An Acoustic Study of Speech Rhythm in Taiwan English

Hua-Li Jian

Department of Foreign Languages and Literature
National Cheng Kung University, Tainan, Taiwan
hljian@mail.ncku.edu.tw

Abstract
American English and Taiwan English have been found to exhibit different rhythmic patterns. American English is often described as a stress-timed language, while it has been suggested that Taiwan English is a syllable-timed language. This paper addresses the differences between these two varieties of English through an acoustic study. In particular the F1/F2 formant space is investigated. The results show that reduced vowels in American English are more concentrated in the F1/F2 formant space than reduced vowels in Taiwan English, which are more dispersed.

1. Introduction
A vast body of literature exist addressing the rhythm and timing in British English (BE) and American English (AE), which are both known to be stressed-timed. However, very little has been written about the rhythm and timing of the Taiwan variety of English (TE), although some studies have addressed other South Asian varieties of English – especially Singaporean English (SE) and Malaysian English (ME) since English is one of the official languages in Singapore (see for example the studies by Low, Grabe and Nolan [11], Brown [6], Platt and Weber [15], Tay [21] and Torque [22]). All these studies describe Singaporean English as being syllable-timed and also staccato sounding. In this study we investigate Taiwan English from the same viewpoint. English is not an official language in Taiwan though everyone is now taught English from an early age. Taiwan English is a distinct variety of English, although American pronunciation is generally taught throughout the island. Factors influencing the Taiwan variety of English include Japanese-style English brought over during the second world war occupation, mainland Chinese English mainly brought over just before 1945, British textbooks used before the introduction of the Kenyon and Knott general American pronunciation system in 1969, persistent defective pronunciation of English in classrooms, and effects from local languages – especially Mandarin and Southern Min (see Chung [7]). Our hypothesis is that the Taiwan English variety shares some characteristics with the Singaporean English due to similar influences such as the Chinese language. As indicated by Jian [9], Taiwan English tends towards syllable-timed. This study aims to further examine the acoustic characteristics of reduced vowels in Taiwan English. This is achieved by the means of an acoustical analysis of speech samples.

2. Background
It has long been assumed that there is near equal interval between stresses in stress-timed languages and that there are near equal intervals between successive syllables in syllable-timed languages (Pike [14]). Most of the early studies were not based on acoustic studies, and Couper-Kuhlen [8] claims that this isochrony is a perceptual phenomenon. Studies also show that isochrony cannot be related to constant inter-stress intervals in stress-timed languages (Strangert [20]). Further, studies show that successive syllable durations vary greatly in syllable-timed languages (Pointon [16]). Other studies show that stress-timed languages are no more regular in inter-stress intervals than syllable-timed languages (Roach, [18]).

1. Introduction
A vast body of literature exist addressing the rhythm and timing in British English (BE) and American English (AE), which are both known to be stressed-timed. However, very little has been written about the rhythm and timing of the Taiwan variety of English (TE), although some studies have addressed other South Asian varieties of English – especially Singaporean English (SE) and Malaysian English (ME) since English is one of the official languages in Singapore (see for example the studies by Low, Grabe and Nolan [11], Brown [6], Platt and Weber [15], Tay [21] and Torque [22]). All these studies describe Singaporean English as being syllable-timed and also staccato sounding. In this study we investigate Taiwan English from the same viewpoint. English is not an official language in Taiwan though everyone is now taught English from an early age. Taiwan English is a distinct variety of English, although American pronunciation is generally taught throughout the island. Factors influencing the Taiwan variety of English include Japanese-style English brought over during the second world war occupation, mainland Chinese English mainly brought over just before 1945, British textbooks used before the introduction of the Kenyon and Knott general American pronunciation system in 1969, persistent defective pronunciation of English in classrooms, and effects from local languages – especially Mandarin and Southern Min (see Chung [7]). Our hypothesis is that the Taiwan English variety shares some characteristics with the Singaporean English due to similar influences such as the Chinese language. As indicated by Jian [9], Taiwan English tends towards syllable-timed. This study aims to further examine the acoustic characteristics of reduced vowels in Taiwan English. This is achieved by the means of an acoustical analysis of speech samples.

2. Background
It has long been assumed that there is near equal interval between stresses in stress-timed languages and that there are near equal intervals between successive syllables in syllable-timed languages (Pike [14]). Most of the early studies were not based on acoustic studies, and Couper-Kuhlen [8] claims that this isochrony is a perceptual phenomenon. Studies also show that isochrony cannot be related to constant inter-stress intervals in stress-timed languages (Strangert [20]). Further, studies show that successive syllable durations vary greatly in syllable-timed languages (Pointon [16]). Other studies show that stress-timed languages are no more regular in inter-stress intervals than syllable-timed languages (Roach, [18]).

Figure 1: Reduction from full vowels (ideal states) to reduced vowels (neutral state).

2.1. Vowel reduction
From an acoustical point of view a reduced vowel can be contrasted to a full vowel from the viewpoint that it has not reached a full or ideal state [2] (see Figure 1). A full vowel can be imagined to be in a unique ideal state, while a reduced vowel is tending towards a neutral or central state. This movement from the ideal full state towards the reduced central region also results in shorter vowel duration and lower overall amplitude [10]. This central point has been termed the neutral-vowel position [19], as it is produced when the vocal tract is in a neutral position. The formants are often more equally spaced than what can be observed for full vowels [12]].

Vowel reduction is often linked to speaking style and stress. Studies have shown that monosyllables pronounced in isolation more often occur further away from the center in the formant space (full) than continuous speech which occur closer to the center in reduced form [19]. Further, the degree of vowel reduction is also shown to be language dependent [10].

2.2. Hypothesis
The objective of this paper is to show that full and reduced vowels can be discriminated based on their acoustical properties and that there is a weaker reduction in Taiwan English vowels than can be found in American English.
In particular, the full and reduced vowels can be discriminated based on their F1 and F2 formant values. Figure 2 (left) illustrates the hypothesis, namely that when the F1 and F2 formants for the syllables are plotted in the F1/F2 plane, the reduced vowels of American English speakers will fall closer to the center of all the F1/F2 values, while the full vowel F1/F2 values for American English subjects are more dispersed relative to the center (greater spread in the F1/F2 plane).

Further, the second hypothesis is that there is an indistinguishable spread in F1/F2 points for the full and reduced vowels in Taiwan English (see Figure 2 (middle)). Finally, the third hypothesis is that when comparing the reduced vowels for American and Taiwan English, the vowels in American English will be more concentrated towards the center than those of Taiwan English, which are more dispersed throughout the F1/F2 plane (greater spread – see Figure 2 (right)).

3. Method

3.1. Subjects

Four Taiwanese English (TE) and four American English (AE) speakers took part in the recording task. All subjects live in Taiwan. The native Taiwanese TE speakers were students at the National Cheng Kung University in Tainan. The native American AE speakers were mostly teachers in the Tainan and Kaohsuing areas in the south of Taiwan. The TE speakers were mostly in their early twenties, while the age of the AE ranged from mid twenties to early forties. None of the subjects reported having any articulation disability.

3.2. Materials

It was decided to adopt material with a similar structure to that described in Low, Grabe and Noland [11]. The TE and AE subjects were asked to read 10 sentences consisting of five full and potentially reduced vowels and five sentences comprising only full vowels. See Table 1 as an example (taken from Low, Grabe and Noland [11]).

Table 1: Example of full and reduced vowels.

<table>
<thead>
<tr>
<th>Full vowel set</th>
<th>John came back through France last Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced vowel set</td>
<td>John was sick of Jack and Sandy</td>
</tr>
</tbody>
</table>

The sentences were shuffled into pseudorandom order and presented to the speakers without context and fillers.

3.3. Recording procedure

The subjects were recorded in an office in the National Cheng Kung University using a Shure KSM32 studio microphone and a portable mixer with a built-in microphone amplifier. The microphone was placed approximately 20 cm in front of the speakers. The subjects were given time to read the material prior to the recording and the subjects were manually informed to repeat utterances when mistakes were made. The material was digitally recorded onto a Minidisk using a stationary Sony deck. The digitally recorded material was subsequently transferred to audio files stored on a personal computer.

3.4. Post-processing and analysis

The speech material comprising 100 utterances of which 50 utterances were produced by the TE speakers, and the remaining 50 were produced by the AE speakers. The digital recording was partitioned so that each audio-file consisted of one sentence. The open source analysis package PRAAT by Boersma and Weenink [4], [5] was used to segment and label the utterances. The spectrogram feature and formant analysis tool in PRAAT was used to select the vowel boundary, which is not a trivial task. See Peterson and Lehiste [13] for a general discussion on segmentation of vowels by the means of spectrograms.

Formant traces were generated by PRAAT and in this study the first (F1) and second (F2) formants are of particular interest. For each syllable represented by a time-series of F1 and F2 values, the means representing the F1 and F2 formant values for the syllable were computed. To be specific,

\[ \bar{F}_i = \frac{1}{e_j - s_j} \sum_{k=s_j}^{e_j} F_i(k) \]  \hspace{1cm} (1)

where \( F_i/k \) is the \( k \)'th time (time) of the \( i \)'th formant and \( s_j \) is the start time of syllable \( j \) and \( e_j \) is the end time of syllable \( j \). This approach is more accurate than previous approximation methods where only the midpoint of the syllable is taken for steady state formant traces and two equally spaced points were taken from syllables without steady state (see [1], [11]). Midpoint F1 and F2 values were also recorded and included for completeness. These F1, F2 pairs make up a vector \( q_i \) in the formant space given by:

\[ q_i = [v_i, u_i] \]  \hspace{1cm} (2)

Where \( v_i \) represents the F1 value for syllable \( i \), and \( u_i \) represents the F2 value for syllable \( i \). The centre of the formant points in the formant space is given by:
where $N$ is the number of syllables. The distance between a point $q_i$ in the formant space and the centre is given by

$$d_i = |q_i - \overline{q}|$$

And for the purpose of comparing the different syllable groups a dispersion measure $D$ is computed as follows (see [1]):

$$D = \frac{1}{n} \sum_{i=1}^{N} (q_i - \overline{q})^2$$

This measure is also known as the mean sum of squares.

PRAAT stores segmentation and formant information in separate text files. A script was therefore written to first extract the start and end positions from the segmentation file, compute the midpoints and then extract the closest F1 and F2 values from the formant files and also the mean formant over the syllable.

4. Results and discussion

Figures 3, 4, 5 and 6 and Table 2 summarise the results of the experiments. Figures 3 and 4 shows the data using the formant means of the vowel, where the means of the first and second formants are plotted against each other. The horizontal axis represents the first formant and the vertical axis represents the second formant. The scatter plot in Figure 3 shows that the full and reduced vowels for Taiwanese subjects are hard to distinguish, while the full and reduced vowels for the American English subjects follow a recognizable pattern in Figure 4. As illustrated in Figure 4, the formant pairs for the full vowels of the American subjects are spread across a larger area in the graph than the formant pairs for the reduced vowels.

Similar observations can be observed in Figures 5 and 6, which are based on the midpoint of the formant traces, where the horizontal axis represents the first formants and the vertical axis represents the second formants. Again, the scatter plot in Figure 5 shows that it is hard to distinguish between full and reduced vowels for Taiwanese subjects, while it is a distinctive pattern for the American English, where the reduced vowels are concentrated in the center and the full vowels span a larger area as in Figure 6.

When comparing the plots obtained using the mean formants over the syllables and the midpoint formant of the syllables there are no apparent patterns. However, the plots are far from equal.

The diagrams in 3, 4, 5 and 6 only serve as an impressionistic tool for assessing the formant patterns. We therefore also computed dispersion indices for the four groups, i.e. Taiwanese speakers’ full and reduced vowels and American English speakers’ full and reduced vowels. The results are summarized in Table 2. In addition, for each language group the difference in percentage is computed to better demonstrate the difference in using the F1 and F2 mean formant values over each syllable compared to the F1 and F2 formant midpoints over the syllables. Comparing the full and
reduced vowels for the Taiwanese speakers one can see that the difference based on the mean is 9.1 % and 13.7 % for the midpoint-based mean. Then, the difference between the dispersion indices for full and reduced vowels for American English subjects are 36.1 % and 37.7 % using the mean-based values and midpoint-based measurements respectively. For both mean and midpoint-based measurements of American syllables the differences between the full and reduced vowels are distinct, i.e. full vowels are more dispersed than reduced vowels, which are more concentrated around the centre. Further, the dispersion indices for Taiwanese full and reduced vowels are similar to the dispersion index for American English full vowels, while all three are different from the American English reduced vowels. We make two conclusions based on these observations. First, reduced vowels in American English are more concentrated in the formant space than reduced vowels in Taiwan English. Second, the mean-based measurements appear to be better at highlighting such differences than the midpoint-based technique previously described in the literature.

Table 2: Dispersion values for full and reduced vowels and percentage differences for Taiwanese English speakers and American English speakers.

<table>
<thead>
<tr>
<th>Type</th>
<th>Means based</th>
<th>Midpoint based</th>
</tr>
</thead>
<tbody>
<tr>
<td>TW full</td>
<td>118507</td>
<td>220364</td>
</tr>
<tr>
<td>TW reduced</td>
<td>129274</td>
<td>190095</td>
</tr>
<tr>
<td>% difference</td>
<td>-9.1 %</td>
<td>13.7 %</td>
</tr>
<tr>
<td>US full</td>
<td>141784</td>
<td>201180</td>
</tr>
<tr>
<td>US reduced</td>
<td>90567</td>
<td>125421</td>
</tr>
<tr>
<td>% difference</td>
<td>36.1 %</td>
<td>37.7 %</td>
</tr>
</tbody>
</table>

5. Conclusion

In this study the acoustic correlates of rhythm in Taiwan English and American English have been compared by investigating the F1/F2 formant space. The results show that reduced vowels in American English are more concentrated in the F1/F2 formant space than reduced vowels in Taiwan English, which are dispersed at a greater degree. American English is known to be stress-timed, while the acoustic measurements described in this paper again suggests that Taiwan English is not stress timed.

Acknowledgements

This work is supported by the National Science Council of Taiwan grant NSC91-2411-H006-020.

6. References