sino-Tibetan</td>
<td>Ningpo region, Zhenhai</td>
</tr>
<tr>
<td>Zulu</td>
<td>Niger-Congo (Bantu)</td>
<td>South Africa, Zimbabwe, Lesotho</td>
</tr>
</tbody>
</table>
## Appendix B

### DEFINITIONS OF TERMS

Definition of terms for lower vocal tract sounds used in this dissertation

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>lower vocal tract</td>
<td>The part of the vocal tract extending from the larynx up to and bounded by the oropharyngeal and velo-pharyngeal ports: the post-velar part of the vocal tract or the locus of so-called guttural sounds. Encompasses the uvular, pharyngeal, and laryngeal articulatory sites.</td>
</tr>
<tr>
<td>glottalization</td>
<td>Cessation of laryngeal vibration tantamount to glottal stop caused by closure of the glottis through vocal fold adduction and (possible) epilarynx closure, involving, first, ventricular incursion, and then aryepiglottic closure; may refer to secondary articulation (e.g. glottal reinforcement of [p] in [\textipa{th}\textipa{ʔ}] ‘top’) or to insertion of a glottal stop (e.g. in vowel-initial German words such as Apfel [\textipa{ʔApfl}] ‘apple’).</td>
</tr>
<tr>
<td>laryngealization</td>
<td>Constricted phonation ranging from creakiness to harshness depending on other factors such as subglottal pressure and specific laryngeal configuration. Often occurs before or after glottalization and is often a concomitant of pharyngealization.</td>
</tr>
<tr>
<td>pharyngealization</td>
<td>Narrowing of the lower vocal tract, particularly in the hypopharyngeal region. Involves a configuration similar to that of raised larynx voice but tending towards relatively lower pitch. Commonly applied as a label of the secondary articulation in Arabic emphatics and Salish retracted consonants, although the term uvularization is also sometimes applied to these sounds.</td>
</tr>
<tr>
<td>uvularization</td>
<td>Narrowing of the lower vocal tract, particularly in the upper pharynx where the uvula is located. An alternative label (and analysis) applied to Arabic emphatic consonants, which are most commonly regarded as being pharyngealized.</td>
</tr>
</tbody>
</table>
epiglottalization  Secondary articulation involving the epiglottis. May be used to indicate epilarynx vibration on vowels (as implied by terms such as sphincteric or strident voice/phonation).

ventricular incursion  Contact between the ventricular folds and the vocal folds resulting in constricted phonatory quality if there is vocal fold vibration, but it also is a suggested mechanism for arresting the vibratory motion of the vocal folds. Associated with constricted sounds.

aryepiglottic sphincter  A term used to describe epilarynx narrowing particularly at the level of the aryepiglottic folds. Involves approximation or contact between the dorsal surface of the infrahyoid region of the epiglottis (of which the epiglottic tubercle is a landmark) and the aryepiglottic folds (of which the cuneiform tubercles are important landmarks). Overlaps with the idea of laryngeal constriction. Associated with constricted sounds.

laryngeal constriction/sphincter mechanism  A more general term used to describe the mechanism associated with protective closure of the larynx as it is used in the production of lower vocal tract sounds. May imply activity of the aryepiglottic sphincter and/or ventricular incursion, and is associated with constricted sounds.

constricted/unconstricted  Denotes engagement or disengagement of the laryngeal sphincter/constrictor mechanism. Constricted sounds can involve ventricular incursion, aryepiglottic sphinctering, or a combination of those components (i.e. action of the laryngeal constrictor mechanism). Encompasses a wide range of sound categories including glottalization, laryngealization, epiglottalization, pharyngealization, raised larynx voice quality, creakiness, harshness, and growling.

raised larynx voice  A voice quality which saliently features a higher larynx position but involves narrowing of the epilarynx and pharynx. The counterpart of pharyngealization but with relatively higher pitch and involving generally tense or possibly laryngealized phonatory quality.

lowered larynx voice  A voice quality associated with the lengthening of the vocal tract expansion of the pharynx via extreme lowering of the larynx. It is similar in configuration to faucalized voice but with relatively lower pitch.
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>faucalized voice</td>
<td>A voice quality associated with vertical stretching of the pharynx and particularly the faucal pillars (from which the term derives its name) caused by larynx lowering: the sound a person makes when vocalizing while yawning. Parallel in articulation to lowered larynx voice, but with relatively higher pitch.</td>
</tr>
<tr>
<td>creakiness</td>
<td>A phonatory quality with an irregular glottal pulse that generates a laryngeal source of low intensity and low pitch. Sometimes called pulse register or vocal fry.</td>
</tr>
<tr>
<td>harshness</td>
<td>A phonatory quality with an irregular and intense laryngeal source not restricted by pitch range. With only vocal fold vibration it can alternatively be labeled tense voice or pressed voice; if there is epilaryngeal vibration it may alternatively be referred to as growling.</td>
</tr>
<tr>
<td>growling</td>
<td>A general term for a specific type of harshness involving vibration of some part of the epilarynx, such as the aryepiglottic folds, epiglottis, or ventricular folds.</td>
</tr>
<tr>
<td>voice quality</td>
<td>A holistic vocal tract configuration associated with specific acoustic and auditory properties, possibly involving phonatory quality superimposed over segmental and suprasegmental speech content. Can be long term or quasi-permanent in nature or can be localized to particular prosodic structures.</td>
</tr>
<tr>
<td>phonatory quality</td>
<td>A particular mode of laryngeal vibration, most often involving the vocal folds, associated with specific acoustic and auditory properties.</td>
</tr>
</tbody>
</table>
Appendix C

TRANSCRIPTION OF BURKIKHAN AND TPIG AGUL

Re-transcription of data for two dialects of Agul, Burkikhan (female speaker) and Tpig (male speaker), based on recordings by Kodzasov (“Aghul,” n.d.). Transcriptions of these data obtained from the UCLA Phonetics Archive (marked UCLA) are juxtaposed against those of the author (Moisik). It must be stated that, while both recordings are poor quality, the Tpig data are especially poor. Judging from the Nyquist frequency, it seems that the sample rates are 22100 Hz for the Burkikhan data and 11050 Hz for the Tpig data. In the cases claimed to have epilaryngeal vibration, it was audibly obvious, visually evident in the waveform, and typically corresponds with subharmonics in the frequency domain. It is possible, however, that epilaryngeal vibration may occur in other tokens, but a very conservative approach was taken to transcription of the data on account of its overall poor quality.

The rows are organized according to the UCLA transcription for Burkikhan then by context (vowel and syllable position). The Moisik transcriptions signify the following general realizations: [h] is a very noisy voiceless epilaryngeal fricative; [ʃ] is a voiced epilaryngeal fricative similar to that which occurs in Morley Stoney (only rarely is it an approximant); [ɦ] is a voiceless epilaryngeal trill (possibly aryepiglottic or epiglottal); [ʢ] is a voiced epilaryngeal trill (again, possibly aryepiglottic or epiglottal) usually occurring with significant co-articulatory leakage onto adjacent vowels; [ʔ] is a voiceless epilaryngeal stop often causing creakiness on preceding vowels and almost always associated with harshness on following vowels, often with concomitant epilaryngeal

Transcription of epilaryngeal trills as [ɦ] and [ʢ] is done following Esling (2010).
vibration. The Notes column includes the following details: ʰ = voiceless epilaryngeal trill; ʕ = voiced epilaryngeal trill; Ʉ = creaky voice; Ʉ = harsh voice without (obvious) epilaryngeal vibration; Ʉ ʰ = harsh voice with obvious epilaryngeal vibration (growling); Ʉ ʕ = pharyngealization/raised larynx voice quality; Ʉ = whispery phonatory quality; ʕ = “tight” voiced approximant-fricative; ? = uncertainty.

<table>
<thead>
<tr>
<th>English Gloss</th>
<th>UCLA Burikhan</th>
<th>Tpig</th>
<th>Moisik Burikhan</th>
<th>Tpig</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>sheep</td>
<td>ha</td>
<td>hub</td>
<td>ha</td>
<td>hub</td>
<td>ʕ, Ʉ?</td>
</tr>
<tr>
<td>wolf</td>
<td>hač</td>
<td>nuč</td>
<td>haʃ</td>
<td>huʃ</td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>ġad</td>
<td>ġan</td>
<td>ġad</td>
<td>ġan</td>
<td></td>
</tr>
<tr>
<td>courtyard</td>
<td>han</td>
<td>ġan</td>
<td>han</td>
<td>ġan</td>
<td>ġan</td>
</tr>
<tr>
<td>cow</td>
<td>hani</td>
<td>huni</td>
<td>hani</td>
<td>huni</td>
<td>ġan</td>
</tr>
<tr>
<td>village</td>
<td>har</td>
<td>hur</td>
<td>har</td>
<td>hur</td>
<td></td>
</tr>
<tr>
<td>flour</td>
<td>har</td>
<td>ġur</td>
<td>har</td>
<td>ġur</td>
<td>ġan</td>
</tr>
<tr>
<td>to know</td>
<td>harxas</td>
<td>ģeβaxas</td>
<td>harxas</td>
<td>ģeβaxas</td>
<td></td>
</tr>
<tr>
<td>gullet</td>
<td>harúd</td>
<td></td>
<td>ha’rúd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>collar</td>
<td>haw</td>
<td>haw</td>
<td>haw</td>
<td>haw</td>
<td></td>
</tr>
<tr>
<td>udder</td>
<td>ġaw</td>
<td>ġaw</td>
<td>ġaw</td>
<td>ġaw</td>
<td></td>
</tr>
<tr>
<td>grandfather</td>
<td>ġadar</td>
<td></td>
<td>ġadar</td>
<td></td>
<td>ġadar</td>
</tr>
<tr>
<td>mill(s)</td>
<td>rah / raḥár</td>
<td>raʃ / raʃar</td>
<td>raʃh:</td>
<td>raʃh:</td>
<td>raʃh:</td>
</tr>
<tr>
<td>to beat</td>
<td>jarḥas</td>
<td>jirḥas</td>
<td>jirḥas</td>
<td>jirḥas</td>
<td>jirḥas</td>
</tr>
<tr>
<td>to thresh</td>
<td>ruḥas</td>
<td>rušas</td>
<td>rušas</td>
<td>rušas</td>
<td>rušas</td>
</tr>
<tr>
<td>to protect</td>
<td>uḥas</td>
<td>uḥas</td>
<td>uḥas</td>
<td>uḥas</td>
<td>uḥas</td>
</tr>
<tr>
<td>shed(s), barn(s)</td>
<td>muḥ / muḥár</td>
<td></td>
<td>muḥ / mu’har</td>
<td>muḥ / mu’har</td>
<td>muḥ / mu’har</td>
</tr>
<tr>
<td>to stand</td>
<td>kārḥas</td>
<td></td>
<td>kārḥas</td>
<td></td>
<td>kārḥas</td>
</tr>
<tr>
<td>apple</td>
<td>hāčh</td>
<td>ħeč</td>
<td>ħeʃ</td>
<td>ħeʃ</td>
<td>ħeʃ</td>
</tr>
<tr>
<td>boiler</td>
<td>hājaq</td>
<td>hūjaq</td>
<td>hējeg</td>
<td>hējeg</td>
<td>hējeg</td>
</tr>
<tr>
<td>horse</td>
<td>hājwan</td>
<td>hājwan</td>
<td>hējwan</td>
<td>hējwan</td>
<td>hējwan</td>
</tr>
<tr>
<td>joint</td>
<td>hālq’a</td>
<td></td>
<td>hēlq’a:</td>
<td>hēlq’a:</td>
<td>hēlq’a:</td>
</tr>
<tr>
<td>grasshopper</td>
<td>hārčaj</td>
<td>hūl</td>
<td>hērʃaj</td>
<td>hērʃaj</td>
<td>hērʃaj</td>
</tr>
<tr>
<td>sea</td>
<td>hūl</td>
<td>harud</td>
<td>hūl</td>
<td>harud</td>
<td>harud</td>
</tr>
<tr>
<td>thick gut</td>
<td>hūrčal</td>
<td>hūf</td>
<td>hūf</td>
<td>hūf</td>
<td>hūf</td>
</tr>
<tr>
<td>hamster</td>
<td>hūf</td>
<td></td>
<td>hūf</td>
<td></td>
<td>hūf</td>
</tr>
<tr>
<td>sharp</td>
<td>mān / mānār</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
</tr>
<tr>
<td>whey(s)</td>
<td>mān / mānār</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
</tr>
<tr>
<td>whey(s)</td>
<td>mān / mānār</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
<td>mēh / mēhɛr</td>
</tr>
<tr>
<td>English</td>
<td>Hebrew</td>
<td>Arabic</td>
<td>Greek</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>caraway seed(s)</td>
<td>הָבָאֵשׁ</td>
<td>דִּבְלְפָא</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>goat(s)</td>
<td>הָנָא</td>
<td>אָד</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>whey(s)</td>
<td>חָנִית</td>
<td>דַּבֵל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>soft</td>
<td>שַׂדּוֹל</td>
<td>דַּבֹּל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>partridge</td>
<td>שַׂדָּל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>potato</td>
<td>שָׂדָל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>שַׂדָּל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sweat</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stomach</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shoulder</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hare</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hunter</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to fall between</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to pass between</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>crest(s)</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>raven</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to enter</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bridge(s)</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>does not come back</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>two</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wet</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>heel</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mouse</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>empty</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>motley</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plug</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cradle</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frog</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>naked</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eagle</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>center(s)</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be quiet</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.Pl/Sg.Imp)</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>horseshoe</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bridle</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>light</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to cry</td>
<td>שַׂדּוֹל</td>
<td>דַּבּוֹל</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Syriac</td>
<td>Transliteration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wide</td>
<td>ܡܪܐܡܐܦ</td>
<td>ʔarʔaf</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deaf person</td>
<td>ܒܝܒܐܒ</td>
<td>ʔâbab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>ܒܝܓܘ</td>
<td>ʔâıc’ in urk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>armpit</td>
<td>ܒܝܓܘ</td>
<td>ʔâbab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dumb, mute</td>
<td>ܒܐܒܐܒ</td>
<td>ʔâbab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>excrement</td>
<td>ܒܐܒܐܒ</td>
<td>ʔâbab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>heavy</td>
<td>ܒܐܒܐܒ</td>
<td>ʔâbab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>night</td>
<td>ܒܐܒܐܒ</td>
<td>ʔâbab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>measure(s)</td>
<td>ܣܐ/ ܣܐእܪ</td>
<td>se:ʔ/ se:ʔer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root(s)</td>
<td>ܡܪܐ/ ܡܪܐܪ</td>
<td>p:ʔ/ p:ʔer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hen</td>
<td>ܡܪܐ/ ܡܪܐܪ</td>
<td>p:ʔ/ p:ʔer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oil</td>
<td>ܟܬܡܢ</td>
<td>ʕalaman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dogrose</td>
<td>ܟܝܒܘ</td>
<td>ʕalaw</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bear</td>
<td>ܒܓܢܝܫ</td>
<td>bâganiš</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nine</td>
<td>ܫܪܐ ܫܪܐ</td>
<td>jarč’u’d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forelock(s)</td>
<td>ܡܐܠܓ</td>
<td>māq małeq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>worm</td>
<td>ܡܠܝܐ</td>
<td>māq małeq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rust</td>
<td>ܡܠܝܐ</td>
<td>māq małeq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td>ܡܠܝܐ</td>
<td>māq małeq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sheet</td>
<td>ܡܠܝܐ</td>
<td>māq małeq</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix D

### LIST OF MATHEMATICAL SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Value</th>
<th>SI Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>speed of sound in air</td>
<td>343</td>
<td>m·s⁻¹</td>
</tr>
<tr>
<td>d</td>
<td>damping coefficients</td>
<td>varies</td>
<td>kg·s⁻¹ (N·s·m⁻¹)</td>
</tr>
<tr>
<td>h</td>
<td>height</td>
<td>varies</td>
<td>m</td>
</tr>
<tr>
<td>l</td>
<td>moment of inertia</td>
<td>varies</td>
<td>kg·m²</td>
</tr>
<tr>
<td>k</td>
<td>stiffness (spring constant)</td>
<td>varies</td>
<td>kg·s⁻² (N·m⁻¹)</td>
</tr>
<tr>
<td>m</td>
<td>mass</td>
<td>varies</td>
<td>kg</td>
</tr>
<tr>
<td>l</td>
<td>length</td>
<td>varies</td>
<td>m</td>
</tr>
<tr>
<td>p</td>
<td>pressure</td>
<td>varies</td>
<td>kg·m⁻¹ s⁻²·(N·m⁻¹, Pa)</td>
</tr>
<tr>
<td>w</td>
<td>width</td>
<td>varies</td>
<td>m</td>
</tr>
<tr>
<td>ζ</td>
<td>damping ratio</td>
<td>varies</td>
<td>dimensionless</td>
</tr>
<tr>
<td>Φ</td>
<td>phase difference</td>
<td>varies</td>
<td>°</td>
</tr>
<tr>
<td>η</td>
<td>coefficient of nonlinearity</td>
<td>varies</td>
<td>dimensionless</td>
</tr>
<tr>
<td>μ</td>
<td>dynamic viscosity of the air</td>
<td>1.86×10⁻⁵</td>
<td>kg·m⁻¹·s⁻¹ (Pa·s)</td>
</tr>
<tr>
<td>ρ</td>
<td>density</td>
<td>varies</td>
<td>kg·m⁻³</td>
</tr>
</tbody>
</table>
# Appendix E

## LIST OF MUSCLES AND MUSCLE ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Muscle Name</th>
<th>Muscle Type/Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>aDG</td>
<td>anterior digastric</td>
<td>mandibular, hyoidal</td>
</tr>
<tr>
<td>AE</td>
<td>aryepiglottic</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>CP</td>
<td>cricopharyngeus</td>
<td>pharyngeal, laryngeal</td>
</tr>
<tr>
<td>CT</td>
<td>cricothyroid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>eTA</td>
<td>external thyroarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>GG</td>
<td>genioglossus</td>
<td>lingual</td>
</tr>
<tr>
<td>GGA</td>
<td>genioglossus anterior</td>
<td>lingual</td>
</tr>
<tr>
<td>GGM</td>
<td>genioglossus middle</td>
<td>lingual</td>
</tr>
<tr>
<td>GGP</td>
<td>genioglossus posterior</td>
<td>lingual</td>
</tr>
<tr>
<td>GH</td>
<td>geniohyoid</td>
<td>mandibular, hyoidal, lingual</td>
</tr>
<tr>
<td>HG</td>
<td>hyoglossus</td>
<td>lingual, hyoidal</td>
</tr>
<tr>
<td>IA</td>
<td>interarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>infrahyoid strap muscles</td>
<td>SH, ST, TH, OH</td>
<td>extrinsic laryngeal, hyoid</td>
</tr>
<tr>
<td>iPc</td>
<td>inferior pharyngeal constrictor</td>
<td>pharyngeal, intrinsic laryngeal</td>
</tr>
<tr>
<td>iTA</td>
<td>internal thyroarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>LCA</td>
<td>lateral cricoarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>lingual floor muscles</td>
<td>MH, GH, aDG</td>
<td>lingual, mandibular, hyoidal</td>
</tr>
<tr>
<td>MH</td>
<td>mylohyoid</td>
<td>mandibular, hyoidal, lingual</td>
</tr>
<tr>
<td>mPC</td>
<td>medial pharyngeal constrictor</td>
<td>pharyngeal, hyoidal, lingual</td>
</tr>
<tr>
<td>OH</td>
<td>omohyoid</td>
<td>hyoidal, mandibular</td>
</tr>
<tr>
<td>oIA</td>
<td>oblique interarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>PCA</td>
<td>posterior cricoarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>pDG</td>
<td>posterior digastric</td>
<td>hyoidal, laryngeal</td>
</tr>
<tr>
<td>pharyngeal constrictors (PC)</td>
<td>iPc, mPC, sPC</td>
<td>pharyngeal, laryngeal, lingual</td>
</tr>
<tr>
<td>pharyngeal levators</td>
<td>PP, SP, SalP, PC</td>
<td>pharyngeal, laryngeal, velic</td>
</tr>
<tr>
<td>PP</td>
<td>palatopharyngeus</td>
<td>pharyngeal, laryngeal, velic</td>
</tr>
<tr>
<td>SalP</td>
<td>salpingopharyngeus</td>
<td>pharyngeal, laryngeal</td>
</tr>
<tr>
<td>SG</td>
<td>styloglossus</td>
<td>lingual</td>
</tr>
<tr>
<td>SH</td>
<td>sternohyoid</td>
<td>hyoidal, laryngeal</td>
</tr>
<tr>
<td>SP</td>
<td>stylopharyngeus</td>
<td>pharyngeal, laryngeal</td>
</tr>
<tr>
<td>sPC</td>
<td>superior pharyngeal constrictor</td>
<td>pharyngeal, velic</td>
</tr>
<tr>
<td>ST</td>
<td>sternothyroid</td>
<td>laryngeal, hyoidal</td>
</tr>
<tr>
<td>suprahypoid strap muscles</td>
<td>MH, GH, aDG, pDG</td>
<td>hyoidal, mandibular, lingual, lar.</td>
</tr>
<tr>
<td>TA</td>
<td>thyroarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>TE</td>
<td>thyroepiglottic</td>
<td>intrinsic laryngeal</td>
</tr>
<tr>
<td>TH</td>
<td>thyrohyoid</td>
<td>laryngeal, hyoidal</td>
</tr>
<tr>
<td>tIA</td>
<td>transverse interarytenoid</td>
<td>intrinsic laryngeal</td>
</tr>
</tbody>
</table>
Appendix F

PHONOLOGICAL EXTENSIONS

Figure 8.1: Potential unconstricted-constricted registers. Diagram simplifications: Brackets = broad possibilities for association (not specific in this abstract case); Gray boxes = synergies (see Figure 6.5). Phonemics/Devices symbols: V = modal register; Y = breathy register; Й = faucalized register; Ъ = creaky register; Ы = harsh register (without epilaryngeal vibration); Y^f = “raised larynx voice” register. Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; vto/vtc = relatively open/close (anterior) vocal tract; dbe = “double-bunching” epilaryngeal configuration; tfr = tongue fronting (forward); tre = tongue retraction (down
and back); tra = tongue raising (up and back); ↑lx/↓lx = raised/lowered larynx. (See Appendix F, Figure 8.2 for a more comprehensive version of the diagram.)

**Figure 8.2:** Cross-height harmony (“ATR”) in the LPP. Grayness of the association lines represents association strength (e.g. [i] is likely to involve greater fronting of the tongue than [e]); Abbreviations: vfo/vfc = vocal fold opening/closure (or abduction/adduction); epc = epilaryngeal constriction; vto/vtc = relatively open/close (anterior) vocal tract; tfr = tongue fronting (forward); tre = tongue retraction (down and back); ↑lx/↓lx = raised/lowered larynx.
Figure 8.3: Linkage of voicing, register, and cross-height harmony in the LPP. To maintain simplicity, only a few potentials are shown (e.g. “ATR” contrast is illustrated with /e/ /ɛ/ vowels). Devices: [˯] = vocal fold vibration (voicing, as in voiced stops); [˯] = vocal fold abduction (voiceless, as in aspirated stops); [ʔ] = vocal fold adduction (glottal stop, glottalization); [ˡ] = pre-phonation (voiceless, as in unaspirated stops); [V] = relatively peripheral vowels; [V̂] = relatively condensed vowels; [Ṽ] = breathy phonation; [V̈] = creaky phonation; [V̄] = harsh phonation (without epilaryngeal vibration); [V̂]/[Ṽ] = relatively low/high F0; [Ṽ]/[V] = relatively fronted/retracted lingual position (+/−ATR vowels); [V̈] = “raised larynx voice quality”/“double-bunching” epilaryngeal configuration applied to vowels; [V̈] = epilaryngeal vibration (“growled” vowels). Abbreviations: vfo/vfc = vocal fold opening/closure (or
abduction/adduction); epc = epilaryngeal constriction; vto/vtc = relatively open/close (anterior) vocal tract; tfr = tongue fronting (forward); tre = tongue retraction (down and back); ↑lx/↓lx = raised/lowered larynx.

Table 8.5: Epilaryngeal vibration in speech sound systems

<table>
<thead>
<tr>
<th>Gross States &amp; Synergistic Relations</th>
<th>Significance</th>
<th>Attested</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>{epv ↔ L}</td>
<td>epilaryngeal vibration biases low F0 (mechanically: increased oscillating mass; acoustically: subharmonics)</td>
<td>!Xóô, Ju’hoansi, N</td>
<td>uu, Bai, Zhenhai Wu</td>
</tr>
<tr>
<td>{epv ↔ H}</td>
<td>high F0 does not tend to occur with epilaryngeal vibration because increased longitudinal tension is in conflict with epilaryngeal stricture</td>
<td>!Xóô, Ju’hoansi, N</td>
<td>uu, Bai, Zhenhai Wu</td>
</tr>
<tr>
<td>{vfo ↔ epv}</td>
<td>increased airflow from vocal fold abduction is helpful in driving epilaryngeal vibration</td>
<td>!Xóô, Zhenhai Wu</td>
<td>7.1.3</td>
</tr>
<tr>
<td>{epc ↔ epv}</td>
<td>epilaryngeal vibration can occur in the context of “pharyngeals”</td>
<td>Iraqi Arabic, Agul</td>
<td>7.1.4</td>
</tr>
</tbody>
</table>
Table 8.6: Epilarynx interaction with the vocal folds

<table>
<thead>
<tr>
<th>Gross States &amp; Synergistic Relations</th>
<th>Significance</th>
<th>Attested</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuum of hypopharyngeal stops</td>
<td>“Glottal constrictions” can be split into two levels giving rise to strong-vs.-weak glottal stops /ʔ/ and /ʔ/ (n.b. not necessarily /ʔ/ vs. /ʔ/)</td>
<td>Jingpo, Gimi, Quianviní Zototec, Dargi (?)</td>
<td>7.2.1</td>
</tr>
<tr>
<td>{↑lx ↔ epc} {tre ↔ epc} {vfc ↔ epc} {vfo ↔ epc} {vfo ↔ ↑lx}</td>
<td>/ʔ/ and /ʔ/ are related by the hypopharyngeal stop continuum; glottal stop has an affinity with pharyngeals (and constricted ejectives) that /h/ does not have</td>
<td>Tigre, Inor, Fula, Harsusi</td>
<td>7.2.2</td>
</tr>
<tr>
<td>continuum of hypopharyngeal stops</td>
<td>Pharyngeals can be allophonically augmented with pharyngeal constriction – a concept that is not expressible without incorporation of the epilarynx; [ɑ] may have more pharyngeal stricture than “pharyngeals” such as [ʔ], which will have more epilaryngeal stricture</td>
<td>Amis</td>
<td>7.2.3</td>
</tr>
<tr>
<td>{vfc ↔ epc} {vfc ↔ ↑lx} {↑lx ↔ epc} {tre ↔ epc}</td>
<td>Sounds which involve more gross-state alignment to pharyngeals are more likely to become pharyngeals; ejectives involve larynx raising but not necessarily epilaryngeal stricture since vocal fold closure is efficient at resisting an overpressure in the supralaryngeal cavity</td>
<td>Southern Wakashan</td>
<td>7.2.4</td>
</tr>
<tr>
<td>{tre ↔ ↑lx} {↑lx ↔ epc} {tre ↔ epc} {tre ↔ vfo} {vfc ↔ epc}</td>
<td>Relatively open vowels, especially [ɑ], can bias epilaryngeal stricture; they are not necessarily produced with epilaryngeal stricture; relatively close vowels are less likely to exhibit this bias</td>
<td>Numerous (Klallam, Malay, Shona, Sliammon, …)</td>
<td>7.2.5</td>
</tr>
<tr>
<td>{↑lx ↔ vfc} {↑lx ↔ epc} {epc ↔ tfr}</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8.7: Epilarynx interaction with the supralaryngeal vocal tract

| Gross States & Synergistic Relations | Significance                                                                 | Language Examples                      | Sectio

| {tre ↔ epc}                        | Pharyngeals can become associated with the “double-bunching” epilaryngeal configuration, which has synergies with retraction and tongue fronting. This is the articulatory basis for phonological patterns that group pharyngeals and palatals; it also explains the “r-colouring” on some pharyngeals because the “double-bunching” configuration is associated with rhotic quality. | Arabic, D’opaasunte, Nxa’a’mxcin, Lak, Bezhta, Lezgi, Tabasaran, Udi, Tsez, Tsakhur, Amis, Even | 7.3.1 |
| {db ↔ tre}                         | Register-like systems tend split into unconstricted and constricted sets, which share a wide range of properties that may include phonatory quality, vowel quality, and overall voice quality; vowel space may be peripheral-condensed or relatively close-open, or the constricted set may be biased towards retraction; it is possible for multiple registers to be instantiated, such as three-way or even four-way systems, like Bor Dinka, where more extreme opposing phonetic properties are employed as a means to differentiate the registers. | | |
| {↑lx ↔ epc}                        | ATR languages represent the most vowel-oriented of the register-like systems; these typically show the same phonetic property biases that other register systems show, such as unconstricted/constricted phonatory correlates with [+/-ATR] vowel sets. While fundamentally some of these systems may have originated as phonatory contrasts, they can reorganize around other articulatory “pivots”, less closely related to epilaryngeal activity. | Akan, Kabiye, Tugen, Kalenjin | 7.3.3 |

7.3.1

7.3.2

7.3.3