Tempo Modulations in English:  
SELECTED PILOT STUDY RESULTS  
Sandra P. Kirkham 
Department of Linguistics 
University of Victoria, Victoria, B.C. Canada  
skirkham@uvic.ca

ABSTRACT
This paper provides a discussion of selected results of a pilot study founding a tempo project. This project potentially determines the phonetic locations for altering the speech rate for English synthetic speech in order to replicate a proposed natural pattern of tempo modulations based upon phrasal foci. A male educated speaker read Canadian English sentences at 3 speech rates and with varied focal points plus 1 neutral control, which were recorded and then digitized using Multispeech, Model 3700, version 2.2 by Kay Elemetrics Corp. By comparing the durations of the averaged syllables of the focused sentences to those of the control, the proposed tempo pattern is confirmed. The results of the pilot study are promising and warrant further examination in order to provide a model of speech rate that can be incorporated into synthetic speech creating a more natural product.

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. 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 [7] (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. This pattern was discovered by Fant et al. in Swedish (p.256). In English, examples of such speech modulations are obvious in speeches and theatre performances. 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 the speech tempo project is to model observed tempo modulations in Canadian English read speech for incorporation into a synthetic speech system. This paper discusses some of the preliminary findings of the pilot study for this project carried out in collaboration with STR and funded by the B.C. Science Council. One of the objectives of this research is to verify and validate a pattern of deceleration and acceleration in tempo. Focus is the centre of interest in the phrase affecting duration in the sentence. It has been established that the correlates of focus are voice fundamental frequency and duration (Brown and McGlone [1], Folkins et al. [8], Weismer and Ingrisano [10], and Eady and Cooper [5]). A focused word significantly increases in duration compared to an unfocused word (Cooper et al. [3] and Eady and Cooper [5]). 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 [5] found that focus does not significantly affect other words in long sentences with 11 or 12 syllables (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. [8], Weismer and Ingrisano [10], and Eady et al. [6]). 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 [5], 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 main purpose of the pilot study was to determine whether 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). Not discussed in this paper are companion experiments that were conducted in order to determine the potential sites for altering tempo.

2. METHOD

2.1. Materials
As a source of data for this study, five sentences were designed to provide focus locations as well as phonetic environments for tense and lax vowels, codas, and VOTs for the companion experiments. 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.
2.2. 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.

2.3. 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. 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.

The following method was used to prepare the data to test the tempo pattern hypothesis, which states that 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 [2]) and phrase-final lengthening (see Crystal and House [4]). 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 average syllable durations (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, counteracting the effects of varying sentence lengths. The ASDs for pre-focal and post-focal sections of the sentence were then compared.

2.4. Statistics

The purpose of the pilot study was to look for trends in the data that support the hypotheses of the experiments. 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.

3. RESULTS AND DISCUSSION

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 1. Because of occurrences of reading errors by the speaker, it was not possible to make accurate boundary determinations for all segments. 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>
<thead>
<tr>
<th>Position</th>
<th>Control</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>62.6</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>66.7</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>67.2</td>
</tr>
</tbody>
</table>

Table 1: Pre-focal Region ASDs

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 1). This large increase in the first two regions 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.
As previously mentioned, Cooper et al. [3] 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.

In Figure 2, a comparison based on the results in Table 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 increase in the focal word. The focal word increases to a much greater degree than the pre-focal region in keeping with the findings of Cooper et al. [3] and Eady and Cooper [5]. 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.

### Table 2: Focus Word ASDs

<table>
<thead>
<tr>
<th>Position</th>
<th>Control</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>76.2</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>67.3</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>88.3</td>
</tr>
</tbody>
</table>

With the exception of the three percent 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 the non-focused 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 1 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. While these findings are not conclusive because of the small sample size, they do contradict Eady and Cooper’s [5] observations that non-focused words remain unaffected in duration. 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.

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 determined 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.

A trend of deceleration is evident when one compares the means of these groups, as seen in Figure 3. For the Focal 1 post-focal regions, the ASDs in the focused version show a five percent decrease from the control based on the results in Table 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). There appears to be a correlation between a focus word located nearer the beginning of a phrase and the compression of the remaining part of the phrase as noted by Eady and Cooper [5]. 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 confirming the observations of Folkins et al. [8], Weismer and Ingrisano [10], and Eady et al. [6] although these studies used short sentences.

<table>
<thead>
<tr>
<th>Position</th>
<th>Control</th>
<th></th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>64.0</td>
<td>10.2</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>66.4</td>
<td>20.5</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>51.8</td>
<td>11.9</td>
</tr>
</tbody>
</table>

Table 4: Post-focal Region ASDs

To summarize, there is a clear trend in the data of a tempo pattern in English when considering the differences of means in the data, lending support to results discovered previously for Swedish by Fant et al. [7]. 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.

### 4. CONCLUSIONS

In conclusion, these findings suggest that tempo changes within a phrase and the focus for a change in tempo is the focus. The results of the experiment suggest that there is a deceleration-acceleration pattern evident in focused sentences. 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.

These findings warrant a subsequent study in order to confirm the tempo pattern. The subsequent study incorporates improvements on some aspects of the current study that may have affected the results adversely. This 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. 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.

This emerging pattern of deceleration and acceleration within a focused phrase can potentially be replicated. This is possible given the confirmation of the results of this experiment and those of the companion experiments determining sites for potential rate changes by a subsequent study.

### 5. REFERENCES


