Tempo Modulations in English

by
Sandra Patricia Kirkham

B.A., University of Victoria, 1988, M.A., University of Victoria, 1992

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

DOCTOR OF PHILOSOPHY

in the Department of Linguistics

We accept this dissertation as conforming to the required standard

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

Dr. B.P. Harris, Departmental Member (Department of Linguistics)

Dr. T.M. Hess, Departmental Member (Department of Linguistics)

Dr. P.F. Driessen, Outside Member (Department of Electrical and Computer Engineering)

Dr. Zita McRobbie, External Examiner (Department of Linguistics, Simon Fraser University)

Abstract

The goal of synthetic speech is to provide speech that is both comprehensible and natural sounding. While synthetic speech is drawing nearer to its goal, it has not yet attained a truly natural quality. Naturalness can be improved by incorporating prosodic rules for duration and intonation that are representative of natural speech. While duration models are widely used, they fail to replicate the variations evident in the tempo of natural speech.

This project proposes a model of tempo modulations in English based upon phrasal foci. In order to replicate this pattern, the potential phonetic locations for altering the speech rate of English synthetic speech are explored.

The results of a pilot study based on the readings of one speaker suggested that tempo modulations are predictable and not random, and that they are not expressed as equal expansions and compressions across all syllable constituents. Vowels, onsets, and codas exhibited varying degrees of change. These results motivated a study of the same phenomena in data derived from the readings of multiple speakers.

The data for the main study were derived from two readings of each of five Canadian English sentences. The first reading varied the position of a focused word in the sentence and the second, only the tempo. Sentences that were neutral in terms of focus and tempo were included in both readings to create experimental

controls. The readings were recorded and digitized to provide waveforms for duration measurement.

Comparisons of average durations of focused syllables to the respective controls revealed significant differences given an alpha level of .05, providing evidence that a pattern of tempo modulations can be predicted. This pattern involved expansion and compression within the sentence.

The pattern can be replicated using the results of the investigation of sites for tempo changes. The results reveal that at a fast tempo and a slow tempo, the durations of syllable constituents change significantly from the control at an alpha level of .01. The vowel, particularly one that comprises a syllable, is the primary site for expansion and compression. Stressed vowels have the largest compression, while unstressed vowels have the largest expansion.

The degree of segmental change varies depending on the position of the syllable constituent. In stressed CVC syllables, codas and then onsets exhibit lessening degrees of compression. The reverse is true for expansion, and the degree of change for these constituents is less than that for compression. However, only stops in these positions show a significant change from the control. It appears that expansions and compressions of segments are ranked according to syllable constituency.

These ranked expansions and compressions of syllable constituents can be incorporated into an existing duration model for synthetic speech in order to replicate the observed pattern of tempo modulations in English. This tempo pattern provides variation at a sentential level and is an improvement over rules for emphasis that are specific to the emphasized word or part thereof. The pattern is expressed by duration rules, and the addition of the criterion for syllable constituency increases the natural distribution of changes in tempo provided a model to bring synthetic speech closer to the natural goal.

Examiners:

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

Dr. B.P. Harris, Departmental Member (Department of Linguistics)

Dr. T.M. Hess, Departmental Member (Department of Linguistics)

Dr. P.F. Driessen, Outside Member (Department of Electrical and Computer Engineering)

Dr. Zita McRobbie, External Examiner (Department of Linguistics, Simon Fraser University)

Contents

Acknowledgements

I would like to take this opportunity to thank those who have contributed to this project. A collaboration with STR marked the beginnings of the project. I would like to thank Craig Dickson for giving me this opportunity and for providing resources for the completion of the pilot study. At the conception stage of the project, Craig Dickson and Dr. Stephen Eady of STR provided valuable insight in the technological considerations of the pilot study and its design. In addition, Dr. Eady offered assistive feedback of the results of this initial study.

I would like to express my appreciation to Dr. John Esling for supervising the project and for his assistance in the preparation of the dissertation. I am obliged to Dr. Barbara Harris for her careful editing and moral support and to Dr. Peter Driessen for his sincere interest in the project and his assistance in procuring resources. I would especially like to thank Dr. Thom Hess for his time and patience as well as his helpful comments on the initial drafts.

The statistical analysis of the main study would not have been possible without the assistance of Dr. Michael Hunter and Dr. Robert M. Dummer both of whom volunteered their time and expertise in making recommendations for the statistical analysis of the data.

I would also like to take this opportunity to acknowledge the British Columbia Science Council for providing funding for the pilot study phase of the project through the Great Award and the University of Victoria for providing funding through the University of Victoria Fellowship.

A final thank you to my family and friends who encouraged and supported me throughout the duration of this degree.

Dedication

In dedication to David and Emmanuel

CHAPTER 1 Introduction

Tempo is the rate of an activity such as speech or music. It serves to organize the activity in terms of duration. In speech, tempo is determined by the duration of a unit of speech. It constrains other temporal phenomena, such as rhythm. [1]

As in music, tempo in speech can be modulated to accent whole phrases. This modulation takes the form of changes in the duration of a phrase or other speech unit. Tempo modulations have been observed in Swedish, where phrases that outline a story in read speech are decelerated for accent, as noted by Fant, Kruckenberg, and Nord (1991c, pp. 251-56). However, this means of accentuation in speech also applies within phrases. The focal point of a phrase is accented by a preceding deceleration of tempo. Fant et al. also discovered this pattern in Swedish (p.256). In English, we find speeches and theatre performance[2] examples of such

1.Rhythm organizes speech into patterns of long and short or accented and unaccented syllables.

speech modulations, which have not yet been modelled. A further example of the correlation between tempo changes and accented semantic content occurs in cinema. However, in this realm, tempo is defined as motion and shot length. In a movie, dramatic story sections and events are signaled by a unique tempo (Adams et al., 2000).

Previous studies describe the various articulation and speech rates[3] of English and other languages, for example, those of Osser and Peng (1964), Gay (1981), Miller (1981), Vaane (1982), den Os (1985), Kohler (1986), Rietveld and Gussenhoven (1987), and Levelt (1989). While there are many models of speech rate in English, it is evident that further research is necessary to create a model of the modulation of tempo in English for articulation rate that can be incorporated into a rule-based synthetic speech system. The primary goal of this project is to model observed tempo modulations in Canadian English read speech for incorporation into a synthetic speech system.

A quantitative description of tempo modulations has important implications for both linguistics and industry. As a contribution to the field of linguistics, a generalization of tempo in English brings us closer to developing a language universal for the temporal organization of speech. Laver (1994) notes that rhythm is a highly

2. Theatrical speech exploits the prosodic elements of language to its fullest, in an effort to place importance or non-importance on specific elements of a text. In terms of tempo, important ideas are emphasized by deceleration and others are de-emphasized by acceleration.

3. The articulation rate describes the tempo of an utterance excluding pauses. The speaking rate describes the overall rate of performance including pauses.

complex phenomenon, which evades an adequate account because of the interference of variables in timing, and differences in accent, style, voice, tone, and rate of speech (p. 526). Therefore, a model of tempo modulation can lessen the complexity of analyzing rhythm by providing an accurate control measure. A model of tempo will also prove useful in the field of language acquisition, by providing language learners with valuable timing information to assist them in attaining native-like proficiency.

In industry, many advances have been made in improving the natural quality of synthetic speech. However, many synthetic speech systems still sound quite unnatural in terms of temporal qualities such as tempo and rhythm. If these could be predicted and replicated, synthetic speech would come closer to its natural target. The challenge lies in this prediction and replication. Tempo must be described quantitatively in order for any replication to be possible. This description can then be translated into the form of an algorithm, which can in turn be incorporated into synthetic speech software for various synthetic speech applications, such as text-to-speech systems.

There are two secondary objectives to this research. The first is to verify and validate a pattern of deceleration and acceleration. Focus is the centre of interest in the phrase affecting duration in the sentence. It has been established that the correlates of focus are fundamental voice frequency and duration (Brown and McGlone, 1974; Folkins et al., 1975; Weismer and Ingrisano, 1979; Eady & Cooper, 1986).

A focused word significantly increases in duration compared to an unfocused word (Cooper et al., 1985, Eady & Cooper, 1986). However, there is some debate as to whether or not the effects of focus remain confined to the focused word or extend to the rest of the phrase. In terms of duration, Eady and Cooper found that focus does not significantly affect other words in long sentences with 11 or 12 syllables (1986, p. 411-412). However, they did note that there is a slight trend for duration to decrease following a focus word located near the beginning of a sentence. In shorter sentences of other studies, this trend was determined significant (Folkins et al., 1975; Weismer and Ingrisano, 1979; Eady et al., 1986).

Therefore, there is some evidence of acceleration in tempo following the focus word in short English sentences. It has also been established that focus can cause the tempo of the focal word to decelerate. However, given the findings of Eady and Cooper (1986), this deceleration is restricted to the focal word. It has not been observed that deceleration occurs in the other words preceding the focus. Whether this portion of the sentence is affected by focus has not been shown. There is some evidence of this deceleration–acceleration tempo pattern in English, but it is not conclusive. An experiment designed specifically to test this pattern is necessary.

The second objective is to determine how this tempo pattern manifests itself phonetically, that is, what elements of the sentence are affected by the changes in timing. If the tempo pattern is to be replicated in synthetic speech, tempo-change sites that produce the desired deceleration-acceleration pattern for the model need to be designated. In order for acceleration to occur, the affected segments must compress, while deceleration involves the expansion of segments. In compression, a simple reduction in duration equally across all phonetic segments does not occur according to Gay (1981). He argues that a vowel is the primary site for compression or expansion (p. 150). At the level of the syllable, Campbell and Isard (1991) found contrasting results. Codas showed more compression than the nucleus or the onset in long syllables (p. 4). Given these findings, a model that equally expands or compresses all focus-affected segments would not be predicted to reflect natural speech accurately. The vowel or nucleus and the coda appear to be the two most likely candidates as sites for timing changes. Further testing is necessary in order to determine under what conditions the vowel and the coda are primary sites for tempo change.

This project involved two phases: the pilot study and the main study. In the literature review in Chapter 2, varying approaches to temporal problems are discussed. A description of the pilot study and the results are discussed in Chapter 3. These results provide the motivation for the second phase, the main study, which consists of two experiments. A detailed methodology for the experiments is discussed in Chapter 4. The first experiment tests for the deceleration–acceleration tempo pattern and the second tests for the potential timing-change sites.

When one designs experiments to determine the tempo pattern and sites for the required timing changes, one must consider potential factors that can contribute to variation in the data. Whiteside (1996) suggests that there is some evidence for speaker gender differences, i.e. women read at a slower rate and display greater variability in sentence duration than men do (p. 38). An obvious solution to overcome this variability is to control for gender. In both studies, common gender as well as dialect and educational background were criteria for the speakers, in order that the incidence of potential speech variations might be reduced.

In addition to gender differences, there also appear to be differences in tempo variations among speakers as Cedergren and Perreault (1994) concluded in their study (p.1089). Although data provided by a single speaker may show trends that support the hypotheses, the results would show a bias of that individual's tempo patterns. Campbell (1990) also recognizes individual variation in speech rate. In an account of speech rate changes, he notes that an inherent tempo alternation in the speaker's presentation of the text may account for some of the error of prediction of his model (p. 75). By examining the speech of multiple speakers, this bias is eliminated.

This project is based upon samples of English data derived from the recordings of male speakers only, because of the potential for a gender distinction in tempo. The data sample was obtained from the recordings of the read speech of one speaker for the pilot study and five speakers for the main study. The text used for the readings was carefully designed to ensure that the sentences were as homogeneous as possible allowing for valid comparisons of focus regions, vowels, and syllabic constituents.

Read speech and spontaneous speech differ in terms of complexity in variation of duration. The increased complexity of spontaneous speech obscures potential generalizations. Therefore, read speech was chosen as a data source. The size of this sample was sufficient to provide trends of tempo modulations and of potential sites for timing changes. The trends observed in the pilot study motivated the main study. Because multiple speakers provided the data, the sample size increased considerably and presented a rich data set for statistical analysis.

The experiments for the pilot and main studies all have the same design. All recordings are digitized to supply both a graphic and numeric means of analyzing the speech data. The resulting waveforms are used for the manual segmentation of temporal units for the tempo pattern experiment and segments for the rate-change-site experiment. There is a problematic factor inherent in this process, as any duration measurements rely on the notion of boundaries. The determination of these boundaries can be quite arbitrary when they define language units necessary to describe tempo, such as syllables[4] or in the case of timing-change sites, sound segments. There appears to be no solution to the arbitrary nature of boundary determination. However, the establishment of definite criteria for unit boundaries and the maintenance of a strict consistency in the determination of the location of

4. Syllables show prominence depending on whether it is the word or the rhythmic foot that is emphasized (Kohler, 1991, p. 259). Therefore, the syllable is considered the minimal unit for speech rate.

the boundary provide some resolution. Consequently, the accuracy of segmentation is improved.

During the segmentation process, the location of pauses was marked and the duration calculated. In synthetic speech, pauses warrant a separate description. Therefore, pause durations were subtracted from the measurements for all syllables and sentences.

Once segmentation was completed, measurements were obtained for the comparison groups of each experiment. For the tempo-pattern experiment, average syllable durations (ASDs) of the region preceding and including the focus word and of those following the focus were compared to a control to determine the proposed deceleration-acceleration pattern. An additional comparison was made in order to observe the effects of focus on the syllables preceding the focus word.

For the site experiment, vowel durations at slow and fast rates of speech were compared to the control. In addition, the durations of syllable constituents, onset, nucleus, and coda for both speech rates are also compared to the control to observe the effects of rate change on duration. All comparisons were statistically evaluated. The results of the tempo pattern experiment are described in Chapter 5 and are discussed and analyzed in Chapter 6. This chapter also includes a description of a basic model that can be incorporated into an algorithm for an existing speech synthesis system, based upon the generalizations observed in these experiments. In conclusion, Chapter 7 provides a summary of the results of the main study.

CHAPTER 2 Literature Review

In this chapter, duration models are discussed, followed by a description of some of the prevalent theories of variation in speech rate. It has been found that variation in speech rate is patterned, and that sentential focus might well be an impetus for a change in tempo. In addition to sentential focus, stress, semantics, and grammar also play a role in the realization of speech rate, as do the pauses within an utterance. Variables such as gender, dialect, idiolect, and pre-pausal lengthening can also affect tempo and must be considered. After a tempo pattern in English has been established, potential locations of a change of speech rate, such as the vowel and the coda, require examination in order that this pattern may be replicated in synthetic speech. A brief discussion of some perceptual issues such as the perception of tempo and the affect of pitch concludes the chapter.

Duration Models

Duration models are of two types; they are either rule-based, such as those proposed by Klatt (1979), Horne (1988), and Campbell (1990), or are corpora-driven models, examples of which include the work of Takeda, Sagisaka, and Kuwabara (1989), and Pitrelli (1990). Many rule-based systems use the model developed by Klatt. It predicts segment durations for American English speech from a set of inherent and minimum default durations. Each phoneme has an inherent duration for a neutral mode of speech and a minimum default duration, which is the minimum duration to which the phoneme can be theoretically reduced. Syntactic and co-articulatory effects are modeled by a set of rules that assign a duration value for the individual phonemes. In addition to pause insertion, these effects include the shortening and lengthening rules shown in Table 2.1.

Table 2.1
Phonemic Shortening and Lengthening Rules[a]

Shortening	Lengthening
Non-phrase-final	Clause-final
Non-word-final	Emphasis
Polysyllabic	Postvocalic voicing
Non-initial consonant	Aspirated plosives
Unstressed	
Clusters	

a. (Klatt, 1 979)

Notice that emphasized phonemes are subject to the lengthening rule. This rule applies to the emphasized vowel only and assigns an increase to the vowel's duration, and follows the results of studies by Bolinger (1972), Carlson and Granström (1973) and Umeda (1975). The inherent difficulty with this approach is that it does not take into account any effects of emphasis or focus that extend beyond the emphasized syllable as previously discussed. In addition, the model does not account for de-emphasis. De-emphasis has been shown to increase tempo in Swedish, French, and English, as noted by Fant et al. (1991a) in a study of temporal patterns of stress in these languages.

The structural and phonetic contexts of the phoneme determine what rules apply. Each rule carries a percentage value that is incorporated into the following duration formula that summarizes the model (Klatt, 1979, p. 294):

$$DUR = \frac{[(INHDUR - MINDUR) \times PRCNT]}{100} + MINDUR \quad (1)$$

The inherent duration of the segment is the value for INHDUR and the minimum duration of a stressed segment constitutes MINDUR. PRCNT is the percentage change established by rule application. The default for PRCNT is 100%. Rule application is a cumulative process. That is, the value of 100 is decreased or increased by each rule that applies.

Based upon Klatt's algorithm for American speech, Campbell (1990) presents a method of quantifying and recording changes in speech rate for British English read speech. He adds a new rule because of the observed overproduction of syllables in phrase initial position. The Phrase Initial Shortening rule is proposed to shorten phrase initial segments by 85% (p. 74).

In addition to rule-based models, corpora-derived models for speech synthesis have been developed by Takeda, Sagisaka, & Kuwabara (1989), and Pitrelli (1990). These models achieve similar results by statistically analyzing speech corpora and modelling the results.

In a review of both types of duration models, Carlson (1991) notes that the standard deviation for the error after the application of these models is 25 ms. He proposes that this degree of error may be attributed to the assumptions of the models. Firstly, stress is given a limited number of levels, i.e. stressed, reduced, or unstressed, and secondly, speech tempo variation within the clause or sentence is not addressed. Carlson and Granström (1986) compared the duration prediction to natural speech and found that the prediction error is a function of time. Carlson (1991) notes that these results could be interpreted as tempo change inside the phrase or the sentence, which has to some degree been accounted for by phrase-final or word-final lengthening rules for phonemes (p. 244). Both Carlson (1991) and Campbell (1990) suggest that parts of speech may account for some of the error. Campbell notes that parts of speech are not yet incorporated into the rules of his model, which would account for some of the error of prediction of fit (p. 77). According to Carlson, different duration rules could apply to affixes and bases, as well as to differing form classes, i.e. noun phrases vs. prepositional phrases (p. 244). Kaika, Takeda, and Sagisaka (1990) included parts of speech in their study and found that the duration of function words and content words were in fact different. However, Carlson argues that the syntactic position of the words may have elicited the difference.

In addition, the large error in fit may be explained by focus, as the effects of focus include changes in the duration of other elements in the phrase or sentence. Local duration models that concentrate on the individual segment fail to account for phrasal variation prompted by the intent of the speaker to emphasize an element in the phrase. Intonational models that change the pitch and duration of the tonic may account for the focus of the phrase, but the remaining elements of the sentence must also be considered. In a comparative perceptual study on synthetically vs. naturally generated intonation, Terken (1993) proposes an algorithm that includes pre-pausal lengthening and focal word lengthening. However, in this study the effect of focus does not transcend the boundaries of the focal word. Horne (1988) also proposes a rule-based model that account for focus; however, it is based on pitch changes rather than on changes in duration.[5]

A duration model that incorporates the effects of focus beyond the focused word could account for internal tempo change in the phrase or sentence as suggested by Carlson and Granström (1986) and, consequently, could improve the predictive power of the model, resulting in more natural sounding speech.

5.In this model, the stressed vowel with the most prominent value in the phrase determines the F_0 peak with the additional application of a 256-ms. frame for F_0 scope, which is centered on the prominent vowel.

Tempo Alternations

In an attempt to create a more natural sounding synthetic speech product, the tempo of an utterance must vary as it does in natural speech. In an analysis of Swedish text reading concentrating on the timing of vowels and consonants, syllables, interstress intervals and pauses, Fant and Kruckenberg noted that speed alternations add a degree of naturalness to the reading (1996, p. 4). This opinion is shared by Rietveld and Gussenhoven (1987) and Eefting (1988).

Several studies have shown that the rate of speech within a sentence is not constant. Variations do occur and can result in a decrease in syllable duration as noted in an early study by Kelly and Steer (1949). Miller et al. (1984) also found that the articulation rate in spontaneous speech varies greatly. In a comparison of runs of speech, a run being a stretch of pause-free speech, the authors noted that syllable durations change between runs. In a response to an interview question, speakers do not gradually accelerate or decelerate their speech but rather they alternately change their rate of articulation resulting in a multiple high-low alternation pattern in syllable rate (p. 221). However, because of the method of organizing the data into runs, which may or may not account for sentence boundaries, it is difficult to determine whether or not there is a variable such as a focused word that is responsible for the change in rate.

In a similar analysis of articulation rate and duration in read English speech by

Crystal and House (1990), this same high-low alternation was found, and this variability manifests itself in patterns that appear to be identical for different speakers. The variability of articulation rate was not random, but rather each run exhibited an alternating speech rate. Similar results were found by Eefting (1988), who notes that this variability is an inherent quality in natural speech. However, Eefting assumes that this variability is random and is not deliberate, as do Crystal and House, who observed that this variation may be a predictable function of the syllable complexity of a run (p. 106). In addition, stress characteristics are shown to be basic to ASD variability.

One form of this alternating pattern of speech rate is evident in a pilot study of speech and pause timing in different modes and tempi of Swedish read speech. In a study in 1991c, Fant et al. noted that the pattern involves only one deceleration followed by one acceleration in terms of average phoneme duration. A similar result was found earlier by Brubaker (1972) in a study of average speech rates of sentences in read monographs. Apparently, speech accelerates within sentences as it also does near the end of a paragraph.

Working in the larger context of the paragraph, Koopmans-van Beinum and van Donzel (1996) found that paragraph boundaries affect the rate of speech. The initial segment after a paragraph boundary shows a deceleration compared to those segments in intermediate position. They also found a correlation between focal points in the text exhibiting a slower rate and explanatory and commentary passages bearing faster tempi. Koiso, Shimojima, and Katagiri (1998) found related results in a study of dynamic speech rates as contextualization cues in Japanese. The opening of information is marked by a deceleration in speech rate, and the absence of an information opening is accompanied by an acceleration in tempo (p. 348).

Miller et al. (1984), Crystal and House (1990), and Eefting (1988), organizing their data according to runs of speech, have disproved the notion set forth by Goldman-Eisler (1968) that accounted for variation in rate as a result of pausing strategies. It was noted that the rate of articulation in terms of words per second does not seem to vary. As previously mentioned, the drawback with this method is the absence of potential sentence boundaries, particularly in the analysis by Miller et al. of spontaneous speech. Speakers may or may not employ a pause to indicate a sentence boundary. Given that a sentence could have one or more pauses, the speech rate of that sentence may alternate one or more times.

Focus as an Impetus for Tempo Alternations

Fant and Kruckenberg (1996) suggested that focus might be the impetus for speech rate variation. Any deviations in normal predicted durations are actually reductions and expansions around and within focal regions that are inclined to cancel each other within a sentence (p. 4). The high-low alternation pattern noted by

Miller et al. (1984), Crystal and House (1990), and Eefting (1988) could perhaps explain these observed reductions and expansions.

As mentioned in the introductory chapter, a word that is focused is significantly longer (Cooper et al., 1985). In addition, there are apparently no definitive studies that can confirm the extent of focus effects. It appears to be well established that the focused word experiences an expansion in duration. In addition, the portion of the sentence following the focused word undergoes a reduction as was observed by Fant et al. and Brubaker. However, it has not been established whether the portion of the sentence preceding the focused word also exhibits an expansion in duration or whether the expansions noted in studies such as Fant and Kruckenberg (1996) are solely the result of the focused word expansion.

As well as textual focal points, emphatic accents also have a lengthening effect. In developing a new method for automatic detection of syllable nuclei, Pfitzinger, Burger and Heid (1996) found that emphatic accents in spontaneous German increase the distances between nuclei (p. 4).

The affects of focus are not restricted to the domain of duration. Pitch is also affected by an emphasized element within an utterance. In Greek, Galanis et al. (1996b) noted that a correlation exists between the focus and an increase in F_0. The lowest value for F_0 occurred in region following the focus.

The Stress and Tempo Interaction

According to Crystal and House (1990), variability is a function of the stress and syllable characteristics of the utterance. For example, in fast tempo runs, they found that the proportion of stressed syllables and phones decreases as the tempo of the run accelerates, and in slow tempo runs, the proportion of stressed syllables or phones in the utterance increases. In addition, when comparing the number of phones per syllable for slow runs and for analogous fast runs, Crystal and House found that the average number of phones per syllable is fewer for the fast runs (p. 107).

Although stressed syllables affect the rate of a sentence, the reverse effect was not observed by Peterson and Lehiste (1960) in a previous study designed to describe the factors that condition the duration of syllable nuclei in spontaneous English speech. They discovered that variation in rate is found to have a negligible effect on the durations of syllable nuclei. Minimal variations were evident within an utterance, on the order of 3 to 5% of the total utterance. They concluded that variation in the rate of an utterance has little effect on the duration of the stressed syllables in the utterance (p. 699).

However, in a later study using a distinct reading mode, Fant and Kruckenberg found that stressed syllables increased more in duration than the unstressed syllables, which remained stable (1996, p. 4). This trend held for both lower and normal tempi. Fast speech provoked a greater reduction of unstressed syllables than of stressed syllables (p. 4).

The same effect was found by Port (1981) in his study of the combinatory effects of timing factors in American English read speech. An increase in the tempo of stressed syllables in relation to the remainder of the utterance resulted in less of a reduction in the stressed syllable compared to unstressed syllables. In addition, for a fast tempo, a neutralization of contrasts in duration occurred for segmental features partially because of articulatory limitations (p. 271). According to Crystal and House (1990), stress and syllable characteristics are not the only variables contributing to observed alternation in speech rate. Overall tempo, productive fluency, reading ability, and dramatic proficiency may provide minor variations (p. 107). In addition to the stress and syllable characteristic variables contributing to speech rate variation as proposed by Crystal and House (1990), variation may be explained in terms of semantics and grammar.

Semantics and Grammar

Looking at correlations between semantic properties and speech rate, Koopmans-van Beinum and van Donzel (1996) found that speech rate variations are related to global and or local information structures in spontaneous Dutch discourse. In 60% of the runs that immediately follow a paragraph boundary, the

ASDs were larger than the median. In addition, the first quartile of a story, the main topic of the story, exhibits runs with lower numbers of syllables, while the last quartile (where in this case the topic is expanded) shows higher numbers of syllables (p. 4).

In addition to a semantic analysis of the texts, Fant et al. (1991c) conclude that most of the variations in the tempo of the text can also be explained by a grammatical analysis, i.e. major clause boundaries (p. 256). For example, non-content words are associated with a faster rate than their content counterparts. This correlation was also noted in a later work by Fant and Kruckenberg (1996). In this study, local tempo appears to be influenced by the substantiality of content words and the consequent stresses within the text considering average segment durations within sentences or phrases (p. 4). A similar conclusion was observed in a description of durational models by Carlson (1991), who commented on the variation found in local speech tempo. Typically, there is a 25-ms. standard deviation in these models. In an effort to account for this variation, the author notes that segment duration could be correlated to parts of speech. For example, function words and content words contrast in terms of duration. In addition, syntactic function appears to play a role as well. In contrast, Umeda and Wedmore (1994) find that accented syllables in Japanese and English spontaneous speech are not always those that are content words or those that are lexically or phrasally stressed (p. 1097).

Pauses and Tempo

Pauses are affected by changes in speech rate. According to Fant et al. (1991c), pauses increase as the tempo of a sentence decreases and in distinct reading mode, pauses are double the length of those of a normal reading (p. 252), a finding similar to that of Grosjean (1979) in a study of timing in American Sign Language and English. An increase in pauses is largely responsible for the lower rate in the distinct mode (p. 252). This conflicts directly with Kohler's suggestion that pauses do not have an independent effect on speech tempo (1986, p. 137).

Inversely, does a change in articulation rate or a change in pausing by increasing the duration of the pause or quantity result in a change in speech rate? Conflicting results exist. Grosjean found that at a faster tempo, the speaker would alter articulation time more than pause time. This stands contrary to Bellugi and Fischer's (1972) findings for spontaneous speech. The rate of articulation also appears more influential than pause time according to Grosjean & Lane (1976) in a study determining how the listener perceives tempo. Previously, Lane & Grosjean (1973) found that when speakers double their rate of speech, pausing occurs half as often. Pause time then affects the global rate of speech (p. 4). However, in production studies of spectrographic analyses of tempo changes in Japanese, Vietnamese, Korean, and English, Han (1966) found that pauses are the most important factor in speech tempo (p. 73). Neither articulation rate nor pausing strategies alone account for speech rate variation; instead, a complex relationship exists between the two.

The studies discussed show that speech rate is not haphazard, but is patterned. These patterns seem to reflect a high-low alternation in speech rate, with focus or emphasis as a potential impetus for the alternation. In English, the tempo pattern within an utterance has not been established conclusively.

Variables affecting Duration

As previously mentioned, experimental design must consider potential factors that can contribute to variation in the data, such as gender differences (p. 5) and individual variation.

Individual variation is also found in spontaneous speech. Tempo appears to be speaker specific, as was found by Cedergren and Perreault (1994) in their attempt to predict syllable timing in spontaneous Montreal French. They proposed a speech rate normalization to aid their prediction of syllable timing. Apparently, normalization is necessary because syllable timing is found to be constrained by speech rate and prosodic organization (p. 1089). Speaking rate was calculated by taking the ratio of segment duration within the intonational phrase to the average duration of the corresponding segments.

The emotions and the dialect of the speaker also appear to affect speaking rate. A correlation has been made between emotions and speech rate in Scherer's (1996) review of recent research on the affective dimension of speech analysis. Speech rate is one of many acoustic parameters for emotion that was proposed, and it was discovered that a fast tempo correlates with joy and a slow tempo with sadness (p. 2). However, in Greek, joy is correlated with a decrease in tempo as is anger and fear (Galanis et al., 1996a). In addition, an increase in tempo is associated with grief. However, emphasized syllables or those in sentence final position appear to be exempt from such changes in duration (p. 1228).

In a study of the effects of gender and dialect on speech rate of read English speech from the TIMIT corpus, Byrd (1992) found that there is a significant effect of regional dialect on speaking rate. Two calibrated sentences read by 420 speakers were chosen from the corpus. These readings provided the data for the analysis of the speaker-dependent variables of gender and regional dialect, and the results showed that regional dialect had an effect on speech rate (p. 594).

Thus far the emotions, regional dialect, and idiolect of a speaker have been shown as factors contributing to variation in tempo. An additional factor is the type of text that the speaker reads. Speakers have been noted to adopt different speaking styles given different styles of text such as children's stories, news stories, weather, etc. In English, there is a much wider range of speaking rates for the ten different text types than in either Dutch or French in a study by Fackrell et al. (2000). The weather and frequently asked questions mark the highest and lowest boundaries of the range respectively (p. 5/6).

2.1 *Pre-pausal Lengthening*

In an analysis of speech rate, phrase final or pre-pausal lengthening also presents variation in measurements of duration. In Swedish read speech, Fant et al. (1991c) found that major clause boundaries affect speech tempo. Syllables in phrase final or pre-pausal position are lengthened (p. 256). In English, Nakatani, O'Connor, and Aston (1981) investigated the rhythm in reiterant speech. The study revealed that phrase final lengthening occurs in syllables preceding a phrase boundary, independently of stress location (p. 102).[6] In a study of American English read speech, Wightman et al. (1992) found that lengthening is relative to the perceived size of the boundary and that it occurs within the rhyme of the syllable (p.1716). Similarly, Campbell and Isard (1991) observed that final lengthening occurs in peak and coda position (p. 43). In addition, there are degrees of pre-pausal lengthening related to prosodic constituents. For example, an increasing degree of pre-boundary lengthening can signal four levels of prosodic constituents: the word, the accentual phrase, the intermediate phrase, and the intonational phrase (Wightman et al., 1992, p. 1716). In Japanese, Hayamizu and Tanaka (1994), using a statistical model for the recognition of speech rhythm,

6. However, Nakatani, O'Connor, and Aston (1981) did find a case where a penultimate stressed syllable exhibited a small degree of lengthening when the phrase final syllable is not stressed. They also found that stressed syllables in word-final position are lengthened more than those with no stress.

examine temporal patterns in read sentences. Average mora durations within utterances showed patterns of pre-pausal lengthening (p. 200).

It is interesting to note that the listener may not perceive pre-pausal lengthening. Hoequist (1983) describes "the order effect", where syllables that are towards the end of a syllable sequence are heard as shorter than they actually are.

Several studies confirm that in English as well as in other languages, pre-pausal or phrase-final lengthening is a variable that affects duration. Consequently, it is important to consider this phenomenon when accounting for tempo variations. According to Carlson (1991), the incorporation of word, phrase, or clause-final lengthening rules into duration models can partially account for the standard deviation of the models in his study.

Locations of Speech Rate Change

Given that tempo is patterned, how does the pattern manifest itself phonetically? Do the speech rate changes affect all phones of the phrase equally or are some locations more elastic than others? Hoequist and Kohler (1986) note that expansion and compression of the speech signal is not manifested as a proportional change at the segmental level (p. 7). However, in an experiment comparing prosodically structured segments with sequences of syllables in which all acoustic segments were proportionately changed in duration, there was no difference between sequences created to model natural acoustic segments and the equally proportioned sequences (Hoequist and Kohler, 1986, p. 38). Port (1977) found that no temporal structuring is evident when tempo is slowed; however, an increase in rate results in restructuring (p. 71).

2.2 *The Vowel as a Site for Tempo Change*

Crystal and House (1990) found some evidence of vowel compression in the stressed syllables of their analysis of read American English. However, there was not the strong evidence of compression that was found previously by Han (1966) for Japanese, Vietnamese, Korean, and English in determining what locations a speaker exploits to change tempo in read speech. By comparing duration measurements, he discovered that there is a higher rate of reduction in stressed vowels when tempo is accelerated in English. For example, if a phrase compresses by 60% the vowel can compress by 45% (p. 78). The same result was not found in unstressed environments. Han observed that unstressed sounds do not undergo a significant change when the tempo is accelerated (p. 78). However, these studies are not conclusive as they are based upon small corpora.

In an investigation into the quantitative characteristics of qualitatively similar English and Finnish vowels under different tempo conditions, Marjomaa (1983, p. 46) also found that rate changes affected the duration of the vowel. The stressed vowels under study were embedded into a standard syllable template within frame sentences. These sentences were read at three rates of speech: slow, normal, and fast. The vowels were then segmented and measured. Statistical results showed that changing the tempo creates a duration difference in the stressed vowel (p. 46), that of expansion and compression (p. 41). Expansion has also been observed by Kessinger and Blumstein (1998) in their investigation of the effects of speaking rate on speech production and perception. They found that as speech slowed, both voiced onset time and vowel durations expand (p.125).

Similar findings are noted by Gay (1981) in his paper arguing against the theory that changes in speech rate are a result of a horizontal time compression mechanism. The horizontal time compression mechanism implies that reduction in duration for a fast speech rate occurs equally across all phonetic segments. In addition, the timing function causes all changes in the dynamic properties of speech movements, and co-articulation remains constant across any change in speaking rate (p. 1). Gay notes that changes in rate for American English are more likely to occur at the vowel as opposed to the consonant (p. 150), although it is not clear whether these changes occur for all vowels regardless of stress. However, vowel-undershoot, or incomplete articulation, occurs during a fast rate of speech (p. 152), which could account for the reduction in duration. Nooteboom (1991) counters this finding proposing that vowel shortening due to increased tempo does not lead to a vowel reduction (p. 234), a perspective that Kohler (1991) finds unjustified, as the data for this study was not a comprehensive enough language sample for such a categorical exclusion (p. 261).

In determining the effect of tempo on stressed and unstressed syllables, Peterson and Lehiste (1960) found that the duration of stressed syllables experienced less change than the duration of unstressed syllables with a faster tempo (p. 699). Port (1981) found similar results in his study of the combinatory effect of linguistic timing factors in American English. However, he found that the vowel and the consonant of a VC syllable are shortened in a stressed syllable, but to a lesser extent than in the unstressed portions of the sentence. (p. 271). If the vowel is the primary site of tempo change as Gay suggested, then Peterson and Lehiste's and Port's findings contradict those of Han (1966) and Marjomaa (1983, who proposed that stressed vowels are the primary site for change.

These studies show that vowels can be considered a location for speech rate change in American English, as well as in Finnish. However, no studies have been done on Canadian English, although one would suspect similar results. In addition, it does not appear to be conclusive that only stressed vowels undergo this change. There is an inconsistency in previous studies regarding stress and vocalic changes given different rates of speech. Further study is therefore necessary to determine whether only stressed vowels undergo duration changes when speech rate is altered in Canadian English.

While vowels may vary in duration with changes in rate, they can undergo different degrees of change dependent on whether or not they are tense or lax.

According to Records (1982), long or tense vowels reduce in duration more than their lax counterparts (p. 6) a tendency confirmed by anecdotal observations of comparisons of the durations for tense and lax vowels in the pilot study of this experiment.

2.3 *The Coda as a Site for Tempo Change*

Vowels are not alone in being affected by rate changes, as noted in Port's (1981) study. Gay (1981) found that post-vocalic consonant closure duration reduced proportionally more than prevocalic consonant duration for fast speech (p. 150). This implies that within a syllable, a coda undergoes a greater degree of rate change than does the onset. Again, the question of stress arises. It is not clear if this statement is based upon all consonants regardless of stress.

A similar result is discovered in a later study of read British English examining to what extent normalized phoneme durations taken from sentences in the corpora is uniform throughout the syllable. Campbell and Isard (1991) compared the relative compression and expansion of different segments with regard to their position in the syllable and the phrase. They found that at a fast rate of speech more compression was evident for the coda in long syllables than the onset or the nucleus (p. 4). Long syllables are those syllables whose average of the normalized values of the syllable's constituents is greater than one. The fact that the nucleus, i.e. the vocalic element, shows less compression than the coda contradicts Han and

Gay's findings, who propose that the vowel is the primary site for an increase in tempo. Also, there appears to be less shortening in the stressed syllable than in the sentence as a whole (p. v).

An advantage of Campbell and Isard's experimental design is that they account for syllable length. By comparing syllables of similar types, the comparison becomes more robust. However, a further distinction of syllable types, i.e. CV, CVC, V, and VC may result in a more accurate distinction. In addition, the results may differ by factoring for stress.

Codas are not the only consonants affected by rate changes. It has been observed by Wayland et al. (1994) in a perceptual experiment on American English, that stop consonants within an open syllable show longer voice-onset-time (VOT) values, a 44% increase in the mean, when the rate of speech is slow (p. 2699). The experiment compared the effects of syllable-level and sentence-level speaking rate on phonetic perception where informants judged the goodness of target syllable VOTs that varied in length, in isolation and within frame sentences.

Taking prior studies into consideration, it would appear that both vowels and codas are potential locations for speech rate changes; however, further research is warranted to confirm the results. Because stress has well-established effects on duration, it must be considered a factor in further studies. In addition, a formal analysis of the effects of tenseness in the vowel may be necessary to determine the validity of this refinement of taking the vowel as a site.

When considering codas as site for compression, a greater degree of distinction of syllable types may improve the value of the comparison between syllabic constituents, in addition to factoring for stress. A further refinement of consonant sites such as codas is possible if the effects of speech rate changes on VOT in speech production are established.

The Perception of Speech Rate

The perception of tempo is an important consideration in proposing a model for tempo alternations. A model based on observational work should be tested in order to ensure that the listener can perceive the resultant changes in speech. This section provides a brief discussion of some of the findings of perceptual studies, while highlighting the factors that would be important in developing a model.

2.4 *The Listener's Tempo Judgements*

In a perception test of Dutch, English, French, Spanish, and Moroccan Arabic texts read at three different speech rates, Vaane (1982) found that listeners are not influenced by their knowledge of the language in judging rates of speech. She suggests that listeners receive their main cues for tempo detection from the temporal features of the speech signal as opposed to lexical information (p. 146). A similar result was discovered by den Os (1985) in a study of Dutch and Italian text and listeners; however, it was found that a lack of fundamental frequency information in a foreign language made the rate judgements more difficult In addition, prosodic information such as the orthographic syllable, the phonetic syllable, and phonetic segments can provide rate information (p. 132). For example, this prosodic information showed a high correlation with the perceived rate of normal and monotone Dutch and Italian utterances.

Miller and Grosjean (1981) investigated listener adjustment for articulation rate and pause rate when the listener processes phonetically pertinent information. Their findings showed that changes in articulation rate have a greater effect on phonetic judgements of /b-p/ distinctions than do changes in pause rate (p. 211). Articulation rate was also discovered to be the most important variable in a listener's determination of global speech rate in an experiment of read English by Grosjean and Lane (1976). In their model, speaking rate is a function of its components, articulation rate, duration, and frequency of pauses, all of which are considered to be independent (p. 538).

Given that articulation rate and duration are, in part, responsible for speaking rate, what degree of change in the components of an utterance is necessary for a listener to determine a change in rate? Hoequist and Kohler (1986) discovered that a 17-ms. change within an acoustic segment or syllable is the minimum for the determination of a tempo difference (p. 39).[7] Although the minimum change for the determination of a tempo change is 17-ms., a difference of 60-ms. per syllable establishes a change in tempo (p. 13). Taking these results into account, any tempo changes in the model should range between 17 to 60 ms.

2.5 *Tempo Perception and Pitch*

Listeners perceive changes in duration and changes in F_0 in a task-dependent manner (Hoequist & Kohler, 1986, p. 42). In experiments determining the listener's ability to divide sequences of syllables into groups based upon F_0 and durational differences between groups, duration and pitch change are proven organizational cues (p. 44). The widest possible domain for a listener to establish a rate of speaking was suggested to be the utterance, and was, in some cases, due to the order effect and the effects of utterance structuring (p. 26).

Den Os (1985) found that fundamental frequency information is important to rate judgments, as did Kohler (1986) in his model of speech rate perception of read German speech. Kohler proposed that, in the first instance, fundamental frequency peaks signal feet (the principle temporal organizer suggested by Hoequist (1984)), and then duration structuring within the foot determines tempo.

According to Hoequist and Kohler (1986), a foot structure that is based on duration causes it to sound slower only at the level of the foot and not beyond (p.

7. Hoequist and Kohler (1986) found the change to be primarily within the quasi periodic portion of the signal.

49). Pitch structuring also causes a perceived slowing, however there is no difference in the effects between local and global structuring (p. 50). In fact, any structuring of an utterance may cause a perceived slowing (p.52).

In addition to Kohler, Rietveld and Gussenhoven (1987) found a correlation between pitch level and perceived speech rate. Kohler correlates a high fundamental frequency with a fast tempo and a low fundamental frequency with a slow tempo, supporting a previous experiment for perceived speech tempo based on non-speech stimuli by Hoequist and Kohler (1986). In a major study of the perception of speech rate in German, Hoequist (1984) found that this effect is intensified when there is a change in the pitch contour of the syllable regardless of the direction (p. 164). Kohler found that in slow utterances, fundamental frequency might drop across the consonantal periphery instead of inside the nucleus. In addition, fundamental frequency movement cues that occur in utterance final position decrease or increase in speech rate respectively (p. 134). The proposed hierarchy of variables in the perception of speech rate is then duration, followed by overall fundamental frequency, followed by fundamental frequency movement.

However, the perception of duration is not without its complications. Hoequist (1984) made a very interesting observation on the perception of durational differences. Given a difference in duration between two identical syllables, listeners perceive a difference between the two syllables but not one of a durational nature. Conversely, a non-durational difference between these syllables such as a falling pitch is perceived as a durational change (p. 161).

In sum, the duration models discussed do not account for tempo pattern in English, particularly one that decelerates the focused word and the pre-focal portion of the sentence and accelerates the remainder. With regard to the replication of this pattern, the vowel and the coda have both been identified as potential locations for tempo change. It is important to note that changes in tempo need to meet perceptual criteria. Because of their effect on tempo, controls for pauses will be included in the study. In addition, an experiment must account for all variables, such as gender, dialect, idiolect, and pre-pausal lengthening, as they all have been shown to have effects on duration.

CHAPTER 3 *Pilot Study*

The purpose of the pilot study was to determine if the previously described deceleration-acceleration tempo pattern is evident in English. In addition, multiple sites for potential rate changes in English were tested: vowels and the codas of syllables. In addition, the tenseness of specific vowels and the VOT of stop consonants were also tested as possible future refinements to the model. However, these sites will not be discussed in this chapter. By testing whether or not rate changes affect the duration of these sites, potential locations for sources of speech rate variation can be determined. A script was purposely designed with sentences that provide these phonetic occurrences and numerous focus locations.

When previous studies of vowels (Peterson & Lehiste, 1960; Han, 1966; Port, 1980; Marjomaa, 1983), and codas or consonants (Gay, 1981; Campbell & Isard, 1991) are considered, it becomes evident that both vowels and codas are potential locations for speech rate changes; however, further research is warranted to confirm the results.

The purpose of the pilot study was to determine whether or not the data could demonstrate that the sentence section following the phrase-initial syllable is decelerated and the section following the focused word minus the phrase-final syllable is accelerated (the tempo hypothesis). To determine the sites of speech rate change necessary to replicate this speech rate pattern, the following hypotheses were tested:

- Vowels will increase or decrease in duration as rate increases or decreases respectively. As previously mentioned, factoring for stress is required in order to ensure an accurate comparison, in addition to resolving the conflict apparent in previous studies regarding stress.

- The coda compresses more than the onset or peak of the syllable at a fast rate. To increase the accuracy of the test of this hypothesis, syllables are marked for stress. In addition, they are categorized and the resulting categories are used as a basis for comparison.

Because the proposed tempo pattern involves both acceleration and deceleration, the inverse of these hypotheses will also be tested to increase the possibilities of both compression and expansion of vowels and syllable constituents. Canadian English and American English have many similarities and it would be expected that the results obtained in the studies of American English could be substantially replicated in Canadian English.

Method

3.1 *Materials*

As a source of data for this study, five sentences were designed to provide phonetic environments for tense and lax vowels, codas and VOTs, and to provide focus locations (the sentences are listed in Appendix A). In addition, voiced and unvoiced segments were alternated in the design of the sentence as much as possible to make the distinction between segments more apparent. This provides a clearer demarcation for the segments, which increases the accuracy of segmentation.

3.2 *Participant and Procedure*

A male participant, a native English-speaking professional with a post-secondary education, read the sentences. He was unaware of the hypotheses of the study. The speaker was asked to provide, in a soundproof room, seven readings of each sentence three times to increase the sample size. The first and fourth readings were at a normal rate of speech with neutral focus, which provided the control conditions for both the site location and sentence pattern experiments. The second and third readings were read at slow and fast rates respectively, and the fifth, sixth, and seventh readings were read emphasizing three specified focal points in the sentence.

The 105 sentences were recorded with a Sony digital audiotape (DAT) recorder, and were then digitized using Multispeech, Model 3700 version 2.2 by Kay Elemetrics Corp. at 20,000 samples per second. Spectrograms were then generated for each sampled data file.

3.3 *Segmentation*

After the sentences were digitized, sentence boundaries were marked at the point of amplitude change from a constant, low energy reading. In addition, all pauses were marked and their durations were subtracted from any segment, syllable, or sentence measurements. The boundaries between segments were determined from waveform readings. The change from unvoiced to voiced or vice versa at the zero crossing constitutes a boundary, which is verified by the spectrogram.[8] In the few cases where there was no voicing alternation, the spectrogram provided the necessary information for segmentation. Auditory checks were used as required in addition to spectrographic verification. All measurements were derived from the time values (in seconds) of these boundaries, and were tabulated using Microsoft Excel 97 SR-2.

8.The point at which the wave crosses the zero amplitude line on the waveform display is the zero crossing.

Separate Experiments

The specific details for the tempo pattern, vowel, and coda experiments are described below. A hypothesis statement precedes the description for each of the three experiments.

3.4 *Tempo Pattern Experiment*

Tempo Pattern Hypothesis:

The sentence section following the phrase-initial syllable up to and including the focus word is decelerated, and the section following minus the phrase final syllable is accelerated.

In measuring sentence length, the initial and final syllables were excluded because of sentence boundary effects, i.e. phrase-initial shortening (see Campbell, 1990) and phrase-final lengthening (see Crystal & House, 1988). Therefore, sentence length equals the duration from the zero crossing of the onset of the second syllable to the zero crossing of the offset of the penultimate syllable.

The zero crossing of the start of the onset or nucleus segment and the zero crossing at the end of the nucleus or coda are the syllable's boundaries. These boundaries were marked for stress determined both by ear and by the shape of the waveforms. The focused syllables were also designated. The pre-focal section of the sentence consisted of all syllables following the phrase-initial syllable up to and including those of the focus word. The post-focal section included the remaining syllables except the ultimate syllable.

After determining the sentence length, syllable boundaries, and pre-focal and post-focal sections of the sentences, the ASDs were calculated for latter. The total duration of the pre-focal and post-focal sections was divided by the number of syllables of the same section. These ASD measurements were then normalized using a ratio of syllable duration to sentence duration, countering the effects of varying sentence lengths. The average syllable durations for pre-focal and post-focal sections of the sentence were then compared.

3.5 *Vowel Experiment*

Vowel Hypothesis:

Vowels will increase or decrease in duration as rate increases or decreases.

Without stress as a factor, the durations of vowels at slow and fast rates were compared to the vowels of the control sentences first. Then both stressed and unstressed vowels were compared at the different tempi within their stress categories.

3.6 *Syllable Constituency Experiment*

The Coda Hypothesis:

The coda compresses more than the onset or peak of the syllable at a fast rate.

The syllables were first divided into the syllable types provided by the primitive syllable inventory of vowel (V), vowel and consonant (VC), consonant and vowel (CV), and consonant, vowel, and consonant (CVC). Within these groups, the durations of the syllable's onsets, peaks, and codas were compared to those of the control sentences for slow and fast rates. Three comparisons were run for the three variables for stress: no stress distinction, stressed and unstressed.

Statistics

The purpose of the pilot study was to look for trends in the data that support the hypotheses. Because of the sample size, no conclusive generalizations could be drawn from the data sets; however, tendencies did become evident. SigmaStat for Windows version 2.03 was used to calculate descriptive statistics. The results of these statistics were used for analysis.

Results and Discussion

In the following section, a description and discussion of the experimental results for the tempo pattern experiment are considered. In addition, both experiments for potential sites for speech rate change will be discussed separately.

3.7 *Focal Regions*

Pre-focal Regions

The first section of the hypothesis claims that the initial portion of the sentence up to and including the focal word exhibits a deceleration. For all sentences, comparisons were made between the ASDs of each of the three pre-focal regions and the control, the corresponding regions in the unfocused sentences based on the results in Table 3.1. Because of reading errors by the speaker, it was not possible to make accurate boundary determinations. In these instances, the data were not included in the sample, resulting in missing values for the data set. In this analysis, there were 16 cases missing, resulting in a total of 164 out of a possible 180 data points.

Table 3.1
Pre-focal Region ASDs

	Control			Focus		
Position	*n*	*M*	*SD*	*n*	*M*	*SD*
1	13	62.6	8.8	15	81.5	20.6
2	13	66.7	71.9	14	77.0	16.5
3	13	67.2	10.2	14	69.3	12.6

The ASDs for all pre-focal regions show increases in duration descending in value as the pre-focal region extends towards the end of the sentence. The ASDs increase 30% from the control for Focus 1 regions, 15% for Focus 2, and only 3%

for Focus 3 (see Figure 3.1). This large increase clearly supports the deceleration portion of the hypothesis. For Focus 1, with the focal word located near the beginning of the sentence, an expansion in the ASDs is evident before and including the focused word. The ASDs for Focus 2 exhibit the same expansion pattern, but to a lesser degree. Possibly because of the sample size, the extremely high standard deviation for the control warrants a cautious interpretation of the results (*SD* = 71.9). The increases in ASDs of the Focus 3 region are minimal and do not lend significant support to the hypothesis.

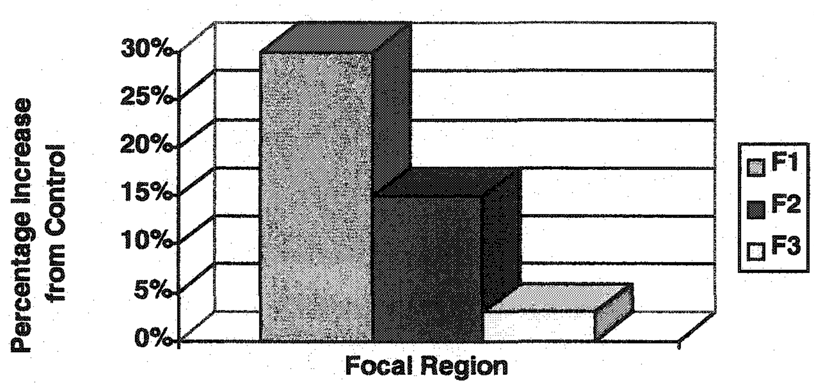

Figure 3.1.
Comparison of the Increases in the Means of ASDs for Pre-focal Regions

As previously mentioned, Cooper et al. (1985) established that focus increases the duration of the focal word. However, they found no other effects in the phrase from this focus. One would expect then, that the percentage increase evident in the ASDs of the pre-focal region would be equal to the increase of the focal word given there should be no other focal effects in the phrase.

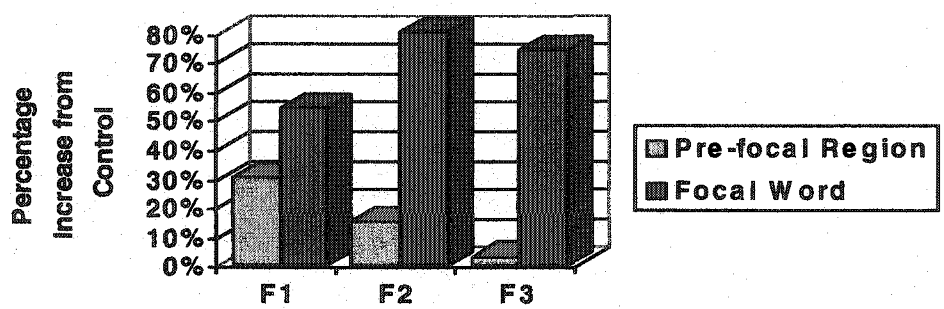

Figure 3.2.
Comparison of the Increases in the Means of ASDs for Pre-focal Regions and the Respective Focus Words

In Figure 3.2, a comparison based on the results in Table 3.2 is drawn between the increases apparent in both the pre-focal region including the focal word and the isolated focal word. Out of a possible two hundred and seventy data points, 208 comprised the data sample for analyzing the focus and the region preceding it. These data are highly variable given the standard deviations. Clearly, there is a difference between the increase in the focal region and the focal word. The focal word increases to a much greater degree than the pre-focal region. It appears that the syllables other than the focal word must be decreasing in duration in order to account for the smaller increase in the pre-focal region.

This tendency for the non-focal syllables in the pre-focal section to compress is borne out in the data as seen in Table 3.3. With the exception of the two percent

Table 3.2
Focus Word ASDs

	Control			Focus		
Position	*n*	*M*	*SD*	*n*	*M*	*SD*
1	13	76.2	22.7	15	117.7	52.8
2	13	67.3	24.4	14	121.4	39.6
3	13	88.3	25.4	14	153.9	55.4

increase in the Focal 1 pre-focal region, there are 9% and 12% decreases for the second and third pre-focal regions respectively. It is clear that there must be some compression occurring in this portion of the sentence in order for the focal word to increase to a so much greater degree than the Focal 1 pre-focal region. Considering that the standard deviations for the control and the Focal 2 pre-focal region are large (*SD* = 15.9, 19.0 respectively), a two percent increase is insignificant. In these data, a trend of compression in the non-focused pre-focal syllables is the apparent expansion of the effects of focus beyond the focal word. In addition, the pattern of this compression compensates for the pattern of expansion: as the ASDs for the pre-focal regions increase, the ASDs for the non-focused syllables in the same region decrease. This is evident in the pre-focal regions given the three positions of the focus in the sentence.

Post-focal Regions

The second part of the hypothesis claims that the region of the sentence following the focal word is accelerated. To test this claim, differences must be deter-

Table 3.3
Pre-focal Region minus Focus ASDs

	Control			Focus		
Position	*n*	*M*	*SD*	*n*	*M*	*SD*
1	13	53.6	15.9	15	55.5	19.0
2	13	66.9	11.0	14	61.0	10.5
3	13	64.4	7.9	14	56.9	6.8

mined between the ASDs of the post-focal regions and those of the control. Comparisons were run between the ASDs of each of the three post-focal regions and the ASDs of the control.

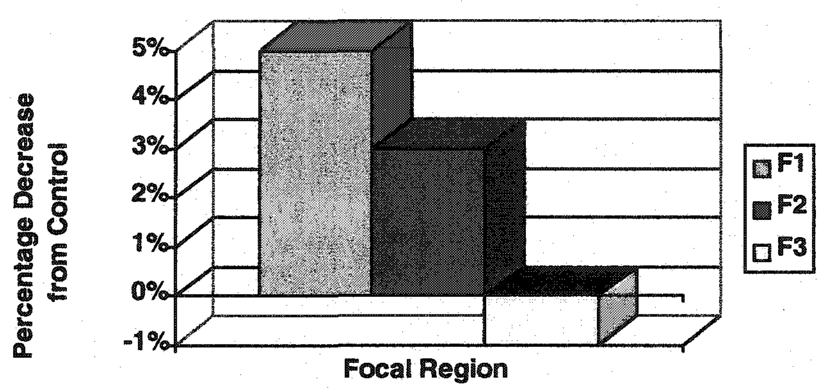

Figure 3.3.
Decreases in the Means of ASDs for Post-focal Regions

A trend of deceleration is evident when one compares the means of these groups, as seen in Figure 3.3. For the Focal 1 post-focal regions, the ASDs in the focused versions show a five percent decrease from the control based on the results in Table 3.4. A marginally smaller decrease is evident in the comparison for Focal 2 post-focal regions, at three percent. However, no decrease is evident in the third focal region. Instead, a one percent increase is apparent, but this difference between the control and the Focal 3 post-focal region cannot be considered significant given the large standard deviations of these groups (*SD* = 11.9, 15.9, respectively). It is interesting to note that the declining trend in the changes of ASDs that was observed in the pre-focal region comparisons is not repeated here. The data more closely replicates the changes evident in the ASDs of the unfocused syllables in the pre-focal regions. Apparently, the post-focal region of the sentence does accelerate to a small degree.

Table 3.4
Post-focal Region ASDs

	Control			Focus		
Position	*n*	*M*	*SD*	*n*	*M*	*SD*
1	13	64.0	10.2	15	61.1	12.7
2	13	66.4	20.5	14	64.7	23.6
3	13	51.8	11.9	14	52.5	15.9

To summarize, there is a clear trend in the data of a pattern in tempo when considering the differences of means in the data, lending support to results discovered previously for Swedish by Fant et al. (1991c). Tempo alternations of one deceleration followed by one acceleration occurred within the sentences of this data. Of interest is the apparent compression found in non-focal syllables in the pre-focal sections of the sentences. It appears that when the duration of the focal word increases, the rest of the sentence compensates by compression in an effort to maintain some predetermined overall sentence duration.

In addition, the location of the focal word seems to affect the degree of deceleration in the pre-focal regions. If the focal word is in the beginning of the sentence, the degree of expansion is prominent; however, as the focus moves from the middle to the end of the sentence the degree of expansion lessens. It seems that the natural die-off of the sentence affects the focus as well. The tempo pattern described in the hypothesis is evident in these data. Where exactly this expansion and the compression of syllables preceding and following the focus occurs must be determined.

3.8 *Vowel Duration*

This experiment considers the vowel as a potential site for speech rate change. There were a total of 496 data points out of a possible 576 for the data set. The reduction in data points is due to reading errors similar to those that occurred during the focus readings. The difference in means for both groups shows the expected expansion in the slow tempo comparisons and the expected compression in those for fast tempo. Given the results in Table 3.5, expansion is evident in the increases of the means for vowels regardless of a stress distinction with a 104%

increase over the control. The large standard deviations for the slow data may be explained by the variable nature of slow speech. However, the compression is minimal, there being only a three percent decrease from normal.

Table 3.5
Vowel Durations by Stress and Rate

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	149	78.1	32.1	170	80.9	32.6	177	165.0	80.7
Unstressed	69	58.2	23.7	74	60.3	24.3	75	134.8	72.2
Stressed	80	95.3	28.2	96	96.8	29.1	102	187.1	79.6

When the data are segregated by stress, the mean for vowels at a slow rate is considerably larger than the mean for the control (Slow: 187.1 ms., Control: 96.8 ms.) as seen in Figure 3.4. There is a 93% increase in stressed vowel duration at a slow speech rate. Therefore, given this sample of data, one can infer that stressed vowels do appear to expand in duration as rate increases, and that they expand to a large degree.

The same degree of change does not hold for the comparisons of stressed vowels to the control at a fast rate of speech. The mean for stressed vowels at a fast rate is smaller than the mean for the control (Fast: 95.3 ms., Control: 96.8 ms.) as seen in Figure 3.4. However, at a fast rate of speech, there is only a two percent

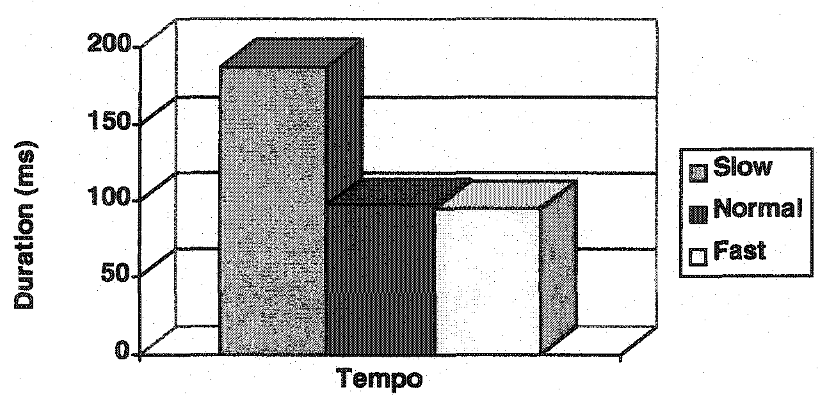

Figure 3.4.
Rate Comparison of the Means for Stressed Vowel Duration

decrease from normal in the duration of stressed vowels. Given the results, stressed vowels show only minimal compression at a fast rate of speech.

Individual speech rates might account for this lack of difference. The participant's normal overall speech rate might be fast resulting in less differentiation between normal and fast tempi. This is evident when comparing the means of the sentences in Table 3.6. For this participant, the difference between the sentences at normal and fast rates of speech is only 16% compared to the 70% increase for a slow tempo. In addition, the sample size is probably not large enough to reflect accurately the potential difference between normal and fast.

Like stressed vowels, unstressed vowels also reflect expansion, but the degree of expansion is much larger (Slow: 134.8 ms., Control: 60.3 ms.). At a slow tempo, the mean for unstressed vowels is much larger, showing a 124% increase over that

Table 3.6
Sentence Durations

	Speech Rate		
	Fast	Normal	Slow
n	13	13	14
M	2407.4	2875.7	4894.2
SD	813.7	414.4	543.9

for the control (as seen in Figure 3.5). In stressed vowels, this increase in expansion is 31 points less at 93%.

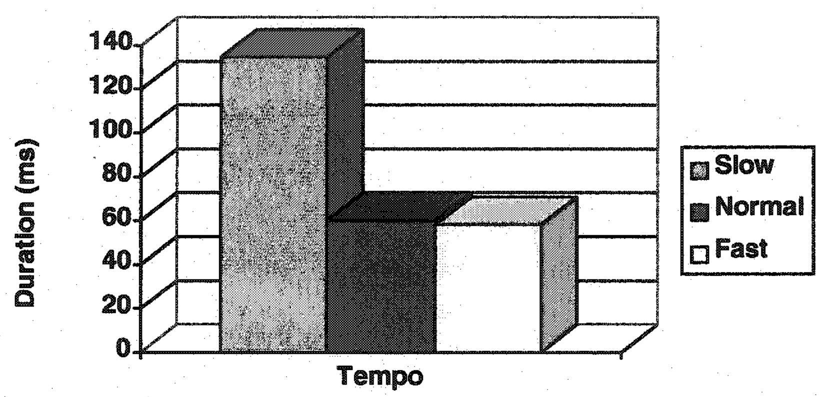

Figure 3.5.
Rate Comparison of the Means for Unstressed Vowel Duration

While unstressed vowels may show a higher degree of expansion, there is little difference in compression between stressed and unstressed vowels. As seen in Figure 3.5, unstressed vowels decrease in duration by three percent from the control (Fast: 58.2 ms., Control: 60.3 ms.), only one point more than the compression found in stressed vowels.

In summary, the data suggest that unstressed vowels do appear to increase in duration as rate decreases and to a much greater degree than do stressed vowels. Therefore, the hypothesis that vowels will increase in duration as rate decreases can be accepted. This increase is considerably larger than the increase found overall in the sentences. Slow sentences expand by 70%, while stressed and unstressed vowels expand by an additional 23% and 54% respectively. Clearly, vowels expand at a slow rate of speech, but do so at the durational expense of other elements in the sentence.

The fact that stressed vowels show expansion supports Marjomaa's findings of expansion in stressed vowels (1983). While the present study suggests that vowels could be a potential site for a tempo change from normal to slow, it provides a further refinement of the stress factor. In expansion, unstressed vowels appear to be the primary site of tempo change leaving stressed vowels as secondary.

Although the data support the notion of vowel expansion at a slow tempo, there is little evidence supporting the notion of compression. As rate increases, stressed and unstressed vowels decrease insignificantly in duration; the hypothesis that vowels will decrease in duration as rate increases cannot be accepted. However, a degree of compression is evident. The degree of compression is not comparable to the decreases apparent in the fast sentences. The fast sentences decrease by 16%, while the stressed and unstressed vowels show decreases of 2% and 3%

respectively. Clearly, additional compression must occur at other locations in the sentence.

Nevertheless, this study does not conclusively support previous findings of compression (Han (1966), Port (1980), and Marjomaa (1983). The results would need to be statistically significant to provide conclusive confirmation of vowel compression at a fast rate of speech. It is possible that a larger sample using multiple speakers may produce different results, as this would generate data with differing ranges among the three tempi.

However, there does appear to be a greater degree of compression in unstressed syllables than in stressed, which concurs with Peterson and Lehiste's findings, and Port's conclusions that unstressed syllables compress more than stressed syllables. The findings in this experiment do not present a definitive resolution of the conflict apparent in previous research regarding which group compresses to a greater degree, stressed or unstressed vowels, but it does provide weak support for the latter.

3.9 *Syllable Constituency*

Given the findings of the previous experiments, one could suggest that vowels have the potential to be a site for tempo modulations; however, consonants within the syllable also exhibit expansion and compression patterns. All the means and the corresponding standard deviations, from which all expansions or compressions are derived, are reported in Appendix B. The data sample consisted of 957 data points from a possible 1089 because of reading errors.

Similar to the findings of the previous experiment, at a fast rate of speech vowels do not show a distinctive difference between the means of syllable constituents and those of vowels in the control sentences across all syllable types. However, the coda and onset consonants compared do show sizeable differences from the control.

Overall, at a fast rate of speech, codas decrease five percent more than onsets and 16% more than peaks (Table 3.7). The results lend support for the hypothesis that the coda compresses more than the onset or peak of the syllable for a fast speech rate. In addition, peaks appear to be more static for this direction of tempo change, decreasing by only four percent, while onsets are at midpoint decreasing 11% more than peaks.

Table 3.7
Percentage Decrease in Total Syllable Constituents Comparing a Fast Tempo to the Control

	Syllable Constituents		
Stress	Onsets	Peaks	Codas
Combined	15%	4%	20%
Unstressed	15%	5%	28%
Stressed	32%	3%	15%

It was observed previously that compression is likely to occur elsewhere in the sentence to account for the vowels not maintaining the same degree of compression predicted by sentence compression. In part, these results can resolve this difference. Codas compress by 20%, which is greater than the 16% compression of the sentence. Clearly, there are other durational compensations occurring within the sentence, as the degrees of compression in onset, codas, and vowels do not equal the total sentential change in duration.

The results of the investigation of the comparative duration of syllable constituents support the previous findings by Campbell and Isard (1991) that at a fast rate of speech more compression is evident in the coda than in the onset or in the nucleus. However, a higher degree of compression was found in codas of VC and CVC syllables than in either onsets or vowels of other syllable types (as seen in Figure 3.6). In addition, codas in VC syllables increase nine percent more than those in CVC syllables. The compression of onsets also shows considerable variation between syllable types, 21% for CV syllables and 11% for CVC syllables. Compression is higher for codas in VC syllables, and for onsets, compression is the highest in CV syllables. However, the compression for codas in VC syllables remains higher than that for onsets in CV syllables, a result that supports the hypothesis. It becomes clear from these results that making a distinction for syllable types in the analysis of codas provides a more accurate representation of the compression process.

Codas have been shown to compress to a greater degree than do the other syl-

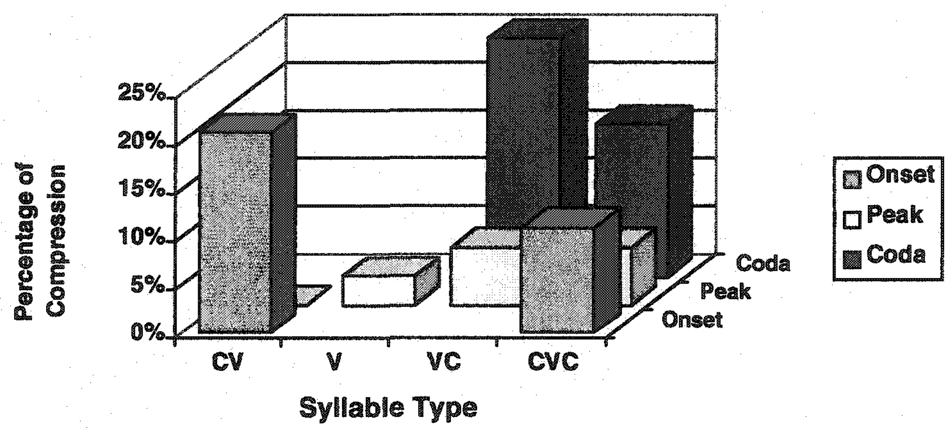

Figure 3.6.
Comparison of the Percentage of Compression for Syllable Constituents across Syllable Types

labic constituents, and this still holds for the unstressed data; unstressed codas compress 13% more than onsets and 23% more than peaks (refer to Table 3.7). However, the same is not true for stressed segments. If a coda is stressed, it compresses less than onsets but more than peaks. Stressed codas show 17% less compression than onsets and 12% more than peaks.

When the unstressed codas are further categorized by syllable types this pattern remains. If an unstressed coda is in a VC syllable, it does compress more than onsets or peaks across all syllable types (Figure 3.7). The VC coda compresses by 31%, significantly more than the other syllable constituents across syllable types do. At a 22% decrease from the control, CV onsets show the second highest degree of compression.

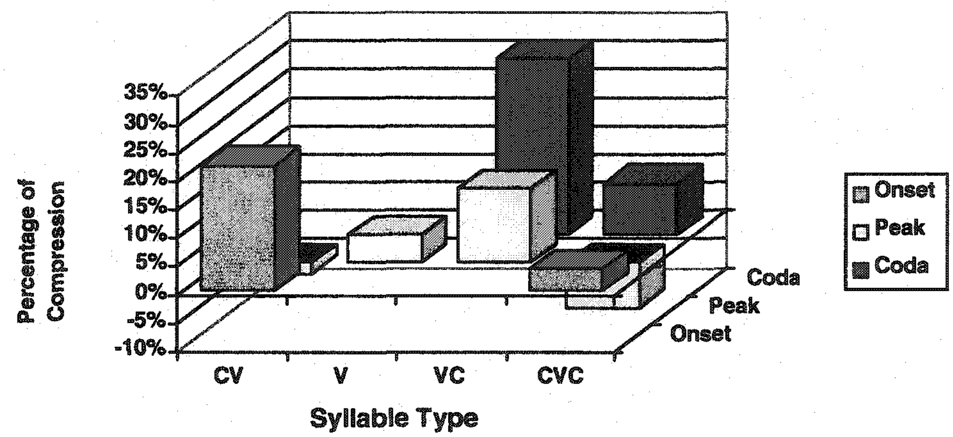

Figure 3.7.
Comparison of the Percentage of Compression for Unstressed Syllable Constituents across Syllable Types

In the stressed data shown in Figure 3.8, onsets in CV syllables show the greatest degree of compression at 20% followed closely by codas in CVC syllables at 16%. However, the overall degree of compression for all syllable constituents across all syllable types is less than that previously observed in the unstressed data. The results lend support to the hypothesis only if the factors for stress and syllable type are included. Unstressed codas in VC syllables compress the most followed by unstressed onsets in CV syllables.

When one considers data that are not categorized by stress, support is not found for the inverse hypothesis. All comparisons for the means of syllable constituents at a slow tempo with those of the control show an expected increase in duration. However, the results do not support the inverse hypothesis. At a slow rate

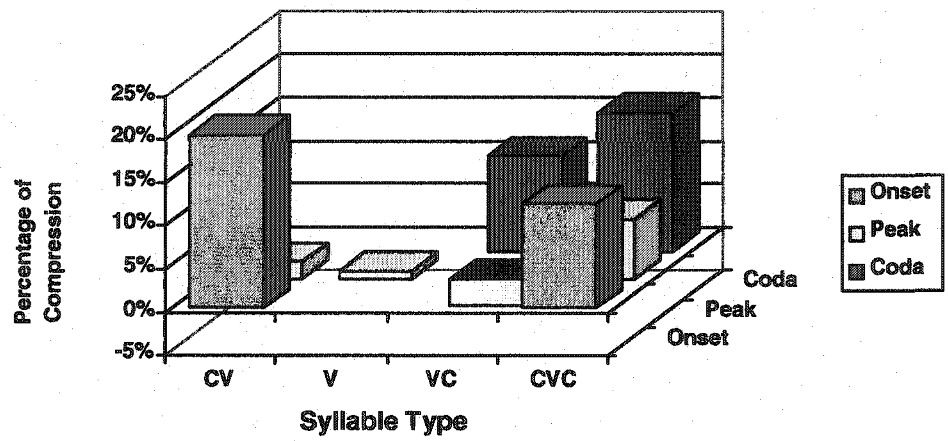

Figure 3.8.
Comparison of the Percentage of Compression for Stressed Syllable Constituents across Syllable Types

of speech, the mean for peaks increases 29% more than that for codas and 66% more than that for onsets as shown in Table 3.8. Codas show an expansion of 75%, 37% more than the expansion of onsets.

Table 3.8
Percentage Decrease in Total Syllable Constituents Comparing a fast Tempo to the Control

	Syllable Constituents		
Stress	Onsets	Peaks	Codas
Combined	38%	104%	75%
Unstressed	38%	127%	36%
Stressed	12%	90%	101%

However, when stress is considered, the results change dramatically. If the coda is stressed, it increases 11% more than peaks and 89% more than onsets. Conversely, the unstressed codas show the smallest degree of change at 36%, two percent less than the change for onsets and 91% less than that for peaks. If the data are separated using a stress factor, support is evident for the inverse hypothesis that stressed codas expand more than onsets or peaks.

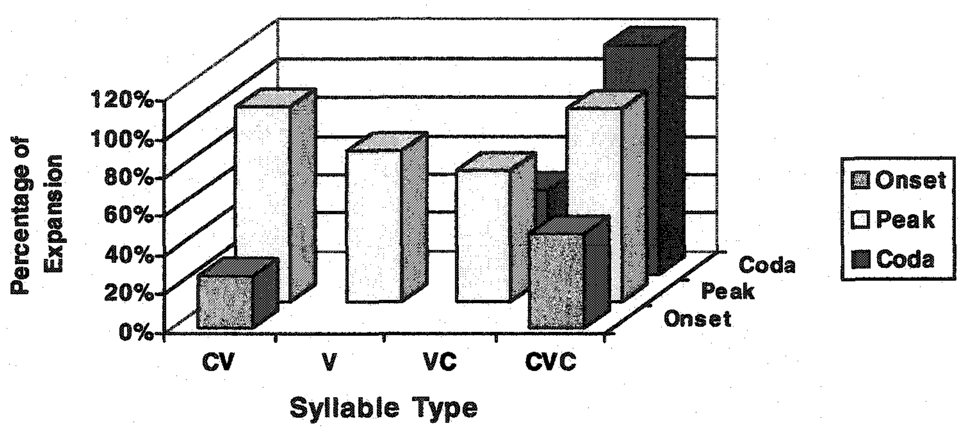

Figure 3.9.
Comparison of the Percentage of Expansion for Stressed Syllable Constituents across Syllable Types

In addition, this hypothesis holds when the data are categorized for syllable type. In Figure 3.9, it is clear that stressed codas in CVC syllables expand more than onsets or peaks showing a 119% expansion, which is greater than the expansion of peaks and onsets across all syllable types. While the hierarchy in the expansion of syllable constituents in CVC syllables is similar for unstressed codas, the degree of expansion is much less as seen in Figure 3.10. In both stress cases, CVC codas expand more than peaks, which, in turn, expand more than onsets; however, the stress factor significantly increases the degree of expansion.

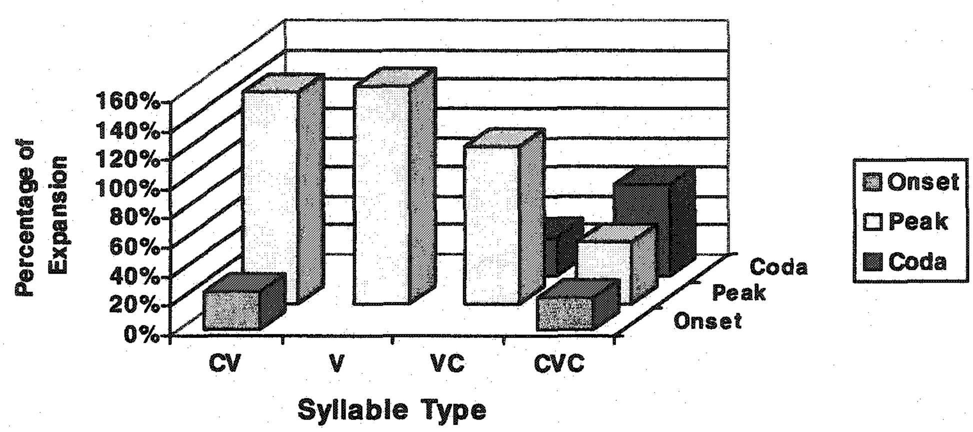

Figure 3.10.
Comparison of the Percentage of Expansion for Unstressed Syllable Constituents across Syllable Types

In the unstressed data, the highest degree of expansion occurs for vowels in unstressed V syllables at 149% closely followed by those in CV syllables at 145%. The common denominator between these two syllable types is that both have the peaks in a syllable-final position. The absence of a coda in these syllables may encourage a greater degree of elasticity in changes in duration.

To summarize, both the hypothesis that codas compress more than onsets or peaks and the inverse hypothesis find support from these data. In support of the findings of Campbell and Isard (1991), the coda in VC syllables will compress to a greater degree than the onset or the peak if it is unstressed. Inversely, a stressed coda in a CVC syllable will expand to a greater degree than the onset or the peak. However, in all cases the standard deviations for the syllable constituent groups compared are high (Appendix B). The sizes of these groups are quite small, which could account for the large dispersion and at a slow tempo; the dispersion is even greater as a result of the inherent variability of slow speech. Therefore, a larger sample size is necessary for robust comparisons.

Conclusion

In conclusion, the results of the experiments suggest that there is a deceleration-acceleration pattern evident in focused sentences. In addition, the modulations in tempo are not manifest in all segments, but are evident to varying degrees in the syllabic constituents.

The hypothesis that pre-focal regions expand and post-focal regions compress finds support in these data; however, the results are not conclusive because of the large dispersion in the data. If the sample size were increased, statistical support might be found. However, a strong tendency for expansion of the pre-focal region is evident. Pre-focal regions of the sentence do expand, particularly if the focal word is located near the beginning of the phrase. Clearly, location is important for the degree of expansion in the pre-focal region. In addition, this expansion appears to occur at the durational expense of other syllables in the region. There is some suggestion that the non-focal syllables of the sentence actually compress.

Clearly, vowels are a potential site for rate variation. The hypothesis that vowels increase or decrease as speech rate decreases or increases is born out in the data. However, only expansions show strong evidence, with unstressed vowels expanding more than stressed vowels. The degree of compression apparent in vowels at a fast rate of speech falls short compared to that of the sentence as a whole.

The minimal differences evidenced in vowel compression may be a result of the lack of distinction between the normal and fast sentences because of the limitation of the sample being provided by one speaker. By including speech samples from multiple speakers, this lack of distinction could be eliminated, thereby ensuring a more accurate representation of tempi. However, the data do show that unstressed vowels do compress more than stressed vowels.

While compression in vowels did not give significant results in this study, a strong tendency for codas to compress given criteria for stress and syllable type was evident. The hypothesis that codas compress more than either onsets or peaks holds. The data show that unstressed codas in VC syllables reveal the greatest degree of change. In terms of expansion, only codas in stressed CVC syllables exhibit a higher increase than onsets, or peaks, while unstressed peaks of short syllables hold the greatest degree of expansion.

Some aspects of the study may have affected the results adversely. In the first place, the study represents the speech of only one individual and therefore, the sample size is small, as previously mentioned, and the sample may be biased by the fundamental tempo patterns of this participant. The same generalizations may or may not hold true in a similar study of multiple speakers, but the results will be more conclusive.

Secondly, manner of articulation might be a consideration for the comparisons. For example, fricatives tend to be more variable than stops, which may introduce an element of variation that can obscure the duration effects of tempo. Therefore, accounting for manner of articulation may increase the accuracy of the study. In addition, the syntactic structure and the number of syllables in the sentence are similar but are not identical. Clearly, sentences as homogeneous as possible would increase the value of the comparison. The results of this study provided a good foundation for the main study using multiple speakers.

CHAPTER 4 Methodology

In the same manner as the pilot study, the main study is organized into two sections. The first section consists of the primary experiment that tests for the deceleration-acceleration tempo pattern based on word focus. The second section is comprised of two experiments that investigate the potential sites for tempo change within the sentence. This chapter describes the materials and methods employed for all experiments in both sections.

Participants

As previously mentioned, the variables gender, dialect, and idiolect have the potential to affect tempo. Therefore, the design of the experiment should control for these factors. In order to counter the effects of individual variation in speech rate, five speakers were chosen to read the script. An odd number of speakers was chosen in order to avoid ties in comparisons. So that gender and dialect effects could be avoided, all speakers were male[9] and spoke the same Vancouver Island dialect of Canadian English. In addition to these variables, the level of education of the speaker has the potential to affect speech. Therefore, it was necessary to ensure that the speakers shared a common post-secondary level of education. In addition, speakers ranged from 35 to 60 years of age.

A further experimental control is essential to remove bias from the results. All speakers were required to be naïve to the hypotheses of the study. If the speakers were aware of the desired language response, they could alter their speech either consciously or subconsciously, thereby skewing the results. To ensure that the results were unbiased, all speakers of the study were unaware of the hypotheses.

Materials

As a data source for this study, five sentences were designed to provide focus locations and phonetic environments for vowels and codas. In order for the comparisons of the effects of the changes in focus and tempo to be accurate, the sentences needed to be as homogeneous as possible. Thus, unwanted variances in the data can be prevented. All sentences are declarative and in the active voice. Additionally, the sentences are identical syntactically and exhibit the same sequence of form-class words. For example, the final sentence of the script, '*Cautious, I seek a*

9.The choice of gender for this study was arbitrary.

sheet atop a tippy ketch' reflects the following syntactic pattern:

[Adverb [Pronoun [Verb [Article Noun] [Preposition [Article Adjective Noun]]]]]

As seen in Table 4.1, the pattern consists of 12 syllables comprised of 24 or 25 segments. An additional feature of the sentences is voicing alternation, which I will discuss presently. This feature severely restricted the potential candidates for the adverb position.[10] As a result, a small variation was necessary in the number of segments of the first syllable of the adverb resulting in the first, third, and fifth sentences having three segments and the second and fourth sentences having two segments. Therefore, two of the five sentences have a total of 24 segments and the others 25; the first syllable accounts for the discrepancy. Any potential variation incurred by the disparity in the number of segments is avoided by the exclusion of the first syllable because of possible boundary affects.

Table 4.1
Segment and Syllable Distribution for Sentences

	Number of Segments	
Form-class	Syllable 1	Syllable 2
Adverb	2 or 3	2
Pronoun	1	
Verb	3	
Article	1	
Noun	3	
Preposition	1	3

10.For example, only a few adverbs fit the length and voicing criteria and even fewer meet the semantic requirements of the sentence.

Table 4.1
Segment and Syllable Distribution for Sentences

Form-class	Number of Segments	
	Syllable 1	Syllable 2
Article	1	
Adjective	3	1
Noun	3	
Sentence Total for 12 Syllables	24 or 25	

As previously mentioned, an additional feature of the sentences is a voicing alternation that results in a greater distinction between segments. This provides a clearer demarcation for the segments, thus increasing the accuracy of segmentation. Based upon periodicity, voiced segments have distinctive waveforms compared to unvoiced segments. Voiced segments are periodic, while those that are unvoiced are aperiodic. Periodic sounds exhibit a repeating waveform that is not evident in aperiodic sounds. This distinction is apparent in the voicing sequence of the last three segments of the name Acacia, /e/ /ʃ/ /ə/, which is voiced-unvoiced-voiced (Figure 4.1). The vertical black lines in the figure represent segmentation boundaries. Despite careful design, there was one case where there was an unavoidable potential for two voiced segments to be adjacent in the third sentence of the script for the main study in Appendix A, 'Sheepish, I teach a cat atop a cushy seat'. In this sentence, there is a sequence 'a-t-a' of 'cat atop', and the 't' of cat has the potential to become a flap changing to a voiced segment depending on the idiolect of the speaker. In this case, segmentation is by no means impossible; it merely requires additional criteria for the segmentation process.

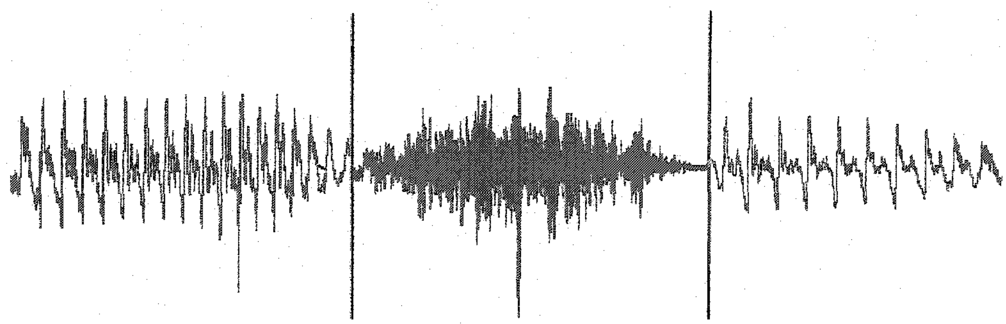

Figure 4.1.
Example Waveform of a Voiced-unvoiced-voiced Sequence

In addition to homogeneity and voicing alternation, spacing was required between the focused words. In order to avoid any potential adjacency effects, at least one word separated the focused words.

Procedure

4.1 *Recording*

Each speaker was asked to complete two reading tasks, one varying the tempi of the script and the second, varying the focus position. Before the speakers were asked to read the script for the recording, they were given the opportunity to practice. This avoided the higher likelihood of error possible when someone is asked to sight-read unknown material. When a speaker was comfortable with the script, he was asked to read the five sentences repeating each sentence three times for one set. The two additional repetitions of each sentence increased the sample size and were necessary because of the inherent variability of the data.

The first tempo task involved three sets with each set being read at a different tempo: normal, slow, and fast. The speaker was asked to read as naturally as possible for the normal rate of speech to create a control group. Readings at the slow rate and fast rate of speech required the speaker to read as closely as possible to his error threshold. In other words, he would read the script as slowly or as quickly as possible without error. All of the sentences recorded were well-formed sentences and were not affected by this threshold.

The second focus task was comprised of four sets. To create a control group for the focus experiment, the first set was a neutral reading. The remaining three sets were read with the focus on one specified word in three different locations in the sentence: near the beginning of the sentence, approximately mid-point, and near the end.

In order to ensure an uncorrupted sound signal, the readings were recorded with a Sony digital audiotape (DAT) recorder in a soundproof room for a total of 225 sentences for the tempo task and 300 sentences for the focus task. All controllable conditions, such as the recording location, microphone, and recording equipment were identical for all five recording sessions. The recordings were then digitized using Multispeech, Model 3700 version 2.2[11] at 20,000 samples per second creating data files of waveforms for the segmentation process. In addition, spectrograms were generated for each sampled data file as a secondary source of information for segmentation.

4.2 *Segmentation*

The segmentation process involved separating the waveforms into predetermined subunits in order to calculate durations. The subunits required for this study are sentences and pauses for all experiments, syllables and focal regions for the focus experiment, and segments for the site experiments. I alone performed the segmentation of the data in order to maintain consistency.

The initial stage of segmentation determines the sentential unit. Sentence boundaries were marked at the point of change in amplitude from a constant, low energy reading. The speakers inserted a pause between each sentence which assisted in the segmentation of sentences. The boundaries between subunits were determined from the waveform readings.

Syllable boundaries were determined from the start of the zero crossing of the beginning of the onset or nucleus segment to the zero crossing of the end of nucleus or coda. On the same principle, the boundaries for focal regions are established using the zero crossing of the beginning of the initial segment of the first syllable of the focal region to the zero crossing of the end of the final syllable of

11.Multispeech was created by STR and distributed by Kay Elemetrics Corporation.

the region.

The zero crossing of a voicing alternation constitutes a segment boundary as seen in Figure 4.1. To ensure that the boundary was accurately positioned, it was verified using the information provided by the spectrogram. In the few cases where there is no voicing alternation, the spectrogram provides the necessary information for the segmentation of all subunits. Auditory checks were used as required in addition to spectrographic verification.

The speech-processing software provided a timing value in seconds for each boundary. In order to calculate the measurements for each experiment, these boundary values were imported into Microsoft Excel 97 SR-2. Data management tasks, such as converting duration values from seconds into milliseconds and computations, were carried out using this software.

As previously discussed, pauses are distinctly affected by changes in tempo and are not the focus of this study. Therefore, it was also necessary to mark the pause boundaries and calculate the pause duration measurements. These measurements were then subtracted from the respective sentence measurements.

Separate Experiments

The specific details of each experiment designed to test the tempo pattern hypothesis and the hypotheses for the sites of tempo changes follow.

4.3 *Section I: The Tempo Pattern Experiment*

The first objective of my research is to confirm the pattern of deceleration and acceleration in English. As previously discussed, focus has been shown to cause the tempo of the focal word to decelerate for emphasis. In addition, there is some evidence in the pilot study that indicates that this deceleration is restricted to the focal word. The results of the pilot study also suggest that the expansion of the focal word is at the expense of the words preceding the focus, which undergo compression. In addition, acceleration following the focal word has been observed. Therefore, the pattern is as follows: deceleration occurs in the focal word, while acceleration takes place in the region of the sentence preceding the focus and in the remainder of the sentence.

Both the pattern observed by Fant et al. (1991c) and the pattern observed in the pilot study were tested in this experiment resulting in two separate tests, I and II. Experiment I compares the average syllable durations (ASDs)[12] in pre-focal position inclusive and post-focal position. The pre-focal inclusive section includes all words up to and including the focal word and the post-focal section consists of the words following. Experiment II contrasts the ASDs in pre-focal exclusive position, which excludes the focal word, the ASDs of the focal word, and those in post-focal position.

12.The syllable unit was determined following the well-argued position adopted by Campbell (1990) that the syllable is the best choice for a unit for speech rate measurement (p. 69).

Because boundary effects had to be taken into consideration, each test required two sets of data. As previously discussed, sentence constituents in phrase-final position are subject to lengthening. In addition, constituents in phrase-initial position are also susceptible to changes in duration, that of shortening (Campbell, 1990). As a consequence of these boundary effects, the initial and final syllables of the sentence cannot be reliable indicators of tempo changes. Therefore, in one set of data the initial and final syllables were omitted from the focal section calculations, and in the second no special measures were taken to control for boundary effects in order to determine if the effects of tempo transcend those of boundaries.

Obtaining the two sets of data for Experiment I involved determining the durations of both the edited (in which the initial and final syllables were removed) and the unedited focal regions and calculating the averages for the syllable durations. The edited focal regions were determined using the zero crossing of the beginning of the initial segment of the second syllable of the focal region to the zero crossing of the end of the penultimate syllable of the region. After the measurements for the edited focal regions were calculated, the ASDs for the pre-focal and post-focal regions were computed by dividing the duration of the focal regions by the number of syllables in that region. There was no difference in the calculations for ASDs in either pre-focal or post-focal position. The same process was repeated for the second set of data using the unedited focal region durations. As a result of these calculations, each set had a total of 375 focal ASDs for Experiment

I.

The method for calculating the ASDs for the control sentences was the same, but additional calculations were necessary. To review, there were a total of four readings for each sentence. One was a control and the remaining three varied in terms of focus. Each focus word was in a different location in the sentence. For each location, the comparable focal region had to be calculated for the control sentence. This resulted in three pre-focal ASDs and three post-focal ASDs for a total of six calculated for each control sentence instead of the two for the focused sentences. This process added an additional 375 ASDs as controls for the 375 focal ASDs for a total of 750 data points.

The ASDs for Experiment II involved the same procedure as that used for Experiment I, except for the difference in the focal regions. Experiment II has three focal regions: pre-focal exclusive excluding the focal word, the focal word, and post-focal. Consequently, three ASDs were calculated, one for each region. The total data points for Experiment II were 1,200.

Even when one speaker reads a sentence three times, the durations of the sentences will vary, which in turn affects the durations of the constituents of the sentence. This can prove problematic when comparing two syllables from different sentences. The effects are even greater given different sentences and different speakers regardless of sentential homogeneity. In order to counter the effects of varying sentence lengths, all ASD measurements were normalized using a ratio of syllable duration to sentence duration.

4.4 *Section II: Sites for Tempo Modification*

The second objective of the study was to determine how the tempo pattern manifests itself phonetically. In order for acceleration to occur, the affected segments must compress, while expansion of the segments results in a decelerated tempo. As previously discussed, the vowel and the coda appear to be the two most likely candidates as sites for timing changes. As a basis for both site experiments, it was necessary to verify that the sentences read at fast and slow tempi were, in fact, faster or slower than the normal readings. Therefore, the durations of sentences read at fast and slow tempi were compared to those of the normal readings.

The Vowel as a Site for Tempo Modification

The purpose of the vowel experiment was to confirm the results found in the pilot study that vowels increase or decrease in duration as tempo decreases or increases, respectively. From the total of 5,535 segments derived from the tempo readings, 2,700 segments were vowels. The duration measurements of these segments were isolated and comprised the data set for the experiment. In order to ensure that the only variability introduced into the experiment was tempo change, controls were established for all other factors that could induce changes in duration.

It is well established that stress has a lengthening effect on syllables and, as previously discussed, segments preceding a phrase boundary or a pause are also subject to lengthening. In addition, the duration of a vowel may be affected by the syllable structure. Clearly, these four factors must be controls in the study. For both site experiments, each segment was marked for stress: those in stressed syllables were marked as stressed and those remaining, as unstressed. Segments preceding pauses and phrase boundaries were controlled separately, each being marked with the appropriate features, pre-pausal or adjacent to phrase-boundary.

Each syllable has a different underlying structure. Therefore, segments potentially interact differently within different syllable structures. For example, segments in syllable-initial, syllable-middle, or syllable-final position may exhibit different degrees of durational effects. Therefore, if segments are compared, it is important for accuracy to compare segments within the same syllable structure.

Three components, the onset, the peak, and the coda can define syllable structure. Onsets in English can consist of one to three consonants. However, in this study no consonant clusters were used, in order to avoid the inherent complication of clusters. A syllabic segment defines the peak of the syllable. In this study, all syllabic peaks are vocalic. In English, one to four consonants may comprise the coda of the syllable, but only single consonant codas were used. For the control, syllable type, (the primitive syllable inventory of vowel (V), vowel and consonant (VC), consonant and vowel (CV), and consonant, vowel, and consonant (CVC)) was again used for categorizing the syllables.

To summarize, there were a total of 2,700 data points for these experiments. Each was marked for stress, pre-pausal position, phrase-boundary adjacency, and syllable type.

The Coda as a Site for Tempo Modification

The purpose of the coda experiment was to discover if the coda remained a viable site for tempo change, particularly the coda in VC syllables as found in the pilot study. At a fast tempo, evidence for the duration of codas compressing more than that for onsets or peaks has been shown in both the pilot study and the study by Campbell and Isard (1991). In addition, the pilot study showed the tendency for codas to expand more than onsets or peaks in stressed CVC syllables. Therefore, the expansion of codas was also tested in this experiment. The data set consisted of a total of 5,535 duration measurements derived from the tempo readings.

In addition to the controls for stress, pre-pausal position, phrase-boundary adjacency, and syllable type, controls for manner of articulation and syllable type were included in this experiment. In addition to the control of syllable type, a categorization for syllable constituency was required. This feature defined each segment as an onset, vowel, or coda, where all peaks are vowels.

The manners of articulation evident in the data were stops, fricatives, and affricates. Changes in tempo can prompt changes in the manner of articulation. For example, stop or fricatives can become approximates. Although this change can occur, it is does not affect the accuracy of the comparisons in the study, as it is the duration of the consonant at fast and slow tempi that is compared to the control. Each manner of articulation has the potential to respond differently to changes in tempo. For example, the duration of a stop may be more restricted than a fricative or affricate in terms of the ability to expand. This difference may obscure the effects of tempo change and therefore, a control for manner of articulation was necessary.

In review, the hypotheses to be statistically tested for the tempo pattern experiments are outlined in Table 4.2.

Table 4.2
Hypotheses for the Tempo Pattern Experiment

Experiment	Hypotheses
Experiment I	The ASDs of the pre-focal inclusive region are greater than the ASDs of the control
Experiment II	The ASDs of the pre-focal exclusive region minus the focal word are less than the ASDs of the control
	The ASDs of the focused word are greater than the ASDs of the control.
Experiment I & II	The ASDs in the post-focal region are less than the ASDs of the control.

Table 4.3 describes the hypotheses for the potential sites for tempo change.

Table 4.3
Hypotheses for the Experiments for the Potential Sites for Tempo Change

Experiment	Hypotheses
Vowel	The vowel increases or decreases in duration as tempo decreases or increases, respectively.
Syllable Constituents	The coda compresses more than the onset or peak of the syllable at a fast rate.
	The coda expands more than the onset or peak of the syllable at a slow rate.

Statistics

Two software programs were employed for the statistical analysis of the three separate data sets, SYSTAT version 10 and the student version of JMP distributed by SAS. As a foundation for analysis, basic statistics were calculated for the data.

4.5 *Tempo Pattern Experiment*

The variables for the tempo pattern experiment are focus having three values, 1, 2, or 3 and focal position having four values in Experiment I, pre-focal inclusive and post-focal and the corresponding controls. In Experiment II, there are six values, pre-focal exclusive, focus, and post-focal in addition to the corresponding controls. All comparisons are made in terms of ASDs. In Experiment I, the values for pre-focal and post-focal are compared to the controls for each focus position, and the same was done for Experiment II with the addition of the focus comparison to the control. The differences generated by these comparisons were analyzed using a type of split plot ANOVA to eliminate the effects of multiple speakers. Because there are more than two different groups of comparisons being made, an ANOVA is the most appropriate choice for the analysis of the data.

Because of the removal of the initial and ultimate syllables in Experiment I, thereby reducing the potential 1,200 data points to 750, the experimental design becomes unbalanced. This is not the case in Experiment II, in which all data points are included. However, this unbalance does not present a serious problem for the analysis, as the software automatically invokes the general linear model (GLM) procedure for Experiment I.

4.6 *The Tempo Site Experiments*

As in the Tempo Pattern experiment, the experiments determining potential sites for tempo change employed a type of split plot ANOVA. A split plot analysis was necessary because of the dependency introduced into the data by having multiple speakers. The vowel data include 2,700 data points and four factors, stress, pre-pausal position, phrase-boundary adjacency, and syllable type. The first three factors have *yes* or *no* values and the last factor, syllable type had the values, V, VC, CV, and CVC. In addition, the data are balanced because each sentence had 12 vowels. The segment durations of the vowels at slow and fast rates were then compared to the vowels of the normal, or control sentences in order to test the hypothesis.

As previously mentioned, 5,535 data points comprise the data for the coda site experiment. However, because the number of segments per sentence was not the same - 24 segments in sentences two and four and 25 segments in sentences one, three, and five - the data set was unbalanced. As discussed earlier, this does not pose a problem for the analysis.

For each of the 5,535 segments, the four factors used for the vowel data set were used in addition to Manner, and Syllable Constituency. Manner had the values, Fricative, Affricate, or Stop, and Onset, while Vowel, or Coda were the values for Syllable Constituency.

Initially, the durations of syllable constituents at slow and fast tempi were compared to the controls to establish a difference. The next stage involved calculating the degree of change in the durations between fast and slow tempi and the controls. The resulting values of onset, peak, and coda differences in duration could then be compared in order to test the hypothesis.

CHAPTER 5 Results

In this chapter, the results are described for the tempo pattern experiments and for the two experiments establishing the potential sites for speech rate, vowel and syllable constituency. Preceding these discussions are brief descriptions of the analyses of the sentences used for both parts of the study.

As described previously, the data for the study were analyzed using a split-plot ANOVA analysis. However, before the data could be analyzed, a natural log transformation was necessary. The data were found to have a positive skew, typical of most data measuring quantities. Consequently, it was necessary to transform all measurements for statistical analysis to satisfy the requirement of normal distribution. For consistency, all results are expressed in terms of the natural log of the original measurements.

Tempo Pattern Experiment

The sentences used for the tempo pattern experiment exhibited variation in duration for both the sentence and the position of the focal word. An alpha level of .05 was used in all statistical tests for this experiment. The sentence effect on duration was significant, F (4, 16) = 4.23, p = .016. A Tukey HSD analysis revealed that there was a significant difference between sentences 2 and 5, p = .022. Although the remaining sentences were not significantly different from one another, Sentence 2 (M = 8.01, SD = .18) was shorter in duration than Sentence 5 (M = 8.08, SD = .22). As previously mentioned, the mean values are based on the natural log of the original measurements. Descriptive statistics of sentential durations by sentence and focus including medians of the original measurements can be found in Tables C.1 and C.2 of Appendix C.

The treatment factor, focus also showed significant differences, F (3, 12) = 7.13, p = .005. Sentences with a mid-location of the focus word, Focus 2 and those with a final-position focus word, Focus 3 were significantly different from the control sentences, N (Tukey a: p = .019 and p = .005 respectively). Focus 2 sentences (M = 8.07, SD = .03) and Focus 3 sentences (M = 8.10, SD = .21) are longer than the control sentences (M = 7.93, SD = .17). The differences observed for the effects of Sentence and Focus on sentence duration provide evidence for normalizing the ASDs by sentence length to neutralize the effects of varying sentence lengths.

Before the analysis, a preliminary test was performed to determine if a sentence boundary did, in fact, significantly affect the duration of the initial and final syllable of the sentences for this experiment. Results showed that the initial syllable did not significantly change in duration for the different focus readings, F (5, 20) = .016, p = .973. The means for the initial and final syllable durations are found in Table 5.1.

Table 5.1
Durations of Initial and Final Syllables

	$n = 75$			
	Initial		Final	
	M	*SD*	*M*	*SD*
N	5.59	.23	6.14	.22
1	5.59	.27	6.08	.19
2	5.60	.28	6.09	.21
3	5.61	.26	6.27	.27

Although the durations of the final syllable did show significant variation between the focus readings, F (5, 20) = 3.22, p = .027, there were no significant differences found between the three focus locations and the control. The probability values for Tukey *a* comparisons between the control and each of the three focus readings were $p = .871$, $p = .903$, and $p = .200$.

These findings provide evidence that the initial and final syllables of the sentences used for this experiment did not undergo any significant changes in duration as a result of being adjacent to a sentence boundary. Consequently, the data sample used for the tempo pattern experiment was unedited, i.e. the initial and final syllables of the sentences were not removed from the calculations for the focal regions.

The results of the tempo pattern experiment are divided into two parts, Experiment I and Experiment II. Experiment I analyzed the pre-focal inclusive region including the focal word as one unit, while Experiment II analyzed it as two separate units, the pre-focal exclusive region and the focal word.

5.1 *Experiment I*

Pre-focal Inclusive Region

For the pre-focal inclusive and post-focal regions for Experiment I and the pre-focal exclusive, focus, and post-focal regions for Experiment II, variances between the ASDs of the regions for the three focus positions and those of the corresponding regions in the control sentences were calculated based on the means of the transformed ASDs located in Tables C.3 through C. 6. in Appendix C. The medians for the non-transformed data are also included in the tables. For the pre-focal inclusive regions of Experiment I, the ANOVA revealed that the ASD variances were significantly different from zero for two of the three focus positions, Focus 1 and 2, $F(1, 74) = 163.74, p = .000$ and $F(1, 74) = 174.92, p = .000$ respectively. The means of these variances reflect an increase in duration from the control. The means of the variances for both the pre-focal inclusive and post-focal regions are represented in Table 5.2. However, the variances did not significantly differ from zero for ASDs in the pre-focal region of Focus 3 sentences, $F(1, 74) = 1.03$, $p = .314$. In addition, no increase in duration was evident for ASDs in this focus position. The mean of this variance showed a marginal decrease in duration compared to the control.

Table 5.2
Pre-focal Inclusive and Post-focal Region Variances of ASDs

	$n = 75$			
	Pre-focal		Post-focal	
Focus	*M*	*SD*	*M*	*SD*
1	.129	.08	-.085	.06
2	.086	.06	-.102	.07
3	-.005	.04	.012	.16

Post-focal Region

Post-focal and pre-focal inclusive regions exhibited similar results. In the post-focal regions, the variances for ASDs in Focus 1 and 2 positions were significantly different from zero, $F(1, 74) = 171.72, p = .000$ and $F(1, 74) = 171.10, p = .000$ respectively. As indicated in Table 5.2, the means of the variances for ASDs in this focus position showed a decrease in duration from the control. As was found in those in the pre-focal inclusive regions, the variance for Focus 3 was not significantly different from zero, F (1, 74) = .44, p = .510. Inverse to the findings for pre-focal inclusive regions, ASDs in this position increased minimally from the control.

In summary, the difference between ASDs in focus positions 1 and 2 and ASDs of the corresponding control region was significantly different from zero for both pre-focal inclusive and post-focal regions. These differences show an increase in duration of ASDs in the pre-focal inclusive region and a decrease in duration in the post-focal region. When the focus is located near the end of the sentence, the variance means are not significantly different for both regions. Additionally, with focus located in this position, the means show a slight decrease in duration for pre-focal inclusive regions and an increase for post-focal regions.

5.2 *Experiment II*

As previously noted, this experiment analyses the pre-focal region as two distinct parts: the region preceding the focus word and the focus word itself. The results of each region will be described separately.

Pre-focal Exclusive Region

For each focus position, the variance in the region preceding the focus word was significantly different from zero. Results from the ANOVA for Focus 1 through 3 were F (1, 74) = 428.15, p = .000, F (1, 74) = 414.92, p = .000, and F (1,

74) = 445.83, *p* = .000 in order. In all cases, the differences for the ASDs in the pre-focal exclusive region and the control showed decreases in duration, as shown in the second column of Table 5.3.

Table 5.3
ASDs of Pre-focal Exclusive Regions

	n = 75			
	Pre-focal		Focus Word	
Focus	*M*	*SD*	*M*	*SD*
1	-.085	.11	.481	.18
2	-.072	.10	.471	.23
3	-.079	.10	.271	.21

Focus Word

Comparable to the findings for the pre-focal inclusive regions in Experiment I, the means of the variances for the focus word differed from zero significantly for all focus positions. The ANOVA results for Focus positions 1 through 3 were *F* (1, 74) = 542.94, *p* = .000, *F* (1, 74) = 314.44, *p* = .000, and *F* (1, 74) = 124.36, *p* = .000, respectively. In all three positions, the variance means showed an increase in duration, as shown in Table 5.3.

Post-focal Region

The results for the post-focal regions for Experiment II are the same as those for Experiment I. The data and analyses of the post-focal regions for both tests were the same; only the data sets for the regions preceding and including the focus word differed for the two tests.

Contrary to the findings in Experiment I, the variance means for the ASDs in the pre-focal exclusive regions and the focus word in all focal positions exhibited measurable differences. The means decreased in pre-focal exclusive position and increased for the focus word. While significant differences in the direction of a decrease in duration were found for the ASDs in Focus 1 and 2 positions, the increase for the ASDs in Focus 3 position was insignificant.

Sites for Rate Change Experiments

In the data for the vowel and syllable constituency experiments to establish sites for changes in speech rate, there were only eight occurrences of missing data because of errors in reading the script by Speaker 1. The mean replaced these data to ensure a balanced data set.

With an alpha level of .05, the statistical analysis of the sentences used for the site experiments showed significant variation in duration for the treatment factor, Rate, $F(2, 8) = 31.87, p = .000$. The results confirmed intuition regarding sentence rate that sentences spoken at a slow tempo are shorter than those at a normal tempo and sentences spoken at a fast tempo are longer. Sentences spoken at a normal speech rate ($M = 8.01$, $SD = .16$) were significantly longer in duration than those spoken at a fast rate (*M* = 7.75, *SD* = .07), *p* = .032, and significantly shorter in duration than those at a slow rate (*M* = 8.41, *SD* = .25), *p* = .004 according to the Tukey *a*. The blocking factors, sentence and repetition, did not show significant effects on sentence duration; therefore, the sentences were sufficiently homogeneous for investigating the treatment factor. Descriptive statistics including the means of the transformed data and the medians of the actual measurements for these tests are presented in Tables D.1 and D.2 of Appendix D. An alpha level .01 was used for all analyses in the following experiments.

5.3 *Vowel Duration*

The results of the vowel experiment show that speech rate does effect a change in the duration of vowels. For the three speech rates, the durations of vowels are significantly different, *F* (2,5) = 230.35, *p* = .000. According to the Tukey *a*, vowels at a fast rate of speech (*M* = 4.18, *SD* = .46) show a significant decrease in duration from those of the control (*M* = 4.44, *SD* = .48), *p* = .001. However, while vowels at a slow rate of speech (*M* = 4.76, *SD* = .43) do show an increase in duration, this change does not reach a significant level, *p* = .018. For Experiment I and II, the medians of the original vowel durations can be found in Tables D.3 and D.4 of Appendix D.

When factors for syllable type, stress, and sentence-boundary adjacency were included in the analysis, vowels at both fast and slow rates of speech showed significant differences from the control. An interaction between rate, syllable type, stress, and sentence boundary adjacency was evident, and when considering these effects, significant differences were evident in the duration of vowels, $F(2,10) = 21.50$, $p = .000$. As was seen in the previous analysis of vowels, the means for vowels at a normal speech rate are shorter than those at a slow rate and longer than those at a fast rate, as shown in Table 5.4. All possible combinations of values for factors, such as stress and sentence-boundary adjacency are not represented in this data. Therefore, only the combinations of effects that were evident in the data are represented in this and other tables in this chapter.

Table 5.4
Vowel Duration by Rate, Syllable Type, Stress, and Sentence Boundary Adjacency

		Speech Rate					
		Fast		Control		Slow	
Syllable	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CVC							
Stressed							
SBA[a]	120	4.61	.31	4.71	.31	4.83	.30
Non-SBA	300	4.34	.35	4.66	.43	4.88	.45
VC							
Stressed							
SBA	30	4.64	.27	4.77	.24	5.02	.26
Unstressed							
Non-SBA	75	4.29	.26	4.36	.29	4.60	.30
V							
Stressed							

Table 5.4
Vowel Duration by Rate, Syllable Type, Stress, and Sentence Boundary Adjacency

		Speech Rate					
		Fast		Control		Slow	
Syllable	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Non-SBA	75	4.35	.34	4.73	.31	4.99	.37
Unstressed							
Non-SBA	300	3.73	.31	4.02	.37	4.56[b]	.40

a. SB = Sentence Boundary Adjacent.
b. $n = 293$

In addition, there were only seven cases of vowels in a prepausal position. All cases were identical in terms of the variables for this analysis; they were unstressed vowels comprising a syllable that was not adjacent to a sentence boundary. The mean duration for the prepausal vowels was 5.04 with a standard deviation of .09. The pause occurred following the penultimate syllable in Sentence 3 in all repetitions for speakers 2 and 5 and in the last repetition for Speaker 4. Coincidentally, all seven cases were read at a slow rate of speech, and as a consequence, no comparisons between speech rates were possible. Therefore, these data were extracted from the sample, and any effects of prepausal position were not included in the results.

Given the effects of stress and sentence boundary adjacency, roughly half of the comparisons compiled in Table 5.5 provided statistically significant results. Significant comparisons are marked in grey in this and other tables of post-hoc results in the chapter.

Table 5.5
Tukey a Results for Speech Rate Comparisons of Vowel Durations

	Speech Rate	
Features	Fast-Control	Slow-Control
CVC		
Stressed		
SBA	.529	.370
Non-SBA	.000	.000
VC		
Stressed		
SBA	.941	.268
Unstressed		
Non-SBA	.974	.037
V		
Stressed		
Non-SBA	.001	.023
Unstressed		
Non-SBA	.000	.000

As was found in the analysis of the rate factor, comparisons of vowel durations between a fast rate of speech and the control showed significant differences, but in this analysis only half of the comparisons were significantly different. Vowels that constitute a V syllable show significant differences in duration regardless of stress. In addition, vowels in a stressed CVC syllable that are not adjacent to a sentence boundary showed a significant difference in duration. The durations of vowels in VC syllables were not significantly different for either of the two speech rate comparisons.

Slightly fewer instances of significant differences were evident in duration comparisons between a slow speech rate and the control. Similar to the fast-control comparisons, only vowels in CVC syllables that are not sentence-boundary adjacent showed significant differences in duration. However, not all vowels in V syllables had durations significantly different from the control. Only unstressed vowels exhibited a significant difference.

In summary, at a fast rate of speech, the means of vowel durations decreased from the control, and at a slow rate, they increased. Overall, only the decreases were significant. However, when factors for syllable type, stress, and sentence boundary adjacency were considered, increases in duration also became significant for specific cases. Vowels in CVC syllables showed significant decreases and increases in duration when not adjacent to a sentence boundary, as did those comprising an unstressed syllable. Stressed vowels in V syllables also exhibited significant decreases in duration. While significant decreases or increases were evident in CVC and V syllables, none were found in VC syllables.

5.4 *Syllable Constituency*

When vowels were not isolated, but rather were analyzed with consonants, the overall results changed dramatically. The interaction of rate and syllable constituency showed significant differences in the durations of segments, $F(2,6) = 29.17$, $p = .001$. In isolation, vowels showed significant decreases, but not increases. However, when vowels were in the context of the syllable, the reverse was true; vowel duration exhibited significant increases, $p = .008$, but not decreases, $p = .047$, from the control given the Tukey *a* results. These increases and decreases are evident in the means provided in Table 5.6. In this table, the descriptive statistics of the vowels are repeated to enable comparisons between syllable constituents.

Table 5.6
Duration of Syllable Constituents by Rate

		Speech Rate					
		Fast		Control		Slow	
Constituents	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Onsets	420	4.57	.24	4.81	.27	5.07	.40
Vowels	900	4.18	.46	4.44	.48	4.76	.43
Codas	525	4.54	.52	4.75	.51	5.03	.50

Rate comparisons of the duration of onsets showed significant increases and decreases from the control, $p = .007$ for both comparisons. While vowels and onsets showed significant differences from the control, codas did not for both decreases and increases in duration, $p = 1.000$ and $p = .820$ respectively.

However, when the features, syllable type, stress, manner, prepausal position, and sentence-boundary adjacency were included in the analysis, all syllable constituents including codas were affected by rate change depending on the feature profile. As was evident in the means for vowels in Table 5.4, the means for onsets and codas show a decrease from the control at a fast rate of speech, and an increase at a slow rate, Tables 5.7, 5.10, and 5.11. The medians of the onset durations are found in Table D.5 of Appendix D.

Table 5.7
Duration of Onsets by Rate, Syllable Type, Stress, Manner, Prepausal Position, and Sentence-boundary Adjacency

		Speech Rate					
		Fast		Control		Slow	
Type	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CVC							
Stop							
Stressed							
Non-PP[a]							
SBA	60	4.58	.16	4.82	.21	5.45	.55
Non-SBA	255	4.51	.18	4.78	.23	4.98	.27
Fricative							
Stressed							
Non-PP							
SBA	60	4.69	.44	4.86	.44	5.03	.52
Non-SBA	45	4.70	.12	4.94	.21	5.20	.55

a. PP = Prepausal Position

In rate comparisons for onsets, the durations of fricatives are not significantly affected by changes in speech rate, as shown in Table 5.8. However, stops show significant increases and decreases when they are not adjacent to a sentence boundary. When a segment is not in prepausal position but is adjacent to a sentence boundary, such as an onset in this data, it is, in effect, in sentence-initial position. Segments that are both prepausal and sentence-boundary adjacent are in sentence-final position. Onsets that are sentence-initial also show significant increases in duration.

Table 5.8
Tukey a Results for Speech Rate Comparisons of Onset Durations

	Speech Rate	
	Fast-Control	Slow-Control
CVC		
Stop		
Stressed		
Non-PP		
SBA	.746	.000
Non-SBA	.000	.004
Fricative		
Stressed		
Non-PP		
SBA	.998	.998
Non-SBA	.927	.860

The descriptive statistics for vowels have been previously discussed in Table 5.4 of Section 5.3, which also included a description of the seven prepausal vowels. When the vowels were analyzed in combination with other syllable constituents, the results of post-hoc comparisons were more conservative, as shown in Table 5.9. However, the same pattern of significant changes in duration found in Table 5.5 of the vowel experiment held with the vowels in context.

Table 5.9
Tukey a Results for Speech Rate Comparisons of Vowel Durations

	Speech Rate	
Features	Fast-Control	Slow-Control
CVC		
Stressed		
SBA	1.000	.999
Non-SBA	.000	.000
VC		
Stressed		
SBA	1.000	.992
Unstressed		
Non-SBA	1.000	.481
V		
Stressed		
Non-SBA	.002	.306
Unstressed		
Non-SBA	.000	.000

For all feature profiles, codas decreased or increased in duration from the control when the rate of speech became faster or slower respectively, as seen in Tables 5.10 and 5.11. In Appendix D, the medians of the original measurements of codas in VC and CVC syllables are presented in Tables D.6 and D.7 respectively.

Table 5.10

Duration of Codas in CVC Syllables by Rate, Stress, Manner, Prepausal, and Sentence Boundary Adjacency

	Speech Rate								
	Fast			Control			Slow		
Type	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Stop									
Stressed									
PP									
SBA	56	5.21	.33	60	5.35	.38	60	5.46	.44
Non-SBA	1	4.33	.00	32	4.62	.26	108	5.05	.34
Non-PP									
SBA	34	4.36	.32	30	4.36	.28	30	4.60	.25
Non-SBA	194	4.15	.35	163	4.42	.38	87	4.51	.33
Fricative									
Stressed									
PP									
Non-SBA				3	4.48	.12	16	5.08	.36
Non-PP									
SBA	15	4.52	.07	15	4.64	.16	15	4.82	.18
Non-SBA	60	4.45	.18	57	4.58	.20	44	4.79	.27
Affricate									
Stressed									
PP									
SBA	15	5.66	.25	15	5.76	.36	15	5.85	.23
Non-SBA				6	4.77	.15	18	5.26	.28

Table 5.10
Duration of Codas in CVC Syllables by Rate, Stress, Manner, Prepausal, and Sentence Boundary Adjacency

	Speech Rate								
	Fast			Control			Slow		
Type	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Non-PP									
Non-SBA	45	4.53	.15	39	4.81	.21	27	4.86	.25

Table 5.11
Duration of Codas in VC Syllables by Rate, Stress, Manner, Prepausal, and Sentence Boundary Adjacency

	Speech Rate								
	Fast			Control			Slow		
Type	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Stop									
Stressed									
Non-PP									
SBA	15	4.28	.10	15	4.39	.18	15	4.65	.30
Fricative									
Unstressed									
PP									
Non-SBA	55	5.13	.20	72	5.24	.40	75	5.58	.24
Non-PP									
Non-SBA	20	4.86	.19	3	5.19	.63			
Stressed									
Non-PP									
SBA	15	4.48	.15	15	4.56	.16	15	4.73	.16

Although the means for the durations of codas showed increases and decreases from the control, there was only one case that was statistically significant (Tables 5.12 and 5.13). Codas in stressed CVC syllables that are not in prepausal position and are not adjacent to a sentence boundary significantly decreased in duration. The insignificant results derived from groups with small n-values are marked with an asterisk. These groups are not sentence-boundary adjacent and are primarily prepausal with the exception of fricatives in VC syllables that are not prepausal. In addition, fricatives and affricates that are adjacent to a sentence boundary have small representations.

Table 5.12
Tukey a Results for Speech Rate Comparisons of Coda Durations in CVC Syllables

	Speech Rate	
Type	Fast-Control	Slow-Control
Stop		
Stressed		
PP		
SBA	1.000	1.000
Non-SBA	.017*	1.000
Non-PP		
SBA	1.000	.996
Non-SBA	.001	.945
Fricative		
Stressed		
PP		
Non-SBA		1.000*

Table 5.12
Tukey a Results for Speech Rate Comparisons of Coda Durations in CVC Syllables

	Speech Rate	
Type	Fast-Control	Slow-Control
Non-PP		
SBA	1.000	1.000
Non-SBA	1.000	.995
Affricate		
Stressed		
PP		
SBA	1.000	1.000
Non-SBA		1.000*
Non-PP		
Non-SBA	.886	1.000

Table 5.13
Tukey a Results for Speech Rate Comparisons of Coda Durations in VC Syllables

	Speech Rate	
Type	Fast-Control	Slow-Control
Stop		
Stressed		
Non-PP		
SBA	1.000	1.000
Fricative		
Unstressed		
PP		
Non-SBA	1.000	.020

Table 5.13
Tukey a Results for Speech Rate Comparisons of Coda Durations in VC Syllables

	Speech Rate	
Type	Fast-Control	Slow-Control
Non-PP		
Non-SBA	1.000*	
Stressed		
Non-PP		
SBA	1.000	1.000

In general, syllable constituents did show significant differences in duration when the rate of speech changed. No significant changes in the duration of syllable constituents were evident in VC syllables. In CVC syllables, durations of onset and vowels showed both significant increases and decreases, while codas showed only significant decreases. Durations also changed significantly in both directions for vowels comprising a syllable, decreasing regardless of stress and only increasing in unstressed syllables. The changes in onsets and codas did not include fricatives and affricates, only stops; and stop onsets were the only constituents to exhibit significant differences adjacent to a sentence boundary. This boundary was sentence initial. In addition, there were no incidences of any significant changes in duration for constituents in prepausal position.

CHAPTER 6 Discussion

The results of this study provide evidence that a tempo pattern does exist in English, and that this pattern is organized around the focused word. This pattern can potentially be replicated by employing the results of the experiment determining sites for tempo change in an algorithm. In addition, further insights have been gained as to the effects of focus on other constituents in the sentence, as well as the effects of changes in tempo on syllabic constituents. This chapter discusses the results of the experiments described in the previous chapter and provides an example of a potential application within an existing algorithm for speech rate that has been described in the Literature Review (Klatt, 1979).

Tempo Pattern Experiment

The purpose of the tempo pattern experiment was to provide evidence that changes in tempo are organized around the focused word. The first pattern to be tested was the basic deceleration-acceleration pattern observed by Fant et al.

(1991c). Table 1 reviews the hypotheses presented in the chapter describing the methodology. The deceleration pattern is reflected in the first hypothesis and acceleration in the second. To separate the effects of focus on those constituents that precede the focused word and the focus word itself, the second and third hypotheses were included in the analysis.

Table 6.1
Hypotheses for the Tempo Pattern Experiment

Experiment	Hypotheses
Experiment I	The ASDs of the pre-focal inclusive region are greater than the ASDs of the control
Experiment II	The ASDs of the pre-focal exclusive region minus the focal word are less than the ASDs of the control
	The ASDs of the focused word are greater than the ASDs of the control.
Experiment I & II	The ASDs in the post-focal region are less than the ASDs of the control.

6.1 *Experiment I*

Pre-focal Inclusive Region

In the pre-focal inclusive region, the variance means were significantly different from zero in two of the three foci positions, Focus 1, where the focus word is located near the beginning of the sentence, and Focus 2, where the focus word was closer to the end. Apparent in Table 5.2 of Chapter 5, Results, both means were positive and therefore reflect an increase in duration from the control. As a result, the hypothesis that the ASDs of the pre-focal inclusive region are greater than those of the control can be accepted only if the focused word is located in either Focus 1 or Focus 2 position. ASDs do not appear to change in duration when the focus is near the end of the sentence.

Table B.3 of Appendix B lists the means of the transformed data for the ASDs in pre-focal inclusive regions. Given these means and the medians of the non-transformed data, there was a greater increase in ASDs from the control in Focus 1 than in Focus 2. The means of the transformed ASDs in Focus 1 show a six percent increase over those in Focus 2. A very similar relationship was also apparent in the non-transformed data. The medians of Focus 1 ASDs show a 13% increase over the control while the medians of Focus 2 ASDs show an 8% increase.

These results confirm the findings of the pilot study, but with a smaller degree of change. Referring to Figure 3.1 of the pilot study, it is clear that the increases in the means decrease as the focus word nears the end of the sentence. To review, there was a 30%, 15%, and 3% increase in ASDs for Focus 1 through Focus 3, respectively.

Post-focal Region

In terms of significant changes in duration from zero, post-focal and pre-focal inclusive regions exhibited similar results. The variances for ASDs in Focus 1 and Focus 2 positions were significantly different from zero, and in both cases the variance means exhibited a decrease from the control. Therefore, the hypothesis that the ASDs in the post-focal region are less than those of the control can be accepted if the focus is initial or medial, but not if it is near the end of the sentence.

The descending pattern of degree of change found in the pre-focal inclusive region is not evident in the post-focal region. Considering the means of the transformed data in Table B.4 in Appendix B, ASDs in Focus 2 decreased from the control only two percent more than those in Focus 1, while the medians of the non-transformed data held the same degree of increase for both foci positions, eight percent. Similar to the results of the pilot study (Figure 3.3), the position of the focus word appears to have little effect on the degree of compression in ASDs in the post-focal region.

Compression found in the post-focal region upholds previous findings by Folkins et al. (1975) and Weismer and Ingrisano (1979). For example, Weismer and Ingrisano found that segment durations decreased following a focused word located near the beginning of a sentence. These findings were based upon data consisting of five-syllable sentences, so there was no opportunity to test for sentence-medial focus; therefore, a valid comparison between studies is not possible on this basis. The results of the present study support and expand these findings by introducing a medial focus and increasing sentence length (the data of the present study were derived from 12 syllable sentences).

However, the fact that compression also occurs in longer sentences refutes the suggestion put forth by Eady and Cooper that shorter utterances may bear more pronounced effects of focus (1986, p. 412). This suggestion was a result of an attempt to resolve the differences in results between their study and those of Folkins et al. (1975) and Weismer and Ingrisano (1979). In their 1986 study and in Eady et al. (1986), it was identified that non-focus words were not significantly affected by focus, although a slight trend for post-focal compression was noted.

The lack of post-focal compression in the most recent study may be a result of experimental design. Eady and Cooper (1986) studied the effects of focus on both statements and questions. Of the 144 sentences, 72 were statements. In the present study, 300 statement sentences were analyzed. A larger sample may reduce the possibility of a Type II error: previously insignificant results may prove significant given a larger sample size. The post-focal results are replicated for Experiment II.

From the results of this study, it can be inferred that a deceleration-acceleration tempo pattern based upon a focused word in sentence-initial or sentence-medial position, exists in English. This supports observations by Fant et al. (1991c) of one deceleration followed by one acceleration in Swedish. The ASDs in Figure 6.1 suggest an inverse proportional relationship between the focus position and the focus region. The more the pre-focal inclusive ASDs increase the less the post-focal ASDs decrease. However, the differences between both focus positions in post-focal region ASDs is minimal.

This pattern suggests that the deceleration is a result of either the focal word

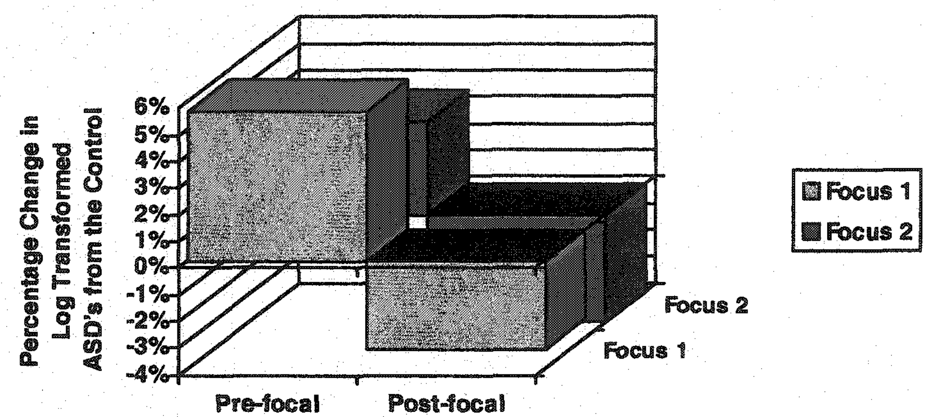

Figure 6.1.
Comparison of Changes in ASDs by Focal Region and Focus Position

increasing as suggested by Eady et al. (1986) and Eady and Cooper (1986), the pre-focal syllables increasing in addition to the focal word, or the focal word increasing to a large extent accompanied by the compression of the pre-focal exclusive syllables as observed in the pilot study (Table 3.2). The latter was confirmed in Experiment II.

6.2 *Experiment II*

Pre-focal Exclusive Region

Similar to the results for the pre-focal inclusive regions in Experiment I, variances in ASDs were significantly different from zero for Focus 1 and 2. However, in this case, the negative means of the variances reflect a decrease from the control in ASDs (Table 5.3 of Results). Unlike in Experiment 1, the pre-focal exclusive variances for ASDs in Focus 3 sentences were also significantly different from zero. As a result, the hypotheses that the ASDs of the pre-focal exclusive region minus the focal word are less than the ASDs of the control can be accepted when the focus is in sentence-initial, medial, and the end portion of the sentence. Therefore, there is evidence that pre-focus exclusive syllables do accelerate in rate, contrasting again with the findings of Eady et al. (1986) and Eady and Cooper (1986) discussed in the previous section.

In terms of the degree of change in ASDs from the control, the compression results are comparable for the post-focal regions in Experiment I and the pre-focal regions exclusive in Experiment II. Transformed ASDs decreased in duration from 3.3% and 4% for the post-focal region, and 2.8%, and 3.1%, and 3.5% for the pre-focal exclusive region, Foci 2, 3, and 1, in that order. For Foci 1 and 2, an inverse relationship in the degree of change in ASDs exists between the pre-focal results for Experiment I and II. In Experiment I, Focus 1 exhibited the highest degree of change, but in Experiment II, Focus 2 holds the highest degree of change with Focus 1 having the lowest. According to the medians in Table B.5 of Appendix B, the non-transformed data displays the degree of compression in descending order from Focus 1 through 3, 9%, 8%, and 5%, respectively.

Compression of pre-focal exclusive syllables was also evident in the pilot study. However, the distribution of the degree of compression over the focal regions differed. In Table 3.3 of the pilot study, the degree of compression in the means of ASDs increased as the focus neared the end of the sentence with the exception of Focus 1, which exhibited expansion.

Focus Word

In light of the observed compression of the pre-focal exclusive syllables, the focus word ASDs must expand beyond the amount necessary to compensate for this compression to create an increase in the ASDs of the pre-focal inclusive region that includes the focal word. This expansion is obtained in the results. For all focus positions, the means of the variances for the focus word increased significantly from zero (Table 5.3 of Chapter 5, Results). Consequently, the hypothesis that the ASDs of the focused word are greater than the ASDs of the control can be accepted. This supports findings that the focus word significantly increases in duration as described by Eady and Cooper (1986) and Cooper et al. (1985, 1986). However, in Eady and Cooper, it was only the initial focus word that exhibited significant changes in duration.

The means reflect a large degree of change for the transformed ASDs. For Foci 1 through 3, ASDs increase from the control 22.4%, 22.7%, and 10% respectively (Table B.6 of Appendix B). Note that the increase of ASDs in Focus 3 are less than half of those in Focus 1 and Focus 2. This fact combined with the degree of compression of the pre-focus exclusive syllables resulted in the insignificant increase for the pre-focal inclusive region which included the focus word in Focus 3. Therefore, it becomes obvious that pre-focus exclusive syllables and the focus syllables must be analyzed separately in order to avoid the effects of one cancelling out the effects of the other, as was found in Focus 3 of this study.

The increase in the medians of the ASDs reflects the same descending order in the degree of change as was found in the pre-focus syllables. In Foci 1 through 3, focus syllables increased 65%, 62%, and 23%, respectively.

Contrary to the findings in Experiment I, the variance means for the ASDs in the pre-focal exclusive regions and the focus word in all focal positions exhibited a difference. The means decreased in pre-focal exclusive position and increased for the focus word. While significant differences in the direction of a decrease in duration were found for the ASDs in Focus 1 and 2 positions, an insignificant increase was evident for the ASDs in Focus 3 position of the post-focal region.

Although a deceleration-acceleration pattern is evident in Experiment I, these results cannot be accepted in light of the findings of Experiment II. The existence of pre-focus compression changes the previously described tempo pattern to one consisting of one acceleration followed by one deceleration, and concludes with another acceleration using the focused word as the locus for change as represented in Figure 2. The only caveat is that if the focus is near the end of the sentence, the final acceleration does not take place. This acceleration-deceleration alternation is

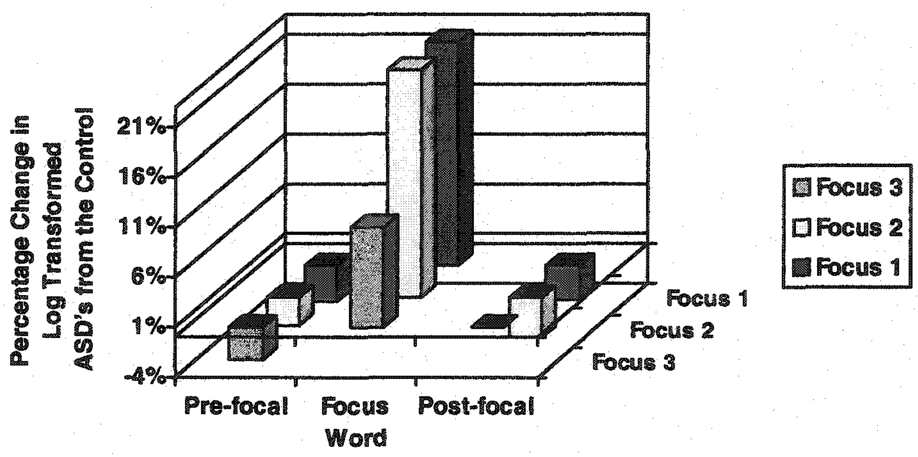

Figure 6.2.
Comparison of Changes in ASDs by Focal Region Including the Focus Word and Focus Position

similar to the alternations in duration observed by Miller et al. (1984) and Crystal and House (1990). However, their studies were not concerned with focus.

To summarize, the focus words exhibit a marked increase in duration when the focus is in sentence-initial or sentence-medial position. However, when the focus nears the end of the sentence, there is roughly a 60% reduction in the expansion of the focus compared to the other focus positions. Pre-focal and post-focal compression is approximately a 3.3% decrease in the transformed durations of the non-focus syllables of the sentence. If the focus word is not in sentence-initial or sentence-medial position, no post-focal compression occurs.

Sites for Rate Change Experiments

The results of the pilot study suggested that unstressed vowels comprising a syllable and stressed codas in CVC syllables are the primary sites for expansion. The highest degree of compression occurred for codas in VC syllables followed by onsets in CV syllables. The main study revealed different results. Vowels appear to be the primary site for rate change for both expansion and compression followed by onsets in the case of expansion and codas and then onsets in the case of compression.

Thus a pattern for tempo change has been established for English. The results for this part of the study provide the specific locations for changing speech rate in order to replicate the tempo pattern. Table 6.2 reviews the hypotheses for the site experiments. The hypothesis for the first experiment establishes the vowel as a viable site for tempo change. This hypothesis is placed within the context of the syllable by the second and third hypotheses which attempt to rank the degree of tempo change. The outline for the discussion echoes the presentation of the results.

Table 6.2
Hypotheses for the Experiments for the Potential Sites for Tempo Change

Experiment	Hypotheses
Vowel	The vowel increases or decreases in duration as tempo decreases or increases, respectively.
Syllable Constituents	The coda compresses more than the onset or peak of the syllable at a fast rate.
	The coda expands more than the onset or peak of the syllable at a slow rate.

6.3 *Vowel Duration*

When tempo changes, compression and expansion is evident in vowels as seen in the relationships of the means in Table 5.4 of Chapter 5, Results. However, expansion is not significant unless the vowels are defined in terms of syllable type, stress, and sentence-boundary adjacency. The transformed durations of vowels exhibited quite large significant increases from the control when the vowel comprised an unstressed syllable, 13.4%. While the degree of increase was considerably less for vowels in stressed CVC syllables, 4.7%, the increase was significant. The non-transformed data held respective increases of 76% and 33% in the medians. The expansion aspect of the hypothesis that vowels increase in duration as tempo decreases can be accepted if the additional criteria of syllable type, stress, and sentence-boundary adjacency is included. For example, unstressed V and stressed CVC vowels increase in duration as tempo decreases.

Marjomaa (1983) showed that there were significant differences in the duration of vowels within CV syllables when the tempo was changed. Although CV syllables were not used in the present study, the observed expansion of stressed CVC vowels expands the scope of the previous findings of Marjomaa.

Confirming the results of the pilot study, this study indicates that unstressed vowels expand more than stressed vowels. This could be interpreted as a conflict with the findings of Crystal and House (1990), who discovered that, in comparisons of vowel durations of slow and fast speakers, stressed vowels were longer than unstressed. Although the design of their study did take syllable type into account, these data represent vowel measurements of only two speakers, which is problematic for providing generalizations about English. In addition, there were no comparisons of vowel durations at either slow or fast tempi with a control. Therefore, it is not clear what the effects are on duration of each tempo.

Additionally, the larger increase in unstressed vowels compared to stressed vowels reflects the tendency for unstressed syllables to increase in duration more than stressed syllables, as observed by Peterson and Lehiste (1960) and Port (1980). Given that there may be a maximum duration for segments and or syllables, stressed constituents, which are longer by virtue of being stressed may have less room to expand than unstressed constituents. This may explain the smaller increases of stressed vowels.

While little evidence of compression was revealed in the pilot study, the main study revealed significant differences for the decreases in general vowel duration.

The mean of the transformed vowel durations decreased 5.9% from the control, and medians of the non-transformed durations 20%. With the further refinement of syllable type, stress, and sentence-boundary adjacency, vowels in stressed and unstressed V syllables and stressed CVC syllables significantly decreased in duration from the control by 8.0%, 7.2%, and 6.9% respectively (28%, 24%, and 27% for the medians of the non-transformed data).

The hypothesis that the vowel decreases in duration as tempo increases can be accepted in light of the general vowel results, providing additional support for the results of Marjomaa (1983). With the additional specification for syllable type, stress, and sentence-boundary adjacency, the hypothesis still holds for vowels in stressed and unstressed V syllables and those in stressed syllables. Similar to the results for expansions, compressions show the largest degree of change for vowels in V syllables, followed by those in CVC syllables. The one distinction between the two is that both stressed and unstressed vowels in V syllables change significantly, stressed vowels having the highest degree of change.

The higher ranking of degree of change for stressed vowels supports previous findings by Han (1966). Also examining the effects of tempo, Han determined that stressed vowels are reduced the most when tempo increased. Contrary to the current findings, unstressed vowels did not exhibit significant differences (p. 78).

Assuming there is a minimum duration for syllables, CVC syllables, which in this study carried more semantic weight than V syllables, may as a result be less elastic. As a consequence, CVC syllables compress less than V syllables.

Although differences in vowel duration were found for V and CVC syllables, significant differences for VC vowels did not occur. This may be a result of the restricted context of this type of syllable. VC vowels occur in the second syllable of the adverb for all sentences. It is clear that rate does not effect significant changes in duration for this position but, given other positions, such as those provided in the pilot study, the changes in duration could be significant.

There were no cases where the duration of sentence-boundary adjacent vowels significantly interacted with changes in tempo. Assuming that vowels undergo phrase-final lengthening, this implies that the same degree of lengthening applies for both fast and slow speech rates.

In summary, vowels do expand and compress as tempo changes, but in these data the change is restricted to vowels in V and CVC syllables (Figure 6.3). Only vowels in unstressed V syllables show a marked expansion. In addition, tempo changes did not induce any variation in vowel duration if the syllable was adjacent to a sentence boundary.

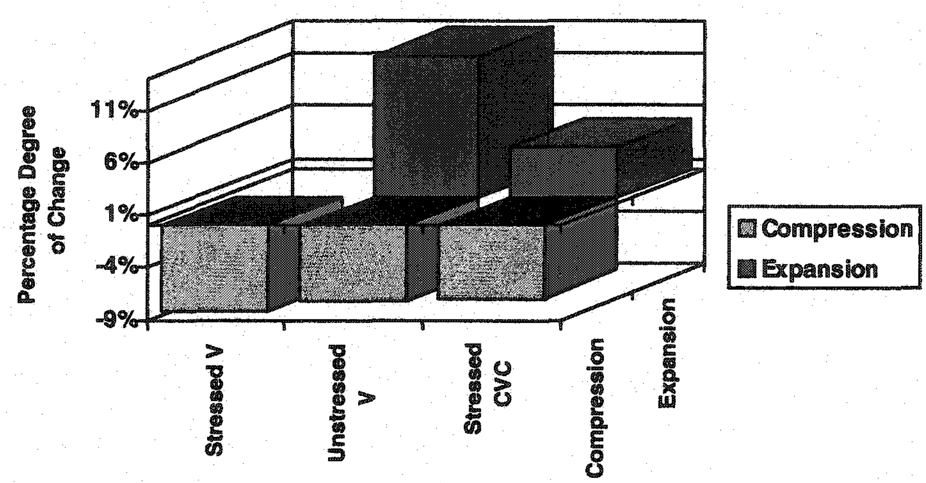

Figure 6.3.
Comparison of Percentage Changes in Vowel Duration from the Control

6.4 *Syllable Constituency*

The distribution of the significant changes in duration of syllable constituents provides potential sites for implementing tempo change, in addition to contributing to the understanding of the effects of tempo on language. The overall analysis of vowels identified by the syllables to which they belong revealed that expansion but not compression in vowel duration was significant. However, more useful results were obtained when syllable and segment features were included in the analysis. Within the context of the syllable, the results for vowels maintained (Chapter 5, Results, Table 5.5). However, not only vowels were affected by tempo changes.

As was described in the preceding chapter, all syllable constituents were affected by rate change depending on the feature profile.

For onsets, only stops and not fricatives showed significant changes from the control (Chapter 5, Results, Table 5.8). Han (1966) also found evidence that fricatives were resistant to changes in tempo (p. 76). Stops in onset position appear to compress more than they expand, 5.7% and 4.3% for respective changes in the means (25% and 21% for the medians of the non-transformed data (Appendix C, Table C.5). These onsets are in stressed CVC syllables that are not in prepausal position. However, when they are adjacent to a sentence boundary, they expand markedly at 13.1% (87% in the non-transformed data) providing evidence that tempo interacts with the initial sentence boundary.

Both expansion and compression of codas were evident in the means of the transformed data; however, there was only one case that was significant (Chapter 5, Results, Table 5.12). Stops in the coda position of CVC syllables that are not prepausal and are not adjacent to a sentence boundary compress significantly for a 6.1% change in the means of the transformed duration (16% for medians of the non-transformed data (Appendix C, Table C.6)). Codas with the same feature profile with the exception of those in prepausal position have the potential for compressing significantly. There are other cases where the *n*-value is small. This combined with a .017 probability value being very close to the .01 alpha level suggests the possibility of a Type II error. Consequently, stops in the coda position of syllables that are prepausal could reach a significant level of compression given a larger group size - a subject for further research. The absence of significant changes in fricatives in coda position further supports the findings of Han (1966). It should also be noted that affricates in coda position did not show significant changes. As was revealed in the vowel experiment, constituents of VC syllables do exhibit significant changes in duration when tempo changes. Like vowels, codas did not exhibit significant changes in duration near a sentence boundary or in prepausal position.

Given that onsets, vowels, and codas demonstrate significant expansions and compressions, the degrees of change require comparison to establish whether or not the hypothesis can be accepted or rejected. In terms of compression, the hypothesis that the coda compresses more than the onset or peak of the syllable at a fast rate cannot be accepted. Consider the relationships between syllable constituents in Figure 6.4; it is clear that vowels or peaks show the greatest degree of compression in CVC syllables, followed by stop codas and finally stop onsets.

With the exception of the difference in peak behaviour, onsets and codas maintain a similar relationship to that observed in the pilot study. Stop codas compress more than onsets. However, that peaks compress more than codas stands in contrast to the findings of Campbell and Isard (1991). They discovered the reverse that codas compress more than peaks as was also found in the pilot study for stressed constituents. However, their findings and those of the pilot study were

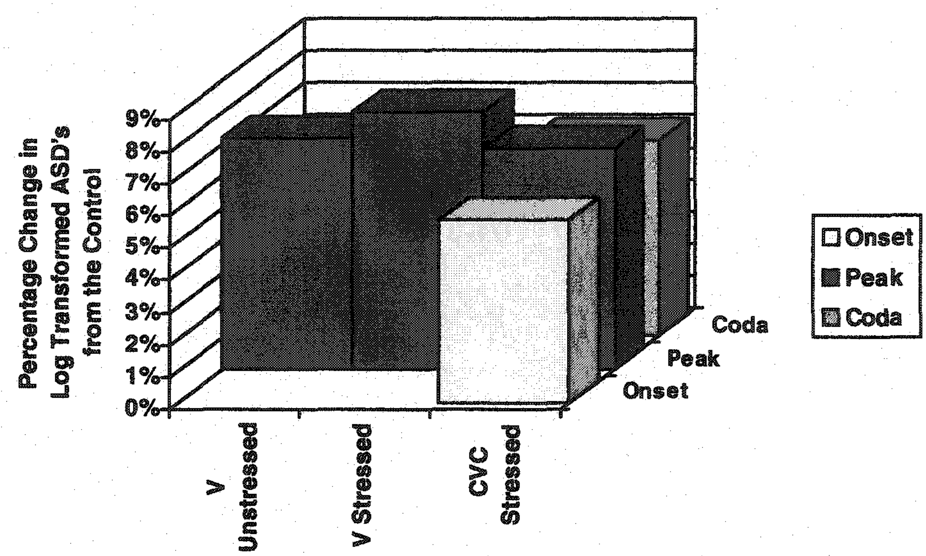

Figure 6.4.
Comparisons of Degrees of Compression in Syllable Constituents

based on data provided by one male speaker unlike the present study which is based on the data of five speakers. As a result, the findings of Campbell and Isard may be a feature of the idiolect of this speaker.

The pilot study indicated that codas in stressed CVC syllables expanded more than either peaks or onsets. However, there was no evidence of this in the main study, in which incidences of significant expansion did not occur. As a result, the hypothesis that the coda expands more than the onset or peak of the syllable at a slow rate cannot be accepted.

Expansion was evident for both vowels and stop onsets, and a comparison of degrees of change provides valuable information for a model. In Figure 6.5, peaks in V syllables expand more than onsets regardless of sentence-boundary adjacency. However, within stressed CVC syllables, stop onsets increase more than peaks.

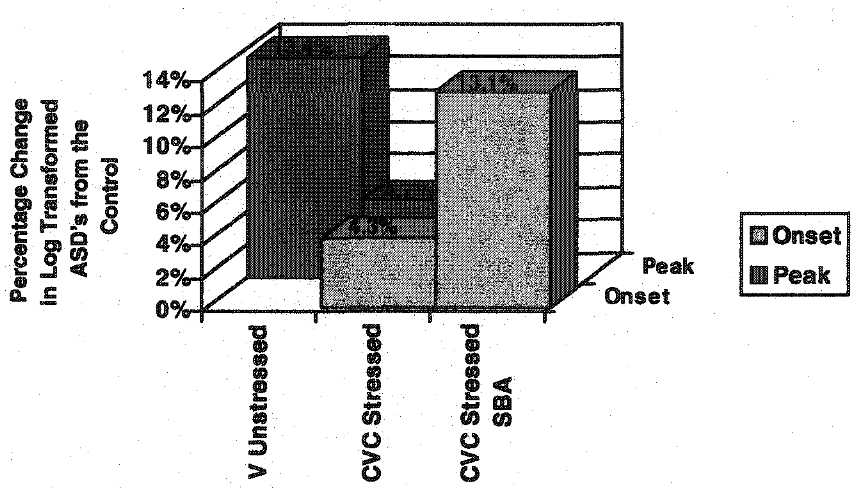

Figure 3.5.
Comparisons of Degrees of Expansion in Syllable Constituents

To summarize the results of the experiments designed to determine sites for tempo changes, vowels, particularly those that constitute a full syllable, appear to be the primary site for expansion and compression. Stressed vowels have the largest compression, while unstressed vowels have the largest expansion. In stressed CVC syllables, codas and then onsets show lessening degrees of compression. This order is reversed for expansion, and the degree of change for these constituents is less than that for compression. Regarding onsets and codas, only stops appear to undergo significant changes. In addition, sentence-boundary adjacency appeared to affect only the expansion of onsets in stressed CVC syllables when tempo decreased. No other elements of the sentence were affected by the initial boundary.

The study shows that fricatives and affricates are resistant to rate changes. However, it was not possible to have all manners of articulation represented in the data. It is not clear how nasals, laterals, etc., react to changes in tempo, or how constituents of CV syllables react. Therefore, this area requires further research.

The results of both the tempo pattern and the site experiments provide both a pattern of tempo in English and a description of the potential compression and expansion sites for replicating the tempo pattern. While these sites do not provide a complete representation of all possible feature combinations, they do indicate that tempo changes involve complex distributions and do not involve a simple change in duration without regard for constituency type, syllable type, stress, etc.

The findings of this study need to be adapted to an existing duration model for a text-to-speech system in order to validate perceptually the proposed generalisations. However, an example of an application can easily be provided using the algorithm proposed by Klatt (1979). The Klatt duration model has become a benchmark according to Carlson (1991, 243), and being rule-based can easily be adapted to reflect the generalizations provided by this study.

As outlined previously in Table 2.1 (p. 10), the rules for the model include both shortening and lengthening rules. Emphasis is included in the lengthening rule. This rule specifies that an emphasized vowel is lengthened by 140% (p. 295). This rule can be easily adapted to reflect the results of this study if one interprets emphasis and focus to mean the same thing.

The tempo pattern experiment provided further evidence that focus words increase in duration, and the results of the site experiments revealed that tempo change does not affect a single constituent, i.e. the vowel, as suggested by the emphasis rule. Therefore, a more accurate representation of speech may be achieved by incorporating the affected members of the syllable into the rule as well as the location of the focused word within the sentence. In addition, the effects of focus on the remainder of the sentence need to be accounted for.

For the sake of illustration, only the percentage changes in the medians of the non-transformed data are used in the following application. Instead of the vowel bearing all of the expansion, lengthening should be shared between the onset and the vowel, according to this study.

If the focus is in sentence-initial or sentence-medial position, the stressed syllable should increase by approximately 64%. As an example, assuming that the stressed syllable is a CVC syllable, the stop onsets should expand by 21% and vowels by 33%. A proportional increase for these constituents of 26% and 38% would be necessary in order to achieve the 64% increase in the focus. The percentage increases are by no means absolute. It is the relationship between syllable constituents that is important. This relationship can easily be applied to the Klatt model.

The degree of change for the focus needs to be modified if it is located near the end of the sentence. In this position, the duration of the focused syllable increases roughly 60% less than syllables in other positions. However, this would bring the degree of change down to 4%. An important consideration is the degree of change, as the change must meet or exceed the perceptual threshold for rate changes. Hoequist and Kohler (1986) suggest that a minimum threshold for perceptual tasks involving tempo changes of 17 ms. per syllable is necessary for determining a change in tempo. The average syllable length for unfocused syllables in the tempo pattern experiment is 247 ms. Therefore, at least a 7% change in the duration of the syllable is necessary for the change to be perceived. Given this perceptual threshold, the focus near sentence-final position is required to increase 57% less than a focus in other positions. This translates into a 3% and 4% expansion in stop onsets and vowels, respectively, in the focused syllable.

Given that there is an expansion for the focus, there should be a compression of the non-focused syllables of the sentence if the focus is sentence-initial or sentence-medial. While the degree of compression is small at approximately 8%, it exceeds the perceptual threshold. Therefore, all non-focused syllables should be compressed by this amount. To achieve the compression, the durations of the constituents of stressed syllables would have to be altered. Vowels that are not in VC syllables would compress the most, followed by stop onsets and then by stop codas. Unfortunately, the ranking of onsets and codas was altered in the log transformation, creating a problem in generalizing the ranking of these constituents. The incorporation of both ranking possibilities within a duration model should be tested to determine which one obtains the most natural results. For the sake of the following example, they are considered of equal ranking. Comparing the medians for compression, onsets were compressed by 25% and codas by 36% less than the onsets. In the transformed data, the codas compressed by 6.1% and the onsets by 7% less. Considering the small difference in degree of change for the transformed data and the fact that the overall compression is 8%, onsets and codas will be considered to have the same degree of change for the purposes of this example. Therefore, vowels that are not in VC syllables will compress by 3.2%, leaving stop onsets and codas to each compress by 2.4% to achieve the overall 8% compression for the syllable.

This generalization closely mimics the observations of a controlled set of English speech provided by multiple speakers, which therefore has the potential to provide a natural quality for synthetic speech. Without testing, it is unclear what amount of detail in the generalization, i.e. manner of articulation, is necessary to provide a natural quality. Therefore, it should be noted that no definitive claims can be made regarding the details and the actual degree of compression or expansion without perceptual testing for naturalness.

CHAPTER 7 Conclusion

The primary goal of this project was to provide a duration model of English tempo, and thus, the first objective was to verify a tempo pattern for English. Secondly, it was necessary to determine the potential sites of rate change in order to be able to replicate the inherent compressions and expansions of the pattern.

Overview

Based upon the results of this study, a pattern of tempo alternations exists in English where there is a focus in the sentence. However, the pattern did not involve one deceleration followed by one acceleration as has been observed in Swedish (Fant et al., 1991). Instead, it was observed that tempo alternates three times, beginning with an acceleration, followed by a deceleration, and concluding with another acceleration. The locus for a change in tempo is the focus. The focus bears the highest semantic prominence of the sentence for the speaker and, consequently, exhibits a large expansion in duration to signal this importance. The lack of prominence also triggers a change in duration for the non-focused syllables of the sentence. This de-emphasis serves to further highlight the focus. However, the degree of compression is relatively much less than that of the expansion of the focus. When the two accelerations in the sentence are compared, the degree of compression preceding and following the focus is quite similar.

The degree of change in the focus and in unfocused syllables does not remain constant when the location of the focus changes. It appears that when the focus is near the end of the sentence, the degree of emphasis is considerably reduced to the point that no de-emphasis occurs in the syllables following the focus. When the focus is located near the beginning or the mid-point of the sentence, the degree of change differs minimally. Clearly, the tempo pattern is more complex than a simple deceleration and acceleration.

These findings also contribute to the research on the effects of focus. They confirm that the focus does significantly increase in duration and that post-focal syllables undergo compression. Evidence is also provided for the phenomenon of pre-focal compression.

These results suggest potential locations for replicating the pattern of accelerations and decelerations. The primary site for the necessary expansions and compressions is the vowel given two conditions, syllable constituency and stress. When a vowel comprises a syllable it appears to be most vulnerable to changes in tempo. If this syllable is stressed, the vowel exhibits the greatest compression, and when it is unstressed the greatest expansion.

The most noteworthy information provided by the tempo site experiment is that the vowel is not the sole location for change. Rather, tempo changes are distributed amongst the other constituents of the syllable. This distribution is not equal. In stressed CVC syllables, codas and then onsets show lessening degrees of compression; however, this order is reversed for expansion. In addition, the degree of change for the expanded constituents is less than that for compression.

However, not all onsets and codas exhibit tempo change. In the present study, it was demonstrated that fricatives and affricates are resistant to change. Unlike fricatives and affricates, stops show more elasticity when tempo changes. The duration of stops appears to expand or compress significantly. Therefore, in terms of consonants, stops are potential sites for tempo change. In addition, all constituents of the syllable do not seem to be affected by a sentence boundary. Only when the syllable constituent is an onset in a stressed CVC syllable, then significant changes in duration are apparent in that the onset expanded.

The tempo site experiment provides further insights in the research of the interaction between changes in tempo and phonetic segments that carry cues of rhythm. In addition, the results can be translated into generalizations of segmental expansions and compressions that are required to reproduce the tempo pattern. These generalizations can be adapted into an existing rule-based model of duration such as the one proposed by Klatt (1979) for applications such as text-to-speech systems.

Further Studies

The most obvious area for further research is a perceptual test of the results. The generalizations provided by this study need to be incorporated into a working duration model of a text-to speech system, so that the product of the resultant expansions and compressions can be assessed for naturalness. The results of a perceptual test may involve changes in the degrees of deceleration of the focus and acceleration of the non-focus syllables in a sentence, as well as in expansion and compression of syllable constituents in order to meet the requirements of naturalness.

The data for this study were derived from male speakers and as a consequence, the results are only valid for one gender. Further research is warranted using both male and female speakers to investigate the effects of gender on the proposed generalizations. Given the observed tendency for women to read slower and to modify sentential duration (Whiteside, 1996), there is the possibility for varying results in the generalizations. Instead of attempting to resolve these differences, creating gender-specific duration models may provide a potential solution. The present study could be replicated using female speakers to create a model for the voice of a woman in synthetic speech.

The generalization of potential sites for changes in tempo is not comprehensive, as it was not possible to include all possible feature combinations of stress, manner, and syllable type in this study. A more robust generalization would require a much larger sample that provided sufficient group sizes for each possible feature profile so that potential differences between the groups could be determined. This may or may not affect the ranking of the degree of expansions and compressions amongst syllable constituents. Although the generalization may not be comprehensive, it does provide for a varied distribution of changes in tempo within a sentence or portion thereof. This feature is an improvement over other approaches in that the variety replicates the variety found in natural speech, thereby increasing the naturalness of the synthetic product.

Conclusion

The proposed model for English tempo provides variation at two levels: sentential and segmental. The sentential level compresses non-focused syllables and expands focused syllables creating an alternating pattern of accelerations and a deceleration. Incorporating the alternating compressions and expansion into existing models of duration is a potential improvement over a rule-based model for tempo that proposes one standard rate of speech for an utterance or is restricted to altering the duration of emphasized words and not the de-emphasized syllables in the sentence. At the segmental level, segments within an accelerated region undergo different degrees of compression dependent upon periodicity and manner of articulation. If these features are available in the duration models used for the synthetic speech system such as those fashioned after the model proposed by Klatt (1979), the parameters for the durations of these segments can be modified to reflect the proposed tempo model. The anticipated result of such modifications is a greater degree of naturalness in the speech product. Therefore, the incorporation of this tempo pattern into a synthetic speech system, such as text-to-speech has the potential to provide a more natural target given any necessary modifications provided by perceptual testing.

Bibliography

Adams, B., Dorai, C., &Venkatesh, S. (2000). Towards automatic extraction of expressive elements from motion pictures: tempo. *IEEE International Conference on Multimedia and Expo*, 2, 641-644.

Allen, GD. (1973). Segmental timing control in speech production. *Journal of Phonetics*, *1*, 219-237.

Bellugi, U. & Fischer, S. (1972). A comparison of Sign Language and spoken language. *Cognition*, 173-200.

Bolinger, D. (1972). Accent is predictable (If you're a mind-reader). *Language, 48* (3), 633-644.

Black A.W. & Taylor, P. (1994). Assigning intonation elements and prosodic phrasing for English speech synthesis from high level linguistic input. *International Congress on Spoken Language Processing*, *1* (4), 715-718.

Brown, W.S. & McGlone, R.E. (1974). Aerodynamic and acoustic study of stress in sentence productions. *The Journal of the Acoustical Society of America*, *56*, 971-974.

Brubaker, R.S. (1972). Rate and pause characteristics of oral reading. *Journal of Psychological Research, 1,* 141-147.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *The Journal of the Acoustical Society of America, 92* (1), 593-596.

Campbell, N. (1990). Measuring speech-rate in the spoken English Corpus. In J. Aarts & W. Meijs (Eds.), *Theory and Practice in Corpus Linguistics* 61-81. Atlanta, GA: Rodopi.

Campbell, W.N. & Isard, S.D. (1991). Segment durations in a syllable frame. *Journal of Phonetics, 19*, 37-47.

Carlson, R. (1991). Duration models in use. *Proceedings of the XIIth International Congress of Phonetic Sciences, 1* (5), 243-246.

Carlson, R. & Granström, B. (1973). Word Accent, emphatic stress, and syntax in a synthesis by rule scheme for Swedish. *STL-QPSR 2-3,* 31-36.

Carlson, R. & Granström, B. (1986). A search for durational rules in a real-speech data base. *Phonetica 43,* 141-154.

Cedergren, H.J., & Perreault, H. (1994). Speech rate & syllable timing in spontaneous speech. *International Congress on Spoken Language Processing, 1* (4), 1087-1090.

Cooper, W.E., Eady, S. J., & Mueller, P.R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America, 77* (6), 2142-2156.

Crystal, T.H. & House, A.S. (1988). Segmental durations in connected speech signals: Current results. *The Journal of the Acoustical Society of America, 83* (4), 1553-1573.

Crystal, T.H. & House, A.S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America, 88* (1), 101-112.

Cummins, F. & Port, R.F. (1996). Rhythmic Constraints on English Stress Timing. *International Congress on Spoken Language Processing* [CD-ROM]. File: A437.pdf.

Eady, S.J. & Cooper, W.E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America, 80* (2), 402-415.

Eady, S.J., Cooper, W.E., Klouda, G.V. Mueller, P.R., & Lotts, D. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech, 29* (3), 233-251.

Eefting, W. (1988). Temporal variations in natural speech: Some explorations. *Proc. Speech '88: 7th FASE Symposium, Edinburgh,* 503-507.

Fackrell, J., Vereecken, H., Martens, J., & Van Coile, B. (2000). The variation of prosody with text type. *IEEE Seminar on State of the Art in Speech Synthesis, 5/* 1-5/9.

Fant, G (1991). Units of temporal organization. Stress Groups versus Syllables and Words. *Proceedings of the XIIth International Congress of Phonetic Sciences, 1* (5), 247-250.

Fant, G., & Kruckenberg, A. (1996). On the quantal nature of speech timing. *International Congress on Spoken Language Processing* [CD-ROM]. File: A749.pdf.

Fant, G., Kruckenberg, A. & Nord, L. (1991a). Language specific patterns of prosodic and segmental structures in Swedish, French and English. In *Proceedings of 12th International Congress of Phonetic Sciences,* 4, Aix-en-Provence, 118-121.

Fant, G., Kruckenberg, A., & Nord, L. (1991b). Some observations on tempo and speaking style in Swedish text reading. Paper presented at the ESCA workshop on the phonetics and phonology of speaking styles, Barcelona, Spain.

Fant, G., Kruckenberg, A., & Nord, L. (1991c). Temporal organisation & rhythm in Swedish. *Proceedings of the XIIth International Congress on Spoken Language Processing, 1* (4), 251-256.

Fant, G., Kruckenberg, A., & Nord, L. (1992). Prediction of syllable duration, speech rate and tempo. *International Congress on Spoken Language Processing, 1,* 667-670.

Folkins, J.W., Miller, C.J., & Minifie, F.D. (1975). Rhythm and syllable timing in phrase level stress patterning. *Journal of Speech and Hearing Research, 18,* 739-753.

Galanis, D., Darsinos, V., & Kokkinakis, G. (1996a). Investigating emotional speech parameters fro speech synthesis. *Proceedings of the Third IEEE Conference on Electronics, Circuits, and Systems, 2,* 1227-1230.

Galanis, D., Darsinos, V., & Kokkinakis, G. (1996b). Modeling of intonation bearing emphasis for TTS-synthesis of Greek dialogues. *International Congress on Spoken Language Processing, 3,* 1357-1360.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica, 38,* 148-158.

Goldman-Eisler, F. (1968). Experiments in spontaneous speech. *Psycolinguistics.* Academic Press: New York.

Grosjean, F. (1979). A study of timing in a manual and a spoken language: ASL and English. *Journal of Psycholinguistic Research, 8,* 379-405.

Grosjean, F. & Lane, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance, 2* (4), 538-543.

Grover, C. & Terken, J. (1994). Rhythmic constraints in durational control. *International Congress on Spoken Language Processing, 1* (4), 363-366.

Han, M. S. (1966). Acoustic-phonetic study on speech tempo. Papers delivered at the Second World Congress of Phoneticians, XII, 26-31.

Hayamizu, S. & Tanaka, K. (1994). Statistical modeling and recognition of rhythm in speech. *International Congress on Spoken Language Processing, 1* (4), 199-202.

Hoequist, C.E. (1983). Syllable duration in stress, syllable and more-timed languages. *Phonetica, 40,* 203-237.

Hoequist, C.E. (1984). Parameters of speech rate perception: Further results. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel,* 149-191.

Hoequist, C.E. & Kohler, K.J. (1986). Summary of speech rate perception research at Kiel. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel, 22,* 7-28.

Horne, M. (1988). Towards a quantified, focus-based model for synthesizing English speech. *Lingua, 75,* 25-54.

Kaika, N., Takeda, K. & Sagisaka, Y. (1990). Statistical analysis for segmental duration rules in Japanese speech synthesis. *Proceedings of the International Congress on Spoken Language Processing.* Kobe, Japan.

Kelly, J. & Steer, M. (1949). Revised concept of rate. *Journal of Speech and Hearing Disorders,* 222-227.

Kessinger, R.H. & Blumstein, S.E. (1998). Effects of speaking rate on voice-onset time and vowel production: some implications for perception studies. *Journal of Phonetics, 26,* 117-128.

Klatt, D.H. (1979). Synthesis by rule of segmental durations in English sentences. B. Lindbolm & S. Öhman (Eds.) *Frontiers of Speech Communication Research.* (pp.287-299). Academic Press: New York.

Kohler, K.J. (1986). Parameters of speech rate perception in German words and sentences: duration, F_0 movement, and F_0 level. *Language and Speech, 29,* 115-139.

Kohler, K.J. (1991). Isochrony, units of rhythmic organization and speech rate. *Proceedings of the XIIth International Congress of Phonetic Sciences, 1* (5), 257-261.

Koiso, H., Shimojima, A. & Katagiri, Y. (1998). Collaborative signalling of informational structures by dynamic speech rate. *Language and Speech, 41* (3-4) 323-350.

Koopmans-Van Beinum, F.J. & Donzell, M.E. (1996). Relationship between discourse structure and dynamic speech rate. *International Congress on Spoken Language Processing* [CD-ROM].

Kumar, S.R.R. & Yegnanarayana, B. (1989). Significance of durational knowledge for speech synthesis system in an Indian language. *TENCON'89. Fourth IEEE Region 10 International Conference,* 486-489.

Lane, H. & Grosjean, F. (1973). Perception of reading rate by listeners and speakers. *Journal of Experimental Psychology, 97,* 141-147.

Laver, J. (1994). *Principles of phonetics.* Cambridge: Cambridge University Press.

Levelt, W.J.M. (1989). *Speaking: from Intention to Articulation.* Cambridge MA: Massachusetts Institute of Technology Press.

Marjomaa, I. (1983). Speech rate and quantity of vowels in English and Finnish. *Proceedings from the Second Nordic Conference for English Studies,* 39-47.

McAngus-Todd, N.P. & Brown, G.J. (1994). A computational model of prosody perception. *International Congress on Spoken Language Processing, 1* (4), 127-130.

Miller, J. (1981). Some effects of speaking rate in phonetic perception. *Phonetica, 38,* 159-181.

Miller, J.L. & Grosjean, F. (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance, 7* (1), 208-215.

Miller, J., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41,* 215-225.

Min, C., Shinan, L., Hongyan, S., Lin, H., & Dinghua, G. (1996). The control of juncture and prosody in Chinese TTS system. *3rd International Conference on Signal Processing, 1,* 725-728.

Mobius, B. & Von Santen, J. (1996). Modeling segmental duration in German text-to-speech synthesis. *International Congress on Spoken Language Processing, 4,* 2395-2398.

Nakatani, L.H., O'Connor, K.D., & Aston, C. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38,* 84-106.

Nooteboom, S.G. (1991). Some observations on the temporal organisation and rhythm of speech. *Proceedings of the XIIth International Congress of Phonetic Sciences, 1* (5), 228-237.

Os, E. den (1985). Perception and speech rate of Dutch and Italian utterances.

Phonetica, 42, 124-134.

Osser, H., & Peng, F. (1964). A cross-linguistic study of speech rate. *Language and Speech, 7,* 120-125.

Pan, N., Jen, W., Yu, S., Yu, M., Huang, S., & Wu, M. (2000). Prosody model in a Mandarin text-to-speech system based on a hierarchical approach. *IEEE International Conference on Multimedia and Expo, 1,* 448-451.

Peterson, G.E. & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America, 32* (6), 693-703.

Pfitzinger, H., Burger, S., & Heid, S. (1996). Syllable detection in read and spontaneous speech. *International Congress on Spoken Language Processing* [CD-ROM].

Pitrelli, J. (1990). *Hierarchical modeling of phoneme duration: application to speech recognition.* Doctoral Dissertation, MIT, Cambridge Mass, USA.

Port, R.F. (1977). *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words.* Doctoral dissertation, Indiana University.

Port, R. F. (1981). Linguistic timing factors in combination. *The Journal of the Acoustical Society of America, 69* (1), 262-274.

Rambally, G.K. & Stelly, J.R., II. (1991). A text-to-speech environment: applications in education. *IEEE Proceedings of Southeastern '91, 2,* 1158-1161.

Records, M. (1982). *Durational Differences with Variations in Natural Speech Rate in Children.* Unpublished doctoral dissertation, University of Minnesota.

Rietveld, A.C.M., & Gussenhoven, C. (1987). Perceived speech rate and intonation. *Journal of Phonetics, 15,* 273-285.

Sakurai, N., Mochida, T., Kobayashi, T., & Shirai, K. (1994). Generation of prosody in speech synthesis using large speech data-base. *International Congress on Spoken Language Processing, 1* (4), 747-750.

Scherer, K.R. (1996). Adding the affective dimension: A new look in speech analysis and synthesis. *International Congress on Spoken Language Processing* [CD-ROM].

Takeda, K., Sagisaka, Y., & Kuwabara, H. (1989). On sentence-level factors governing segmental duration in Japanese. *Journal of Acoustical Society of America, 86* (6)

Terken, J.M.B. (1993). Human synthetic intonation: A case study. In J. van Hueven-Vincent & C.W. Louis (Eds.), *Analysis and speech: Strategic research towards high-quality speech generation.* (pp. 241-259). Berlin, Germany: Mouton de Gruyter.

Umeda, N. (1975). Vowel duration in American English. Journal of the Acoustical Society of America, 58, 434-445.

Umeda, N. & Wedmore, T. (1994). A rhythm theory for spontaneous speech: The role of vowel amplitude in the rhythmic hierarchy. *International Congress on Spoken Language Processing, 1* (4), 1095-1098.

Vaane, E. (1982). Subjective estimation of speech rate. *Phonetica, 39,* 136-149.

Wang, R., Liu, Q., & Tang, D. (1996). A new Chinese text-to-speech system with high naturalness. *International Congress on Spoken Language Processing, 3,* 1441-1444.

Wayland, S.C., Miller, J.L., & Volaitis, L.E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America, 95* (5, Pt. 1), 2694-2701.

Weismer, G. & Ingrisano, D. (1979). Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate. *Journal of Speech and Hearing Research, 22,* 516-533.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P.J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91* (3), 1707-1717.

Whiteside, S.P. (1996). Temporal-based acoustic-phonetic patterns in read speech. *Journal of the International Phonetic Association, 26* (1), 23-40.

APPENDIX A

Scripts

Pilot Study

1. Officious, *Acacia* cut each of *us* a piece of *okay* coffee cake.
2. Up a *high* peak, *I* toss a *cat* off a face.
3. Peckish, I *sauté* half a *cup* of *icy* fat opah.[13]
4. Atop a *tacky* shop, apish *Akako* took a *cuff* off a coat.
5. Sheepish, I *seek* a *cache* of *'ice'* atop a quay.

Note. The italicized words are those that were individually emphasized for the readings for the focus experiment.

13. Opah is otherwise known as moonfish.

Main Study

1. Sickish, I *catch* a *tick* atop a *poppy* cake.

2. Oafish, I *toss* a *pack* atop a *tufa*[14] peak.

3. Sheepish, I *teach* a *cat* atop a *cushy* seat.

4. Uppish, I *tack* a *sash* atop a *peachy* coat.

5. Cautious, I *seek* a *sheet* atop a *tippy* ketch.

14. Tufa is a porous rock formed around mineral springs.

APPENDIX B

Pilot Study: Descriptive Statistics

Table B.1.
Durations of Syllable Constituents by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Constituent	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Onsets	72	94.1	26.3	82	111.1	23.6	86	153.2	32.0
Peaks	147	77.9	31.9	166	81.1	32.9	175	165.5	81.0
Codas	69	77.8	24.5	78	97.0	33.0	82	169.6	78.4

Table B.2.
Durations of Unstressed Syllable Constituents by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Constituent	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Onsets	72	94.1	26.3	82	111.1	23.6	86	153.2	32.0
Peaks	73	59.7	23.2	82	63.0	25.4	87	143.0	76.0
Codas	24	81.2	29.9	27	112.2	45.1	28	152.2	84.9

Table B.3.
Durations of Stressed Syllable Constituents by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Constituent	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Onsets	52	97.9	21.9	59	144.6	23.3	63	162.2	28.0
Peaks	74	95.9	29.0	84	98.7	29.7	88	187.6	80.0
Codas	45	75.9	21.1	51	89.0	21.3	54	178.6	74.0

Table B.4.
Durations of Onsets in CV Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	33	87.8	31.7	36	110.5	25.2	37	140.0	27.2
Unstressed	16	82.9	35.9	17	105.7	22.1	17	133.4	31.6
Stressed	17	92.4	27.6	19	114.8	27.5	20	145.7	22.0

Table B.5.
Durations of Onsets in CVC Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	39	99.4	19.5	46	111.6	22.6	49	163.1	32.0
Unstressed	4	89.0	29.1	6	92.6	23.2	6	114.0	14.7
Stressed	35	100.6	18.4	40	114.4	21.4	43	170.0	27.3

Table B.6.
Durations of Peaks in V Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	20	84.2	34.0	23	102.3	22.7	23	128.3	29.1
Unstressed	35	51.2	23.5	38	53.9	22.2	42	134.2	46.3
Stressed	13	115.5	31.6	14	116.5	27.6	14	207.7	64.7

Table B.7.
Durations of Peaks in VC Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	30	82.6	38.0	32	87.7	41.1	33	165.1	79.3
Unstressed	20	61.4	17.3	21	70.2	25.9	22	146.2	76.4
Stressed	10	124.8	32.5	11	121.0	45.1	11	202.7	74.5

Table B.8.
Durations of Peaks in CV Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	30	86.1	21.2	36	86.0	25.9	37	188.0	123.7
Unstressed	14	73.8	20.6	17	72.2	27.8	17	176.6	125.3
Stressed	16	96.8	15.4	19	98.4	16.5	20	197.6	124.7

Table B.9.
Durations of Peaks in CVC Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	39	79.5	21.1	46	84.3	24.1	49	163.5	57.9
Unstressed	4	75.2	30.3	6	69.7	22.3	6	98.8	17.3
Stressed	35	80.0	20.3	40	86.4	23.8	43	172.5	55.8

Table B.10.
Durations of Codas in VC Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	30	80.3	22.7	32	106.6	36.0	33	141.5	73.6
Unstressed	20	75.1	25.7	21	109.3	43.7	22	139.3	88.1
Stressed	10	90.6	9.2	11	101.6	12.6	11	145.7	31.8

Table B.11.
Durations of Codas in CVC Syllables by Rate (ms.)

	Speech Rate								
	Fast			Normal			Slow		
Stress	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Combined	39	75.8	25.9	46	90.4	29.8	49	188.5	76.5
Unstressed	4	111.3	35.2	6	122.6	52.9	6	199.5	54.4
Stressed	35	71.8	21.8	40	85.5	21.9	43	187.0	79.4

APPENDIX C

Tempo Pattern Experiment: Descriptive Statistics

Sentence Analysis

Table C.1.
Durations of Focus Sentences by Sentence (ms.)

	$n = 60$		
		ln	
Sentence	*Mdn*	*M*	*SD*
1	2907.33	8.02	.17
2	2868.83	8.01	.18
3	3046.19	8.06	.19
4	2993.22	8.02	.18
5	2993.06	8.08	.22

Table C.2.
Durations of Focus Sentences by Focus (ms.)

	$n = 75$		
		ln	
Focus	*Mdn*	*M*	*SD*
N	2760.39	7.93	.17
1	2984.03	8.03	.15

Table C.2.
Durations of Focus Sentences by Focus (ms.)

	$n = 75$		
		ln	
Focus	*Mdn*	*M*	*SD*
2	3093.58	8.07	.18
3	3149.21	8.10	.21

Experiment I

Table C.3.
Normalized ASDs for the Pre-focal Inclusive Regions

	$n = 75$						
		ln				ln	
Focus	*Mdn*	*M*	*SD*	Control	*Mdn*	*M*	*SD*
1	.108	-2.223	.07	1	.096	-2.356	.08
2	.096	-2.344	.04	2	.089	-2.430	.05
3	.074	-2.608	.04	3	.074	-2.603	.03

Table C.4.
Normalized ASDs for the Post-focal Regions

	$n = 75$						
		ln				ln	
Focus	*Mdn*	*M*	*SD*	Control	*Mdn*	*M*	*SD*
1	.071	-2.647	.05	1	.077	-2.562	.04
2	.071	-2.652	.06	2	.077	-2.550	.05
3	.187	-1.683	.17	3	.183	-1.695	.13

Experiment II

Table C.5.
Normalized ASDs for the Pre-focal Exclusive Regions

	n = 75						
		ln				ln	
Focus	*Mdn*	*M*	*SD*	Control	*Mdn*	*M*	*SD*
1	.080	-2.525	.10	1	.088	-2.440	.11
2	.074	-2.603	.08	2	.080	-2.531	.08
3	.072	-2.659	.12	3	.076	-2.580	.04

Table C.6.
Normalized ASDs for the Focus Word

	n = 75						
		ln				ln	
Focus	*Mdn*	*M*	*SD*	Control	*Mdn*	*M*	*SD*
1	.190	-1.666	.15	1	.115	-2.148	.12
2	.203	-1.593	.12	2	.125	-2.062	.21
3	.081	-2.449	.23	3	.066	-2.720	.11

APPENDIX D

Site Experiment: Descriptive Statistics

Sentence Analysis

Table D.1.
Durations of Site Sentences by Sentence (ms.)

	$n = 45$		
		ln	
Sentence	*Mdn*	*M*	*SD*
1	3057.78	8.06	.31
2	2774.78	8.02	.32
3	2904.09	8.06	.31
4	2784.97	8.05	.32
5	2929.10	8.09	.36

Table D.2.
Durations of Site Sentences by Rate (ms.)

	$n = 75$		
		ln	
Rate	*Mdn*	*M*	*SD*
Fast	2314.53	7.75	.07

Table D.2.
Durations of Site Sentences by Rate (ms.)

	n = 75		
		ln	
Rate	*Mdn*	*M*	*SD*
Control	3037.80	8.01	.16
Slow	4898.78	8.40	.25

Vowel

Table D.3.
Vowel Durations by Rate (ms.)

	n = 900		
		ln	
Rate	*Mdn*	*M*	*SD*
Fast	66.46	4.18	.46
Control	83.42	4.44	.48
Slow	119.58	4.76	.43

Table D.4.
Vowel Durations by Rate, Syllable Type, Stress, and Sentence Boundary Adjacency (ms.)

	Speech Rate		
	Fast	Control	Slow
Syllable	*Mdn*	*Mdn*	*Mdn*
CVC			
Stressed			
SBA[a]	106.48	119.56	126.95
Non-SBA	78.65	108.00	143.83
VC			
Stressed			
SBA	108.06	120.05	153.50
Unstressed			
Non-SBA	68.76	77.46	102.50
V			
Stressed			
Non-SBA	80.81	113.01	156.68
Unstressed			
Non-SBA	42.82	56.63	99.47[b]

a. SB = Sentence Boundary Adjacent.

b. All other entries in this table are for vowels that are not in prepausal position. The prepausal counterpart for this entry had a mean of 155.92 ms. Being the only prepausal vowels, this group was edited from the data.

Syllable Constituency

Table D.5.
Onset Durations by Rate, Syllable Type, Manner, Stress, Prepausal Position, and Sentence Boundary Adjacency (ms.)

	Speech Rate		
	Fast	Control	Slow
Syllable	*Mdn*	*Mdn*	*Mdn*
CVC			
Stop			
Stressed			
Non-PP[a]			
SBA	97.75	127.43	238.33
Non-SBA	91.91	122.18	148.15
Fricative			
Stressed			
Non-PP			
SBA	119.56	143.42	165.38
Non-SBA	110.25	142.52	187.28

a. PP = Prepausal Position

Table D.6.
Coda Durations in CVC Syllables by Rate, Manner, Stress, Prepausal Position, and Sentence Boundary Adjacency (ms.)

	Speech Rate		
	Fast	Control	Slow
Syllable	*Mdn*	*Mdn*	*Mdn*
Stop			
Stressed			
PP			
SBA	188.77	202.26	213.24
Non-SBA	76.07	103.90	146.39

Table D.6.
Coda Durations in CVC Syllables by Rate, Manner, Stress, Prepausal Position, and Sentence Boundary Adjacency (ms.)

	Speech Rate		
	Fast	Control	Slow
Syllable	*Mdn*	*Mdn*	*Mdn*
Non-PP			
SBA	73.24	83.56	97.81
Non-SBA	70.09	83.29	87.36
Fricative			
Stressed			
PP			
Non-SBA		87.33	168.03
Non-PP			
SBA	90.57	98.85	121.07
Non-SBA	86.60	98.74	117.82
Affricate			
Stressed			
PP			
SBA	300.29	328.86	365.13
Non-SBA		111.65	190.55
Non-PP			
Non-SBA	93.27	125.18	126.81

Table D.7.
Coda Durations in VC Syllables by Rate, Manner, Stress, Prepausal Position, and Sentence Boundary Adjacency (ms.)

	Speech Rate		
	Fast	Control	Slow
Syllable	*Mdn*	*Mdn*	*Mdn*
Stop			
Stressed			
Non-PP			
SBA	71.83	81.86	93.03
Fricative			
Unstressed			
PP			
Non-SBA	176.76	179.00	256.87
Non-PP			
Non-SBA	131.14	202.98	
Stressed			
Non-PP			
SBA	89.65	96.55	114.74

Glossary

This glossary provides brief definitions of the most important recurrent terms used in the text.

active voice In a sentence that is in active voice, the subject of the sentence performs the action of the verb.

affricate An affricate is a complex segment comprised of a stop closely followed by a fricative. The initial and final sounds of *church* and *judge* are affricates.

coda A member of the syllable. The coda consists of all segments that follow the peak of the syllable.

CV A type of syllable that is comprised of an initial consonant followed by a vowel.

CVC A type of syllable that is comprised of a medial vowel that is preceded and followed by consonants.

declarative A type of sentence that is a statement and not a question.

dialect A dialect is a variant of a language that is spoken by a group of people who share the same geographical habitat or social setting or exist in the same historical period.

foot A metrical unit consisting of a group of syllables, one of which is usually stressed.

fricative A consonant that is produced by constricting the flow of air in the vocal tract so as to create turbulence and air friction noise. The initial sounds of *fee* and *sea* are fricatives.

idiolect An idiolect is a variety of language that is unique to an individual speaker.

intonation Using pitch differences, intonation serves to organize an utterance by grouping strings of syllables into phrases and phrases into sentences.

nucleus The nucleus of a syllable is the peak of a syllable.

onset A member of the syllable. The coda consists of all segments that precede the peak of the syllable.

peak The most prominent part of the syllable. The peak is usually a resonant. A resonant is a sound made by shaping the vocal tract using an unobstructed flow of air and includes vowels and some consonants.

stop A consonant that is produced by an interruption in the outflow of air from the lungs (or in rare cases, the inflow of air). The initial sounds of *key*, *tea*, and *pea* are stops.

stressed A feature of the syllable. In comparing syllables, one syllable that is more prominent than the others is described as being stressed.

syllable An easily recognized unit of speech that is uttered without interruption. The syllable centers on one prominent segment, usually a vowel and can be preceded or followed by other less prominent segments.

V A type of syllable that is comprised of a vowel.

VC A type of syllable that is comprised of a vowel followed by a consonant.

voiced A feature of a sound that is produced with accompanying vocal cord vibration.