The Punctuation and Intonation of Parentheticals

by

Christel Bodenbender
B.Sc., University of Victoria, 1999

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

MASTER OF ARTS

in the Department of Linguistics

We accept this thesis as conforming
to the required standard

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

Dr. Tadao Miyamoto, Departmental Member (Department of Linguistics)

Dr. Suzanne Urbanczyk, Departmental Member (Department of Linguistics)

Dr. Peter F. Driessen, External Examiner (Department of Electrical and Computer Engineering, University of Victoria)

Supervisor: Dr. John H. Esling

ABSTRACT

From a historical perspective, punctuation marks are often assumed to be present in a text only to represent some of the phonetic structure of the spoken form of that text. It has been argued recently that punctuation today is a linguistic system that not only represents some of the phonetic sentence structure but also syntactic as well as semantic sentence structures. One case in point is the observation that the semantic difference in differently punctuated parenthetical phrases is not reflected in the intonation contour. This study provides the acoustic evidence for this observation and makes recommendations to achieve natural-sounding text-to-speech output for English parentheticals.

The experiment conducted for this study involved three male and three female native speakers of Canadian English reading aloud a set of 20 sentences with parenthetical and non-parenthetical phrases. These sentences were analyzed with respect to acoustic characteristics due to differences in punctuation as well as due to differences between parenthetical and non-parenthetical phrases.

A number of conclusions were drawn based on the results of the experiment: (1) a difference in punctuation, although entailing a semantic difference, is not reflected in the intonation pattern; (2) in contrast to the general understanding that parenthetical phrases are lower-leveled and narrower in pitch range than the surrounding sentence, this study shows that it is not the parenthetical phrase itself that is implemented differently from its non-parenthetical counterpart; rather, the phrase that precedes the parenthetical

exhibits a lower baseline and with that a wider pitch range than the corresponding phrase in a non-parenthetical sentence; (3) sentences with two adjacent parenthetical phrases or one embedded in the other exhibit the same pattern for the parenthetical-preceding phrase as the sentences in (2) above and a narrowed pitch range for the parenthetical phrases that are not in the final position of the sequence of parentheticals; (4) no pausing pattern could be found; (5) the characteristics found for parenthetical phrases can be implemented in synthesized speech through the use of SABLE speech markup as part of the SABLE speech synthesis system.

This is the first time that the connection between punctuation and intonation in parenthetical sentences has been investigated; it is also the first look at sentences with more than one parenthetical phrase. This study contributes to our understanding of the intonation of parenthetical phrases in English and their implementation in text-to-speech systems, by providing an analysis of their acoustic characteristics.

Examiners:

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

Dr. Tadao Miyamoto, Departmental Member (Department of Linguistics)

Dr. Suzanne Urbanczyk, Departmental Member (Department of Linguistics)

Dr. Peter F. Driessen, External Examiner (Department of Electrical and Computer Engineering, University of Victoria)

TABLE OF CONTENTS

Chapter One INTRODUCTION 1
1.1 Purpose of this study 1
1.2 Research questions 6
1.3 Limitations of the study 7
1.4 Outline 8

Chapter Two PARENTHETICALS REVIEWED 10
2.1 Definition 10
2.2 Punctuation 12
2.3 Intonation 15
2.4 Text-to-speech synthesis 19

Chapter Three METHODOLOGY 29

Chapter Four ANALYSIS 38
4.1 Pitch 38
4.1.1 Non-parentheticals 38
4.1.2 Single-parentheticals 44
4.1.2.1 Comma-parentheticals 44
4.1.2.2 Dash-parentheticals 46
4.1.2.3 Bracket-parentheticals 48
4.1.2.4 Comparison of differently punctuated parentheticals 51
4.1.3 Two-parentheticals 53
4.1.3.1 Nested two-parentheticals 53
4.1.3.2 Sequential two-parentheticals 55
4.1.3.3 Comparison between two-parentheticals and single-parentheticals 57
4.1.4 Comparison between parentheticals and non-parentheticals 60
4.1.4.1 Toplines 62
4.1.4.2 Baselines 63
4.1.4.3 Pitch ranges 67
4.1.4.4 Conclusion 71
4.2 Pauses 72
4.2.1 Sentences with one medial phrase 73
4.2.2 Two-parentheticals 75
4.2.1 Conclusion 77

Chapter Five SYNTHESIZING PARENTHETICALS 78
5.1 H-L-based encoding 79
5.2 Phonetic encoding 81
5.3 Encoding using markup 83
5.3.1 Single-parentheticals 84
5.3.2 Multi-parentheticals 86
5.3.2.1 Baseline comparison 89
5.3.2.1.1 Female baseline 90
5.3.2.1.2 Male baseline 94
5.3.2.1.3 Baseline summary 97

5.3.2.2 Pitch range comparison ... 98
5.3.2.2.1 Female pitch range ... 99
5.3.2.2.2 Male pitch range ... 101
5.3.2.2.3 Pitch range summary ... 103
5.3.2.3 Synthesizing multi-parentheticals ... 103
5.4 Summary ... 105

Chapter Six CONCLUSION ... 106
6.1 Findings ... 106
6.2 Future studies ... 109

REFERENCES ... 112

APPENDIX A Parenthetical background ... 119
APPENDIX B Set of sentences ... 121
APPENDIX C Female frequency data ... 123
APPENDIX D Male frequency data ... 126
APPENDIX E Pause data ... 129

FIGURES

Fig. 1.1 Stylized representation of the pitch contour of a sentence containing a parenthetical 2

Fig. 2.1 A model for using ACSS and SABLE 26

Fig. 3.1 A pitch contour and its corresponding topline/baseline representation 32
Fig. 3.2 Comparison of 70-500 Hz and 140-300 Hz analysis pitch ranges for females .. 34
Fig. 3.3 Comparison of 70-500 Hz and 75-225 Hz analysis pitch ranges for males 35

Fig. 4.1 Toplines and baselines for female and male unpunctuated non-parenthetical sentences 39
Fig. 4.2 Pitch ranges in female and male unpunctuated non-parenthetical sentences 40
Fig. 4.3 Toplines and baselines for female and male punctuated non-parenthetical sentences 41
Fig. 4.4 Pitch ranges in female and male punctuated non-parenthetical sentences 43
Fig. 4.5 Toplines and baselines for female and male comma-punctuated parenthetical sentences 45
Fig. 4.6 Pitch ranges in female and male comma-punctuated parenthetical sentences ... 46
Fig. 4.7 Toplines and baselines for female and male dash-punctuated parenthetical sentences 47
Fig. 4.8 Pitch ranges in female and male dash-punctuated parenthetical sentences 48
Fig. 4.9 Toplines and baselines for female and male bracket-punctuated parenthetical sentences 49
Fig. 4.10 Pitch ranges in female and male bracket-punctuated parenthetical sentences . 50
Fig. 4.11 Toplines and baselines for female and male nested two-parenthetical sentences 54
Fig. 4.12 Pitch ranges in female and male nested two-parenthetical sentences 55
Fig. 4.13 Toplines and baselines for female and male sequential two-parenthetical sentences 56
Fig. 4.14 Pitch ranges in female and male sequential two-parenthetical sentences 57
Fig. 4.15 Comparison between fr(a) pitch trough values in non-parenthetical sentence 2 and parenthetical sentence 4 66
Fig. 4.16 Female pitch ranges of medial and final phrases in relation to their respective initial phrases 69
Fig. 4.17 Male pitch ranges of medial and final phrases in relation to their respective initial phrases 69
Fig. 4.18 Female pause length at different boundaries 73
Fig. 4.19 Male pause length at different boundaries 75
Fig. 4.20 Pauses in sentences with two parentheticals 76

Fig. 5.1 Female pitch range trend in sequential and nested two-parentheticals 87
Fig. 5.2 Male pitch range trend in sequential and nested two-parentheticals 88

Fig. 5.3 Development of the non-parenthetical baseline over the course of the sentences for both females and males 89

Fig. 5.4 The effect of the use of different multipliers on the non-parenthetical female baseline function 92
Fig. 5.5 Development of the non-parenthetical pitch range over the course of the sentences for both females and males..98

TABLES

Table 4.1 Non-parenthetical toplines and baselines are unaffected by the presence or absence of punctuation...42
Table 4.2 Parenthetical toplines, baselines and pitch ranges are unaffected by the use of different punctuation marks..51
Table 4.3 Comparison of the location of corresponding segments within different parenthetical sentences for females 58
Table 4.4 Comparison of the location of corresponding segments within different parenthetical sentences for males 59
Table 4.5 Female topline comparison...62
Table 4.6 Male topline comparison 62
Table 4.7 Female baseline comparison...63
Table 4.8 Male baseline comparison 64
Table 4.9 Female pitch range comparison 67
Table 4.10 Male pitch range comparison...68

Table 5.1 Results for the preliminary function for the non-parenthetical female baseline 91
Table 5.2 The three closest-fit non-parenthetical female baseline functions 92
Table 5.3 Difference between female sequential two-parenthetical sentences and corresponding non-parenthetical sentences 93
Table 5.4 Difference between female nested two-parenthetical sentences and corresponding non-parenthetical sentences 94
Table 5.5 The three closest-fit male baseline functions...95
Table 5.6 Difference between male sequential two-parenthetical sentences and corresponding non-parenthetical sentences 96
Table 5.7 Difference between male nested two-parenthetical sentences and corresponding non-parenthetical sentences 97
Table 5.8 Difference between female sequential and nested two-parenthetical sentences and corresponding non-parenthetical sentences 100
Table 5.9 The three closest-fit male pitch range functions...101
Table 5.10 Difference between male sequential and nested two-parenthetical sentences and corresponding non-parenthetical sentences 102

ACKNOWLEDGEMENTS

This thesis could not have been written without the help of Dr. John Esling who took me under his supervisory wings. Thanks also go to my committee members – Dr. Suzanne Urbanczyk for greatly improving my academic research and writing skills, and Dr. Tadao Miyamoto for providing me with the necessary background in acoustic phonetics. Thank you to Dr. Peter Driessen for agreeing to be my external examiner.

Special thanks go to Dr. Ewa Czaykowska-Higgins for introducing me to phonetics when I came to the University of Victoria as an exchange student many years ago and for a great amount of help when I entered the MA program years later. Thanks also to Dr. Leslie Saxon, Dr. Hua Lin, Dr. Thomas Hukari and the other faculty members in the Department of Linguistics for their support and encouragement. Thank you to Greg Newton for technical support as well as to the six experiment participants.

I wish to thank the 2001/2002 team of the student newspaper *the Martlet* for a fun year and the inspiration for a great topic. Furthermore, thanks go to all the amazing fellow grad students, as well as my parents, Eddie, Andrea, Alex, Lori, the UVSS Women's Centre collective and all my friends and family.

Chapter One

INTRODUCTION

1.1 Purpose of this study

This thesis investigates whether different punctuation marks used for marking parentheticals, i.e. commas, dashes and brackets, correspond to different manifestations in prosody. Nunberg (1990) claimed – based on an informal investigation – that different interpretations of a sentence, resulting from a use of different punctuation marks, have no correspondence in prosody. However, no evidence supporting this claim has been provided yet.

Parentheticals consist of words, phrases or sentences that are inserted into a sentence to provide additional explanatory or commentary information. Their nature is that of a digression that is semantically related but not semantically essential to the sentences they are in.

For an acoustic study of parentheticals, pitch is the primary prosodic feature to be investigated as parentheticals in English are found to exhibit a lower-leveled and narrower pitch range than the surrounding sentence constituents (Bolinger, 1989, p. 186; Cruttenden, 1997; Crystal, 1969; Grosz & Hirschberg, 1992; Kutik et al., 1983; O'Shaughnessy, 1990; Wichmann, 2000). Pitch refers to the perceptual sensation of the frequency of vocal fold vibrations. The frequency of the vibrations is also referred to as the fundamental frequency (F0), and its characteristics allow us, for instance, to distinguish between male and female voices, with females exhibiting a higher-level pitch range than males. A stylized graphic of the lower-level and narrower pitch range for parentheticals is displayed in Figure 1.1.

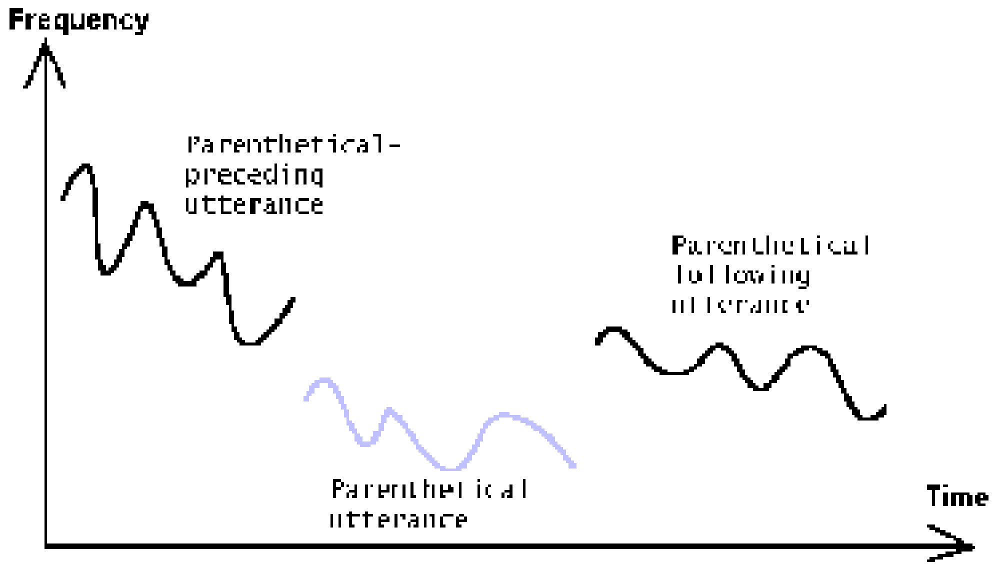


Figure 1.1. Stylized representation of the pitch contour of a sentence containing a parenthetical.

Figure 1.1 also indicates the declination of pitch height over the length of a sentence. Declination has often been regarded as being related to the decline in transglottal air pressure as the speaker uses up the air in the lungs (Cruttenden, 1997). Experiments in perception have shown that a declining series of pitch peaks is actually perceived as being of the same height (Cruttenden, 1997). That is, to express the same degree of prominence, "a peak does not have to be as high later in the sentence as it was earlier" (Pierrehumbert, 1981, p. 987).

Work on the acoustics of parentheticals has been conducted by Kutik et al. (1983). This is the only experimental acoustic study on parentheticals that I could find, and Wichmann (2000) confirms that this is the only study she could find as well. The experiment in Kutik et al. indicates that there is a clear intonational boundary at both ends of the parenthetical. In unison with Bolinger (1986), Cruttenden (1997) and Crystal (1969), Kutik et al. find that the parenthetical is characterized by a drop in pitch range and pitch level at its start and a rise back to the pitch level and range of the sentence it is embedded in at its end, as shown in Figure 1.1 above.

Kutik et al., however, did not investigate sentences with two parentheticals next to each other, as in (1a), or one embedded in another, as in (1b).

(1) a. We saw the movie (which had been banned in Boston) – Jane insisted on going – but were unimpressed. (Nunberg, 1990, p. 34)

b. We saw the movie – Jane (who knows the director) insisted on going – but were unimpressed.

The question surrounding two-parentheticals is whether one parenthetical phrase is lower and narrower in pitch range than the other or whether they are the same. If they are different, then the goal is to identify a pattern in pitch behavior for multi-parenthetical constructions. This furthers our understanding of parentheticals and closes one of the research gaps found in Kutik et al.

Furthermore, Kutik et al. only investigated comma-enclosed parentheticals. Thus, they did not look at the prosodic correlates of other parenthetical punctuation, such as brackets and dashes. According to many style guides, such as *Merriam-Webster's Guide to Punctuation and Style* (Merriam-Webster [MW], 2001) and the Canadian Press's *CP Stylebook* (Canadian Press [CP], 1984), commas, dashes and brackets are used in that order to indicate an increasing level of digression of a parenthetical element from the rest of the sentence. The choice of punctuation marks determines how the reader interprets the parenthetical – a choice, as Nunberg (1990) suggests, that is not reflected in pitch differences.

Historically, there is a link between punctuation and intonation. Much of 19th-century writing, for example, features instances of punctuation that do not follow grammatically based rules (Chafe, 1987b). “The fashion was to create punctuation units that were very much like the intonation units of speech” (Chafe, 1987a, p. 4). In the past reading aloud was in fashion. Thus, writers used punctuation marks like “stage directions for effective oral presentation” (Chafe, 1987b, p. 6). This study’s investigation that the punctuation difference for parentheticals is not a prosodic difference aims at providing evidence that, today, punctuation is a linguistic system that goes far beyond representing prosody. This provides support for Nunberg’s study on punctuation as a linguistic system in its own right (Nunberg, 1990) and subsequent research (Bayraktar et al., 1998; Briscoe, 1996; Carroll et al., 1998; Doran, 2000; Jones, 1994b; Reed & Long, 1997; Sampson, 1992; Say & Akman, 1998; White, 1995). Like spelling, which is the orthographic representation of vowels and consonants, punctuation is a system that is used cross-linguistically for the orthographic representation of prosodic, syntactic as well as semantic information.

Investigating how parentheticals are prosodically implemented also aims at providing acoustical details to enhance the naturalness of text-to-speech synthesis as well as the performance of automatic speech recognition. Examples that involve automatic speech recognition are airline or train reservations over the phone. However, to reduce the scope of this study, I have focused on text-to-speech synthesis only.

Text-to-speech technology allows one to convert an electronic text directly into speech (O’Gara, n.d.). Its uses reach from reading out what is displayed on the computer screen to a visually impaired person, as well as having the synthesizer speak a text that is entered by a person with speech difficulties, such as Steven Hawking, to helping a person to learn a different language (Childers et al., 1989; O'Gara, n.d.).

For these applications, the synthesized speech will be able to achieve more naturalness when it is known how an encountered parenthetical is to be implemented acoustically. To do that, the system has to be able to identify a phrase as parenthetical. Thus, a prediction method for parentheticals is needed (Klatt, 1987). Nonetheless, most text-to-speech systems ignore sentence-internal punctuation (Edgington et al., 1996a; Flach, 1999). Flach's study shows that out of 18 investigated text-to-speech systems only three incorporate punctuation as a parameter, although research in parsing has shown that attention to punctuation can significantly improve the performance of text parsing (Jones, 1994b; Briscoe, 1994). The problem with the punctuation of parentheticals is that commas, dashes and brackets are also used to mark other structures. To identify a parenthetical as such, one has to *understand* the sentence, which is something that machines are still lacking. "The human process of reading text out aloud . . . cannot be accomplished without some understanding of the text on the part of the reader" (Tatham & Lewis, 1992, p. 450). Text-to-speech synthesizers, however, do not understand what they say, as a satisfactory model of language understanding has yet to be developed (Childers et al., 1998; Hunt, 2000; Tatham & Lewis, 1992).

Since understanding is important for the naturalness of speech generation, this thesis investigates methods of text annotation that allow integrating the knowledge about higher-level discourse structures into the text. Many researchers work on the use of tags and markup in the document to improve the naturalness of synthetic speech (Flach, 1999; Hitzeman et al., 1999; Hunt, 2000; Mertens, 2002; Möhler & Mayer, 2002, 2001; Pierrehumbert, 1981; Sproat et al., 1998; Sproat & Raman, 1999; Taylor, 2000). The tags are either provided by the author (human or machine) or inserted by subsequent labeling. However, to make text-to-speech synthesis available to non-linguist users, tags should be based on naming the structure to be annotated (e.g. <PARENTHETICAL>), rather than using detailed phonetic tags (e.g. <PITCH RANGE="-20%">).

Many text-to-speech uses, such as reading out emails to the visually impaired, have to be performed immediately when they are encountered. Therefore, this thesis focuses on finding text-to-speech methods that avoid cumbersome and lengthy labeling procedures by a third person to prepare the text for speech synthesis, such as Grosz and Hirschberg (1992) and Syrdal et al. (2001). Rather, the system has to provide the author with the tools to easily insert intuitive (and with that user-friendly) structure tags upon text creation.

To be precise about the markup and acoustic implementation of parentheticals, it is important to know how differently punctuated parentheticals are to be treated. If there is a prosodic difference between differently punctuated parentheticals as well as multi-parentheticals, then tagging and implementation should reflect this. Therefore, this thesis determines pitch specifications for parentheticals and identifies a user-friendly method to synthesize parentheticals with the acoustic specifications found in this study.

1.2 Research questions

The research questions of this study aim at investigating the connection between punctuation and intonation for parentheticals, the prosodic characteristics of two connected parenthetical constructions and implications of these findings for text-to-speech synthesis. Specifically, the questions are:

(a) Is the difference in the punctuation of parentheticals reflected in intonation or pausing?

(b) If there is a difference, what is its nature?

(c) How do parenthetical phrases acoustically differ from non-parenthetical phrases?

(d) What is the effect on intonation or pausing when a parenthetical is next to a parenthetical?

(e) What effect on intonation or pausing has the embedding of a parenthetical within a parenthetical?

(f) How can the findings of the acoustic study be integrated in text-to-speech synthesis to improve the naturalness of synthesized speech?

The experiment in this study consists of reading aloud a set of 20 sentences by six participants, which are recorded for subsequent acoustical analysis.

1.3 Limitations of the study

Although reading aloud is not identical to naturally spoken language (Blauw, 1992; Chafe,1987a; Daly & Zue, 1992), it has been shown that speakers that read aloud tend to translate the reading of the text into the same prosodic constraints that are used in natural speech, such as using short intonation units that do not correspond to the rather longer punctuation units (Chafe, 1987b). Chafe concludes that reading aloud can be useful for an investigation of how different punctuation marks are prosodically interpreted. Cruttenden (1997) states that the wooden style that informants tend to use when reading in an experimental setting is a result of the decontextualized environment in an experiment – as opposed to the situationality in natural speech. Cruttenden concludes that intonation patterns in experimental settings represent neutral intonation patterns.

Therefore, reading aloud provides the means to factor out environmental and speaker-related influences, such as emotional attachment to a statement, and enables one to isolate what is supposed to be tested. For these reasons reading aloud has been used as an experimental means by many researchers, such as Chafe (1987a), Clark (1999), Hill and Murray (2000) and Kutik et al. (1983), and is also used in this study.

Furthermore, investigating parentheticals in spontaneous speech allows no control over type and utterance location of parentheticals. In fact, without control even large amounts of speech data may not contain a single parenthetical phrase. A further problem of using spontaneous speech instead of read speech is, as Wichmann (2000) has pointed out, that a parenthetical is a parenthetical because of the way it is acoustically treated. That is, only when the expected acoustic cues for parentheticals are present, the phrase in question can be identified as parenthetical in spontaneous speech, since no written version of the utterance exists. Thus, what is supposed to be tested is at the same time the only means of distinguishing parentheticals from other phrases. An experiment of this nature is circular and is therefore not useful to gain further insights into the intonation of parentheticals.

The data obtained by the experiment in this study is based on six participants only. With such a rather small number of participants, it is difficult to cancel out all idiosyncratic effects and results have to be seen as preliminary. However, this study is not exceptional with regard to this limitation as it is common in acoustic research to use small numbers of participants – usually between two and 10 – such as seven in Kutik et al. (1983) and seven in Grosz and Hirschberg (1992) as well.

It is beyond the scope of this study to investigate all possible parenthetical constructions, instead relative-clause parentheticals are primarily looked at. Relative clauses are chosen because as clauses they exhibit an internal structure that, unlike one-word adverbial parentheticals, allows embedding of further parentheticals. Investigating parentheticals on a broader range has to be left to future research.

There might be the danger that the experiment participants get into a routine when reading a set of similar sentences. This constitutes a problem for an experiment that investigates a difference in intonation for sentences that differ only in punctuation – although none of the exact same sentences (disregarding punctuation) are presented immediately next to each other. This is the reason why six non-parenthetical sentences were dispersed throughout the set of sentences. Their function is to avoid the manifestation of a routine.

1.4 Outline

This thesis reports on a current study on the punctuation and intonation of parenthetical phrases in English, with a focus on applying the acoustic findings in text-to-speech synthesis. The thesis contains six parts. Chapter two provides the theoretical background on which this study is based. Chapter three describes the experiment and the method of analysis used in this study. The acoustic analysis of the experiment is presented in chapter four. This includes discussing the findings and answering research questions (a) to (e). Chapter five presents the implementation of the findings in text-to-speech systems and, with that, answers research question (f). The last chapter summarizes the thesis and makes recommendations for future studies. Furthermore, it discusses the contributions of this thesis to the understanding of the relationship between punctuation and intonation as well as the acoustic characteristics of sentences containing parentheticals and their implementation in text-to-speech synthesis.

Chapter Two

PARENTHETICALS REVIEWED

The purpose of this chapter is to review the literature relevant to the study of a correlation between punctuation and intonation for parentheticals, and implications for text-to-speech synthesis. The review begins with a definition of parentheticals, continues with a discussion of punctuation and intonation with respect to parentheticals, and concludes with a discussion of the treatment of parentheticals in text-to-speech synthesis.

2.1 Definition

Dictionaries and scientific papers present a multitude of definitions of what parentheticals are. The definitions of dictionaries and style guides usually concentrate on a writing-based definition, such as definitions provided in *Gage Canadian dictionary* (Avis et al., 1983) and *Merriam-Webster's Guide to Punctuation and Style* (MW, 2001).

The Gage dictionary calls them *parenthesis* and defines *parenthesis* as “a word, phrase, or sentence, inserted within a sentence to explain or qualify something, and usually set off by brackets[1], commas or dashes. A parenthesis is not grammatically essential to the sentence it is in.” (Avis et al., 1983, p. 823)

[1] Different sources use different labels to denote the punctuation marks “(” and “)”. Style guides usually call these ‘parentheses,’ while the GAGE dictionary calls them ‘brackets.’ Linguistic sources, such as Nunberg (1990), also call them ‘brackets’ to avoid confusion with the linguistic structure ‘parenthesis.’ This thesis adopts the linguistic labeling method.

MW (2001) defines parenthetical elements as explanatory or modifying words, phrases or sentences inserted in a passage. They are set off by brackets, commas or dashes. Examples are (MW, 2001, p. 333):

(2) a) A ruling by the FCC (*Federal Communications Commission*). . . .

b) All of us, *to tell the truth*, were amazed.

c) The examiner chose – *goodness knows why* – to ignore it.

Similarly, in scientific papers researchers remark upon the formal independence of the inserted clause from the main clause (Altenberg, 1987), since parentheticals "are semantically unimportant in the context in which they occur" (Meyer, 1987, p. 66). With respect to intonation, a "parenthesis interrupts the prosodic flow of the frame utterance" (Bolinger, 1989, p. 185), primarily through a lower and narrower pitch range than the surrounding sentence contour (Bolinger, 1989; Cruttenden, 1997; Crystal, 1969; Grosz & Hirschberg 1992; Kutik et al. 1983, O'Shaughnessy, 1990; Wichmann 2000). At the end of the parenthetical, there is a reset to the pitch range and level of the frame utterance, i.e. the sentence continues as if there were no parenthetical inserted.

In conclusion, a parenthetical can be any grammatical structure from a word to a sentence and provides additional explanatory information to the frame sentence or expresses an opinion. It is set off by punctuation or, in speech, by a lower and narrower pitch range than the surrounding sentence contour. When a parenthetical is removed from a sentence, the sentence stays fully intact with respect to semantics, syntax and prosody.

The coinciding boundary marking of parentheticals through punctuation and intonation leads to the question whether different punctuation marks around parentheticals correspond to different intonation patterns. To answer that question is one of the goals of this study.

2.2 Punctuation

In writing, parentheticals are set off by either commas, dashes or brackets. Hence, different punctuation marks function as visually marking the boundary of parentheticals. Historically, punctuation emerged as an indicator of prosody in written language but evolved over the centuries into its modern form of marking a set of prosodic, syntactic and semantic boundaries (Meyer, 1987; Chafe, 1987a; Nunberg, 1990). Its popular reputation is that punctuation "is arbitrary, unmotivated, and governed by rules that make no particular sense" (Chafe, 1987b, p. 1). Hence the treatment of punctuation has been left to style guides and not seen worthy of linguistic investigation. That view has changed recently as researchers such as Meyer (1987) and Nunberg (1990) have been pointing out that we shouldn't just know how to punctuate but also why we punctuate that way. The prescriptive treatment of punctuation in style guides and printers' manuals does not provide the answer to that question. Instead, a descriptive, linguistic treatment is needed.

Meyer (1987) provides a survey of the American usage of punctuation for English. He investigates the relationship of punctuation to syntax, semantics and prosody and lays out a hierarchy for punctuation marks. This hierarchy categorizes punctuation marks into different levels according to the nature of the grammatical units that they set off. For instance, the period, question mark and exclamation mark are members of the same level, since the grammatical unit they set off is the sentence, while the comma is assigned to a different level as it only sets off grammatical units below the sentence level. However, although Meyer calls for a formalized grammar of punctuation usage, he does not provide one in his book. This has subsequently been undertaken by Nunberg (1990). Nunberg lays out rules for punctuation in different environments and in relation to each other. One of the rules involves the promotion of comma to semicolon when items containing commas are conjoined (Nunberg, 1990, p. 44):

(3) Among the speakers were Jon; Ed; Rachel, a linguist; and Shirley.

Meyer's and Nunberg's publications were the starting point for more intensive linguistic research on punctuation. Nunberg's rules have been built on, commented on and improved by subsequent research (Bayraktar et al., 1998; Doran, 2000; Say & Akman, 1998, 1996; White, 1995). Especially, researchers on natural language processing have been extending on this research (Briscoe, 1994, 1996; Carroll et al., 1998; Jones, 1994a, 1994b; Reed & Long, 1997; Say & Akman, 1997). For example, Jones (1994b) and Briscoe (1994) report that the performance of text parsers is greatly improved if a text is punctuated, as compared to an unpunctuated text. This shows that punctuation is not just included in a text because a style guide prescribes it but because it helps the reader understand a text. This makes it an important part of written language and, as many of the researchers point out, a thorough investigation into the theory of punctuation is needed.

Chafe (1987a) reports that often punctuation is not viewed as a linguistic system in its own right, because punctuation is assumed to merely closely reflect the prosodic boundaries of spoken language. However, a comma cannot be inserted between a subject and a predicate, as in (4a), although a pause at the comma might seem natural (Bolinger, 1975; Chafe, 1987a; Hill & Murray 2000). From a grammatical point of view, that makes sense since there is also no comma in the same sentence with a shorter subject, as in (4b).

However, because there is no pause after the subject, the absence of the comma in (4b) is unquestioned.

(4) a. *The man over there in the corner, is obviously drunk. (Quirk et al., 1985, p. 1619)

b. *The man, is obviously drunk.

Thus, punctuation is not automatically forced by prosody nor is its use restricted to locations that are prosodic boundaries. Rather, while punctuation captures some of the writers prosodic intent, it is also placed at grammatical boundaries that are not at the same time prosodic ones, such as the comma in (Chafe, 1987a, p. 6):

(5) . . . red, white and blue. . . .

The intonational differences involved in setting off parentheticals by bracket, dash or comma have not been investigated yet, such as in:

(6) a. We saw the movie, which had been banned in Boston, but were unimpressed.

b. We saw the movie – which had been banned in Boston – but were unimpressed.

c. We saw the movie (which had been banned in Boston) but were unimpressed.

Note that it is assumed that it is known which movie is talked about. Hence, the parenthetical relative clause in (6a) is non-restrictive. If it were restrictive, i.e. the phrase after *movie* defines what movie is talked about, there would be no comma between *movie* and *which*.

As discussed in Chapter One, style guides such as *Merriam-Webster's Guide to punctuation and style* (2001) and the *CP Stylebook* (1984) note that the choice of punctuation is not arbitrary but reflects the intention of the author with regard to how the parenthetical information relates to the rest of the sentence. For instance, dashes are used to set off parenthetical elements that are "more digressive than elements set off with commas but less digressive than elements set off by parentheses" (MW, 2001: 26).

The aim of this thesis is to provide evidence that there is no distinctive acoustic and perceptual difference for these punctuation marks when they are read aloud, which is what Nunberg (1990) predicts but has never been proven. As a consequence, this thesis lays out the acoustical nature of parentheticals as they need to be implemented by text-to-speech synthesis systems. The following section reviews the current state of knowledge about the intonation of parentheticals before its integration into text-to-speech synthesis is discussed in section 2.4.

2.3 Intonation

Parentheticals are a sentence structure phenomenon which convey explanatory or commentary information. They are often used in speech to insert an additional thought and are clearly perceived as such by the listener. Hence, they exhibit distinctive prosodic characteristics.

Intonation is “the sound pattern of speech produced by differences in stress and pitch” (Avis et al., 1983, p. 611). Bolinger (1989), Cruttenden (1997) and Crystal (1969) provide comprehensive discussions of intonation. They all report that parentheticals in

English exhibit a lower-leveled and narrower pitch range than the surrounding sentence constituents, but none of these studies include a discussion of an acoustic study to back this up. However, there seems to be a general consensus in the literature that these are the two main features of parentheticals (Grosz & Hirschberg, 1992; Kutik et al., 1983; O'Shaughnessy, 1990; Wichmann, 2000). These prosodic characteristics for parentheticals are not restricted to English but have been found in other languages as well, such as Danish (Hansen, 2002) and for males in French (Fagyal, 2002).

Grosz and Hirschberg (1992) investigated intonational characteristics of discourse structure through an experiment that involved labeling discourse structure in a text. In the experiment, one group was labeling a punctuated text. A second group labeled the same text with all except sentence-final punctuation removed, but they were also supplied with an acoustic recording of the text. The study showed that labeling performance improves when an acoustic recording of the text is provided along with it. Hence, this supports their hypothesis that discourse structure is marked intonationally. Parentheticals were one of the structures they used to measure labeling performance. Through identifying the acoustic cues that the experiment participants used to label a sentence part as parenthetical, Grosz and Hirschberg found that parentheticals are marked intonationally with a compressed pitch range and a decrease in intensity.

Similarly, Wichmann (2000) points out that parentheticals can primarily be identified by the way a word, phrase or clause is prosodically implemented. That is, the means to distinguish a parenthetical from other structures is not inherent in the morphology and syntax of the word, phrase or clause itself. "However, some kinds of structures are more capable of being treated parenthetically than others. These include co-ordinated noun phrases, tag exclamations, adverbials, relative clauses, elliptical clauses, reporting verb groups, and amplificatory phrases" (Wichmann, 2000, p. 95). Hence, it is not part-of-speech or syntax that defines a parenthetical but semantics and corresponding prosodic implementation. Wichmann suggests two possible prosodic characteristics that might be useful for further explorations of parentheticals. The first one is the prosodic coherence of the utterance if the parenthetical element were removed, and the second characteristic is the change in pitch range (Wichmann, 2000: 99). Correspondingly, these are the factors that Kutik et al. (1983) investigated.

In their study on the acoustics of parentheticals, Kutik et al. presented experiment participants with a set of seven sentences that featured a parenthetical construction of increasing length over the course of the text. The examples in (7) show the difference between shortest and longest construction.

(7) Examples of Kutik et al.'s parentheticals (Kutik et al., 1983, p. 1732)

a. shortest: The clock in the church, *it occurred to Clark*, chimed just as he began to talk.

b: longest: The clock in the church, *it never in a million years would have occurred to the absent-minded Clark*, chimed just as he began to talk.

These sentences were read aloud by subjects and recorded. In the subsequent acoustic analysis the researchers were looking at the change of pitch range over the utterances. The F0 contour of a sentence consists of high and low pitch values that are enveloped by topline and baseline. The topline is a derivative of the upper end of the pitch range in an utterance, i.e. it delineates the series of high peaks. The baseline is a derivative of the lower end of the pitch range in an utterance. The effect of parentheticals on overall topline declination and the nature of the topline of the parenthetical were the focus of Kutik et al.'s study. The study shows that the falling topline is interrupted during the insertion of a parenthetical into the sentence and that at the end of the parenthetical interruption, the topline resumes its initial declination pattern. This shows that despite the interruption, the prosodic coherence of the main sentence is not compromised by the presence or absence of a parenthetical, as it was also suggested by Wichmann (2000). Furthermore, Kutik et al.'s research indicates that the parenthetical has its own lower-set topline. The fall to a lower topline at the beginning and then reset to original topline at the end shows that there is a clear acoustic borderline at both ends of the parenthetical – marked by punctuation in the written language. A lower topline results in a low, narrow pitch range for parentheticals, corresponding to what has been stated by Bolinger (1986), Cruttenden (1997), Crystal (1969), Grosz and Hirschberg (1992) and Wichmann (2000). However, Kutik et al. only investigate comma-enclosed parentheticals and do not look at other parenthetical punctuation, such as brackets (parenthesis) and dashes. Furthermore, they don't investigate a parenthetical embedded within a parenthetical and parentheticals side by side, which are further cases that this thesis investigates.

Wichmann (2000) criticizes Kutik et al.'s study with regard to the unnatural length of some of their parenthetical structures and their use of comment clauses only. Consequently, this thesis uses much shorter parenthetical constructions. Furthermore, this thesis investigates mainly relative clauses as an alternative. Relative clauses are chosen because as clauses they exhibit an internal structure that, unlike one-word adverbial parentheticals, allows embedding of further parentheticals. Additionally, it is beyond the scope of this thesis to investigate all possible parenthetical structures.

None of the sources reports on a pausing characteristic at the boundaries of parenthetical phrases. On pausing as a boundary marker, Cruttenden (1997) states that "pause does not always mark intonation boundaries nor are intonation boundaries always marked by pause" (Cruttenden, 1997, p. 32). Nevertheless, whether this general statement on pausing as boundary marker also applies to parentheticals is investigated in this thesis as part of the acoustic analysis in Chapter Four.

The findings of the investigation in this thesis have implications on the relationship between punctuation and intonation, but also on the integration and acoustic implementation of parentheticals in text-to-speech synthesis – which is reviewed in the following section.

2.4 Text-to-speech synthesis

Modeling intonation plays an important part in achieving naturalness for speech generated by text-to-speech synthesis systems. There is a wide range of systems whose methods range from strategies that use dictionaries of speech sounds in conjunction with a set of prosodic rules for producing the F0 contour to providing markup in the text to guide contour generation.

To synthesize parentheticals, the text-to-speech system has to be able to identify a parenthetical as such in a text. Furthermore, the system has to be able to encode acoustic specifications for a detected parenthetical into the prosodic instructions to the synthesizer. The synthesizer is the part of the system that converts the encoded information into acoustic signals, a process that is not a focus of this thesis. Thus, the detection stage and the encoding stage are the focus for the discussion of synthesizing parentheticals with different speech synthesis systems.

There are different ways of sentence structure detection that range from automatic detection through a parser to annotation by hand. Pierrehumbert (1981), Tatham et al. (1999, 1998) and Taylor (2000) use recorded readings of the text as the input that is analyzed with respect to prosodic parameters, such as extent and duration of a rising or falling pitch (Taylor, 2000). These parameters are then modeled into an intonation contour that most closely resembles the intonation contour of the input. Following this, the modeled contour is resynthesized. These speech synthesis systems are developed for carrying out research on modeling intonation contours. Comparing the original recording and the synthesized output is useful for investigating where our knowledge about pitch contours is still insufficient to make the synthesized signal identical to the recorded one. Since these systems cannot use text as the input, they are not true *text*-to-speech systems as this study is looking for.

Dutoit (1997), Edgington et al. (1996a, 1996b) and Westall et al. (1998) provide a description of the procedures involved in most commercial text-to-speech systems. In these systems, a preprocessor first identifies the individual words and the end of a sentence, processes punctuation marks, such as periods involved in abbreviations, recognizes acronyms and converts numbers into words. It also removes any sentence-internal punctuation. Thus, the presence of sentence-internal punctuation in the input text is neglected in the following text analysis. The preprocessing task is called text normalization. Preprocessing is followed by the syntactic parse. With the help of a dictionary, the system identifies the part-of-speech of each word and uses the sequence of words to derive a structural analysis of the sentence with respect to syntax. Algorithms are used to predict the most likely prosodic structure based on the phrase structures identified by the syntactic analysis. The pronunciation of the segments of each word is achieved through a set of letter-to-sound rules or, for more frequent words, the entire word can be stored in a dictionary. Some systems also use syllable dictionaries to capture inter-syllabic transitions from one speech sound to the next in less frequent words.

Parentheticals exhibit no unique syntactic structure. Hence, they cannot be identified by the syntactic parser of these text-to-speech systems. As a consequence, these systems do not synthesize parentheticals differently from non-parenthetical phrases. This makes commercial text-to-speech systems unsuitable for incorporating the findings of this study.

The dependency parser (Lindström et al., 1996) improves the syntactic investigation of a text by identifying head-modifier relations in addition to part of speech. This provides a better method to determine the position and prominence of pitch accents than simple part-of-speech tagging as performed by the commercial text-to-speech systems discussed above. Pitch accents are the pitch peaks of an intonation contour. They indicate prominent syllables and words (Cruttenden, 1997). Similarly, Hirschberg (1995) introduces decision-tree algorithms in addition to part-of-speech tags to predict pitch accents. The decision tree takes focus and function/content word distinction into account, as well as whether a thought is new or has already been introduced previously. However, Hirschberg’s and Lindström’s methods both do not improve the detection of parentheticals, since parentheticals neither feature a particular part-of-speech combination nor a particular pitch accent trend.

The same criticism applies to Wang and Hirschberg (1995), who present a method for predicting intonational boundaries from text using decision trees that use part-of-speech information, syntactic constituency as well as predicted pitch accents. These trees are primarily based on likelihood decisions, such as that phrase boundaries can rarely be found after function words. Taking only syntactic phrase structure and pitch accent structure into account, these decision trees are not able to register whether a boundary-enclosed phrase is a parenthetical or not. What is needed for the detection of parentheticals by text parsers are parsing methods that go beyond syntactical analysis.

The SPRUCE (SPeech Response from UnConstrained English) text-to-speech system (Tatham & Lewis, 1996, 1992) uses a parser that performs a syntactic parse as well as a semantic parse. The semantic parse identifies logical relationships between words and between sentences. Based on the syntactic and semantic parses, a system of rules determines the most plausible intonation contour for each sentence. A syllable dictionary provides the phonetic specifications for each syllable as well as some of the words. For the synthesis output, these phonetic specifications are overlaid with the intonation contour. However, as the authors admit, embedded phrases like parentheticals require information that is not available from the input text through the parser (Tatham & Lewis, 1992). The semantic parse of this system is not sophisticated enough to identify higher-level discourse structure to detect parentheticals.

Efforts are undertaken to create parsers that can identify discourse structure such as the rhetorical parser by Marcu (1997, 1998, 1999). There are two algorithms at work in the rhetorical parser. The first parses the text and identifies cue phrases, such as the conjunction *although*, as potential discourse markers. Furthermore, it identifies punctuation marks as discourse markers and, with that, makes use of the fact "that discourse structural information can be inferred from the orthographic cues in [a] text, such as . . . punctuation" (Hirschberg & Nakatani, 1996, p. 286). The second algorithm uses the presence of discourse markers to identify the discourse structure that they entail or introduce. For instance, the words between two dashes are identified as a parenthetical phrase if no sentence-final punctuation is encountered before the second dash. Although this procedure can identify most parentheticals, it disregards the possibility of a bracket or comma-delimited parenthetical phrase embedded within the dash-delimited parenthetical phrase.

The rhetorical parser is not presented as being a component of a text-to-speech system. Although integrating rhetorical parsing into commercial text-to-speech systems could provide the means to improve the naturalness of structures like parentheticals in synthesized speech, no such system exists to date. This and the fact that the rhetorical parser neglects the occurrences of embedded parentheticals make it necessary to look for alternative parenthetical detection methods.

To improve the naturalness of synthesized speech, it is important to get the system to a level that it “understands” what is says. As researchers (Childers et al., 1998; O’Shaughnessy, 1990; Tatham & Lewis, 1992) have pointed out, the lack of understanding of what is said is a key factor in the problem of achieving naturalness. Identifying the discourse structure is an important step to understanding. Since the discourse structure cannot be automatically identified in any existing text-to-speech system to date, it has to be supplied through annotation. By using annotation, the knowledge of discourse structure is delivered with the text to the text-to-speech system. This way a parenthetical phrase can be identified by the system by being appropriately labeled.

One of the goals of this thesis is to find a system that synthesizes an unknown text on the spot. This is important for applications such as a computer reading an email to a blind person. For such a task, it is not feasible to have someone go through every text that is to be synthesized and label it appropriately before it can be supplied to a synthesis system. This is the method though that many research-oriented systems use, because for them, on-the-spot synthesis is not important. An example for this is the text-to-speech system discussed by Pierrehumbert (1981), which involves phonetic annotation. Pierrehumbert points out that in any text-to-speech system, the computer must assign an F0 contour without understanding what the text is saying. Therefore, it is necessary "to design an input which encodes appropriately the knowledge about a sentence" (Pierrehumbert, 1981, p. 986), such as by using annotation or markup. The synthesis program then translates this knowledge with a set of rules into an intonation contour. In Pierrehumbert's model the input to the text-to-speech program "is a string of phonemes, annotated with durations, phrase boundaries, and target levels" (Pierrehumbert, 1981, p. 989). Rather than annotating discourse structure, this model integrates the prosodic consequences of the discourse structure directly into the text through tags that feature phonetic specifications such as pause duration.

This model results in improved naturalness of the synthesized speech but has the disadvantage that it requires a third-person annotator. Furthermore, it is not very efficient for general usage, because the annotator must possess a deep understanding of phonetics to make the required specifications in the annotation. Additionally, the model is not useful for parentheticals, because there is no option for pitch range change within a sentence as it is needed for parentheticals.

There are other methods of annotating, namely annotating higher-level structure instead of specific pitch values. One such method is discussed by Hitzeman et al. (1999). They argue for annotating using linguistic tags, such as "predicative" – meaning that the entity under discussion is predicative. The advantage of linguistic tags over prosodic tags, such as "pitch," is that higher-level linguistic tags allow a synthesizer to prosodically interpret the tags according to its own settings. Hence, linguistic tags are synthesizer independent and require no specific phonetic knowledge. Similarly, Möhler and Mayer (2001, 2002) present the idea of concept-to-speech where discourse structure information, such as *elaboration*, which indicates that a part of the text provides additional detail, is given in the markup. An algorithm converts the discourse structure first to phonological registers and then to pitch range values. Both Hitzeman et al. and Möhler and Mayer propose the use of XML (eXtensible Markup Language) tags. XML is a commonly used Internet standard for marking up structure and meaning in documents (Hunt, 2000). Consequently, XML was the choice to base specific speech markup languages on, such as JSML (Java Speech Markup Language) and SABLE[2].

As with markup standards in general, a browser interprets the document with its markup to determine the proper display of the document. In the case of creating speech output, a voice browser is used. Style sheets that are linked to a document to provide speech style specifics can be retrieved from the network by the browser (Heavener, 2002). For a speech synthesis system that uses speech markup such as JSML or SABLE as input, the voice browser converts the original document, such as an HTML (HyperText Markup Language) document, into a JSML or SABLE document. The resulting document is then used by the synthesizer to generate speech by interpreting the speech markup with the help of an XML processor.

What the voice browser does is interpret all markup into speech-production markup, such as SABLE, using Aural Cascaded Style Sheets (ACSS). ACSS are a set of specifications that define how text enclosed by a particular tag should be interpreted (Sproat & Raman, 1999). For instance, using ACSS the voice browser converts HTML

[2] SABLE is not an acronym but the name is tentative and may be changed at some time.

<H1> Introduction </H1> into SABLE *<PITCH BASE="lowest" RANGE="80%"> <EMPH LEVEL="0.5"> Introduction </EMPH></PITCH>* (EMPH = emphasis). In other words, the voice browser rewrites the document, which then can be interpreted by the text-to-speech system. Figure 2.1 displays the steps for the text-to-speech generation process involving SABLE speech markup.

HTML/XML/SABLE Document + ACSS	-->	Voice/Audio Browser [Converts HTML/XML into SABLE]	-->	Text-to-speech System [Interprets SABLE text]

Figure 2.1. A model for using ACSS and SABLE (Sproat & Raman, 1999, p. 5).

SABLE markup can already be contained in the original document. In the case of structural SABLE tags, these are converted to phonetic SABLE tags. For instance, a structural tag <DIV TYPE="x-tl">, which marks the boundary at the end of a line in a table, can be converted to phonetic markup that marks the presence of the boundary in auditory terms (Sproat & Raman, 1999). Hence, marking the boundary in an abstract way leads, nonetheless, to a specific speech output. In the case of phonetic SABLE markup, such as <PITCH RANGE="-20%">, in the original document, the tag is left as it is.

JSML (Hunt, 2000) provides the means for structural markup, such as marking a sequence of words as a "sentence," but also markup to control the production of synthesized speech, such as providing specific pitch control. With this, JSML provides a system of tags that give professionals the tools to fine-tune intonation contours, as well as tools for a non-professional audience. Thus, it combines Pierrehumbert's annotation of phonetic detail with Hitzeman et al. and Möhler and Mayer's proposal for higher-level annotation. However, there is no higher-level markup available for parentheticals. The only way to markup parentheticals with JSML is to use pitch-specific tags to capture the parenthetical pitch changes.

The advantage of annotation is that it is the author (person or machine) that supplies markup with the text during the text creation process. Thus, the intended discourse structure – which, of course, is known to the author – does not have to be identified later by a parser (as discussed) or different person (Grosz & Hirschberg, 1992; Syrdal & Hirschberg, 2001). This is particularly useful for parentheticals, since authors know which phrases are intended to be parenthetical.

SABLE (Sproat et al., 1998; Sproat & Raman, 1999; "SABLE," n.d.) has been developed as an improved speech markup standard that is based on JSML and STML (Spoken Text Markup Language). Consequently, SABLE is very similar to JSML in the composition and use of tags as well as goals, such as providing a markup standard that is synthesizer independent because the XML processor renders the document synthesizer ready. Like JSML, SABLE provides no specific structure tag for parentheticals but enables annotation of parentheticals through prosodic tags, such as <PITCH BASE="-20%" RANGE="small"> *parenthetical* </PITCH>. BASE refers to the baseline pitch which represents the normal minimum pitch of a sentence and is lowered for parentheticals.

However, if prosodic tags are the only option, then both SABLE and JSML still require prosodic knowledge to markup a parenthetical. A structural tag for parentheticals, such as <parenthesis high/low>, has been proposed by Mertens (2002). Nonetheless, creating a separate tag is not necessary within the SABLE markup scheme. Since SABLE has been designed to be extendible, the value *parenthetical* can be added as a possible value of the attribute TYPE of the DIV (division) element in the SABLE markup scheme. The parenthetical can then generally be tagged as <DIV TYPE="parenthetical"> *parenthetical* </DIV> and intonation specifics come into play when the synthesizer interprets the tag.

Applications of SABLE can be found in the speech production for an automatic teaching agent for children, as discussed by Wouters et al. (1999), as well as the German text-to-speech system MARY (Modular Architecture for Research on speech sYnthesis) (Schröder & Trouvain, 2001).

This thesis aims at investigating the prosodic correlation to differently punctuated parentheticals. The results of the acoustic study are used to propose specifics for an incorporation of parentheticals as part of the structural SABLE markup, as well as accurate implementation by the synthesizer.

Chapter Three

METHODOLOGY

This chapter discusses the experiment setup and method of analysis. The results of the analysis are presented in chapter four.

The purpose of this thesis is to further our understanding of the pronunciation of sentences that contain one or more parentheticals and to discuss how these findings can be integrated into text-to-speech systems. In particular, this study seeks to provide evidence that the use of different punctuation marks does not change the way parentheticals are spoken.

To achieve these goals, an experiment has been conducted in which three female and three male speakers of Canadian English read 20 sentences each. The participants are university students and between 20 and 35 years old. They are volunteers, chosen for their willingness to read aloud 20 sentences which are recorded for an acoustic analysis. Each participant was recorded in a single individual session.

Before the experiment, each participant was given a short text about parentheticals and how their meaning is interpreted differently when different punctuation marks are used. This was done to bring each student to the same level of knowledge about parentheticals. The text is given in Appendix A.

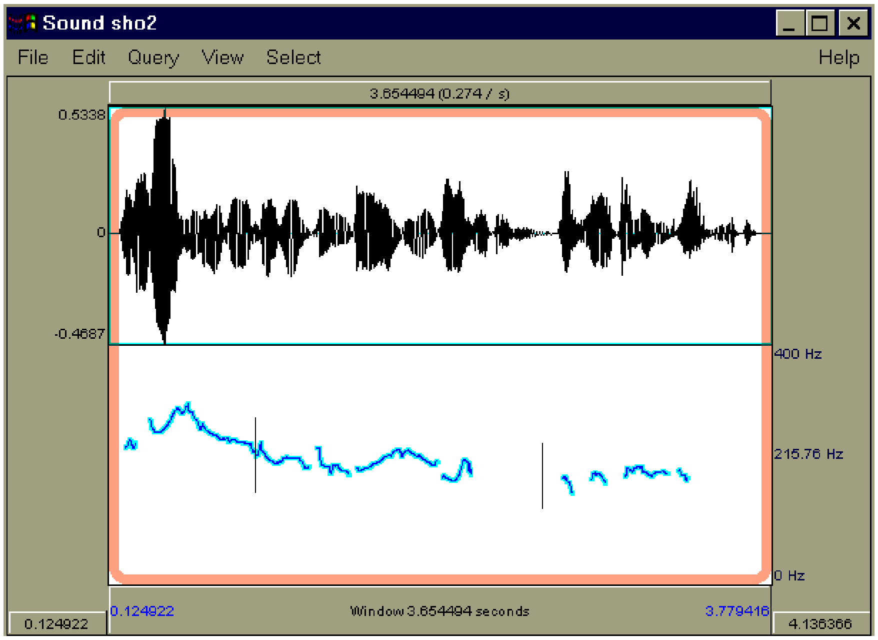
The experiment consisted of reading aloud a set of 20 sentences. The list of sentences is provided in Appendix B. Each sentence contains the frame sentence "We went to the movie but were unimpressed." Following Nunberg (1990), parenthetical and non-parenthetical phrases are inserted into the frame sentence between the first clause (ending with *movie*) and the second clause (beginning with *but*). Thirteen of the 20 sentences contain differently punctuated parenthetical phrases that are inserted into the frame sentence. Additionally to the 13 parenthetical sentences, there is an instance of the frame sentence on its own (to get a control recording of the frame sentence itself) and six sentences that contain non-parenthetical constructions. The non-parenthetical sentences will be used for a direct comparison between parenthetical and non-parenthetical phrases. Furthermore, the non-parenthetical sentences are dispersed throughout the set of sentences to avoid the manifestation of a reading-parentheticals-routine by the participants. Each sentence was read twice to increase the chance of obtaining at least one instance of each sentence that is free of reading errors. For the analysis, only one speech sample of a particular sentence was chosen for every speaker.

The participants read the sentences in an acoustically treated room at the University of Victoria's Phonetics Laboratory. The microphone in the room is connected to a personal computer. The speech samples are digitized onto the computer at 22,050 samples per second, 16-bit, using Cool Edit Pro LE, manufactured by Syntrillium Software Corporation.

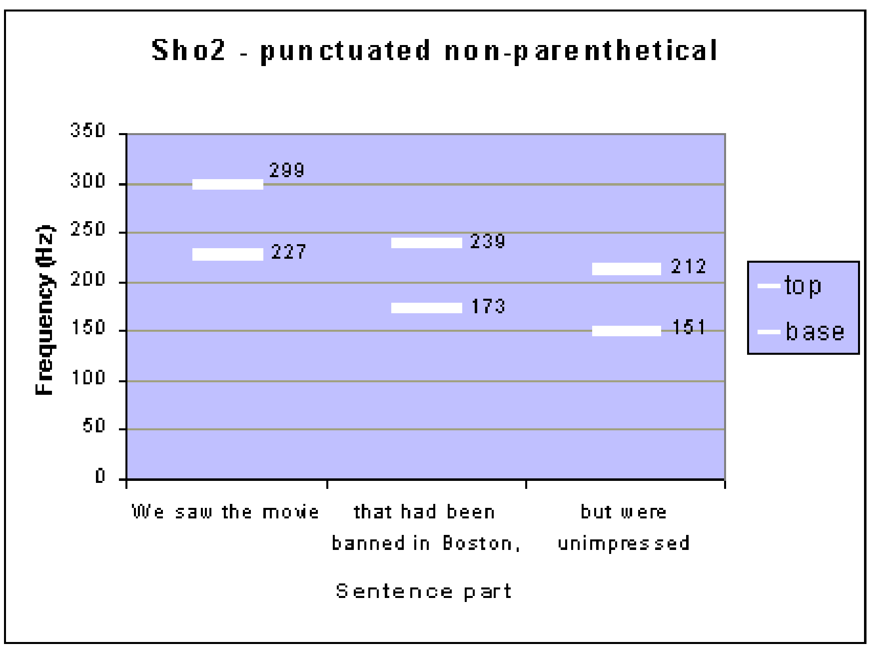
The recordings are analyzed using Praat, version 3.9.9 ((c) 1999-2000 by Paul Boersma and David Weenink). This software allows one to display the intonation contour for a speech signal and it provides options for queries and measurements on the signal, such as maximum and minimum pitch values in a chosen interval. The software cannot identify phrase boundaries itself. Rather, these have to be determined by the person who is carrying out the analysis. They are determined by listening to the sentence and noting at which part of the contour the phrase boundaries occurred. On the screen the phrase can then be encompassed by cursors and the maximum and minimum pitch values of the encompassed interval can be measured using one of the functions in Praat. Hence, this software provides the means to investigate topline (the upper boundary of a phrase, i.e. the maximum pitch value), baseline (the lower boundary, which is based on the minimum pitch value) and pitch range (the difference between maximum and minimum pitch) in a phrase. As discussed above, these are the three acoustic features to look for when comparing parentheticals to non-parentheticals.

How the pitch contour of a sentence translates into topline and baseline is illustrated in Figure 3.1. Part (a) of the figure shows the pitch contour of sentence 2 spoken by the female subject SHO as displayed by Praat. Part (b) shows how this contour translates into phrasal toplines and baselines after the measurements have been converted to a graph using Microsoft Excel (version 97). In this study all pitch values have been rounded to the nearest full Hertz value.

Note that in Figure 3.1 Part (a), as well as all following Praat displays, the temporal alignment of the graphics display is condensed in the lower display. That amplitude peaks (top display) and pitch peaks (bottom display) do not line up are not an inaccuracy of the measurement algorithm but can be explained the nature of Praat's redisplay algorithm. As discussed below, pitch artifacts have been dealt with.



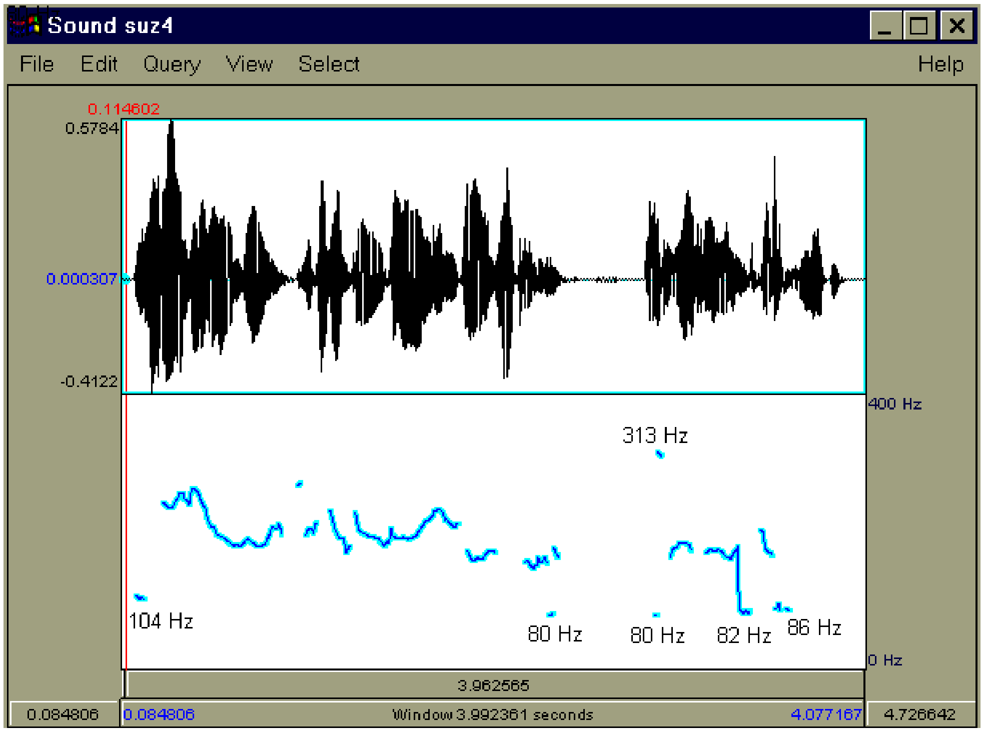
(a) “We saw the movie | that had been banned in Boston, | but were unimpressed.”



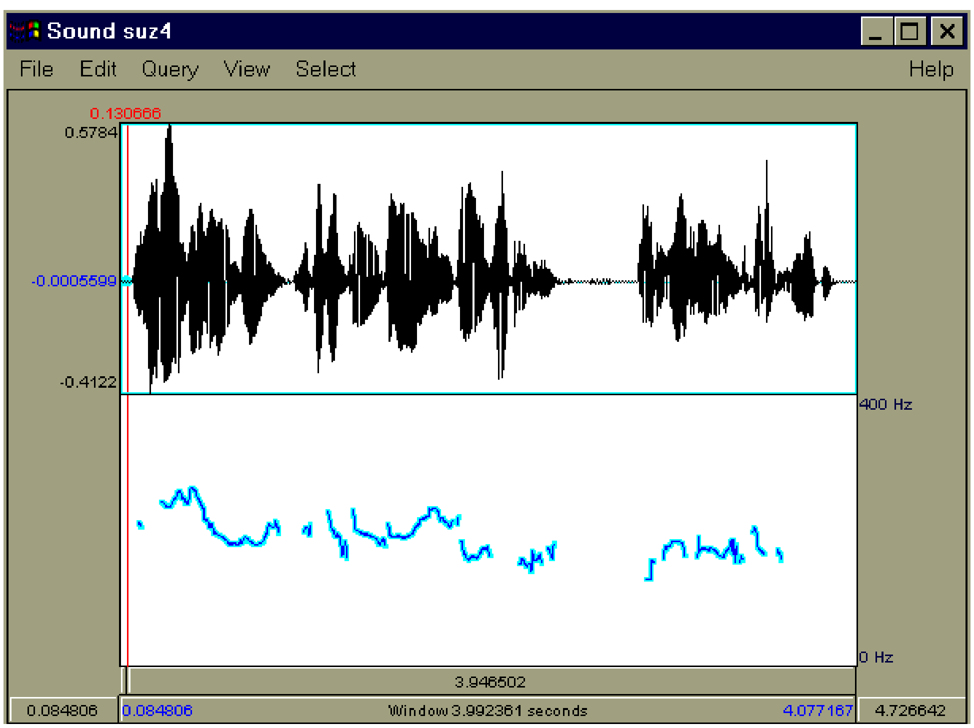
(b) Topline/baseline representation of (a).

Figure 3.1. A pitch contour and its corresponding topline/baseline representation.

Praat default settings were used in all measurements, except for two categories of the pitch settings. Firstly, in the pitch settings window the time steps were increased from 100 to 1000. What this means is that instead of partitioning the speech signal displayed on the screen into 100 units only and extrapolating the pitch contour from that, the signal is partitioned into 1000 units for extrapolation. The default setting of 100 did not provide sufficient accuracy. Secondly, Praat's default setting for pitch range analyses is 70 Hz to 500 Hz. However, due to frequent outliers, a pitch range of 75 Hz to 225 Hz was chosen for males and 140 Hz to 300 Hz for females. Figure 3.2 shows that the default range of 75-500 Hz creates outliers in the female data, while a 140-300 Hz analysis range is able to cut them off for females. Figure 3.3 shows that 75-500 Hz also creates outliers in the male data, while a 75-225 Hz analysis range is able to cut them off for males.

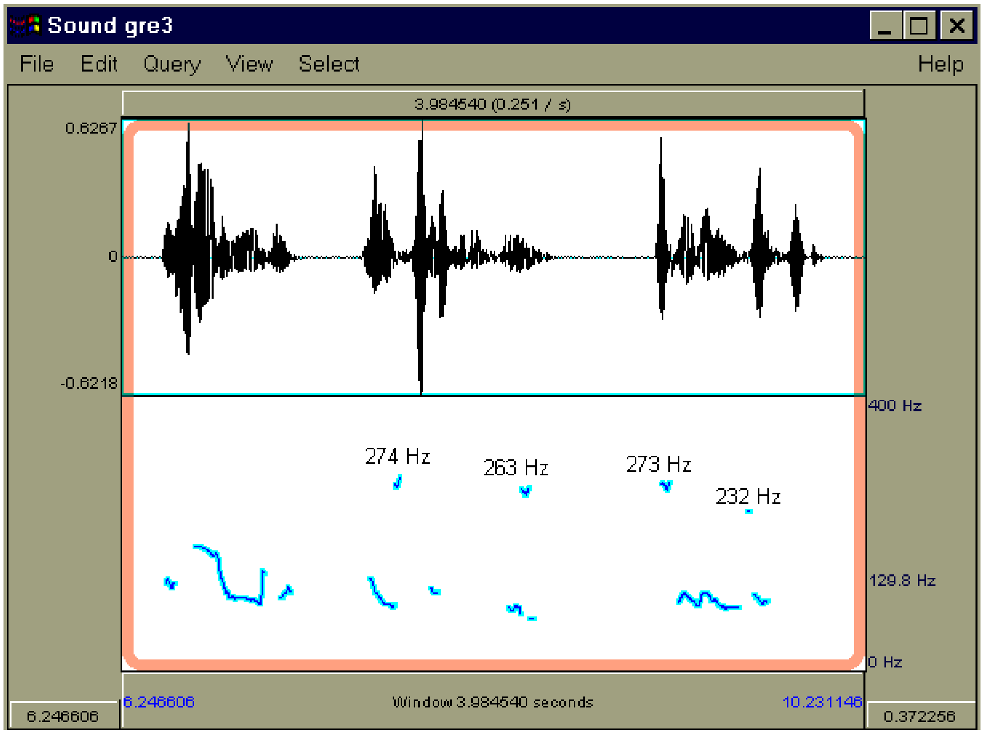


(a) 70-500 Hz analysis range pitch contour of sentence 4 for female subject SUZ

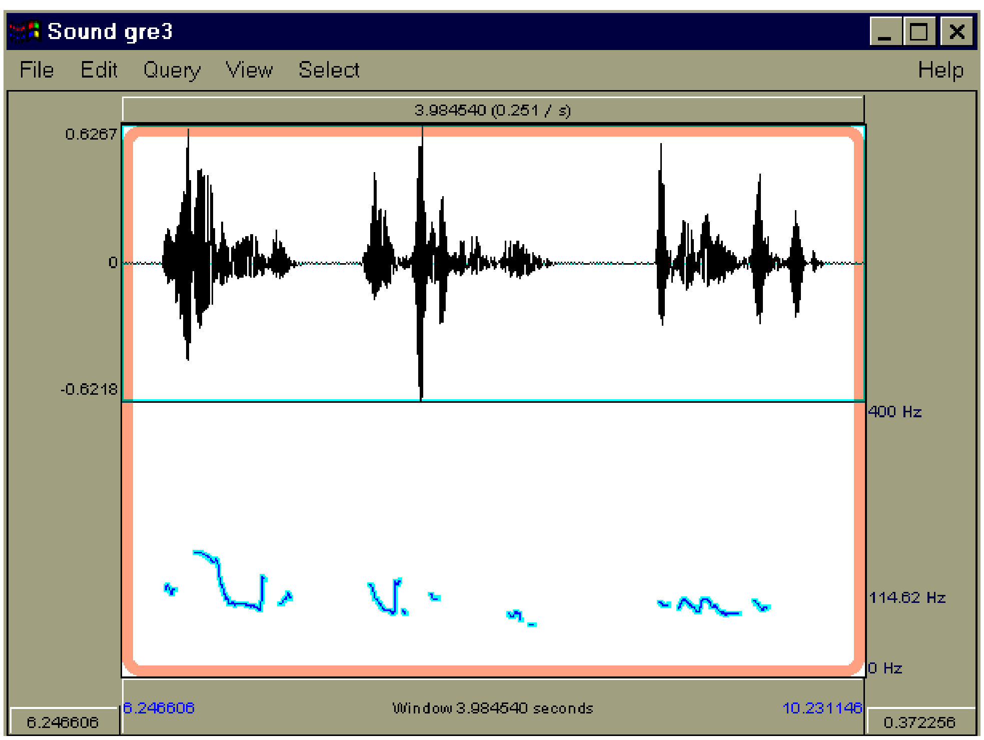


(b) 140-300 Hz analysis range pitch contour of sentence 4 for female subject SUZ

Figure 3.2. Comparison of 70-500 Hz and 140-300 Hz analysis pitch ranges for females. 70-500 Hz is producing outliers, whereas 140-300 Hz does not.



(a) 70-500 Hz analysis range pitch contour of sentence 3 for male subject GRE



(b) 75-225 Hz analysis range pitch contour of sentence 3 for male subject GRE

Figure 3.3. Comparison of 70-500 Hz and 75-225 Hz analysis pitch ranges for males.

70-500 Hz is producing outliers, whereas 75-225 Hz does not.

For a few sentences the chosen analysis ranges still produced outliers. However, choosing a smaller range for all sentences creates distortion effects in many of the other sentences. Thus, different analysis pitch ranges have been used for the following male sentences: AAR's sentence 3 (used: 75 Hz to 190 Hz) and CRA's sentence 2, 17 and 20 (used: 75 Hz to 200 Hz). Furthermore, the data of the parentheticals in AAR's sentences 10 and 15 had to be taken out.

The analysis of topline, baseline and pitch range is discussed in section 1 of chapter four. Throughout the analysis the first part of the frame sentence, i.e. "we saw the movie," has been labeled 'fr(a)' and the second part of the frame sentence, i.e. "but were unimpressed," has been labeled 'fr(b).' The part between fr(a) and fr(b) has either been labeled 'non-par' for non-parenthetical phrases or 'par' when there is only one parenthetical embedded in the frame sentence. In the case of two sequential parentheticals, the first one has been labeled 'par1' and the second 'par2.' In the case of two nested parentheticals, the first part of the outer parenthetical has been labeled 'par1(a),' the second part of the outer parentheticals has been labeled 'par1(b),' and the nested parenthetical has been labeled 'par2.' Hence, for a nested two-parenthetical, such as "We saw the movie – Jane (who knows the director) insisted on going – but were unimpressed," the sequence is: fr(a) / par1(a) / par2 / par1(b) / fr(b).

Although researchers did not report that parentheticals are distinct from other phrases through pausing, this study took the opportunity to also have a look at a possibly distinct pausing pattern for parentheticals. Pause length has been determined by measuring the interval of quietness between phrases using Praat. All pause values have been rounded to the nearest 10th of a second as the difficulty in detecting the exact location of pause boundaries did not allow greater precision. The analysis of the pauses is discussed in section 4.2.

There are further measures that are not reported by previous research but can be investigated by this study, such as amplitude as well as mean pitch and pitch variance. However, rather than exploring further ways of distinguishing parentheticals, the focus of this study is to look at multi-parentheticals as well as the significance of parenthetical punctuation and compare results to previous findings.

Chapter Four

ANALYSIS

This chapter presents the outcome of the experiment. Section 4.1 discusses the analysis results with respect to pitch and section 4.2 presents analysis results with respect to pausing.

4.1 Pitch

Leading to a comparison of parentheticals and non-parentheticals in 4.1.4, this section first discusses non-parenthetical pitch results in 4.1.1, followed by a discussion of single-parentheticals in 4.1.2 and of sentences with two adjacent or nested parentheticals in 4.1.3.

4.1.1 Non-parentheticals

As part of the experiment, six non-parenthetical sentences have been recorded. Of these, sentences 7, 11 and 14 are unpunctuated, whereas sentences 2, 5 and 17 are punctuated with a comma between the medial phrase (non-par) and fr(b).

Figure 4.1 displays the maximum and minimum pitch levels for each unpunctuated non-parenthetical sentence (sentences 7, 11 and 14) for both females and males. An example is sentence 7:

(8) Sentence 7

“We saw the movie on the opening night but were unimpressed.”

Figure 4.1 shows that females generally have a much higher pitch level than males. Furthermore, for both females and males the pitch level declines over the course of the sentence as both topline and baseline decline. While the topline drops relatively evenly over the course of the sentence for females and in a ratio of 2:1 for males, the baseline drops steeply between fr(a) and the medial phrase and then stays almost level between the medial phrase and fr(b). The values displayed throughout the analysis are averaged numbers of all three female speakers and averaged numbers of all three male speakers. In the legends used throughout this analysis, the topline has been named “hi” and the topline “lo.” This has been done to provide more space for the actual graphic.

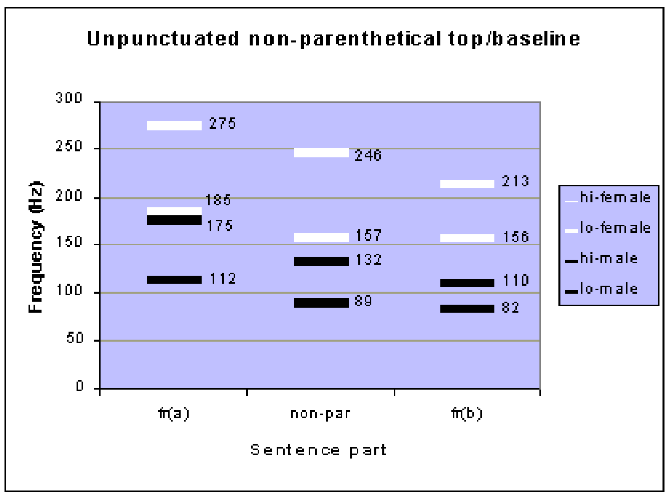


Figure 4.1. Toplines and baselines for female and male unpunctuated non-parenthetical sentences.

Figure 4.2 shows the pitch ranges that result from the differences between the toplines and baselines displayed in Figure 4.1. It can be seen that the male pitch range narrows down over the course of the sentence by first dropping about 20 Hz between fr(a) and the medial phrase and then dropping about 14 Hz from the medial phrase to fr(b). The female pitch range stays about level from fr(a) to the medial phrase but then drops 31 Hz between the medial phrase and fr(b). Hence, female and male pitch range drops about the same over the course of the sentences (female: 32 Hz, male: 34 Hz). However, the distribution of the drops are very different between female and male speakers. Female speakers drop topline and baseline equally for the medial phrase, which results in an unchanged pitch range, whereas males exhibit the steeper drop in pitch range at the first boundary, rather than at the second, where females drop steeply.

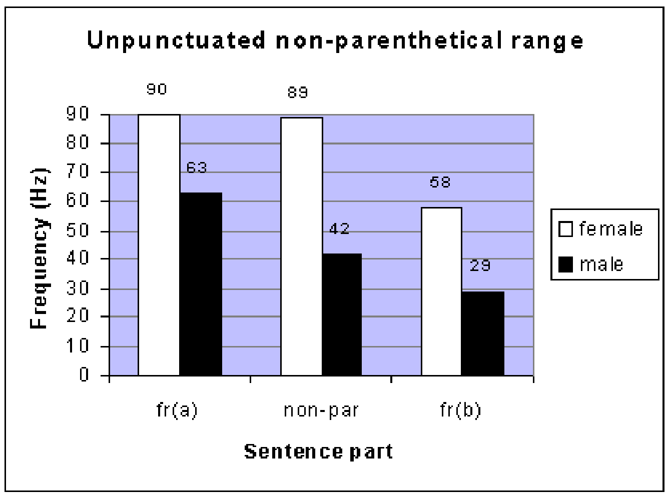


Figure 4.2. Pitch ranges in female and male unpunctuated non-parenthetical sentences.

Figure 4.3 displays the maximum and minimum pitch levels for each punctuated non-parenthetical sentence (sentences 2, 5 and 17) for both females and males. An example is sentence 17:

(9) Sentence 17

“We saw the movie because everyone went, but were unimpressed.”

The pitch numbers for males and females in Figure 4.3 are not significantly different from those in Figure 4.1. Like in the unpunctuated non-parenthetical sentence, the pitch level for both females and males declines over the course of the sentence as both topline and baseline decline. Analogously, the topline drops relatively evenly over the course of the sentence for females and in a ratio about 2:1 for males, the baseline drops steeply between fr(a) and the medial phrase and then stays almost level for fr(b).

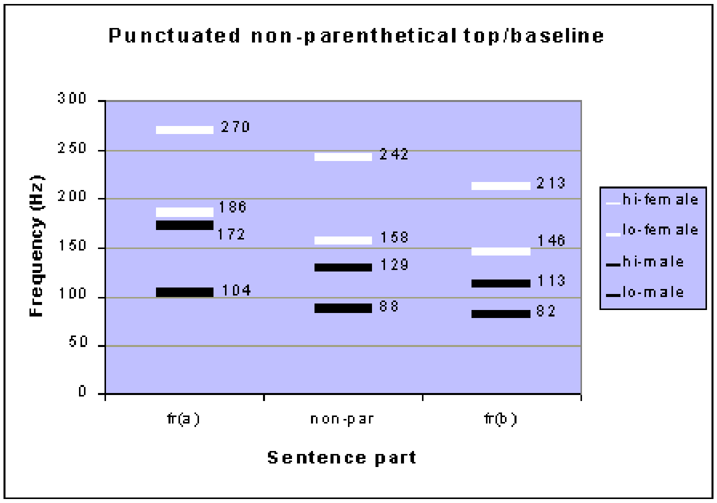


Figure 4.3. Toplines and baselines for female and male punctuated non-parenthetical sentences.

Comparing the pitch values directly to the unpunctuated ones, as in Table 4.1, shows that being punctuated or not makes no significant difference in the non-parenthetical toplines and baselines.

Table 4.1.

Non-parenthetical toplines and baselines are unaffected by the presence or absence of punctuation. All values are in Hertz.

	fr(a)	non-par	fr(b)
female topline			
unpunctuated non-par	275	246	213
punctuated non-par	270	242	213
female baseline			
unpunctuated non-par	185	157	156
punctuated non-par	186	158	146
male topline			
unpunctuated non-par	175	132	110
punctuated non-par	172	129	113
male baseline			
unpunctuated non-par	112	89	82
punctuated non-par	104	88	82

The pitch ranges that result from the differences between the toplines and baselines of the punctuated non-parentheticals are shown in Figure 4.4. The male pitch range narrows down over the course of the sentence by first dropping about 27 Hz between fr(a) and the medial phrase and then dropping about 10 Hz from the medial phrase to fr(b). Like in the unpunctuated non-parenthetical sentences, the female pitch range stays exactly level from fr(a) to the medial phrase, but then drops only 18 Hz between the medial phrase and fr(b). Thus, the male pitch range drops about twice as much over the course of the sentence than the female pitch range (female: 18 Hz, male: 37 Hz).

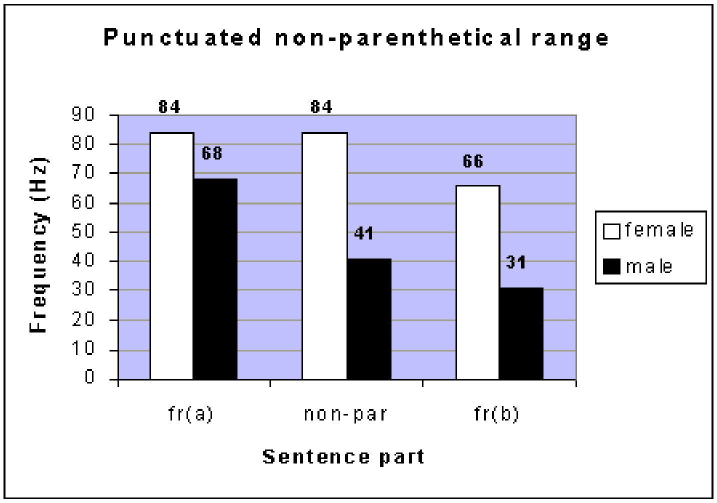


Figure 4.4. Pitch ranges in female and male punctuated non-parenthetical sentences.

In both punctuated and unpunctuated sentences, it can be observed that for males the larger drop occurs between fr(a) and the medial phrase of the word, whereas for females the larger drop occurs from the medial phrase to fr(b). Hence, the distribution of the drops are different between female and male speakers in non-parenthetical sentences. Female speakers maintain the range of fr(a) for the medial phrase. With respect to pitch range, they implement the middle phrases as a continuation of fr(a), whereas male speakers implement them distinctive in range. Whether that is also the case for parenthetical sentences will be shown in the next section.

4.1.2 Single-parentheticals

One of the goals of this study is to determine whether there is a difference in the intonation of parentheticals when they are punctuated differently. The three punctuation marks under investigation are comma, dash and bracket. This section discusses only the sentences that contain one parenthetical phrase. Sentences with two parenthetical phrases are discussed in 4.1.3.

4.1.2.1 Comma-parentheticals

Figure 4.5 displays the maximum and minimum pitch levels for comma-punctuated parenthetical sentences (sentences 4 and 12) for both females and males. An example is sentence 4:

(10) Sentence 4

“We saw the movie, which had been banned in Boston, but were unimpressed.”

For both females and males the pitch level declines over the course of the sentence as both topline and baseline decline. The rate of drop in baseline between fr(a) and the medial phrase is much less in comma-parenthetical sentences than in non-parenthetical sentences. In fact, the baseline for males stays almost level in comma-parentheticals.

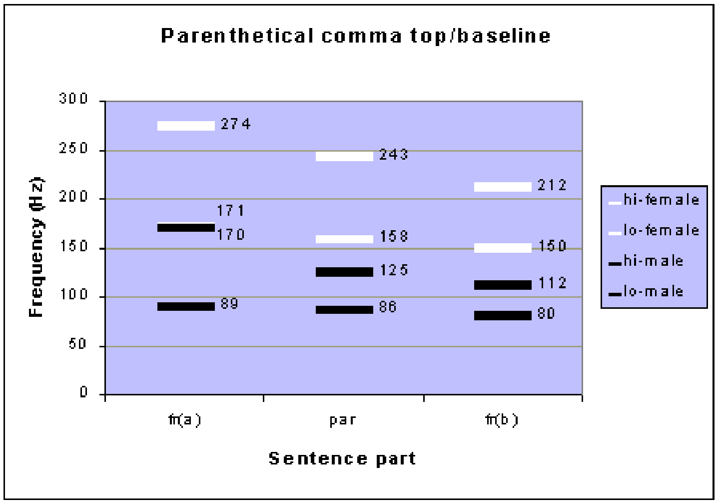


Figure 4.5. Toplines and baselines for female and male comma-punctuated parenthetical sentences.

Figure 4.6 shows the pitch ranges that result from the differences between the toplines and baselines displayed in Figure 4.5. It can be seen that the male pitch range narrows down over the course of the sentence by first dropping about 42 Hz between fr(a) and the medial phrase and then dropping about 7 Hz from the medial phrase to fr(b). The female pitch range narrows down over the course of the sentence by first dropping about 18 Hz between fr(a) and the medial phrase and then dropping about 23 Hz from the medial phrase to fr(b). Hence, male pitch ranges drop more over the course of the sentence (male: 49 Hz, female: 41 Hz). For males the larger drop occurs between fr(a) and the medial phrase, whereas for females the larger drop occurs between the medial phrase and fr(b).

These trends in female/male pitch range change match the trends observed in the non-parenthetical sentences. However, the male pitch range drop between fr(a) and the medial phrase is much steeper in comma-punctuated parentheticals than in non-parentheticals. Furthermore, unlike the pattern exhibited in the non-parentheticals, the female speakers do not maintain the pitch range of fr(a) for the medial phrase, but a pitch range drop occurs.

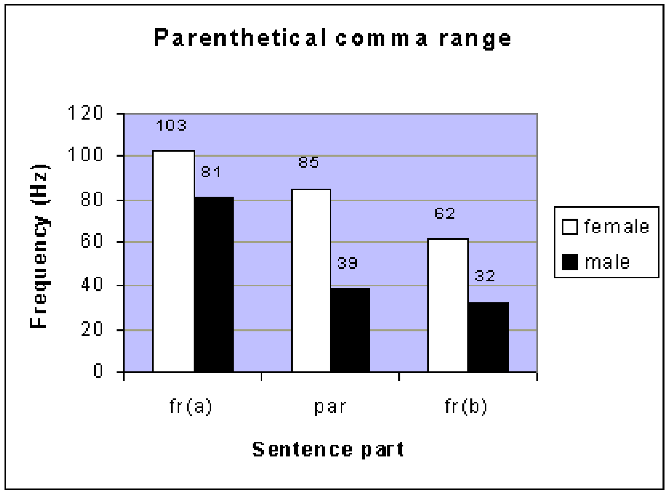


Figure 4.6. Pitch ranges in female and male comma-punctuated parenthetical sentences.

4.1.2.2 Dash-parentheticals

The maximum and minimum pitch levels for dash-punctuated parenthetical sentences (sentences 3 and 13) for both females and males are displayed in Figure 4.7. An example of a dash-punctuated sentence is sentence 13:

(11) Sentence 13

“We saw the movie – which had been banned in Boston – but were unimpressed.”

For both females and males the pitch level declines over the course of the sentence as both topline and baseline decline. The rate of drop in baseline between fr(a) and the medial phrase is much less in dash-parenthetical sentences than in non-parenthetical sentences. Like in comma-parentheticals, the baseline for males stays almost level in dash-parentheticals.

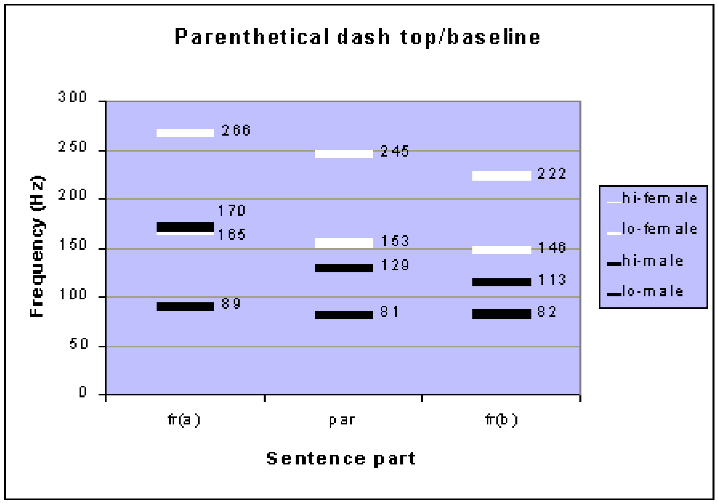


Figure 4.7. Toplines and baselines for female and male dash-punctuated parenthetical sentences.

Figure 4.8 shows the pitch ranges that result from the differences between the toplines and baselines displayed in Figure 4.7. It can be seen that the male pitch range narrows down over the course of the sentence by first dropping about 33 Hz between fr(a) and the medial phrase and then dropping about 17 Hz from the medial phrase to fr(b). The female pitch range narrows down over the course of sentence by first dropping about 9 Hz between fr(a) and the medial phrase and then dropping about 16 Hz from the medial phrase to fr(b). Hence, male pitch ranges drop more over the course of the sentence (male: 50 Hz, female: 25 Hz). For males the larger drop occurs between fr(a) and the medial phrase, whereas for females the larger drop occurs between the medial phrase and fr(b).

These trends in female/male pitch range change match the trends observed in the non-parenthetical and comma-punctuated sentences. However, the pitch range drop between fr(a) and the medial phrase is much steeper in dash-parentheticals and comma-parentheticals than in non-parentheticals. Hence, unlike the pattern exhibited in the non-parentheticals, the female speakers do not maintain the pitch range of fr(a) for the medial phrase.

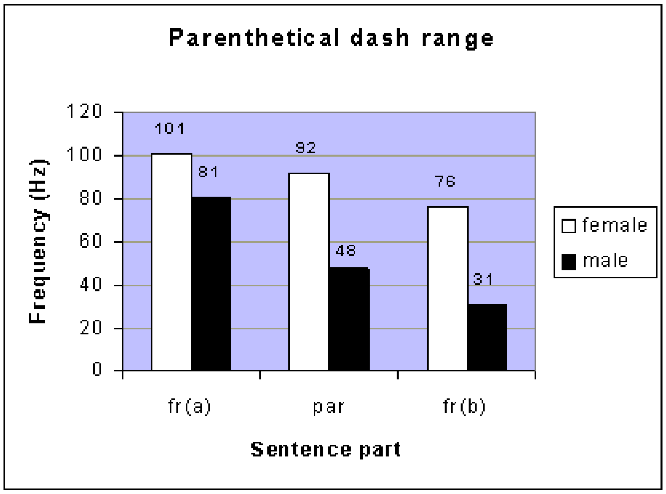


Figure 4.8. Pitch ranges in female and male dash-punctuated parenthetical sentences.

4.1.2.3 Bracket-parentheticals

The maximum and minimum pitch levels for each bracket-punctuated parenthetical sentence (sentences 6 and 9) for both females and males are displayed in Figure 4.9. An example of a bracket-punctuated sentence is sentence 9:

(12) Sentence 9

"We saw the movie (which had been banned in Boston) but were unimpressed."

For both females and males the pitch level declines over the course of the sentence as both topline and baseline decline. The rate of drop in baseline between fr(a) and the medial phrase is smaller in bracket-parenthetical sentences than in non-parenthetical sentences. Like in comma- and dash-parentheticals, the baseline for males stays almost level in bracket-parentheticals.

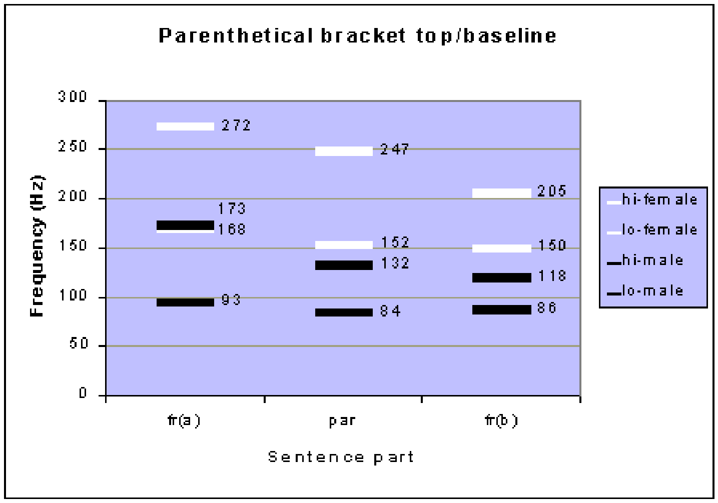


Figure 4.9. Toplines and baselines for female and male bracket-punctuated parenthetical sentences.

Figure 4.10 shows the pitch ranges that result from the differences between the toplines and baselines displayed in Figure 4.9. It can be seen that the male pitch range narrows down over the course of the sentence by first dropping about 32 Hz between fr(a) and the medial phrase and then dropping about 16 Hz from the medial phrase to fr(b). The female pitch range narrows down over the course of the sentence by first dropping about 9 Hz between fr(a) and the medial phrase and then dropping about 40 Hz from the medial phrase to fr(b). Hence, in bracket-punctuated parenthetical sentences, male and female pitch ranges drop about the same over the course of the sentence (male: 48 Hz, female: 49 Hz). This corresponds to the pattern exhibited by unpunctuated non-parentheticals, in which male and female pitch ranges also drop about the same over the course of the sentence.

For males the larger drop occurs between fr(a) and the medial phrase, whereas for females the larger drop occurs between the medial phrase and fr(b). These trends in female/male pitch range change match the trends observed in the non-parenthetical and the other parenthetical sentences. However, the pitch range drop between fr(a) and the medial phrase is much steeper in comma-parentheticals, dash-parentheticals and also bracket-parentheticals than in non-parentheticals. Hence, unlike the pattern exhibited in non-parentheticals, the female speakers do not maintain the pitch range of fr(a) for the medial phrase.

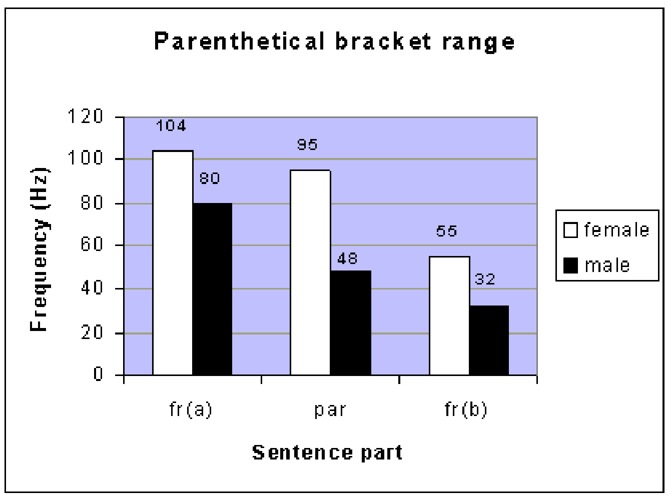


Figure 4.10. Pitch ranges in female and male bracket-punctuated parenthetical sentences.

4.1.2.4 Comparison of differently punctuated parentheticals

Male and female pitch ranges drop about the same amount over the course of sentences that contain bracket-parentheticals, but drop differently in sentences with comma- and dash-parentheticals. However, this does not separate sentences with bracket-parentheticals from the others. As Table 4.2 shows, parenthetical toplines, baselines and pitch ranges are about the same for either punctuation for both females and males.

Table 4.2.

Parenthetical toplines, baselines and pitch ranges are unaffected by the use of different punctuation marks. All values are in Hertz.

	fr(a)	**medial**	**fr(b)**
female topline			
comma-par	274	243	212
dash-par	266	245	222
bracket-par	272	247	205
female baseline			
comma-par	171	158	150
dash-par	165	153	146
bracket-par	168	152	150
female pitch range			
comma-par	103	85	62
dash-par	101	92	76
bracket-par	104	95	55

	fr(a)	medial	fr(b)
male topline			
comma-par	170	125	112
dash-par	170	129	113
bracket-par	173	132	118
male baseline			
comma-par	89	86	80
dash-par	89	81	82
bracket-par	93	84	86
male pitch range			
comma-par	81	39	32
dash-par	81	48	31
bracket-par	80	48	32

Interesting to note is that in all parenthetical and non-parenthetical sentences, the female pitch range drops less between fr(a) and the middle than between the medial phrase and fr(b), whereas for males it is the opposite. For instance, in dash-parentheticals the female pitch range first drops about 9 Hz and then about 16 Hz, whereas the male pitch range first drops about 34 Hz and then about 15 Hz. However, this difference occurs to be a general trend, unrelated to a distinction between differently punctuated parenthetical sentences as well as non-parenthetical sentences. Hence, investigating this gender difference further goes beyond the scope of this study.

The data discussed in section 4.1.2 shows that there is no significant difference between differently punctuated parentheticals. With respect to pitch, this supports Nunberg’s claim that punctuation is more than a graphical reflection of intonation.

As a consequence of the findings in this section, the remainder of this study does not distinguish different punctuation marks in the acoustic analyses discussed.

4.1.3 Two-parentheticals

The previous section showed that the variance in punctuation for parentheticals is not reflected in the spoken version of the sentences. This was predicted by Nunberg (1990) and to provide acoustic evidence for this claim, is one of the goals of this study. Another goal is to investigate how sentences with more than one parenthetical phrase are spoken. In particular, this thesis takes a look at two parentheticals with one nested within the other and two parentheticals where one directly follows the other.

4.1.3.1 Nested two-parentheticals

The maximum and minimum pitch levels for each sentence containing nested parentheticals (sentences 8, 18 and 20) are displayed in Figure 4.11. An example is sentence 18:

(13) Sentence 18 with phrase labels underneath

“We saw the movie – Jane (who knows the director) insisted on going – but were unimpressed.”
| fr(a) | par1(a) | par2 | par1(b) | fr(b) |

As observed before, females exhibit a higher pitch level than males. For both females and males the topline drops steeply (female: 43 Hz, male: 43 Hz) between fr(a) and the first part of the outer parenthetical, par1(a). Following that, the female topline rises towards par1(b) (first about 11 Hz then about 6 Hz) and ends with a drop of 29 Hz for fr(b). The male topline first further drops about 4 Hz for par2, before it rises for par1(b) (about 9 Hz) and ends with a drop of 14 Hz for fr(b). The baseline of par1(a) stays about level with fr(a) for both females and males, then it drops (female: 9 Hz, male: 10 Hz) for the nested parenthetical, par2, stays at about the same level for par1(b), and drops (female and male: 6 Hz) to its lowest level for fr(b).

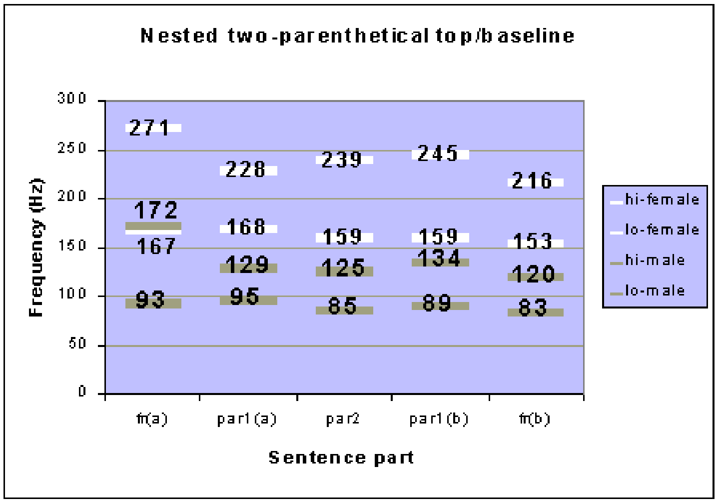


Figure 4.11. Toplines and baselines for female and male nested two-parenthetical sentences.

Displaying the pitch ranges in Figure 4.12 shows a steep drop (females: 44 Hz, males: 45 Hz) between fr(a) and par1(a) – much steeper than the average drop in single-parentheticals (female average: 12 Hz, male average: 35 Hz) – then a smaller rise in both par2 and par1 (b), and finally a drop back to the par1(a) range in fr(b).

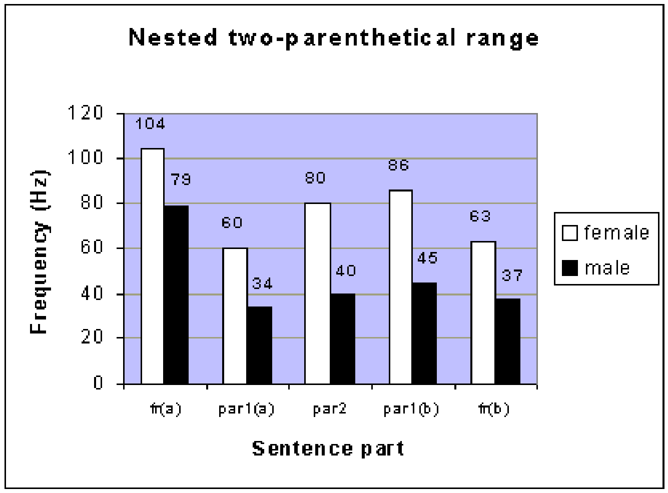


Figure 4.12. Pitch ranges in female and male nested two-parenthetical sentences.

4.1.3.2 Sequential two-parentheticals

The maximum and minimum pitch levels for each sentence containing two parentheticals in sequence (sentences 10 and 15) are displayed in Figure 4.13. An example is sentence 10:

(14) Sentence 10 with phrase labels underneath

“We saw the movie – I heard it was good (it had received much praise) – but were unimpressed.”

| fr(a) | par1 | par2 | fr(b) |

In this case as well, females exhibit a higher pitch level than males. For both females and males the topline drops steeply (female: 31 Hz, male: 43 Hz) between fr(a) and par1. Continuing from there, the topline rises (female: 4 Hz, male: 8 Hz) for par2 and ends with a drop (female: 28 Hz, male: 16 Hz) for fr(b). The female baseline first drops about 4 Hz for par1, then another 3 Hz for par2, before it drops a further 7 Hz to its lowest level for fr(b). The male baseline falls about 7 Hz for par1, then another 3 Hz for par2, but stays at about the same level for fr(b).

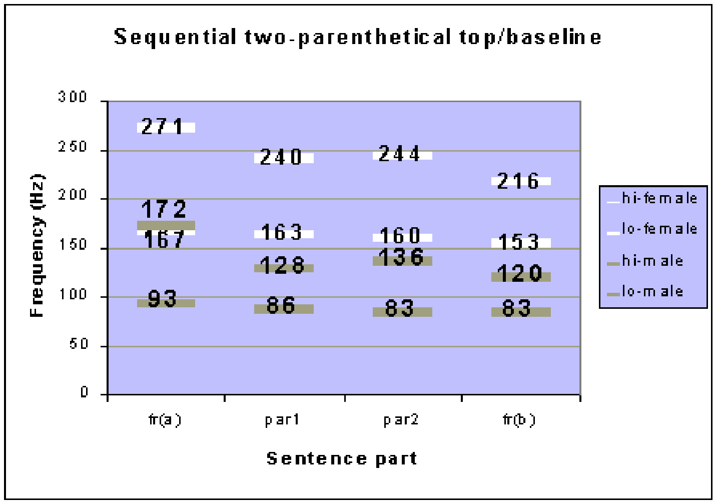


Figure 4.13. Toplines and baselines for female and male sequential two-parenthetical sentences.

Displaying the pitch ranges of the sequential two-parentheticals in Figure 4.14 shows a striking similarity to the pitch ranges of the nested two-parentheticals in Figure 4.12 above, provided that par1(a) is left out of the picture. Not only are fr(a) and fr(b) about the same, but also sequential par1 corresponds to nested par2 and sequential par2 corresponds to nested par1(b). These similarities will be further discussed in the following section, which provides a comparison between single-parentheticals and the two kinds of two-parentheticals investigated in this study.

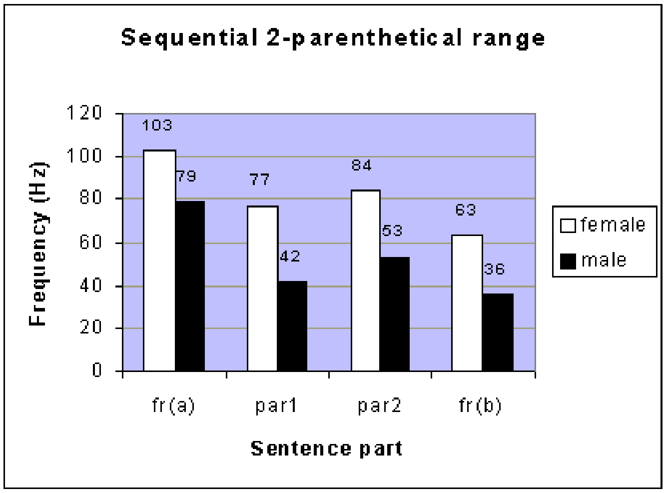


Figure 4.14. Pitch ranges in female and male nested two-parenthetical sentences.

4.1.3.3 Comparison of two-parentheticals and single-parentheticals

When comparing the sequential and nested parentheticals to the single-parentheticals, it can be observed that the more parenthetical constituents that are present, the smaller in pitch range the one that directly follows fr(a). Furthermore, the one directly preceding fr(b) exhibits the same pitch level as the parenthetical phrase in a single-parenthetical and all other parenthetical constituents are of decreasingly lower pitch levels down to the lowest, the phrase that directly follows fr(a). That is, the parenthetical constituent that occurs first over the course of a sentence is also the lowest in comparison to directly following parenthetical constituents.

This relationship is illustrated in Table 4.3. for females and for males in Table 4.4. The tables shows the pitch values of the individual phrases and their respective location within the different parenthetical sentences. It can be seen that phrases at the same relative location to fr(b) exhibit pitch values that are about the same. The pitch values for fr(a) and fr(b) match the values for single-parentheticals.

Since there is no difference between differently punctuated parentheticals, the single-parentheticals have been averaged into one value for this comparison and the same applies to the sequential and the nested parentheticals.

Table 4.3.

Comparison of the location of corresponding parenthetical segments within different parenthetical sentences for females. All values are in Hertz.

	fr(a)	**third closest to fr(b)**	**second closest to fr(b)**	**closest to fr(b)**	**fr(b)**
female topline					
single-par	271			245 (par)	212
sequential-par	271		240 (par1)	244 (par2)	216
nested-par	271	228 (par1a)	239 (par2)	245 (par1b)	216
female baseline					
single-par	168			154 (par)	149
sequential-par	167		163 (par1)	160 (par2)	153
nested-par	167	168 (par1a)	159 (par2)	159 (par1b)	153
female pitch range					
single-par	103			91 (par)	63
sequential-par	103		77 (par1)	84 (par2)	63
nested-par	104	60 (par1a)	80 (par2)	86 (par1b)	63

Table 4.4.

Comparison of the location of corresponding parenthetical segments within different parenthetical sentences for males. All values are in Hertz.

	fr(a)	third closest to fr(b)	second closest to fr(b)	closest to fr(b)	fr(b)
male topline					
single-par	171			129 (par)	115
sequential-par	172		128 (par1)	136 (par2)	120
nested-par	172	129 (par1a)	125 (par2)	134 (par1b)	120
male baseline					
single-par	91			84 (par)	83
sequential-par	93		86 (par1)	83 (par2)	83
nested-par	93	95 (par1a)	85 (par2)	89 (par1b)	83
male pitch range					
single-par	80			45 (par)	32
sequential-par	79		42 (par1)	53 (par2)	36
nested-par	79	34 (par1a)	40 (par2)	45 (par1b)	37

Comparing the columns of the table gives the impression that (i) the more parentheticals are inserted, the narrower is the pitch range of the first one, and (ii) a parenthetical constituent in the same relative position to fr(b) is spoken the same way with respect to topline, baseline and pitch range. This pattern fits nested- and sequential two-parentheticals as well as the single-parentheticals.

This section shows that sentences with nested-parentheticals exhibit similar pitch patterns to sentences with sequential parentheticals as well as sentences with one parenthetical phrase only. In fact, regarding pitch values, the phases of single-parentheticals are a subset of sequential parentheticals that are a subset of nested parentheticals. Hence, single-parenthetical *par* corresponds to sequential *par2* as well as nested *par1(b)*, and sequential *par1* corresponds to nested *par2* (see Table 4.3 and 4.4 above).

This section also shows that parenthetical phrases are not all spoken the same way. When they occur next to each other, the first is lower in pitch level and narrower in range than the directly following one. Thus, in a text-to-speech system parentheticals cannot all be implemented with the exact same specifications. The implementation difference is not necessarily due to the use of different punctuation, as originally suspected, but when two or more parenthetical phrases occur next to each other. Therefore, the text-to-speech system has to be told that each part of a multi-parenthetical – sequential or nested – has to be implemented with different specifications. The exact nature of the specifications for implementing single-parentheticals as well as multi-parentheticals in text-to-speech systems will be discussed in chapter five. However, before discussing text-to-speech implementation, it has to be shown if and how parentheticals differ from non-parentheticals in pitch, which is the focus of the discussion in the following section.

4.1.4 Comparison between parentheticals and non-parentheticals

One of the goals of this thesis is to elaborate on Kutik et al.'s findings that parenthetical phrases are distinct through a lower topline than both the preceding and following phrase. A lower topline and unchanged baseline results in a lower pitch level and narrower pitch range for a parenthetical, which are the characteristics of parentheticals as reported by Bolinger (1989), Cruttenden (1997), Crystal (1969), Grosz and Hirschberg (1992), Kutik et al. (1983), O'Shaughnessy (1990) and Wichmann (2000). The only instance of a parenthetical sentence being implemented by a text-to-speech system was found in SABLE. The markup example implemented the parenthetical not only with a narrower pitch range, but also with a 20 percent lower baseline – instead of Kutik et al.'s lower topline. This section compares parentheticals with non-parentheticals to find out how they exactly differ and with which prosodic specifications parentheticals should be implemented in text-to-speech systems.

The comparison uses only sentences with one parenthetical phrase, because all non-parenthetical sentences also only have one non-parenthetical phrase inserted, i.e. in the data there is no non-parenthetical equivalent for nested and sequential two-parentheticals. Thus, using the data provided by this study, the specifications for multi-parentheticals cannot be established through a comparison to measurements of corresponding non-parentheticals. Instead, the missing non-parenthetical data have to be interpolated from the existing data. This will be part of the discussion about synthesizing parentheticals in chapter five.

As it has been shown in 4.1.1 and 4.1.2, difference in punctuation do not lead to differing intonation patterns for non-parenthetical sentences. The same applies to parenthetical sentences. Therefore, the results of all non-parentheticals have been averaged and all single-parentheticals have been averaged before being compared.

The comparisons in this section start with toplines in 4.1.4.1, then baselines in 4.1.4.2 and resulting pitch ranges in 4.1.4.3. The conclusion is provided in 4.1.4.4.

4.1.4.1 Toplines

Table 4.5 displays averaged topline values for females and Table 4.6 for males. Comparing the non-parenthetical and parenthetical values reveals that they do not differ significantly. For instance, the difference of 4 Hz in the male fr(b) topline is only 1 Hz over the perceptual threshold of 3 Hz (Hess, 1983) – a difference too small to constitute a trend. For both females and males the non-parenthetical and parenthetical sentences show the same topline development over the course of the sentence.

Table 4.5.

Female topline comparison. All values are in Hertz.

	fr(a)	**medial**	**fr(b)**
female topline			
non-par	272	244	214
par	271	245	212
difference in Hertz	1	1	2
difference in %	0%	0%	0%

Table 4.6.

Male topline comparison. All values are in Hertz.

	fr(a)	**medial**	**fr(b)**
male topline			
non-par	174	130	111
par	171	129	115
difference in Hertz	3	1	4
difference in %	2%	1%	4%

Table 4.5 and 4.6 show that there is no difference in topline between parentheticals and non-parentheticals, neither in the medial phrase nor in any other part of the sentence. This contradicts Kutik et al.'s findings.

4.1.4.2 Baselines

Table 4.7 displays averaged baseline values for females. Comparing the fr(b) and medial phrase values for non-parenthetical and parenthetical values reveals that they do not differ significantly in these. Thus, there is no difference between the parenthetical and non-parenthetical sentences in the parenthetical or non-parenthetical medial phrase itself. However, the fr(a) baseline differs significantly (18 Hz) between parentheticals and non-parentheticals. The data show that the fr(a) part of the parenthetical sentences is lowered in comparison to the non-parenthetical sentences. With the topline unchanged, this means that for females the range is increased for fr(a) in parenthetical sentences (see 4.1.4.3).

Table 4.7.

Female baseline comparison. Pitch value deviation from the non-parenthetical pattern is displayed in boldface. All values are in Hertz.

	fr(a)	**medial**	**fr(b)**
female baseline			
non-par	186	157	152
par	**168**	154	149
difference in Hertz	18	3	3
difference in %	10%	2%	2%

Table 4.8 displays averaged baseline values for males. Like in the pattern exhibited by the females, the fr(b) and medial phrase values for non-parenthetical and parenthetical values do not differ significantly in these. Hence, for males as well, there is no difference between the parenthetical and non-parenthetical sentences in the parenthetical or non-parenthetical medial phrase itself. However, the fr(a) baseline differs significantly (17 Hz) between parentheticals and non-parentheticals. The data show that the fr(a) part of the parenthetical sentences is lowered in comparison to the non-parenthetical sentences. With the topline unchanged, this means that for males the range is also increased for fr(a) in parenthetical sentences (see 4.1.4.3).

Table 4.8.

Male baseline comparison. Pitch value deviation from the non-parenthetical pattern is displayed in boldface. All values are in Hertz.

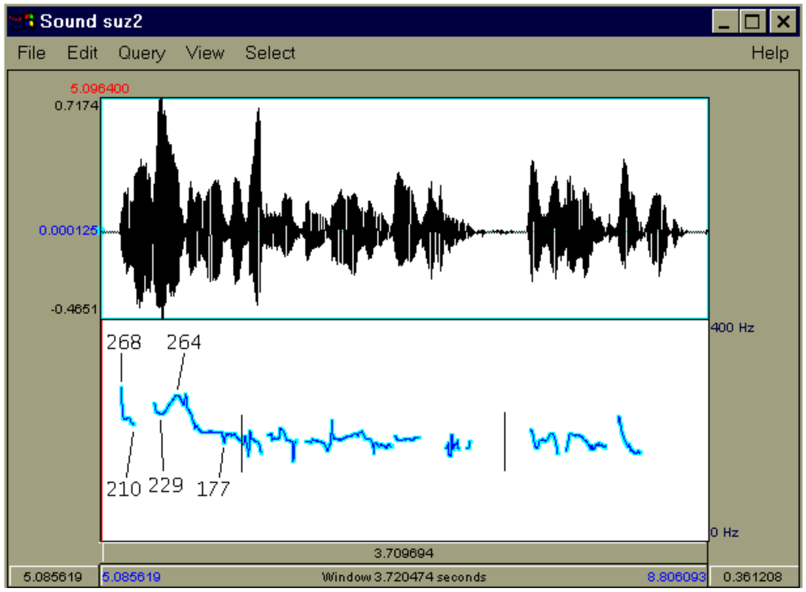
	fr(a)	**medial**	**fr(b)**
male baseline			
non-par	108	88	82
par	**91**	84	83
Difference in Hertz	17	4	1
Difference in %	16%	5%	1%

Table 4.7 and 4.8 show that there is a difference between non-parenthetical and parenthetical baseline, although not in the medial phrase itself but in fr(a). The first part of the frame sentence, fr(a), is implemented with a 17-18 Hz lower baseline for both females and males. This contradicts the implementation in SABLE that specifies a lower baseline for the parenthetical phrase itself instead of the preceding phrase. Furthermore, it

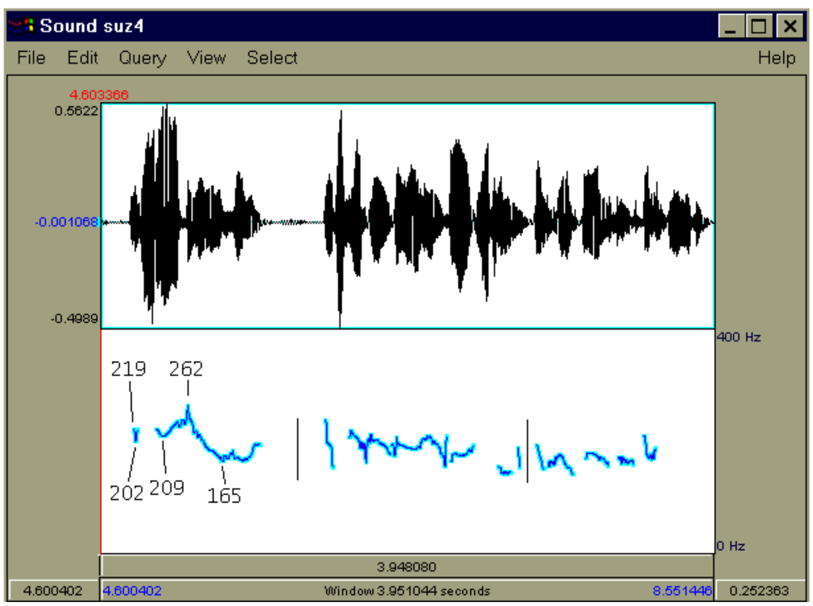
contradicts the statements of Bolinger (1989), Cruttenden (1997) and Crystal (1969) that the parenthetical itself is lower-leveled. By comparing parenthetical phrases to non-parenthetical phrases that occur at the same place in the same surrounding sentence, this study reveals that they differ neither in baseline nor in topline.

A pitch contour is comprised of upward and downward pitch movements, resulting in crests and troughs. The baseline is established by measuring the lowest point of the contour in each sentence part. The finding of a lowered baseline value for fr(a), the parenthetical-preceding phrase, leads to the question whether the troughs of the pitch contour are lowered throughout fr(a), or whether only the trough closest to the medial phrase is lowered. The latter scenario would indicate that only directly before implementing the parenthetical, the speaker realizes that fr(a) needs to exhibit a larger range to indicate that there is a parenthetical following.

As Figure 4.15 illustrates, there is evidence that the bottom values of the pitch contour are lowered throughout fr(a). The figure shows that all fr(a) troughs are at least 8 Hz lower in the parenthetical-containing sentence than in the corresponding non-parenthetical sentence. This shows that from the start of the sentence, fr(a) is implemented with lower baseline values to indicate that the following phrase – which is implemented the same whether parenthetical or not – is a parenthetical.



(a) Sentence 2: We saw the movie | that had been banned in Boston, | but were unimpressed.



(b) Sentence 4: We saw the movie, | which had been banned in Boston, | but were unimpressed.

Figure 4.15. Comparison between fr(a) pitch trough values in non-parenthetical sentence 2 and parenthetical sentence 4. Both are spoken by SUZ. All values in Hertz.

4.1.4.3 Pitch ranges

A comparison of pitch ranges between parenthetical and non-parenthetical sentences is displayed in Table 4.9 for females and Table 4.10 for males. Comparing the fr(b) and medial phrase values for non-parenthetical and parenthetical values reveals that they do not differ significantly in these. Therefore, regarding pitch range as well, there is no difference between the parenthetical and non-parenthetical sentences in the parenthetical or non-parenthetical medial phrase itself. However, due to the lowered fr(a) baseline in the parenthetical sentences, there is a wider pitch range (female: +17 Hz, male: +14 Hz) in fr(a) in parenthetical sentences.

Table 4.9.

Female pitch range comparison. Pitch value deviation from the non-parenthetical pattern is displayed in boldface. All values are in Hertz.

	fr(a)	**medial**	**fr(b)**
female pitch range			
non-par	86	87	62
par	**103**	91	63
difference in Hertz	17	4	1
difference in %	20%	5%	2%

Table 4.10.

Male pitch range comparison. Pitch value deviation from the non-parenthetical pattern is displayed in boldface. All values are in Hertz.

	fr(a)	**medial**	**fr(b)**
male pitch range			
non-par	66	42	29
par	**80**	45	32
difference in Hertz	14	3	3
difference in %	21%	7%	10%

Instead of a narrower pitch range for the medial phrase in parenthetical sentences, the data show a 20-21% wider pitch range in fr(a). On first glance, this contradicts what other researchers found. However, it has to be taken into consideration that this analysis is based on a phrase to phrase comparison between sentences. If the comparison, instead, looks primarily at the phrase differences within a sentence, it shows that the drop in pitch range between fr(a) and the medial phrase is much steeper in parentheticals (female: -12 Hz, male: -35 Hz) than in non-parentheticals (female: +1 Hz, male: -24 Hz).

The following figures display the difference between parenthetical and non-parenthetical medial phrases in relation to fr(a). The initial phrase fr(a) is set to 100 percent for each sentence type, which represents the circumstance that, perceptually, the initial phrase defines the pitch level that the following phrase stands in relation to and, thus, is compared to. Figure 4.16 shows that in relation to their respective initial phrases fr(a) (100%), the female medial parenthetical phrase (88%) is 13 percent lower in pitch range when compared to the medial non-parenthetical phrase (101%). Figure 4.17 shows that in relation to their respective initial phrases fr(a) (100%), the male medial parenthetical phrase (56%) is 7 percent lower in pitch range when compared to the medial non-parenthetical phrase (63%).

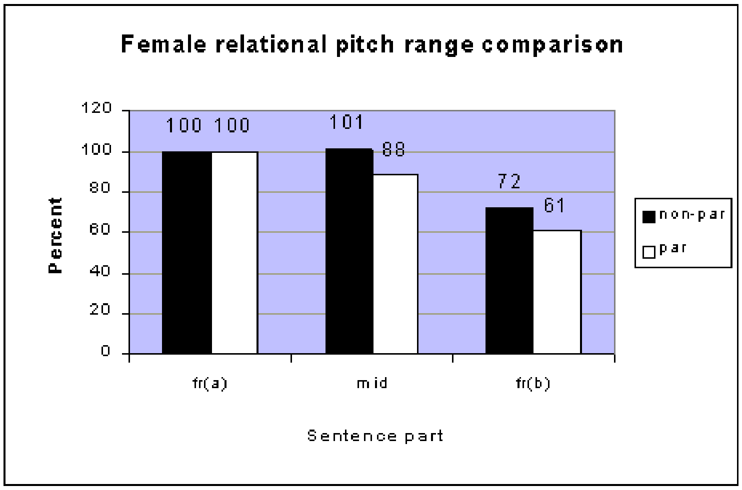


Figure 4.16. Female pitch ranges of medial and final phrases in relation to their respective initial phrases (for each sentence type: fr(a) = 100 %).

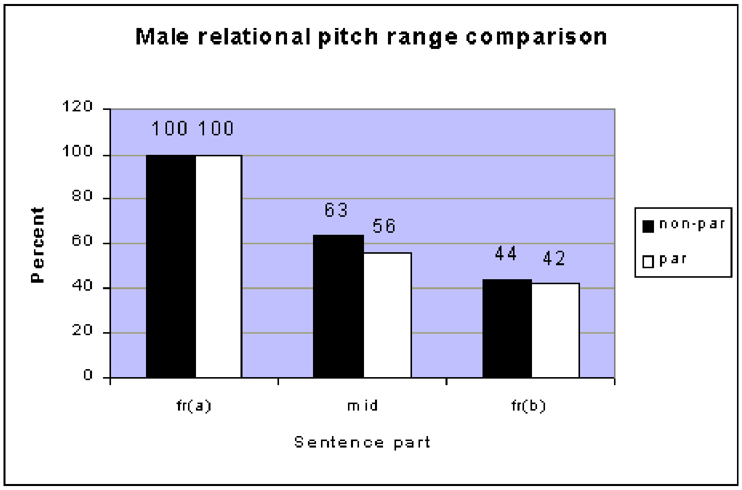


Figure 4.17. Male pitch ranges of medial and final phrases in relation to their respective initial phrases (for each sentence type: fr(a) = 100 %).

Thus, in relation to fr(a) the pitch range is indeed narrower in parenthetical phrases than in non-parenthetical phrases at the same sentence position. However, this observation would fail to see that, physically, it is fr(a) that is changed and not the medial phrase.

It further has to be mentioned that in this study, parentheticals are inserted at the clausal level, i.e. between one clause and another (see (15a)). Kutik et al. (1983), however, inserted the parentheticals at the phrasal level, i.e. between noun phrase and verb phrase within a clause (see (15b)).

(15) Location of insertion of parenthetical element

a. this study – insertion between clauses:

We saw the movie, which was banned in Boston, but were unimpressed.
| fr(a) clause | parenthetical phrase | fr(b) clause |

b. Kutik et al. (1983, p. 1732) – insertion between noun phrase and verb phrase:

The clock in the church, it occurred to Clark, chimed just as he began to talk.
| noun phrase | parenthetical phrase | verb phrase |

Following previous research, there is no evidence for an intonational difference between parentheticals based on the type of syntactic structure that they are inserted in. Comparing Kutik et al. (1983) and this study can reveal intonational differences. This study reveals that the fr(a) baseline is lowered when a parenthetical follows. Kutik et al. find that the topline of the parenthetical is lowered. However, the studies used different research methods. This study compared parenthetical sentences to non-parenthetical sentences, whereas Kutik et al. did not use any non-parenthetical sentences but compared parenthetical sentences of differing length to each other. Hence, whether the different intonational characteristics for parentheticals found in the two studies depend on the type of syntactic structure the parenthetical is inserted in cannot be answered by this study and has to be left to future research.

4.1.4.4 Conclusion

The comparison between parenthetical and non-parenthetical sentences shows that there is no difference between the toplines but a difference in baseline and pitch range in fr(a). The medial phrase itself does physically not differ. Furthermore, variation in punctuation does not induce a pitch difference, as hypothesized by Nunberg (1990).

The fr(a) baseline is 17-18 Hz lower for both females and males in parenthetical sentences. Due to the lowered fr(a) baseline there is a wider pitch range (female: +17 Hz, male: +14 Hz) in fr(a) in parenthetical sentences. Thus, the drop in range from fr(a) to the medial phrase is much larger for parenthetical phrases (female: -12 Hz, male: -35 Hz) than for non-parenthetical phrases (female: +1 Hz, male: -24 Hz). The steeper drop in parentheticals explains why parenthetical phrases are perceived as being smaller in range, although when directly compared to the non-parenthetical phrases, they are not. It is the adjacency to a widened preceding fr(a) that makes the parenthetical itself appear to be narrower. With respect to topline and baseline the parenthetical phrase is implemented the same as a non-parenthetical phrase at the same position. Hence, what has always been claimed to be a characteristic of parenthetical phrases is actually the result of the environment and not of the parenthetical phrase itself. Nonetheless, this does not prove former observations wrong, since the effect of a lowered baseline, and with that widened range for fr(a), results in a perceptually lower pitch level and range in parenthetical phrases.

Comparing parenthetical and non-parenthetical phrases embedded in the same frame sentence reveals that both types of phrases are physically the same, but there is a steeper drop in pitch range between fr(a) and a parenthetical phrase than between fr(a) and a non-parenthetical phrase. This steeper drop is the perceptual cue that identifies a parenthetical and it is caused by a change in fr(a) in the parenthetical sentences. The wider fr(a) pitch range in parenthetical sentences of about 17 Hz for females and 14 Hz for males corresponds to a pitch range increase of about 20 percent for both females and males.

How these findings are to be integrated in text-to-speech systems will be discussed in chapter five.

4.2 Pauses

Although there are no reports that the length of pauses between phrases is different for parentheticals, this possibility has also been investigated. In particular, this study examined if the pausing pattern of parentheticals is different from non-parentheticals and if there is a distinction between differently punctuated parentheticals.

First the pause analysis will look at sentences with one medial phrase in 4.2.1, then the two-parentheticals will be looked at in 4.2.2 and a conclusion follows in 4.2.3.

4.2.1 Sentences with one medial phrase

The pausing length at the phrase boundaries of the non-parenthetical and single-parenthetical sentences for females are displayed in Figure 4.18. The bottom axis of the graph shows the sentence type, starting with unpunctuated non-parentheticals and ending with bracket-punctuated parentheticals. Instead of displaying the names of the sentence types in the graph, a structural representation has been chosen that allows to display the location and type of punctuation. For each sentence type, the white, left bar shows the pause between fr(a) and the medial phrase, while the black, right bar shows the pause between the medial phrase and fr(b).

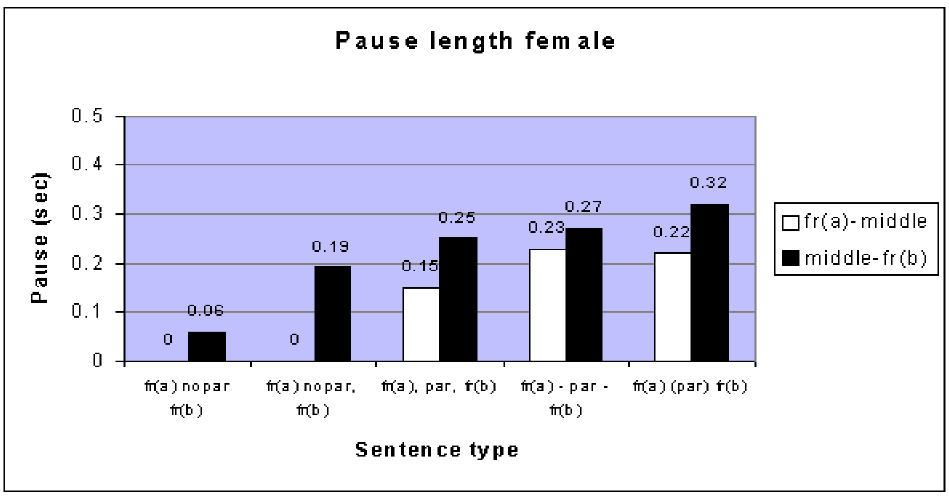


Figure 4.18. Female pause length at different boundaries.

It can be seen that there is a distinction in pausing at a boundary depending on the fact whether there is a punctuation mark at the boundary or not. Boundaries that are not punctuated exhibit no pause. In contrast, boundaries that are punctuated exhibit a pause between 0.15 s and 0.32 s. Of these, pauses at punctuated medial-fr(b) boundaries are longer than pauses at punctuated fr(a)-medial boundaries in the same sentence.

Parenthetical pauses at the medial-fr(b) boundary are somewhat longer than punctuated non-parenthetical pauses (par: 0.25 s to 0.32 s, non-par: 0.19 s). However, the difference between punctuated non-parenthetical and comma- and dash-punctuated parenthetical pauses is less than 0.1 s. Hence, the difference is less than the rounding range (pause values were rounded to the nearest 10th of a second) and, thus, not significant. Similarly, the difference between the differently punctuated parentheticals is less than 0.1 s and therefore not significant as well. Since there is no significant difference between the parentheticals as well as comma- and dash-punctuated parentheticals to non-parentheticals, no significant difference can be established between punctuated parentheticals and non-parentheticals in general. Hence, punctuation triggers pausing, but if the pause occurs at a parenthetical or non-parenthetical boundary makes no difference.

The pausing length at the phrase boundaries of the non-parenthetical and single-parenthetical sentences for males are displayed in Figure 4.19. All boundaries that are not punctuated exhibit no significant pause. Like in Figure 4.18, all boundaries that are punctuated exhibit a pause with a length between 0.25 s and 0.45 s for males. This shows that males generally pause longer than females.

Corresponding to the female pattern, pauses at punctuated medial-fr(b) boundaries are longer than pauses at punctuated fr(a)-medial boundaries in the same sentence. Furthermore, at the medial-fr(b) boundary parenthetical pauses are somewhat longer than punctuated non-parenthetical pauses (par: 0.42 s to 0.45 s, non-par: 0.37 s). However, these differences are also less than 0.1 s and therefore insignificant. That is, for both males and females, there is no significant difference in pausing for differently punctuated parenthetical as well as punctuated non-parenthetical sentences.

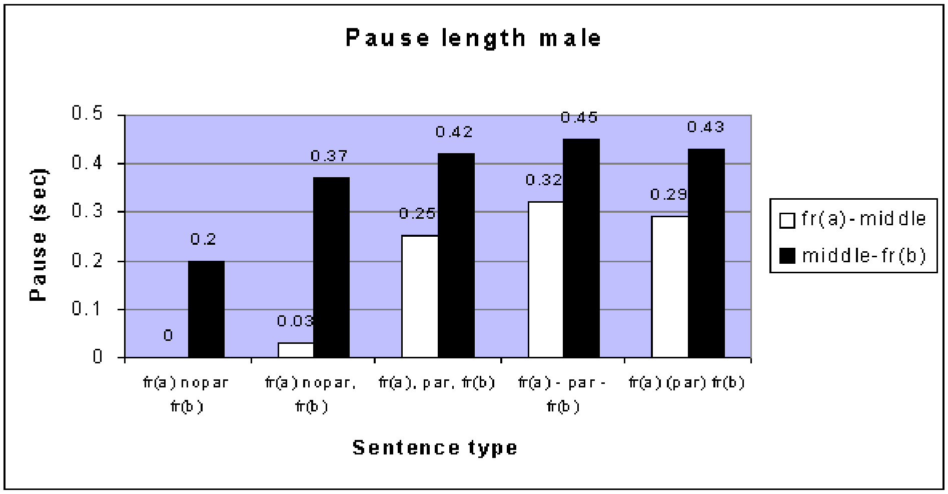


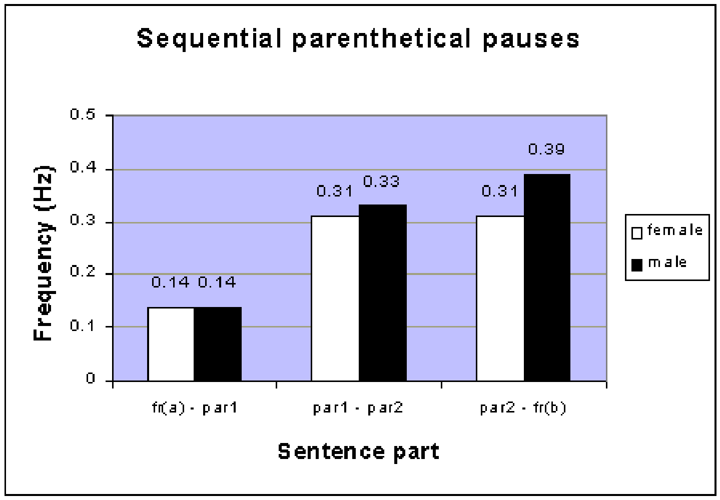
Figure 4.19. Male pause length at different boundaries.

4.2.2 Two-parentheticals

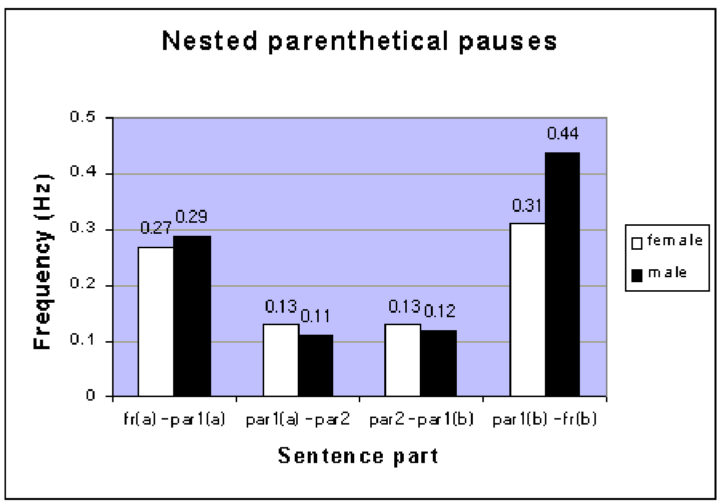
This section discusses the pausing trend in sentences with two sequential or nested parenthetical phrases. As it has been shown in the previous section, there is no difference in non-parenthetical and parenthetical sentences due to differences in punctuation. Thus, the comparison discussed in this section does not differentiate between different punctuation marks but only boundary location.

Figure 4.20 shows the trends of pausing for both males and females at phrase boundaries in sentences with two parenthetical phrases. Part (a) of the figure displays the pausing pattern of sentences with sequential parentheticals and part (b) displays the pausing pattern of sentences with nested parentheticals.

When comparing both types of two-parentheticals, it becomes clear that there is no general trend in pausing between parenthetical phrases. Pauses between nested parenthetical phrases are extremely short (about 0.11 s to 0.13 s), whereas pauses between two sequential parenthetical phrases are long (about 0.31 s to 0.33 s).



(a) Pauses in sentences with sequential parenthetical phrases.



(b) Pauses in sentences with nested parenthetical phrases.

Figure 4.20. Pauses in sentences with two parentheticals.

4.2.3 Conclusion

The examination of pauses in section 4.2.1 and 4.2.2 revealed the following general trends. Firstly, males pause longer than females. Secondly, there is a trend that pauses later in the sentence tend to be longer than earlier ones. The only exceptions to this are nested parentheticals. For these, the parenthetical-internal pauses are significantly shorter than the preceding or following pauses. Thirdly, the presence of a punctuation mark triggers a pause. However, the observed trends are general and not parenthetical-specific. Hence, they do not contribute to the goal of improving the naturalness of parentheticals in text-to-speech systems.

Comparing non-parenthetical sentences with sentences that contain differently punctuated parenthetical phrases shows that there is no difference in pausing based on the type of punctuation or type of phrase. Furthermore, comparing different types of two-parentheticals indicates that there is no general pattern in pausing between parenthetical phrases as some are implemented with short duration, whereas others exhibit long duration. Therefore, the pausing analysis did not reveal a trend for parentheticals that can be used to improve the naturalness of parentheticals in text-to-speech systems.

Chapter Five

SYNTHESIZING PARENTHETICALS

This chapter discusses how the findings of the analysis presented in chapter four can be integrated into text-to-speech systems to improve the naturalness of parentheticals in synthesized speech.

Synthesized speech still lacks a great deal of naturalness, because speech synthesizers do not have all the necessary specifications available to produce speech as natural as a human being. Some of these phonetic specifications derive from higher-level knowledge, such as the occurrence of old versus new information, which is not uniquely identified in the syntax. Therefore, the syntax-based parse of most text-to-speech systems cannot identify these structures. As discussed in chapter two, the methods for identifying parentheticals in a text are the rhetorical parser (Marcu, 1997, 1998, 1999), which automatically detects parentheticals, as well as using markup such as the markup language SABLE (Sproat et al., 1998; Sproat & Raman, 1999). Since rhetorical parsing is not incorporated into any text-to-speech system, it does not solve the thesis goal of presenting a system that is able to detect parentheticals in a text and to subsequently synthesize them with the specifications that result from the acoustic study of this thesis. SABLE, on the other hand, is not a component of a particular text-to-speech system but a method that provides a means of annotation as well as conversion of this annotation into acoustic specifications that are supplied to a synthesizer.

Although SABLE has already been identified as the best method for detecting and synthesizing parentheticals (see chapter two), this chapter also discusses other encoding methods besides markup. The purpose is to provide an overview of how different encoding methods perform with respect to incorporating the acoustic findings of this study. Thus, this helps to point out deficits in the encoding methods of different systems with respect to synthesizing parentheticals.

Translating the findings of chapter four into the encoding schemes of text-to-speech systems is discussed as follows: The first section discusses high tone and low tone-based encoding, the second section discusses encoding using detailed phonetic tags and the third section discusses markup encoding; the last section summarizes this chapter.

5.1 H-L-based encoding

To represent intonation contours orthographically, some speech synthesis systems use a sequence of abstract high (H) and low (L) tones. The pitch movements throughout the contour are encoded through H-L-based boundary tones, pitch accents and a phrase accent (Pierrehumbert, 1980). Examples for these encoding schemes are ToBI (Silverman et al., 1992) and the advanced intonation model for synthesis developed by Tatham et al. (1999, 1998).

Wightmann (2002) argues that, particularly with respect to text-to-speech synthesis, ToBI has outlived its usefulness. Its significant inter-transcriber variance is a major drawback, because the use of differing labels for the same acoustic event by different annotators results in different speech synthesis output.

H-L-based models are concerned about representing the upward and downward pitch movements over the course of the intonation contour but not the ups and downs in pitch range. Except declination, changes in topline, baseline and pitch range cannot be represented in H-L-based pitch encoding schemes. An example of modeling normal non-parenthetical declination with the advanced intonation model of synthesis can be seen in (16). The H at the beginning indicates that the sentence starts on a high pitch level and the L at the end indicates that the sentence finishes with a low pitch level. The internal stress-related structure of the sentence is left out in these examples as this study has not investigated stress. Instead of S for a stressed and U for an unstressed syllable, dots have been used as a placeholder analogous to Tatham et al. (1999, 1998).

(16) Example of non-parentheticals in the advanced intonation model of synthesis

"We saw the movie on the opening night but were unimpressed."

H[...]L

An attempt at representing a parenthetical sentence in the advanced intonation model of synthesis following the findings of chapter four is shown in (17). The first part of the frame sentence shows the largest pitch drop from H to L to represent the larger pitch range of fr(a). As discussed in chapter four, the highest pitch in fr(a) is found towards the beginning, while the lowest is found at the end. The two phrases following fr(a) decrease consistently from a level higher than the end of fr(a) to the lowest level.

(17) Example of parentheticals in the advanced intonation model of synthesis

"We saw the movie, Jane insisted on going, but were unimpressed."

H[............................]L H[...]L

Although this represents the contour of the parenthetical-containing sentence, it lacks the capability to integrate exact pitch specifications for increasing the naturalness of the speech output. Furthermore, as mentioned in section 2.4, this is one of the speech synthesis methods that requires a previous recording of the utterance to derive the contour. Hence, this is not a suitable text-to-speech method for this study.

5.2 Phonetic encoding

Tatham and Lewis's SPRUCE text-to-speech system (1992, 1996) uses the phonetic encoding system of Pierrehumbert (1981) to integrate prosodic structure into the structure of the written text. This means that down to the smallest unit of the letter, each text constituent is tagged with detailed prosodic specifications. An example is provided in (18). The sentence that is prosodically encoded is, *In November the region's weather was unusually dry*. In this encoding method, these symbols have the following meaning:

- * = word boundary;
- % = minor phrase boundary;
- ( ) = pitch target level, e.g. (0.5) results in a pitch target of 50 percent of the possible maximum pitch at this point in the sentence;
- numbers = duration in 100th of a second;
- { } = maximum pitch at start and end, followed by minimum pitch at start and end;
- ax = schwa.

(18) Prosodic annotation example (Pierrehumbert, 1981, p. 989)

Sentence: In November the region's weather was unusually dry.

Phonetic transcription: ihn nowvehmber dhax riyjhenz wehdher waxz axnyuzhaxliy drai

Encoded as: {325 260 195 165}(0.4) SIL 8 ih 4 n 3* n 4 ow 7 v 9 eh 10(1.0) m 8 b 4 er 13% (0.3) SIL 34 *dh 2 ax 2* r 5 iy 7(0.6) jh 10 en 8 z 10* w 3 eh 7 (0.8) dh 4 er 8* w 6 ax 5 z 7* ax 4 n 9 yu 16 (0.8) zh 6 ax 5 l 4 iy 9* d 8 r 5 ai 24 (1.0)%

Prosodic annotation either requires a recording of the sentence – the method used in Pierrehumbert (1981) – or a segment, syllabe or word dictionary that the parser can access and a system of rules for adding prosody. The latter is used in SPRUCE. However, Pierrehumbert's encoding system, and with that SPRUCE as well, does not allow pitch range control for individual phrases, which is needed for the model to be applicable to parentheticals. The lack of pitch range control can be seen in (18) above. The topline and baseline values can only be specified in the curly brackets preceding the sentence, which sets the topline and baseline pitch values for the start of the sentence as well as the end. Thus, there is no mechanism for sentence-internal pitch range control, which is required for synthesizing parentheticals.

5.3 Encoding using markup

Markup-based encoding schemes, such as JSML (Hunt, 2000) or SABLE (Sproat et al., 1998; Sproat & Raman, 1999; "SABLE," n.d.), do not require annotation of complicated phonetic details as discussed in the previous section. Rather, they allow user-friendly annotation that the system is going to interpret phonetically during the synthesizing process. An example for this is to specify that a marked-up text is to be perceived as an invitation by using <invite> text </invite> (Mertens, 2002). As described in chapter two, the voice browser then converts such functional tags into prosodic tags that can be interpreted by the synthesizer. Thus, the system converts the invitation into what that means phonetically and the phonetic markup is then converted into an acoustic output by the synthesizer, which is reinterpreted as an invitation by the hearer. Marking up allows one to easily annotate text without explicit phonetic knowledge as long as the tag is known to the system.

A couple of markup schemes have been discussed in chapter two. Of these, SABLE is the most appropriate because it provides a user-friendly annotation scheme that allows to capture the parameters involved with parentheticals, such as pitch range.

The discussion of how the findings of this study can be integrated into SABLE to synthesize parenthetical sentences is divided into two parts. First in section 5.3.1 the sentences with one parenthetical phrase only are discussed. Then in section 5.3.2 the sentences with more than one parenthetical phrase are discussed.

5.3.1 Single parentheticals

Following the analysis results from chapter four, the phrase preceding the parenthetical exhibits a lower baseline and with that a wider pitch range than an identical phrase that is at the same sentence position but not pre-parenthetical, i.e. not directly followed by a parenthetical. The pitch range increase for females is 17 Hz and that for males is 14 Hz, which corresponds to an increase of 20 percent for both female and males. This can easily be marked up with <PITCH RANGE="+20%"> text </PITCH>.

Although both female and male baselines drop about 17-18 Hz, SABLE does not allow to specify *<PITCH BASE="-18 Hz">. Rather, the drop has to be expressed as a percentage. Unfortunately, with respect to percentages the baseline drop is not identical for females (-10%) and males (-16%). To avoid cumbersome gender-specific annotation, a percentage between 10 and 16 has to be found that works for both genders. Otherwise, if the gender of the speaker specified with <SPEAKER GENDER="male"> at the beginning of the text were changed later, all pre-parenthetical baseline specification would have to be changed by hand as well.

Choosing the median between 10 percent and 16 percent leads to 13 percent. However, a baseline drop of 13 percent entails a 6 Hz larger drop for females (-24 Hz instead of -18 Hz), whereas the male baseline drops about 3 Hz less (-14 Hz instead of -17 Hz). An even distribution of drop deviation is achieved when 12 percent is chosen. Then the female baseline only drops an additional 4 Hz (-22 Hz instead of -18 Hz), whereas the male baseline drops about 4 Hz less (-13 Hz instead of -17 Hz). According to Hess (1983), an F0 difference of 3 Hz or less in natural speech is not noticeable. This means that the deviation of 4 Hz is noticeable. However, the deviation is not significant enough to counterweigh ease of annotation.

Using SABLE's pitch tags, a 12 percent lower baseline and 20 percent wider pitch range for pre-parenthetical fr(a) can be easily marked up as:

(19) Pre-parenthetical pitch markup example

<PITCH BASE="-12%" RANGE="+20%"> We saw the movie, </PITCH> which had been banned in Boston, but were unimpressed.

Nevertheless, pitch tags still require some phonetic knowledge by the author of the text. To avoid this disadvantage, this thesis proposes a new type for the structure tag DIV. The new type is called 'prepar' and marks up pre-parenthetical phrases as in:

(20) Pre-parenthetical division markup example

<DIV TYPE="prepar"> We saw the movie, </DIV> which had been banned in Boston, but were unimpressed.

As discussed in chapter one, the voice browser interprets the marked up document and converts all tags into SABLE speech production markup using style sheets. Thus, the voice browser converts the division markup into the pitch markup as presented in (19). For this to work, the style sheets have to include what the value "prepar" corresponds to in terms of pitch.

This section discussed two possibilities for marking up sentences with parentheticals to improve synthesized speech. The first is marking up the pre-parenthetical phrase with pitch specifications directly. The second option is to mark up the pre-parenthetical phrase with a unique tag and supply the pitch specifications that are associated with this tag through style sheets. Due to ease of annotation, this thesis proposes the latter option as the way of marking up sentences containing one parenthetical phrase.

So far, the tagging of fr(a) as the pre-parenthetical phrase has been discussed, which is all that needs to be marked up in sentences containing one parenthetical phrase only. In sentences with more than one parenthetical phrase adjacent to each other, however, some of the parenthetical phrases have to be marked up as well. The details of this are the focus of the following section.

5.3.2 Multi-parentheticals

This study investigates sentences with two sequential or nested parentheticals. Of these, sequential two-parentheticals have two parenthetical parts. Nested two-parentheticals have three parenthetical parts since the outer parenthetical phrase is broken up into two parts, par1(a) and par1 (b), as in:

(21) Nested two-parenthetical example

“We saw the movie – Jane (who knows the director) insisted on going – but were unimpressed.”

| fr(a) | par1(a) | par2 | par1(b) | fr(b) |

As has been shown in chapter four, the parenthetical phrase directly preceding fr(b) in multi-parenthetical sentences behaves like the parenthetical phrase of single-parenthetical sentences. Furthermore, the trend for each additional parenthetical part in multi-parenthetical sentences is that the further away it is from fr(b), the lower is the topline, higher the baseline and smaller the range. This relationship is displayed in Figure 5.1 for females and Figure 5.2 for males. Each figure includes both sequential and nested values to show the general trend of two-parentheticals. In each figure, the bar represents the pitch range of each sentence part. The pitch range is delimited by the baseline at the bottom and the topline at the top. Since parenthetical phrases in nested and sequential two-parentheticals are not exactly the same, some bars feature short vertical lines at the top or bottom ends to indicate the difference.

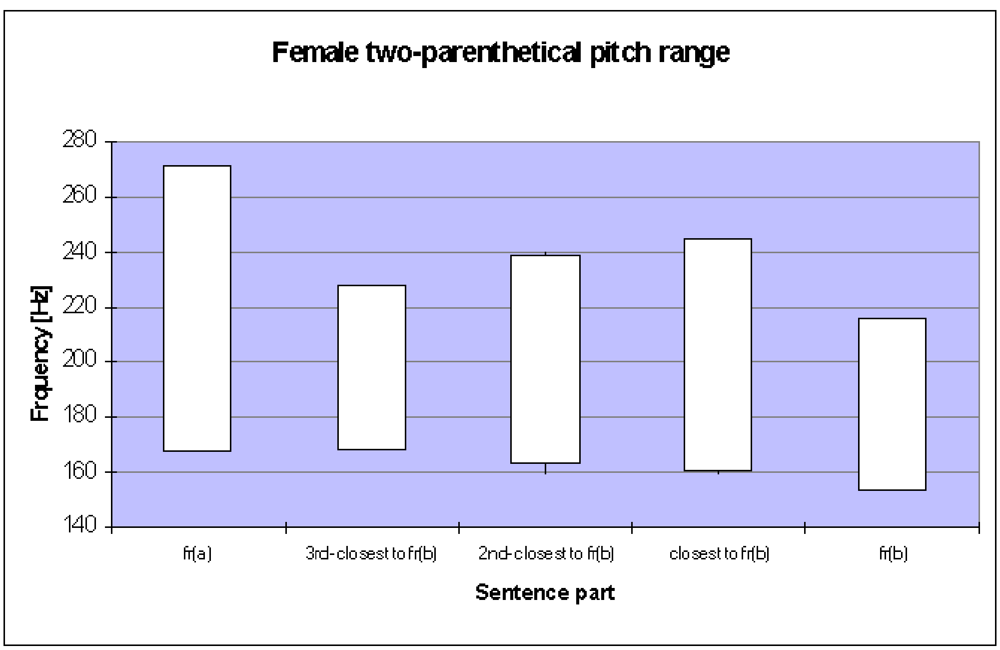


Figure 5.1. Female pitch range trend in sequential and nested two-parenthetical sentences.

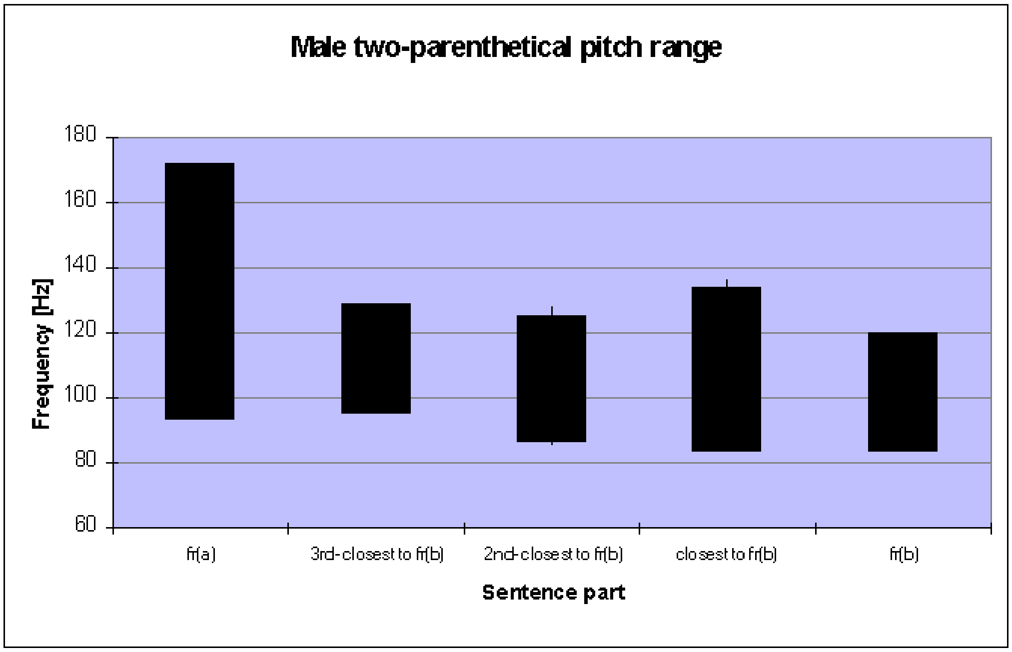


Figure 5.2. Male pitch range trend in sequential and nested two-parenthetical sentences.

The parenthetical part closest to fr(b) does not have to be marked up, because as discussed in chapter four, the parenthetical part directly preceding fr(b) is not acoustically different from a non-parenthetical phrase at the same sentence position. However, the parenthetical parts that are second (penultimate) and third closest to fr(b) (antepenultimate) need to be marked up. The purpose of this section is to introduce pitch specifications as well as appropriate tags for these parenthetical parts.

Unfortunately, all non-parenthetical sentences have only one medial phrase and with that three data points in total, namely fr(a), mid and fr(b). In contrast, sentences with nested parentheticals have five sentence parts and with that five data points. For these, there are no corresponding non-parenthetical sentences included in the set of sentences that was used in the experiment. The need for such sentences was not foreseen prior to the experiment. As a consequence, the analysis in chapter four could not provide a direct comparison between the acoustics of the antepenultimate and penultimate parenthetical and their non-parenthetical counterparts. This means that the specifications required for synthesis need to be interpolated from the existing data.

First in section 5.3.2.1, I will interpolate missing non-parenthetical baseline pitch values that the parenthetical values are compared to. Similarly, in section 5.3.2.2 I will interpolate missing non-parenthetical pitch range values that the parenthetical values are compared to. Section 5.3.2.3 discusses how these findings can be integrated into SABLE markup to improve synthesizing parentheticals.

5.3.2.1 Baseline comparison

The non-parenthetical data that is available shows that the baseline trend is a steep drop between fr(a) and the following non-parenthetical phase, while the drop between the non-parenthetical phrase and fr(b) is much smaller. This relationship is displayed in Figure 5.3.

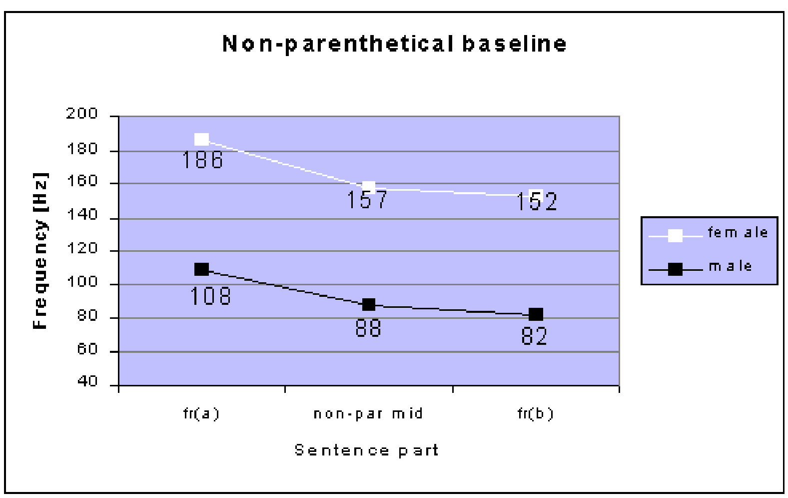


Figure 5.3. Development of the non-parenthetical baseline over the course of the sentences for both females and males.

Figure 5.3 shows that each baseline trend resembles a falling curve that can be decribed by a quadratic equation. The next step is to find the functions that best describe the curves. The missing data points can then be determined through this function. This assumes that the values for fr(a), the ultimate non-parenthetical and fr(b) do not change when inserting further non-parenthetical phrases. The assumptions is based on the observation that the values for fr(a), the ultimate parenthetical and fr(b) also do not change in parenthetical sentences when further parenthetical phrases are inserted.

To find the function, the sentence part labels have to be converted to numbers, since the function requires a numeric input for x. Therefore, on the x-axis the label fr(a) is replaced with “1,” the label non-par mid is replaced with “2” and the label fr(b) is replaced with “3.”

5.3.2.1.1 Female baseline

The female baseline has the data points 186, 157 and 152. Hence, the function has to yield 186 for $x = 1$, 157 for $x = 2$ and 152 for $x = 3$.

The final and lowest data point, 152, represents the displacement of the curve from the x-axis. Hence, for $x = 3$ the function has to yield 152, which is achieved through $y = (3 - x)^2 + 152$. In the function, *x* is the placeholder for the number of the sentence part as the variable and *y* is the placeholder for the resulting pitch value. The function as it is yields:

Table 5.1.

Results for the preliminary function for the non-parenthetical female baseline
[$y = (3 - x)^2 + 152$].

X	1	2	3
Y	156	153	152

What the function is still missing is a multiplier that stretches the curve out to reach up to 186 for x = 1 and 157 for x = 2. Using different values as the multiplier, the results closest to the measured data points are displayed in Table 5.2. Figure 5.4 provides a graphical representation of the differences of the values that the functions yield to the measured data points.

Table 5.2.

The three closest-fit non-parenthetical female baseline functions. All values are in Hertz.

	x = 1	x = 2	x = 3	
$y = 7 * (3 - x)^2 + 152$	180	159	152	
$y = 8 * (3 - x)^2 + 152$	184	160	152	<== best fit!
$y = 9 * (3 - x)^2 + 152$	188	161	152	
measured values	186	157	152	

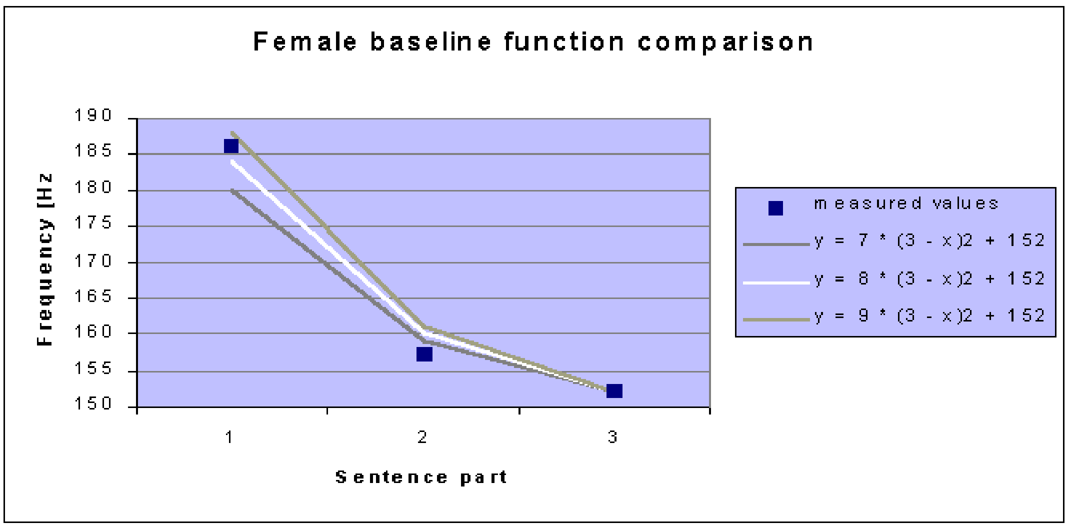


Figure 5.4. The effect of the use of different multipliers on the non-parenthetical female baseline function.

The function that describes the female baseline curve the best is $y = 8 * (3 - x)^2 + 152$. It fits best because no deviation from the measured values is larger than 3 Hz and with that below the perceptual threshold (Hess, 1983). As can be seen in Table 5.2, using the multiplier 7 results in a too-large deviation of 6 Hz for x = 1 and using the multiplier 9 results in a too-large deviation of 4 Hz for x = 2.

The missing data points are determined by entering values between 1 and 2 for x. Sentences with sequential two-parentheticals have four data points. Hence, the non-parenthetical equivalent needs to have four data points as well. This can be achieved by inserting the missing data point between 1 (fr(a)) and 2 (ultimate non-par mid). Since the acoustic analyses in this study do not take phrase length into account but only sequential order, the missing non-parenthetical phrase between 1 and 2 can be assumed to be equidistant 1.5. Plugging 1.5 into the function yields $y = 8 * (3 - 1.5)^2 + 152 = 170$.

The comparison between the female sequential two-parenthetical sentence and the corresponding two-non-parenthetical sentence is displayed in Table 5.3. The table shows that the penultimate parenthetical is 7 Hz (4 percent) lower than the interpolated non-parenthetical phrase at the same position.

Table 5.3.

Difference between female sequential two-parenthetical sentences and corresponding non-parenthetical sentences. Unless marked otherwise, all values are in Hertz.

	fr(a) = 1	**penultimate mid = 1.5**	**ultimate mid = 2**	**fr(b) = 3**
female baseline				
non-parenthetical	184	170	160	152
sequential two-parenthetical	167	163	160	153
difference in Hertz		-7		
difference in %		-4%		

Nested two-parentheticals have five sentence parts and with that two additional data points between 1 and 2, one more than sequential two-parentheticals have. To determine the shape of corresponding sentences with non-parenthetical phrases, the missing two data points in the non-parenthetical sentences can be assumed at equidistant 1.33 and 1.66. Plugging 1.33 into the function yields $y = 8 * (3 - 1.33)^2 + 152 = 174$ and plugging 1.66 into the function yields $y = 8 * (3 - 1.66)^2 + 152 = 166$.

The comparison between the female nested two-parenthetical sentences and the corresponding two-non-parenthetical sentences is displayed in Table 5.4. The table shows that the antepenultimate parenthetical is 6 Hz (3 percent) lower than the interpolated non-parenthetical phrase at the same position and the penultimate parenthetical is 7 Hz (4 percent) lower than its corresponding non-parenthetical phrase.

Table 5.4.

Difference between female nested two-parenthetical sentences and corresponding non-parenthetical sentences. Unless marked otherwise, all values are in Hertz.

	fr(a) = 1	**antepenultimate mid = 1.33**	**penultimate mid = 1.66**	**ultimate mid = 2**	**fr(b) = 3**
female baseline					
non-parenthetical	184	174	166	160	152
nested two-parenthetical	167	168	159	159	153
difference in Hertz		-6	-7		
difference in %		-3%	-4%		

5.3.2.1.2 Male baseline

The male non-parenthetical baseline has the data points 108, 88 and 82. The function that describes this curve most accurately is determined using the same method as for the females. First, the function that yields 82 for x = 3 is determined as $y = (3 - x)^2 +$

82, then the multiplier has to be found. Using different values as the multiplier, the results closest to the measured data points are displayed in Table 5.5.

Table 5.5.

The three closest-fit male baseline functions. All values are in Hertz.

	x = 1	**x = 2**	**x = 3**	
$y = 5 * (3 - x)^2 + 82$	102	87	82	
$y = 6 * (3 - x)^2 + 82$	106	88	82	<== best fit!
$y = 7 * (3 - x)^2 + 82$	110	89	82	
measured values	108	88	82	

The function that describes the male baseline curve the best is $y = 6 * (3 - x)^2 + 82$. It fits best because only x = 1 deviates by about 2 Hz, which is better than what the surrounding two functions yield.

Sequential two-parenthetical sentences feature four data points. Hence, the non-parenthetical equivalent needs to have four data points as well, which can be achieved by inserting the missing data point at 1.5. Plugging 1.5 into the function yields $y = 6 * (3 - 1.5)^2 + 82 = 96$.

The comparison between the male sequential two-parenthetical sentence and the corresponding two-non-parenthetical sentence is displayed in Table 5.6. The table shows that the penultimate parenthetical is 10 Hz (10 percent) lower than the interpolated non-parenthetical phrase at the same position.

Table 5.6.

Difference between male sequential two-parenthetical sentences and corresponding non-parenthetical sentences. Unless marked otherwise, all values are in Hertz.

	fr(a) = 1	**penultimate mid = 1.5**	**ultimate mid = 2**	**fr(b) = 3**
female baseline				
non-parenthetical	106	96	88	82
sequential two-parenthetical	93	86	83	83
difference in Hertz		-10		
difference in %		-10%		

Nested two-parentheticals have five sentence parts and with that two additional data points between 1 and 2, which can be assumed to be at equidistant 1.33 and 1.66. Plugging 1.33 into the function yields $y = 6 * (3 - 1.33)^2 + 82 = 99$ and plugging 1.66 into the function yields $y = 6 * (3 - 1.66)^2 + 82 = 93$.

The comparison between the male nested two-parenthetical sentences and the corresponding two-non-parenthetical sentences is displayed in Table 5.7. The table shows that the antepenultimate parenthetical is 4 Hz (4 percent) lower than the interpolated non-parenthetical phrase at the same position and the penultimate parenthetical is 8 Hz (9 percent) lower than its corresponding non-parenthetical phrase.

Table 5.7.

Difference between male nested two-parenthetical sentences and corresponding non-parenthetical sentences. Unless marked otherwise, all values are in Hertz.

	fr(a) = 1	antepenultimate mid = 1.33	penultimate mid = 1.66	ultimate mid = 2	fr(b) = 3
male baseline					
non-parenthetical	106	99	93	88	82
nested two-parenthetical	93	95	85	89	83
difference in Hertz		-4	-8		
difference in %		-4%	-9%		

5.3.2.1.3 Baseline summary

The data discussed in 5.3.2.1 show that for both female and male baselines the antepenultimate parenthetical is 4-6 Hz (3-4 percent) lower and the penultimate parenthetical is 7-10 Hz (4-9 percent) lower than non-parenthetical phrases at the same sentence position. This means that, on average, the antepenultimate parenthetical has a four percent lowered baseline and the penultimate parenthetical has a seven percent lowered baseline. These are the specifications with regard to baseline with which sentences containing multi-parentheticals have to be implemented in synthesized speech. Still missing are the specifications for pitch range, which are the focus of the following section.

5.3.2.2 Pitch range comparison

Like with the baselines, there are no pitch range data available for sentences with more than one medial non-parenthetical phrase. Thus, the pitch range values for these multi-non-parenthetical sentences have to be interpolated from the available data as well.

Looking at the non-parenthetical data that is available shows that the pitch range trend for females exhibits no drop between fr(a) and the medial non-parenthetical phrase and then there is a steep drop between the medial phrase and fr(b). Opposite to that, the trend for males shows a steeper drop between fr(a) and the medial non-parenthetical phrase than between the medial phrase and fr(b). Hence, the male pitch range trend corresponds to the baseline trends. The pitch range trends are displayed in Figure 5.5.

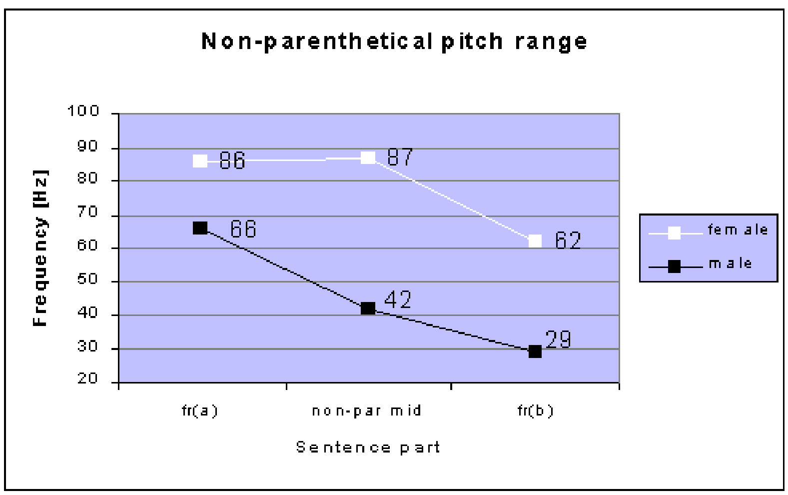


Figure 5.5. Development of the non-parenthetical pitch range over the course of the sentences for both females and males.

5.3.2.2.1 Female pitch range

As can be seen in Figure 5.5. above, females exhibit no pitch range drop between fr(a) and the medial non-parenthetical phrase. Unfortunately, there is no explicit acoustic data available to show how the pitch range progresses over the course of sentences with several medial non-parenthetical phrase. The observation that fr(a) and the medial non-parenthetical phrase have the same pitch range leads to the assumption that for females all non-parenthetical phrases except fr(b) exhibit the same pitch range as fr(a). This differs from the males, who exhibit a pattern of continual pitch range drops over the course of the sentence. However, female maintenance of pitch range over multiple phrases has also been observed in French (Fagyal, 2002). The study by Fagyal shows that females maintain the pitch range width of the preceding phrase also for a following parenthetical. Pitch range width only narrows for the final phrase. The same did not occur in the French male data, since male pitch range has been measured to drop for the parenthetical. Interestingly, the French female pattern is the opposite of what has been observed for females in English. In French, the maintenance of the larger pitch range occurs for parentheticals, whereas non-parenthetical pitch range drops. Besides the opposite patterning, the French data show that an observation of females maintaining pitch range over a set of phrases is not exclusive to this study.

Following the assumption that any additionally inserted non-parenthetical phrases have a pitch range between the value of the preceding and the following phrase, the additional phrases feature the same pitch range of 86-87 Hz for females in English. Thus, there is no interpolation required for female pitch ranges in two-parenthetical sentences.

The comparison between female sequential and nested two-parenthetical sentences is displayed in Table 5.8. The table shows that the antepenultimate parenthetical is 26 Hz (30 percent) lower and the penultimate parenthetical is 7-10 Hz (8-11 percent) lower than the non-parenthetical phrases at the same sentence position.

Table 5.8

Difference between female sequential and nested two-parenthetical sentences and corresponding non-parenthetical sentences. Unless marked otherwise, all values are in Hertz.

	fr(a) = 1	**penultimate mid = 1.5**	**ultimate mid = 2**	**fr(b) = 3**
female pitch range				
non-parenthetical	86	87	87	62
sequential two-parenthetical	103	77	84	63
difference in Hertz		-10		
difference in %		-11%		

	fr(a) = 1	**antepenultimate mid = 1.33**	**penultimate mid = 1.66**	**ultimate mid = 2**	**fr(b) = 3**
female pitch range					
non-parenthetical	86	86	87	87	62
nested two-parenthetical	104	60	80	86	63
difference in Hertz		-26	-7		
difference in %		-30%	-8%		

5.3.2.2.2 Male pitch range

The male non-parenthetical pitch range curve has the data points 66, 42 and 29. The function that yields 29 for $x = 3$ is $y = (3 - x)^2 + 29$. Using different values as the multiplier, the results closest to the measured data points are displayed in Table 5.9.

Table 5.9

The three closest-fit male pitch range functions. All values are in Hertz.

	x = 1	**x = 2**	**x = 3**	
$y = 9 * (3 - x)^2 + 29$	65	38	29	
$y = 10 * (3 - x)^2 + 29$	69	39	29	<== best fit!
$y = 11 * (3 - x)^2 + 29$	73	40	29	
measured values	66	42	29	

The function that describes the male pitch range curve the best is $y = 10 * (3 - x)^2 + 29$. It fits best because no deviation from the measured values is larger than 3 Hz, the perceptual threshold (Hess, 1983).

Sequential two-parenthetical sentences feature four data points. Hence, the non-parenthetical equivalent needs to have four data points as well, which can be achieved by inserting the missing data point at 1.5. Plugging 1.5 into the function yields $y = 10 * (3 - 1.5)^2 + 29 = 52$.

Nested two-parentheticals have five sentence parts and with that two additional data points between 1 and 2. Plugging 1.33 into the function for the antepenultimate medial phrase yields $y = 10 * (3 - 1.33)^2 + 29 = 57$ and plugging 1.66 into the function for the penultimate medial phrase yields $y = 10 * (3 - 1.66)^2 + 29 = 47$.

The comparison between male sequential and nested two-parenthetical sentences is displayed in Table 5.10. The table shows that the antepenultimate parenthetical is 23 Hz (40 percent) lower and the penultimate parenthetical is 7-10 Hz (15-19 percent) lower than the non-parenthetical phrases at the same sentence position.

Table 5.10

Difference between male sequential and nested two-parenthetical sentences and corresponding non-parenthetical sentences. Unless marked otherwise, all values are in Hertz.

	fr(a) = 1	**penultimate mid = 1.5**	**ultimate mid = 2**	**fr(b) = 3**
male pitch range				
non-parenthetical	69	52	39	29
sequential two-parenthetical	79	42	53	36
difference in Hertz		-10		
difference in %		-19%		

	fr(a) = 1	**antepenultimate mid = 1.33**	**penultimate mid = 1.66**	**ultimate mid = 2**	**fr(b) = 3**
male pitch range					
non-parenthetical	69	57	47	39	29
nested two-parenthetical	79	34	40	53	36
difference in Hertz		-23	-7		
difference in %		-40%	-15%		

5.3.2.2.3 Pitch range summary

The data discussed in 5.3.2.2 show that for both female and male pitch ranges the antepenultimate parenthetical is 23-26 Hz (30-40 percent) lower and the penultimate parenthetical is 7-10 Hz (8-19 percent) lower than non-parenthetical phrases at the same sentence position. This means that, on average, the antepenultimate parenthetical has a 35 percent lowered pitch range and the penultimate parenthetical has a 13 percent lowered pitch range. How these specifications and the baseline specifications are used for synthesizing parentheticals is the focus of the next section.

5.3.2.3 Synthesizing multi-parentheticals

The previous sections 5.3.1 and 5.3.2 discuss pitch range and baseline specifications for sentences with multi-parenthetical medial phrases. The specifications are not derived from actual data but through mathematical interpolation from the incomplete existing data. Thus, these specifications have to be regarded as preliminary but nonetheless trend describing. This section discusses the integration of the specifications for multi-parenthetical sentences into the SABLE text-to-speech system.

Like for single parentheticals, there are two ways to mark up these specifications in SABLE. The first is the use of the pitch markup and the second is the use of structure tags for the different kinds of parenthetical phrases.

Translating the baseline and pitch range specifications into pitch markup is exemplified by (22) for sequential parentheticals and by (23) for nested parentheticals.

(22) Sequential parenthetical pitch markup example with sentence 19

fr(a): <PITCH BASE="-12%" RANGE="+20%"> We saw the movie </PITCH>

par1: <PITCH BASE="-7%" RANGE="-13%"> (which had been banned in Boston) </PITCH>

par2: – Jane insisted on going –

fr(b): but were unimpressed.

(23) Nested parenthetical pitch markup example with sentence 18

fr(a): <PITCH BASE="-12%" RANGE="+20%"> We saw the movie </PITCH>

par1(a): <PITCH BASE="-4%" RANGE="-35%"> – Jane </PITCH>

par2: <PITCH BASE="-7%" RANGE="-13%"> (who knows the director) </PITCH>

par1(b): insisted on going –

fr(b): but were unimpressed.

More user-friendly than pitch markup is the introduction of two more structure tags. Again, these tags are expressed as additional types of DIV, namely *antepenultpar* for antepenultimate parentheticals and *penultpar* for penultimate parentheticals. Pitch specifications for the conversion of these values into phonetic markup through the browser are included in the style sheets. Using the same sentences as above results in (24) for sequential parentheticals and (25) for nested parentheticals.

(24) Sequential parenthetical structure markup example with sentence 19

fr(a): <DIV TYPE="prepar"> We saw the movie </DIV>

par1: <DIV TYPE="penultpar"> (which had been banned in Boston) </DIV>

par2: – Jane insisted on going –

fr(b): but were unimpressed.

(25) Nested parenthetical structure markup example with sentence 18

fr(a): <DIV TYPE=“prepar”> We saw the movie </DIV>

par1(a): <DIV TYPE=“antepenultpar”> – Jane </DIV>

par2: <DIV TYPE=“penultpar”> (who knows the director) </DIV>

par1(b): insisted on going –

fr(b): but were unimpressed.

This section shows that sentences containing multi-parenthetical phrases can easily be marked up with structure tags that entail the pitch specifications determined by this study. This helps to improve the naturalness of parentheticals in synthesized speech.

5.4 Summary

Chapter five identifies SABLE as the most useful and user-friendly text-to-speech system for synthesizing sentences that contain parentheticals. Within the SABLE markup scheme, structure tags allow easy mark up of the phrases that require specific implementation. The pitch specifications entailed by the structure tags have been taken from chapter four for single-parentheticals and specifications for multi-parentheticals have been established by this chapter. Furthermore, chapter five provides examples of how the required markup is integrated into sentences. Thus, this chapter provides a guide to improve the naturalness of sentences containing parentheticals in synthesized speech.

Chapter Six

CONCLUSION

This chapter is organized as follows: the findings of this thesis are summarized in section 6.1 and section 6.2 presents limitations of the findings of this study in conjunction with recommendations for future studies.

6.1 Findings

This study investigates the effect of the use of different punctuation marks on the intonation of parenthetical phrases, as well as the phonetic specifications necessary for synthesizing natural-sounding parentheticals.

The notion of lower pitch level and narrower pitch range for parentheticals has been reported in many studies (Bolinger, 1989; Cruttenden, 1997; Crystal, 1969; Grosz & Hirschberg, 1992; O’Shaughnessy, 1990; Wichmann, 2000). However, only one acoustic study on parentheticals (Kutik et al., 1983) could be found that actually provides evidence for such claims. Furthermore, there has been no study on a possible connection between punctuation and intonation for parentheticals, although some punctuation marks, such as question marks, have a direct connection to intonation. Nunberg (1990) claims that, despite some overlap, punctuation is a system that not only deviates from but also goes far beyond intonation. Therefore, it is a linguistic system in its own right.

Reviewing the literature on text-to-speech systems shows that most text-to-speech systems do not provide a mechanism to implement parentheticals distinctively. One of the major challenges for text-to-speech systems, however, is achieving naturalness of the speech synthesis output. This study has identified the text-to-speech system SABLE (Sproat et al., 1998; Sproat & Raman, 1999) as the most promising for producing natural-sounding parentheticals through the use of text markup.

To determine the exact pitch specifications needed to synthesize parentheticals, an experiment has been carried out in which three female and three male native speakers of Canadian English read aloud a set of 20 parenthetical and non-parenthetical sentences. The sentences were recorded and acoustically analyzed with a speech analysis software.

The analysis revealed that the use of different punctuation marks for parentheticals has no acoustic correspondence. This provides evidence for Nunberg’s (1990) claim that punctuation is not directly tied to intonation but is a linguistic system in its own right. The only acoustic effect of punctuation found by this study is that parenthetical punctuation triggers pausing. However, there is no correlation between pause length and type of punctuation.

Following the literature, the distinctiveness of parentheticals is based on a change in maximum (topline) and minimum (baseline) pitch levels, and with that a change in pitch range, over a phrase. This study involved a direct comparison of sentences containing one parenthetical phrase with sentences containing a non-parenthetical phrase at the same position. The comparison revealed that the parenthetical and non-parenthetical phrases do not differ in topline, baseline or pitch range. However, it has been found that the phrase preceding a parenthetical is spoken with a 12 percent lower baseline and – as a consequence of that – a 20 percent wider pitch range than the phrase at the same position in a non-parenthetical sentence. Thus, for synthesizing sentences containing one parenthetical phrase, it is not the parenthetical phrase itself but the preceding phrase that has to be marked up with specifications differing from the expected pitch change trends over the course of a sentence.

In sentences containing two parentheticals in sequence or nested, i.e. one embedded within the other, the preceding non-parenthetical phrase also has to be marked up with the same specifications as in sentences containing just one parenthetical phrase. Additionally, it has been found that in sentences containing more than one parenthetical constituent next to each other, the ultimate, i.e. last, parenthetical constituent behaves like the parenthetical phrase in a one-parenthetical sentence as discussed above. Therefore, the ultimate parenthetical constituent does not need to be marked up. However, the penultimate and antepenultimate parenthetical constituents each feature distinctive pitch specifications. The penultimate parenthetical constituent has been found to have a seven percent lower baseline and a 13 percent narrower pitch range than a non-parenthetical phrase would have at the same position in the sentence. Similarly, the antepenultimate parenthetical constituent has been found to have a four percent lower baseline and a 35 percent narrower pitch range than a non-parenthetical phrase would have at the same position in the sentence. As discussed in chapter five, a variety of tags have been proposed to integrate these findings into the SABLE markup scheme.

Kutik et al.'s findings of a lowered topline for parentheticals could not be replicated by this study. Whereas previous research claims that the acoustic implementation of the parenthetical phrase itself sets it apart from the rest of the sentence, this study has shown that the parenthetical phrase does not differ in pitch level and range from a non-parenthetical phrase at the same sentence position. Rather, it is the preceding phrase that is implemented with a lower baseline and a resulting wider pitch range. This has the effect that, perceptually, there is a steeper pitch range drop between the pre-parenthetical phrase and the following parenthetical than between the corresponding phrases in a non-parenthetical sentence. The steeper pitch range drop explains why previous studies report that parentheticals are lower in pitch level and narrower in range, because in comparison to the preceding phrase in the same sentence, they perceptually are. However, a direct comparison to a non-parenthetical phrase at the same sentence position reveals that it is not the parenthetical itself that is implemented differently to evoke this perception, but the preceding phrase. Hence, the findings of this study do not contradict the general notion about parentheticals as reported elsewhere but complete the picture by providing exact pitch specifications of how the perceptual parenthetical characteristics are achieved.

The research included in this thesis has provided evidence for Nunberg’s claim, as well as new insights into the acoustic implementation of parentheticals. The specifications for parentheticals presented in this thesis are important for implementing sentences containing parentheticals in text-to-speech systems. Improving the naturalness of synthesized speech is still a key point in their development. The findings of this study can be integrated in the SABLE markup scheme. Therefore, this study contributes to research on synthesized speech in speech technology applications.

6.2 Future studies

The present study is the first to analyze multi-parentheticals and investigate how synthesizing parentheticals can be improved. Furthermore, it is also the first to investigate a connection between punctuation and intonation in parentheticals. However, further investigation is required to look into parenthetical issues more comprehensively and include other concerns that this study did not address.

Although text-to-speech pitch specifications for sentences with more than one parenthetical phrase have been established by this study, these specifications require support through a future study. The problem is that the multi-parenthetical sentences investigated in this study have up to three medial parenthetical constituents (five data points over the course of the sentence, including the frame sentence constituents), whereas all non-parenthetical sentences have only one medial phrase and with that three data points. To enable a phrase-to-phrase pitch comparison, pitch values for missing non-parenthetical data points are interpolated using a function that is derived from the existing three data points and that describes the declination of pitch over the course of the sentence. However, to establish a more reliable foundation for multi-parenthetical pitch specifications, future investigators are recommended to record sentences containing multi-parentheticals as well as corresponding sentences containing equivalent non-parenthetical phrases. A direct phrase-to-phrase comparison of measured data will provide more accurate specifications than the interpolation-based comparison that this study used.

Similar trends for female and male speakers were observed in this study. However, only three subjects of each gender were included in this study, which makes the results less significant. Future studies are recommended to recruit a larger number of experiment participants – large enough to make the results representative of the population.

Although a variety of sentences has been used in the experiment, primarily relative-clause parentheticals are considered. To achieve a comprehensive picture of the pitch specifications of parentheticals, other parentheticals, such as adverbial or elliptical clause parentheticals (Wichmann, 2000), should also be looked at in future studies.

Following previous research, there is no evidence for a difference in intonation between parentheticals based on the type of syntactic structure parentheticals are inserted in. Whether the different topline/baseline findings in this study and in Kutik et al. (1983) depend on the circumstance that both studies chose to insert parentheticals into different types of syntactic structure – that is, insertion between clauses (this study) as opposed to insertion between phrases (Kutik et al.) – cannot be answered by this study and has to be left to future research.

This study investigated pausing as an additional acoustic cue to differentiate parentheticals. There are further measures that previous research has not identified as cues for parentheticals but which should be considered for future investigations. These measures are amplitude as well as mean pitch and variance.

REFERENCES

Altenberg, B. (1987). Prosodic patterns in spoken English: Studies in the correlation between prosody and grammar for text-to-speech conversion. *Lund Studies in English 76*. Lund, Sweden: Lund University Press.

Avis, W. S., Drysdale, P. D., Gregg, R. J., Neufeldt, V. E., & Scargill, M. H. (Eds.). (1983). *Gage Canadian dictionary.* Toronto, Canada: Gage Publishing Limited.

Bayraktar, M., Say, B., & Akman, V. (1998). An analysis of English punctuation: The special case of comma. *International Journal of Corpus Linguistics, 3,* 33-57. Retrieved April 2002 from http://cogprints.ecs.soton.ac.uk/archive/00000214/00/ijcl.ps

Blauw, E. (1992). Phonetic differences between read and spontaneous speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 92*, 751- 754.

Bolinger, D. (1975). *Aspects of language.* New York: Harcourt Brace Jovanovich.

Bolinger, D. (1986). *Intonation and its parts.* Stanford, CA: Stanford University Press.

Bolinger, D. (1989). *Intonation and its uses.* Stanford, CA: Stanford University Press.

Briscoe, T. (1994). *Parsing (with) punctuation etc.* (Technical report MLTT-TR-002). Grenoble, France: Rank Xerox Research Centre. Retrieved April 2002 from http://citeseer.nj.nec.com/briscoe94parsing.html

Briscoe, T. (1996). The syntax and semantics of punctuation and its use in interpretation. *Proooceedings of the Association for Computational Linguistics (ACL) Workshop on Punctuation*, 1-8. Retrieved April 2002 from http://www.citeseer.nj.nec.com/briscoe96syntax.html

Carroll, J., Briscoe, T., & Sanfilippo, A. (1998). Parser evaluation: A survey and a new proposal. *Proceedings of the 1st International Conference on Language Resources and Evaluation*, 447-454. Retrieved May 2002 from http://www.cogs.susx.ac.uk/lab/nlp/carroll/abs/98cbs.html

Chafe, W. (1987a). *Punctuation and the prosody of written language.* (Technical Report No.11). Berkeley, CA: Center for the Study of Writing. Retrieved May 2002 from http://www.writingproject.org/downloads/csw/TR11.pdf

Chafe, W. (1987b). *What good is punctuation?* (Occasional Paper No.2). University of California: Center for the Study of Writing. Retrieved May 2002 from http://www.writingproject.org/downloads/csw/OP02.pdf

Childers, D., Cox, R. V., DeMori, R., Furui, S., Juang, B.-H., Mariani, J. J., et al. (1998, May). The past, present, and future of speech processing. *IEEE Signal Processing Magazine*, *15*(3), 24-48.

Clark, R. A. J. (1999). Using prosodic structure to improve pitch range variation in text-to-speech synthesis. *Proceedings of the International Congress of Phonetic Sciences (ICPhS 99)*, 69-72. Retrieved June 2002 from http://archive.ling.ed.ac.uk/documents/disk0/00/00/00/82/index.html

CP Stylebook. (1984). Toronto, Canada: The Canadian Press.

Cruttenden, A. (1997). *Intonation*. New York: Cambridge University Press.

Crystal, D. (1969). *Prosodic systems and intonation in English*. New York: Cambridge University Press.

Daly, N. A., & Zue, V. W. (1992). Statistical and linguistic analyses of F0 in read and spontaneous speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 92*, 763-766.

Doran, C. (2000). Punctuation in a lexicalized grammar. *Proceedings of the 5th International Workshop on Tree Adjoining Grammars and Related Formalism (TAG+5)*. Retrieved April 2002 from http://citeseer.nj.nec.com/439214.html

Dutoit, T. (1997). *An introduction to text-to-speech synthesis*. Boston: Kluwer Academic Publishers.

Edgington, M., Lowry, A., Jackson, P., Breen, A. P., & Minnis, S. (1996a). Overview of current text-to-speech techniques: Part I - text and linguistic analysis. *BT Technical Journal, 14*, 68-83. Retrieved May 2002 from http://www.sc-server1.bt.com/bttj/vol14no1/05.pdf

Edgington, M., Lowry, A., Jackson, P., Breen, A. P., & Minnis, S. (1996b). Overview of current text-to-speech techniques: Part II - prosody and speech generation. *BT Technical Journal, 14*, 84-99. Retrieved May 2002 from http://www.sc-server1.bt.com/bttj/vol14no1/06.pdf

Fagyal, Z. (2002). Tonal template for background information: The scaling of pitch in utterance-medial parentheticals in French. *Proceedings of Speech Prosody 2002*, 279-282. Retrieved February 2003 from http://www.lpl.univ-aix.fr/sp2002/pdf/fagyal.pdf

Flach, G. (1999). *Interface design for speech synthesis systems*.Paper presented at a COST258 meeting in 1999. Retrieved June 2002 from http://www.unil.ch/imm/docs/LAIP/COST_258/Scientific/Proceedings/xml/flach.pdf

Grosz, B., & Hirschberg, J. (1992). Some intonational charactersitics of discourse structure. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 92*, 429-432.

Heavener, M. Q. (2002). *Cascading style sheets for great control*. Retrieved May 2003 from http://www.heavenr.com/stuff/css.html

Hess, W. (1983). *Pitch determination of speech signals.* Heidelberg: Springer Verlag.

Hill, R. L., & Murray, W. S. (2000, March). *Atoning for punctuation: prosody and ambiguity while reading aloud.* Poster session presented at the City University of New York (CUNY) Conference on Human Sentence Processing 2000, La Jolla, CA. Retrieved May 2002 from http://www.dundee.ac.uk/psychology/rlhill/cuny2000.pdf

Hirschberg, J. (1995). Pitch accent in context: predicting intonational prominence from text. *Artificial Intelligence, 63* (1993), 305-340. Retrieved May 2002 from http://www.research.att.com/~julia/papers/aij93.ps

Hirschberg, J., & Nakatani, C. H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96)*, 286-293. Retrieved May 2002 from http://www.research.att.com/~julia/papers/acl96.ps

Hitzeman, J., Black, A. W., Mellish, C., Oberlander, J., Poesio, M., & Taylor, P. (1999). An annotation scheme for concept-to speech synthesis. *Proceedings of the European Workshop on Natural Language Generation*, 59-66. Retrieved May 2002 from http://www.cs.cmu.edu/~awb/papers/NLG99_sole.ps

Hunt, A. (2000). *JSpeech Markup Language*. Retrieved June 2002 from http://www.w3.org/TR/2000/NOTE-jsml-20000605

Jones, B. E. M. (1994a). Exploring the role of punctuation in parsing natural text. *Proceedings of the International Conference on Computational Linguistics (COLING-94)*, 421-425. Retrieved May 2002 from http://arxiv.org/abs/cmp-lg/9505024

Jones, B. E. M. 1994b. *Can punctuation help parsing?* (Acquilex-II Working Paper 29). Cambridge, UK: Cambridge University Computer Laboratory. Retrieved April 2002 from http://www.citeseer.nj.nec.com/jones94can.html

Klatt, D. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America, 82,* 737-793.

Kutik, E, Cooper, W. E., & Boyce, S. (1983). Declination of fundamental frequency in speakers' production of parenthetical and main clauses. *Journal of the Acoustical Society of America, 73*, 1731-1738.

Lindström, A., Bretan, I., & Ljungqvist, M. (1996). Prosody generation in text-to-speech conversion using dependency graphs. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 96, 3* [CD-ROM], 1341-1344. Retrieved June 2002 from http://www.asel.udel.edu/icslp/cdrom/vol3/532/a532.pdf

Marcu, D. (1997). The rhetorical parsing, summarization, and generation of natural language texts (Doctoral dissertation, University of Toronto, 1997). *Dissertation Abstracts International-B, 60/01*, 260. Retrieved October 2002 from http://www.isi.edu/~marcu/papers/phd-thesis.ps.gz

Marcu, D. (1998). A surface-based approach to identifying discourse markers and elementary textual units in unrestricted texts. *Proceedings of the COLING/ACL-98 Workshop on Discourse Relations and Discourse Markers*, 1-7. Retrieved October 2002 from http://www.isi.edu/~marcu/papers/markers-units98.ps

Marcu, D. (1999). A decision-based approach to rhetorical parsing. *Proceedings of the Meeting of the Association for Computational Linguistics (ACL-99)*, 365-372. Retrieved October 2002 from http://acl.ldc.upenn.edu/P/P99/p99-1047.pdf

Merriam-Webster's guide to punctuation and style (2nd ed.). (2001). Springfield, MA: Merriam-Webster Inc.

Mertens, P. (2002). Synthesizing elaborate intonation contours in text-to-speech for French. *Proceedings of Speech Prosody 2002,* 449-502. Retrieved June 2002 from http://bach.arts.kuleuven.ac.be/~piet/papers/sp2002.pdf

Meyer, C. F. (1987). *A linguistic study of American punctuation.* New York: Peter Lang Publishing.

Möhler, G., & Mayer, J. (2001). A discourse model for pitch-range control. *Proceedings of the 4th International Speech Communication Association (ISCA) Tutorial and Research Workshop on Speech Synthesis.* Retrieved June 2002 from http://www.ims.uni-stuttgart.de/phonetik/joerg/papers/ssws01.pdf

Möhler, G., & Mayer, J. (2002). *Pitch-range control based on discourse structure.* Manuscript submitted for publication. Retrieved June 2002 from http://www.ims.uni-stuttgart.de/phonetik/joerg/papers/JSpeechTech2002.pdf

Nunberg, G. (1990). *The linguistics of punctuation.* (CSLI Lecture Notes 18). Stanford, CA: CSLI Publications.

O'Gara, P. (n.d.). *An introduction to text to speech.* Retrieved November 2002 from http://www.scit.wlv.ac.uk/~c9662750/

O'Shaughnessy, D. (1990). Relationship between syntax and prosody for speech synthesis. *Proceedings of the European Speech Communication Association (ESCA) Tutorial on Speech Synthesis (AUTRANS, France)*, 39-42

Pierrehumbert, J. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America, 70*, 985-995.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman Inc.

Reed, C., & Long, D. (1997). *Generating punctuation in written arguments.* (Technical Report RN/97/157/). Department of Computer Science, University College, London, UK. Retrieved April 2002 from http://citeseer.nj.nec.com/reed97generating.html

SABLE: A Synthesis Markup Language (version 1.0). (n.d.). Retrieved June 2002 from http://www.research.att.com/~rws/Sable.v1_0.htm

Sampson, G. (1992). [Review of the book *The linguistics of punctuation*]. *Linguistics*, *30*, 467-475.

Say, B., & Akman, V. (1996). Information-based aspects of punctuation. *Proceedings of the Association for Computational Linguistics Workshop on Punctuation.* Retrieved April 2002 from http://citeseer.nj.nec.com/205712.html

Say, B., & Akman, V. (1997). Current Approaches to Punctuation in Computational Linguistics. *Computers and Humanities, 30*, 457-469.

Say, B., & Akman, V. (1998). Dashes as Typographical Cues for the Information Structure. *Proceedings of the 3rd Conference in Information-Theoretic Approaches to Logic, Language, and Computation (ITALLC '98)*, 209-223. Retrieved April 2002 from http://citeseer.nj.nec.com/say98dashe.html

Schröder, M., & Trouvain, J. (2001). The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *Proceedings of the 4th International Speech Communication Association (ISCA) Tutorial and Research Workshop on Speech Synthesis*, 131-136. Retrieved January 2003 from http://www.ssw4.org/papers/112.pdf

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., et al. (1992). ToBI: A standard for labeling English prosody. *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP)*, 867-870. Retrieved March 2003 from http://ling.osu.edu/~tobi/ame_tobi/Silverman_etal1992.pdf

Sproat, R., Hunt, A., Ostendorf, M., Taylor, P., Black, A., Lenzo, K., & Edgington, M. (1998). SABLE: A standard for TTS markup. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 98*, 1719-1722. Retrieved June 2002 from http://citeseer.nj.nec.com/34259.html

Sproat, R., & Raman, T. V. (1999). *SABLE: An XML-based aural display list for the WWW*. Retrieved June 2002 from http://www.research.att.com/~rws/csssable.html

Syrdal, A. K., & Hirschberg, J. (2001). Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Communication, 33*, 135-151. Retrieved June 2002 from http://www.research.att.com/projects/tts/papers/2000_SpeechCom/spcom.ps

Tatham, M., & Lewis, E. (1992). Prosodic assignment in SPRUCE text-to-speech synthesis. *Proceedings of the Institute of Acoustics*, 447-454. Retrieved May 2002 from http://speech.essex.ac.uk/speech/archive/ss-pros/ss-pros.html

Tatham, M., & Lewis, E. (1996). Improving text-to-speech synthesis. *Proceedings of the Institute of Acoustics*, 35-42. Retrieved May 2002 from http://speech.essex.ac.uk/speech/archive/improvingss/MT-EL.html

Tatham, M., Morton, K., & Lewis, E. (1998). Assignment of intonation in high-level speech synthesiser. *Proceedings of the Institute of Acoustics*, 255-262. Retrieved May 2002 from http://www.essex.ac.uk/speech/pubs/spruce/windermere98/intonation/intassign.html

Tatham, M., Lewis, E., & Morton, K. (1999). An advanced intonation model for synthesis. *Proceedings of the Sixth European Conference on Speech Communication and Technology (Europseech 99)*, 1871-1874. Retrieved May 2002 from http://www.cs.bris.ac.uk/Tools/Reports/Ps/1999-tatham-0.ps.gz

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America, 107*, 1697-1714. Retrieved June 2002 from http://www.cstr.ed.ac.uk/publications/papers/2000/Taylor_2000_a.ps

Wang, M. Q., & Hirschberg, J. (1995). Automatic classification of intonational phrase boundaries. *Computer, Speech and Language, 6* (1992, April), 175-196. Retrieved May 2002 from http://www.research.att.com/~julia/papers/csl92.ps

Westall, F. A., Johnston, R. D., & Lewis, A.V. (1998). Speech technology for telecommunications. In F. A. Westall, R. D. Johnston, & A. V. Lewis (Eds.), *Speech technology for telecommunications* (pp. 1-29). New York: Chapman & Hall.

White, M. (1995). Presenting Punctuation. *Proceedings of the Fifth European Workshop on Natural Language Generation*, 107-125. Retrieved April 2002 from http://citeseer.nj.nec.com/white95presenting.html

Wichmann, A. (2000). *Intonation in text and discourse: Beginnings, middles and ends.* Harlow, England: Pearson Education Limited.

Wightman, C. W. (2002). ToBI or not ToBI? *Proceedings of Speech Prosody 2002.* Retrieved February 2003 from http://www.lpl.univ-aix.fr/sp2002/pdf/wightman.pdf

Wouters, J., Rundle, B., & Mason, M. W. (1999). Authoring tools for speech synthesis using the Sable markup standard. *Proceedings of the Sixth European Conference on Speech Communication and technology (Eurospeech 99)*, 963-966. Retrieved January 2003 from http://tmos.org/tech/papers/eurospeech99.pdf

APPENDIX A

Parenthetical background

[This text was given to each experiment participant prior to the experiment.]

This study uses parentheticals as a means to increase our linguistic knowledge of language. To participate in this study it is important to know a few things about parentheticals.

Parenthetical elements consist of words, phrases or sentences that are inserted into a sentence to provide additional explanatory or commentatory information. A parenthetical element is not grammatically essential to the sentence it is in. When a parenthetical is removed from a sentence, the sentence stays fully intact with respect to semantics and syntax.

Parentheticals are set off by commas, dashes and brackets (also called *parentheses*), such as in (parenthetical elements are italized):

(1) a) All of us, *to tell the truth*, were amazed.

b) The examiner chose – *goodness knows why* – to ignore it.

c) A ruling by the FCC (*Federal Communications Commission*). . . .

Commas are being used to indicate parenthetical elements that are the least digressive from the main sentence (as in (a)), while dashes (as in (b)) are being used to set off parenthetical elements that are more digressive than elements set off with commas but less digressive than elements set off by brackets. Hence, brackets are used to indicate the greatest extent of digressiveness (as in (c)). Brackets are also used instead of commas or dashes when the interruption of the sentence caused by the parenthetical is too great to be indicated by commas or dashes.

What this short introduction to parentheticals intends to show you is that the choice of punctuation determines the semantic relation of the parenthetical to the main sentence. Please keep that in mind when participating in this experiment.

Any more questions about parentheticals?

APPENDIX B

Set of sentences

The set of sentences that were read by every candidate:

1) We saw the movie but were unimpressed.
[for a control recording of the frame sentence]

2) We saw the movie that had been banned in Boston, but were unimpressed.
[the *that* clause is restrictive and integrated in the frame sentence; hence, there is no parenthetical in this sentence]

3) We saw the movie - Jane insisted on going - but were unimpressed.

4) We saw the movie, which had been banned in Boston, but were unimpressed.

5) We saw the movie that received five Oscars, but were unimpressed.
[a looser-upper to prevent the readers from getting into a routine]

6) We saw the movie (which was Jane's idea) but were unimpressed.

7) We saw the movie on the opening night but were unimpressed.
[another looser-upper]

8) We saw the movie, Jane (who knows the director) insisted on going, but were unimpressed.

9) We saw the movie (which had been banned in Boston) but were unimpressed.

10) We saw the movie - I heard it was good (it had received much praise) - but were unimpressed.

11) We saw the movie in a sneek preview but were unimpressed.
[another looser-upper]

12) We saw the movie, Jane insisted on going, but were unimpressed.

13) We saw the movie - which had been banned in Boston - but were unimpressed.

14) We saw the movie in the new movie theater but were unimpressed.
[another looser-upper]

15) We saw the movie, Jane really wanted to go - it had received much praise - but were unimpressed.

16) We saw the movie (Jane insisted on going) but were unimpressed.

17) We saw the movie because everyone went, but were unimpressed.
[the last looser-upper]

18) We saw the movie - Jane (who knows the director) insisted on going - but were unimpressed.

19) We saw the movie (which had been banned in Boston) - Jane insisted on going - but were unimpressed.

20) We saw the movie (Jane - who knows the director - insisted on going) but were unimpressed.

APPENDIX C

Female Frequency Data

Female frequencies [Hz]

	MAR			SHO			SUZ			Average (female)		
	max	min	range	max	min	range	max	min	range	max	min	range
sentence 1												
fr(a)	292	205	87	299	207	92	281	157	124	291	190	101
fr(b)	216	162	54	216	164	52	208	149	59	213	158	55
Sentences with non-parenthetical medial phrase												
sentences 2,5,7,11,14,17												
sentence 2												
fr(a)	255	182	73	299	227	72	268	175	93	274	195	79
medial (comma)	244	154	90	239	173	66	230	146	84	238	158	80
fr(b)	225	145	80	212	151	61	218	151	67	218	149	69
sentence 5												
fr(a)	275	185	90	244	201	43	262	174	88	260	187	74
medial (comma)	243	170	73	220	164	56	236	142	94	233	159	74
fr(b)	209	125	84	199	143	56	211	144	67	206	137	69
sentence 7												
fr(a)	280	175	105	251	189	62	250	173	77	260	179	81
medial	258	150	108	205	154	51	256	165	91	240	156	83
fr(b)	213	148	65	204	157	47	199	148	51	205	151	54
sentence 11												
fr(a)	299	152	147	291	211	80	254	190	64	281	184	97
medial	278	163	115	240	175	65	251	162	89	256	167	90
fr(b)	219	163	56	206	165	41	225	166	59	217	165	52
sentence 14												
fr(a)	298	190	108	265	203	62	286	184	102	283	192	91
medial	254	146	108	230	147	83	239	150	89	241	148	93
fr(b)	233	158	75	209	149	60	213	149	64	218	152	66
sentence 17												
fr(a)	286	169	117	283	183	100	262	178	84	277	177	100
medial (comma)	263	164	99	242	146	96	263	161	102	256	157	99
fr(b)	220	153	67	214	150	64	208	156	52	214	153	61
Sentences with one parenthetical medial phrase												
sentences 3,4,6,9,12,13,16												
sentence 3												
fr(a)	257	169	88	274	176	98	263	160	103	265	168	96
medial dash	261	149	112	212	150	62	262	165	97	245	155	90
fr(b)	242	147	95	220	142	78	223	152	71	222	147	81
sentence 4												
fr(a)	278	181	97	295	160	135	264	179	85	279	173	106
medial comma	262	166	96	243	173	70	234	141	93	246	160	86
fr(b)	208	142	66	201	154	47	220	147	73	210	148	57
sentence 6												
fr(a)	284	174	110	252	179	73	286	162	124	274	172	102
medial bracket	252	161	91	253	157	96	239	152	87	248	157	91
fr(b)	205	160	45	206	167	39	203	150	53	205	159	46

Female frequencies [Hz]												
	MAR			SHO			SUZ			Average (female)		
	max	min	range	max	min	range	max	min	range	max	min	range
sentence 9												
fr(a)	293	171	122	239	154	85	267	166	101	266	164	103
medial bracket	242	165	77	246	140	106	251	140	111	246	148	98
fr(b)	208	141	67	207	163	44	208	141	67	208	148	59
sentence 12												
fr(a)	294	163	131	265	165	100	246	178	68	268	169	100
medial comma	239	145	94	228	162	66	250	158	92	239	155	84
fr(b)	216	159	57	204	158	46	222	141	81	210	159	52
sentence 13												
fr(a)	265	149	116	269	160	109	265	176	89	266	162	105
medial dash	256	141	115	248	154	94	231	158	73	245	151	94
fr(b)	218	149	69	218	140	78	213	146	67	216	145	71
sentence 16												
fr(a)	298	164	134	271	175	96	257	167	90	275	169	107
medial bracket	239	152	87	226	155	71	271	145	126	245	151	95
fr(b)	201	145	56	222	141	81	189	143	46	204	143	61
Sentences with two parenthetical phrases												
sentences 8,10,15,18,19,20												
sentence 8												
fr(a)	292	165	127	276	162	114	253	177	76	274	168	106
outerpar(a) comma	237	178	59	219	158	61	229	167	62	228	168	61
innerpar bracket	238	170	68	239	174	65	280	167	113	252	170	82
outerpar(b) comma	235	164	71	236	152	84	272	164	108	248	160	88
fr(b)	210	154	56	208	165	43	209	154	55	209	160	49
sentence 10												
fr(a)	279	169	110	276	174	102	272	171	101	276	171	104
par1 dash	256	174	82	237	152	85	238	160	78	244	162	82
par2 bracket	239	163	76	219	148	71	243	145	98	234	152	82
fr(b)	219	163	56	244	141	103	211	142	69	225	149	76
sentence 15												
fr(a)	300	151	149	258	179	79	252	175	77	270	168	102
par1 comma	235	158	77	228	167	61	236	171	65	233	165	68
par2 dash	262	162	100	241	185	56	275	152	123	259	166	93
fr(b)	245	151	94	213	173	40	198	153	45	219	159	60
sentence 18												
fr(a)	267	166	101	264	163	101	263	161	102	265	163	101
outerpar(a) dash	225	166	59	237	165	72	213	168	45	225	166	59
innerpar bracket	242	155	87	214	166	48	237	157	80	231	159	72
outerpar(b) dash	276	162	114	244	162	82	255	164	91	258	163	96
fr(b)	219	145	74	208	145	63	225	160	65	217	153	64
sentence 19												
fr(a)	300	150	150	273	183	90	259	165	94	277	166	111
par1 bracket	262	149	113	242	176	66	225	156	69	243	160	83
par2 dash	235	167	68	222	163	59	260	159	101	239	163	76
fr(b)	232	141	91	214	154	60	192	162	30	213	152	60

Female frequencies [Hz]												
	MAR			SHO			SUZ			Average (female)		
	max	min	range	max	min	range	max	min	range	max	min	range
sentence 20												
fr(a)	299	150	149	252	174	78	240	168	72	264	164	100
outerpar(a) bracket	230	172	58	245	172	73	215	165	50	230	170	60
innerpar dash	241	150	91	222	145	77	234	148	86	232	148	85
outerpar(b) bracket	237	153	84	206	156	50	247	152	95	230	154	76
fr(b)	225	159	66	216	145	71	195	144	51	212	149	63

APPENDIX D

Male Frequency Data

Male frequencies [Hz]

	AAR			CRA			GRE			Average (male)		
	max	min	range	max	min	range	max	min	range	max	min	range
sentence 1												
fr(a)	158	92	66	117	87	30	222	98	124	166	92	73
fr(b)	112	75	37	117	81	36	120	81	39	116	79	37
Sentences with non-parenthetical medial phrase												
sentences 2,5,7,11,14,17												
sentence 2												
fr(a)	170	87	83	141	100	41	181	119	62	164	102	62
medial (comma)	125	84	41	126	81	45	130	95	35	127	87	40
fr(b)	101	77	24	108	83	25	127	92	35	112	84	28
sentence 5												
fr(a)	187	114	73	135	96	39	166	117	49	163	109	54
medial (comma)	146	81	65	112	88	24	123	91	32	127	87	40
fr(b)	129	82	47	107	83	24	112	79	33	116	81	35
sentence 7												
fr(a)	174	105	69	128	99	29	193	125	68	165	110	55
medial	125	84	41	110	88	22	123	89	34	119	87	32
fr(b)	108	75	33	101	81	20	113	80	33	107	79	29
sentence 11												
fr(a)	192	123	69	139	95	44	168	119	49	166	112	54
medial	143	81	62	117	77	40	134	88	46	131	82	49
fr(b)	110	80	30	100	78	22	120	75	45	110	78	32
sentence 14												
fr(a)	213	116	97	146	95	51	223	135	88	194	115	79
medial	164	91	73	116	88	28	152	117	35	144	99	45
fr(b)	114	78	36	107	81	26	116	98	18	112	86	27
sentence 17												
fr(a)	225	102	123	130	78	52	212	124	88	189	101	88
medial (comma)	155	91	64	108	82	26	134	95	39	132	89	43
fr(b)	114	74	40	107	81	26	111	91	20	111	82	29
Sentences with one parenthetical medial phrase												
sentences 3,4,6,9,12,13,16												
sentence 3												
fr(a)	170	79	91	131	97	34	182	98	84	161	91	70
medial dash	137	79	58	120	83	37	142	77	65	133	80	53
fr(b)	106	76	30	102	83	19	115	91	24	109	83	24
sentence 4												
fr(a)	179	87	92	127	80	47	195	93	102	167	87	80
medial comma	124	76	48	120	77	43	134	95	39	126	83	43
fr(b)	111	76	35	101	78	23	111	77	34	108	77	29
sentence 6												
fr(a)	190	87	103	142	89	53	188	97	91	173	91	82
medial bracket	147	77	70	110	82	28	132	94	38	130	84	45
fr(b)	110	84	26	108	79	29	106	91	15	108	85	23

Male frequencies [Hz]												
	AAR			CRA			GRE			Average (male)		
	max	min	range	max	min	range	max	min	range	max	min	range
sentence 9												
fr(a)	184	90	94	135	83	52	192	112	80	170	95	75
medial bracket	146	83	63	122	75	47	125	88	37	131	82	49
fr(b)	111	83	28	111	82	29	161	98	63	128	88	40
sentence 12												
fr(a)	187	93	94	143	87	56	191	91	100	174	90	83
medial comma	135	82	53	106	80	26	130	105	25	124	89	35
fr(b)	119	77	42	105	81	24	127	92	35	112	79	33
sentence 13												
fr(a)	204	87	117	133	75	58	200	97	103	179	86	93
medial dash	125	79	46	123	79	44	125	91	34	124	83	41
fr(b)	121	80	41	108	79	29	128	81	47	119	80	39
sentence 16												
fr(a)	203	87	116	131	94	37	194	99	95	176	93	83
medial bracket	141	83	58	115	83	32	148	93	55	135	86	48
fr(b)	128	86	42	113	87	26	110	82	28	117	85	32
Sentences with two parenthetical phrases												
sentences 8,10,15,18,19,20												
sentence 8												
fr(a)	191	89	102	127	86	41	201	120	81	173	98	75
outerpar(a) comma	142	89	53	119	91	28	142	102	40	134	94	40
innerpar bracket	157	82	75	113	82	31	142	99	43	137	88	50
outerpar(b) comma	146	80	66	115	82	33	140	96	44	134	86	48
fr(b)	129	78	51	114	78	36	120	97	23	121	88	30
sentence 10												
fr(a)	195	86	109	146	88	58	168	104	64	170	93	77
par1 dash	n/a	n/a	n/a	130	82	48	122	98	24	126	90	36
par2 bracket	n/a	n/a	n/a	124	79	45	129	95	34	127	87	40
fr(b)	115	83	32	105	79	26	128	98	30	116	87	29
sentence 15												
fr(a)	191	99	92	124	83	41	192	92	100	169	91	78
par1 comma	n/a	n/a	n/a	111	80	31	124	94	30	118	87	31
par2 dash	n/a	n/a	n/a	119	77	42	148	88	60	134	83	51
fr(b)	122	82	40	107	78	29	130	75	55	120	78	41
sentence 18												
fr(a)	203	90	113	130	81	49	184	102	82	172	91	81
outerpar(a) dash	139	96	43	113	98	15	112	95	17	121	96	25
innerpar bracket	117	82	35	116	78	38	115	82	33	116	81	35
outerpar(b) dash	128	81	47	115	82	33	140	124	16	128	96	32
fr(b)	121	81	40	108	86	22	116	76	40	115	81	31
sentence 19												
fr(a)	204	88	116	140	85	55	182	97	85	175	90	85
par1 bracket	133	82	51	140	76	64	125	94	31	133	84	49
par2 dash	147	83	64	118	81	37	144	75	69	136	80	57
fr(b)	108	75	33	109	80	29	157	97	60	125	84	41

Male frequencies [Hz]												
	AAR			CRA			GRE			Average (male)		
	max	min	range	max	min	range	max	min	range	max	min	range
sentence 20												
fr(a)	192	93	99	132	87	45	187	97	90	170	92	78
outerpar(a) bracket	137	92	45	111	89	22	145	104	41	131	95	36
innerpar dash	113	78	35	122	85	37	131	93	38	122	85	37
outerpar(b) bracket	159	84	75	114	84	30	147	89	58	140	86	54
fr(b)	133	79	54	110	84	26	125	96	29	123	86	36

APPENDIX E

Pause Data

Pauses [sec]

	AAR	CRA	GRE	MAR	SHO	SUZ	Av. male	Av. female	Average
Sentence 1									
fr(a) to fr(b)	0.2	0	0	0	0	0	0.1	0.0	0.0
Sentences with non-parenthetical medial phrase									
sentences 2,5,7,11,14,17									
Sentence 2									
fr(a) to medial	0.3	0	0	0	0	0	0.1	0.0	0.1
medial to fr (b) (comma)	0.5	0.4	0.3	0.3	0.1	0.3	0.4	0.2	0.3
Sentence 5									
fr(a) to medial	0	0	0	n/a	0	0	0.0	0.0	0.0
medial to fr (b) (comma)	0.5	0.5	0.3	0.1	0.1	0.1	0.4	0.1	0.3
Sentence 7									
fr(a) to medial	0	0	0	0	0	0	0.0	0.0	0.0
medial to fr (b)	0.7	0.1	0.1	0.2	0.2	0	0.3	0.1	0.2
Sentence 11									
fr(a) to medial	0	0	0	0	0	0	0.0	0.0	0.0
medial to fr (b)	0.2	0.3	0.2	0	0	0.1	0.2	0.0	0.1
Sentence 14									
fr(a) to medial	0	0	0	0	0	0	0.0	0.0	0.0
medial to fr (b)	0.1	0.1	0	0	0	0	0.1	0.0	0.0
Sentence 17									
fr(a) to medial	0	0	0	n/a	0	0	0.0	0.0	0.0
medial to fr (b) (comma)	0.3	0.4	0.1	0.1	0.1	0.5	0.3	0.2	0.3
Sentences with one parenthetical medial phrase									
sentences 3,4,6,9,12,13,16									
Sentence 3									
fr(a) to medial (dash)	0.4	0.4	0.4	0.4	0.1	0.2	0.4	0.2	0.3
medial to fr (b) (dash)	0.4	0.5	0.6	0.2	0.2	0.3	0.5	0.2	0.4

Pauses [sec]									
	AAR	CRA	GRE	MAR	SHO	SUZ	Av. male	Av. female	Average
Sentence 4									
fr(a) to medial (comma)	0.4	0.4	0.1	0.1	0.1	0.1	0.3	0.1	0.2
medial to fr (b) (comma)	0.5	0.4	0.4	0.2	0.2	0.4	0.4	0.3	0.4
Sentence 6									
fr(a) to medial (bracket)	0.5	0.3	0.2	0.1	0.0	0.4	0.3	0.2	0.3
medial to fr (b) (bracket)	0.4	0.6	0.3	0.4	0.2	0.2	0.4	0.3	0.4
Sentence 9									
fr(a) to medial (bracket)	0.3	0.3	0.2	0.4	0.1	0.2	0.3	0.2	0.3
medial to fr (b) (bracket)	0.3	0.6	0.4	0.5	0.3	0.2	0.4	0.3	0.4
Sentence 12									
fr(a) to medial (comma)	0.2	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.2
medial to fr (b) (comma)	0.4	0.4	0.4	0.1	0.2	0.4	0.4	0.2	0.3
Sentence 13									
fr(a) to medial (dash)	0.2	0.3	0.2	0.3	0.2	0.2	0.2	0.2	0.2
medial to fr (b) (dash)	0.5	0.3	0.4	0.3	0.2	0.4	0.4	0.3	0.4
Sentence 16									
fr(a) to medial (bracket)	0.3	0.3	0.2	0.4	0.2	0.2	0.3	0.3	0.3
medial to fr (b) (bracket)	0.3	0.6	0.4	0.5	0.2	0.4	0.4	0.4	0.4
Sentences with two parenthetical medial phrases									
sentences 8,10,15,18,19,20									
Sentence 8									
fr(a) to par1(a) (comma)	0.3	0.2	0.3	0.5	0.2	0.3	0.3	0.3	0.3
par1(a) to par2 (bracket)	0.4	0.2	0.1	0.0	0.3	0.2	0.2	0.2	0.2
par2 to par1(b) (bracket)	0.0	0.2	0.3	0.0	0.1	0.1	0.2	0.1	0.1
par1(b) to fr(b) (comma)	0.4	0.4	0.3	0	0.5	0.4	0.4	0.3	0.3

Pauses [sec]	AAR	CRA	GRE	MAR	SHO	SUZ	Av. male	Av. female	Average
Sentence 10									
fr(a) to par1 (dash)	0.1	0.4	0.2	0.4	0.3	0	0.2	0.2	0.2
par1 to par2 (bracket)	0.4	0.4	0.1	0.3	0.1	0.4	0.3	0.3	0.3
par2 to fr (b) (bracket + dash)	0.2	0.4	0.3	0.5	0.5	0.4	0.3	0.5	0.4
Sentence 15									
fr(a) to par1 (comma)	0	0	0.2	0.2	0.2	0.1	0.1	0.2	0.1
par1 to par2 (dash)	0.3	0.3	0.1	0.4	0.2	0.0	0.2	0.2	0.2
par2 to fr (b) (dash)	0.2	0.4	0.5	0.2	0.1	0.4	0.4	0.2	0.3
Sentence 18									
fr(a) to par1(a) (dash)	0.3	0.4	0.3	0.4	0.2	0.0	0.3	0.2	0.3
par1(a) to par2 (bracket)	0.0	0.2	0.0	0.1	0.1	0.0	0.1	0.1	0.1
par2 to par1(b) (bracket)	0.1	0.2	0.0	0.3	0.1	0.0	0.1	0.1	0.1
par1(b) to fr(b) (dash)	0.5	0.4	0.4	0.2	0.2	0.5	0.4	0.3	0.4
Sentence 19									
fr(a) to par1 (bracket)	0.1	0.2	0.1	0	0.1	0	0.1	0.0	0.1
par1 to par2 (bracket + dash)	0.4	0.7	0.3	0.6	0.4	0.4	0.5	0.5	0.5
par2 to fr (b) (dash)	0.3	0.7	0.5	0.2	0.2	0.2	0.5	0.2	0.4
Sentence 20									
fr(a) to par1(a) (bracket)	0.1	0.4	0.3	0.2	0.3	0.3	0.3	0.3	0.3
par1(a) to par2 (dash)	0.1	0.2	0.0	0.1	0.2	0.0	0.1	0.1	0.1
par2 to par1(b) (dash)	0.0	0.3	0.1	0.0	0.2	0.3	0.1	0.2	0.2
par1(b) to fr(b) (bracket)	0.4	0.6	0.6	0.2	0.3	0.5	0.5	0.3	0.4

<u>VITA</u>

Surname: Bodenbender Given Names: Christel

Place of Birth: Offenbach/Main, Hessen, Germany

<u>Educational Institutions Attended:</u>

Verwaltungsfachhochschule in Wiesbaden 1989 to 1992

Universität Kassel 1992 to 1994

University of Victoria 1994

to 2003

<u>Degrees Awarded:</u>

Dipl.Verwaltungswirt	Verwaltungsfachhochschule in Wiesbaden	1992
B.Sc.	University of Victoria	1999

<u>Publications:</u>

A case of was-w in Hessian: New evidence against scope marking. (2002). *Working Papers of the Linguistic Circle of the University of Victoria*, Vol. 16, 1-11.

Title of Thesis:

The Punctuation and Intonation of Parentheticals

Author _______________________

Christel Bodenbender

August 17, 2003