<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Sentence stress pattern in English produced by Cantonese-English bilingual speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other Contributor(s)</strong></td>
<td>University of Hong Kong</td>
</tr>
<tr>
<td><strong>Author(s)</strong></td>
<td>Tam, Sze-Nga; 譚詩雅</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2008</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/123879">http://hdl.handle.net/10722/123879</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Creative Commons: Attribution 3.0 Hong Kong License</td>
</tr>
</tbody>
</table>
Sentence Stress Pattern in English Produced By Cantonese-English Bilingual Speakers

Tam Sze Nga, Emily

2004196170

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2008
Abstract

This study investigated the stress pattern in English sentences produced by Cantonese-English (C-E) bilingual speakers, and the relationship between their English (L2) proficiency and stress perception. Suprasegmental parameters reflecting stress production including fundamental frequency (F0), vowel duration and intensity were measured from the English sentences produced by forty C-E speakers. Results were compared with those obtained from native American English (A-E) monolingual speakers. The speech samples were also judged by eight raters who were native speakers of American English regarding the placement of stress, degree of stress and naturalness of stress. Results showed that the C-E speakers were able to use F0, vowel duration and intensity to differentiate different stress patterns. A comparison between C-E and A-E speakers revealed that female C-E speakers exhibited consistently higher F0 in stressed words than A-E monolingual speakers. Use of duration and intensity to signal stress was found to be comparable among all speakers. F0 and intensity were found to correlate closely with perceptual judgment and the degree of stress with the naturalness of stress.
Introduction

During verbal communication, suprasegmental features including stress, intonation, duration and juncture are important elements in conveying metalinguistic information such as emotion and identity of the speaker (Raphael, Borden & Harris, 2007). This packet of information appears to be superimposed onto the strings of speech sounds (segments) and speakers are usually unaware of it. However, the way of using suprasegmental features strongly affects the communication effectiveness in daily life. Previous studies have reported deviations in the stress pattern in English produced by individuals who speak English as their second language (L2), as compared with native speakers of English (L1). Such phonetic deviations are generally regarded as accent; the more it is deviated, the greater accent it is perceived. As a result of such phonetic deviations, the effectiveness of communication of L2 speakers can be affected. In fact, stress learning has been considered as one of the major difficulties in L2 acquisition (Cheng, 1968). The present study investigated the difference in the stress pattern associated with English sentences produced by Cantonese-English (C-E) bilingual speakers and native American English (A-E) monolingual speakers expressed in terms of fundamental frequency (F0), vowel duration, and vowel intensity. Moreover, the relationship between the deviation of stress pattern and perception of sentence stress pattern were investigated.

In English, two forms of stress are identified: lexical stress and sentence stress (Raphael et al., 2007). Lexical stress is referred to the pattern of stress within the boundary of a word. For example, the placement of lexical stress of the word “object” determines if it is a noun or a verb (e.g., OBJect vs. obJECT). Sentence stress, on the other hand, acts as a “pointer” which indicates the most important part of information within an utterance. For example, the more important word of the sentence “This is my green book.” changes from “this” to “my” from the sentence “THIS is my green book” to “This is MY green book”.


Previous studies generally found that stressed syllables within a sentence were acoustically correlated to higher fundamental frequency (F0), longer vowel duration and greater vowel intensity (Chen, Robb, Gilbert & Lerman, 2001; Cooper, Eady, & Mueller, 1985; Fry, 1955; Raphael et al., 2007). In a study of 80 adult Mandarin-English bilingual speakers and native speakers of American English reported by Chen and colleagues, both speaker groups exhibited higher average F0, longer vowel durations and greater vowel intensity in stressed words than unstressed words in English (Chen et al., 2001). However, the stress pattern of English sentences produced by speakers of English as L2 has been found to deviate from that by native English speakers (Chen et al., 2001; Chun, 1982, cited in Chen et al., 2001). It was suggested that the stress deviation may be due to the interference of L1 (Mandarin) on the production of L2 (English). This observation was supported by studies of L2 speakers of tonal languages such as Mandarin (Cheng, 1968; Chun, 1982), in which difficulty in acquiring the native stress pattern of English has been reported (Chun, 1982).

Chen et al. (2001) compared the acoustic characteristics associated with English sentence stress produced by Mandarin-English bilingual and English monolingual speakers between stressed and unstressed words. Both groups of speakers were able to perceptually differentiate the stressed and unstressed syllables. Yet, Mandarin female and male speakers produced the stressed English words with a significantly higher F0 and Mandarin male speakers stressed the words with shorter vowel duration when compared with English monolingual speakers (Chen et al., 2001). The elevated F0 exhibited by Mandarin speakers was suggested to be attributed by the interplay between sentence intonation and lexical tones in the accented English. Production of F0 in English sentences was influenced by the intrinsic lexical tones of Mandarin words (Chen et al., 2001). This theory of interplay between sentence intonation and lexical tones was also proposed by Shen (1990) that it may also be an account for the deviation found in tonal L2 language produced by native non-tonal speakers in
her study. She believed that such interplay is the cause of greater pitch fluctuations and increased overall pitch level in French, which is a non-tonal language, produced by native Mandarin speakers.

The shorter vowel duration produced by Mandarin speakers in the study by Chen et al. (2001) was believed to be related to the syllable-timed nature of Mandarin, in which Mandarin syllables occupied relatively constant duration across a sentence. However, English is a stress-timed language in which the unstressed syllables are produced at a much shorter duration when compared with stressed syllable (Chan & Li, 2000). Therefore, difficulty arises for the Mandarin speakers to adjust for syllable length produced in English (Chen et al., 2001).

Intensity measurement of English sentences produced by Mandarin speakers in Chen et al. (2001) was found to be the most comparable with those produced by native English speakers, although Mandarin speakers appeared to produce the stressed and unstressed words at a greater intensity than native English speakers (Chen et al., 2001).

The studies of accented English stress produced by Mandarin-English bilingual speakers provided us important information on stress deviation and the influence by L1. However, accented English stress pattern exhibited by Cantonese-English (C-E) bilingual speakers speaking English as an L2 is lacking. To respond to the knowledge gap, the present study explored the English stress pattern produced by C-E speakers. Objective measurements were obtained from English produced by C-E bilingual speakers, and compared with those obtained from monolingual speakers of American-English (A-E). Based on previous studies, the acoustical differences between productions made by C-E and A-E speakers and between stressed and unstressed syllables were predicted. Firstly, it was predicted that both C-E and A-E speakers should be able to use F0, vowel duration and intensity to differentiate stressed and unstressed words. Moreover, C-E speakers were predicted to produce the stressed words with a higher F0, shorter vowel duration and comparable intensity in stressed syllables.
compared with native English speakers. These predictions were made based on the fact that both Cantonese and Mandarin are tonal languages with interplay effects between intonation and lexical tones and the nature of relatively constant syllable duration across sentence. Therefore, such characteristics of Cantonese would exert an interference effect on the production of English sentences by L2 C-E speakers in a way similar to Mandarin-English bilingual speakers.

According to Fry (1955), both duration and intensity are physical cues for the perception of linguistic stress patterns and duration ratio is a more effective cue. However, Sluijter, Heuven and Pacilly (1997) argued that intensity in high frequency was a strong cue in linguistic stress. The present study also explored the differences in F0, vowel duration and intensity between the English produced by C-E bilingual speakers and A-E monolingual speakers, and the differences may affect listener’s perception of stress. It was predicted that the A-E listeners could correctly differentiate the placement of stress within a sentence produced by C-E bilingual speakers due to their predicted ability to differentiate stress by F0, duration and intensity. Yet, the relevance of each measurement will on the perception was predicted to be different in a way that the duration may be the strongest cue for the stress in perception. Moreover, perceptual degree and naturalness of stress are predicted to be correlated with each other.

Due to interference of L1 on L2 English, it is not uncommon that C-E speakers encounter difficulty in adjusting their accent in speaking English. Chan and Li (2000) stated that one of the pronunciation problems for C-E speakers is to achieve a natural suprasegmental features. From the results of the present study, recommendations could be made regarding how C-E bilingual speakers can improve their English spoken as an L2 by reducing the accentedness in stress production.
Method

Participants

Speakers

Forty (18 male and 22 female) adult C-E bilingual speakers participated in the study. The selection criteria for the bilingual speakers included: (1) the speakers were Cantonese-English (C-E) bilingual speakers between 19 and 27 years of age with a mean age of 23 years.; (2) Cantonese was their first language and they have learned English for over 10 years; (3) they demonstrated proficiency in listening and speaking English; (4) they had obtained at least a grade C in the HKCEE or D in the HKAL or A in the verbal section in the University Entrance Test for Speaker of other Languages; and (5) they were physically healthy and free of speech, language and/or hearing problems. Data were compared with speech samples produced by native monolingual speakers of American English (A-E) provided by Chen and colleagues (Chen et al., 2001).

Listeners

Eight adults who were native speakers of American English served as listeners in the present study. They were randomly selected from the Midwestern region of North America with ages ranged from 20 to 32 years (average age = 24 years) who had no prior training and exposure to Cantonese. The listeners had no reported history of hearing problems and were literate with at least college-level education.

Speech materials

The speech materials used in the study were similar to that used by Chen et al. (2001). Sentence stress pattern was evaluated from participant’s production of the sentence “I bought a cat there”. This sentence was produced in four distinct manners by placing the primary stress on one of four different words (stressed word is in CAPITAL letter and italics): (1) I bought a cat there; (2) I BOUGHT a cat there; (3) I bought a CAT there; and (4) I bought a cat
Each sentence was produced three times and the order at which they were produced was randomized. A total of 12 English sentences were eventually produced by each participant.

**Audio recordings**

All recording took place in a sound-treated booth of the Speech Science Lab. Prior to the actual recording, the speakers were allowed to practice the speech materials as often as they needed. The sentence was accepted if it was perceptually fluent and contained no misarticulation and conformed to the prescribed placement of primary stress. The investigator who was proficient in English monitored the entire recording procedure. The speakers were instructed to produce the sentences at a comfortable conversational loudness level and a normal speaking rate. Speech samples were recorded by using a high-quality microphone (Shure, SM 58) and a preamplification system (M-Audio, MobilePre USB). During the recording, the microphone was placed approximately 10 cm away from the speaker’s mouth. To calibrate for intensity, 10 data points of each of the three pure tones of different intensities (40 dBSPL, 50 dBSPL, & 60 dBSPL) were recorded. The acoustic signals were digitized at 20 kHz and 16 bits/sample by using Praat and then stored in the computer for later analyses.

**Acoustic analysis**

To complete the intensity calibration, the calibration signals corresponding to different intensity levels were retrieved and used to calculate for the calibration equation. The equation was used to calculate the actual intensity of the speech samples.

During the acoustic analysis, each of the recorded sentences was displayed in an amplitude-by-time waveform. The cycle-to-cycle F0 values, intensity values, and vowel duration were measured from the vocalic nucleus of each stressed and unstressed words of the sentences produced by C-E speakers. All measurements were made only from the vowel nucleus of the stressed targets as, according to Fry (1955), the vowel segments were found to
show apparent differences. Similar to the study reported by Chen et al. (2001), to trace the possible changes in F0 and intensity of the vowel, three measurements of F0 and intensity values were obtained from three different locations (beginning, midpoint and end) of each vowel segment. Moreover, in order to minimize the possible coarticulatory influence of pre- and post-vocalic consonants on the vowel segment, the beginning of the vowel was defined as the first 50 ms after the first recognizable glottal cycle. Similarly, the ending of the vowel was defined as the last 50 ms before the last recognizable glottal pulse. The actual intensity level of the segment was calculated by using the intensity values and the calibration equation previously obtained.

The data obtained were compared with that obtained from A-E monolingual speakers reported by Chen et al. (2001). Two measurements for evaluating the sentence stress were used. The first one was the average sentences stress: the average F0, intensities and vowel duration of the stressed word. The second measurement was the across-sentence stress in which the average F0, intensities and vowel durations obtained from stressed compared with unstressed syllables. This latter measurement aimed at evaluating the differentiated stress since stress should be a relative concept with the stressed syllable more prominent than other syllables (Laver, 1994). For instance, according to Chen et al. (2001), for the across sentence stress measurement, F0 for the word “bought” across the four sentences were calculated by the difference between the mean F0 of the stressed \textit{BOUGHT} and the average mean F0 of the unstressed \textit{bought}.

\[
\Delta F0 = F0_{BOUGHT} - \frac{(F0_{bought(1)} + F0_{bought(2)} + F0_{bought(3)})}{3}
\]

Perceptual Experiment

The English sentences produced by C-E speakers were used for the perceptual experiment. The listening experiment was carried out in a sound-treated room. Speech samples were presented to the listeners via high-quality headphones at a comfortable loudness
Upon listening to each speech sample, the listeners were asked to judge for the placement of stress, degree of stress and naturalness of the stressed syllable. The latter two was rated using a 10-point equal-interval scale, with a rating of 0 indicated the least amount of stress and least natural production, and a rating of 10 corresponded to the most amount of stress and the most natural production of stress, respectively. Answer sheets on which the rating scales are printed were provided to the listeners. An inter-stimulus silence of 5 seconds was provided to allow sufficient time to complete the perceptual tasks. Prior to the experiment, the listeners were provided with brief practice exercises in order to acquaint themselves with the procedure and experimental setting.

Reliability Measures

To assess the reliability of listeners’ rating of stress, 10% of the speech samples were randomly selected and presented to the listeners a second time. Correlation coefficient was calculated from the scores obtained from the first and second ratings and was used to describe the reliability of rating. To assess the reliability of acoustic measurements, 5% of the sentences were randomly selected and subject to acoustic measurements a second time by another individual who was experienced in acoustic measurements. Correlation coefficients were calculated and used to indicate intra-rater and inter-rater reliability.

Statistical Analysis

To determine the possible differences in the acoustic measurements between the English speech samples produced by C-E and A-E speakers, a mixed-design factorial analysis was used for each measurement. Specifically, two-way repeated-measure Analyses of Variance (ANOVAs) with language group as the between-subjects variable and placement of stressed words (I, BOUGHT, CAT, THERE) as within-subjects measure were used for each gender group, accompanied by necessary multiple comparisons.
Results

Average sentence stress

The average F0, intensity and vowel duration values for sentence stress produced by C-E and A-E female and male speakers are shown in Tables 1 and 2, respectively.

Table 1. Mean (M), standard deviation (s.d.), and range of F0 (Hz), intensity (dB), and duration (ms) for each stressed word produced by female American English monolingual (A-E) and Cantonese-English (C-E) bilingual speakers.

<table>
<thead>
<tr>
<th>Group</th>
<th>Stressed word</th>
<th>F0 (Hz)</th>
<th>Intensity (dB)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>s.d.</td>
<td>Range</td>
</tr>
<tr>
<td>A-E Female</td>
<td>I</td>
<td>225</td>
<td>35</td>
<td>148 - 303</td>
</tr>
<tr>
<td></td>
<td>BOUGHT</td>
<td>234</td>
<td>44</td>
<td>148 - 332</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>189</td>
<td>34</td>
<td>134 - 491</td>
</tr>
<tr>
<td></td>
<td>THERE</td>
<td>207</td>
<td>36</td>
<td>73 - 285</td>
</tr>
<tr>
<td>C-E Female</td>
<td>I</td>
<td>252</td>
<td>31</td>
<td>187 - 331</td>
</tr>
<tr>
<td></td>
<td>BOUGHT</td>
<td>292</td>
<td>35</td>
<td>165 - 373</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>231</td>
<td>29</td>
<td>174 - 360</td>
</tr>
<tr>
<td></td>
<td>THERE</td>
<td>219</td>
<td>23</td>
<td>110 - 261</td>
</tr>
</tbody>
</table>

Table 2. Mean (M), standard deviation (s.d.), and range values of F0 (Hz), intensity (dB), and duration (ms) for each stressed word produced by male American English (A-E) monolingual and Cantonese-English (C-E) bilingual speakers.

<table>
<thead>
<tr>
<th>Group</th>
<th>Stressed word</th>
<th>F0 (Hz)</th>
<th>Intensity (dB)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>s.d.</td>
<td>Range</td>
</tr>
<tr>
<td>A-E Male</td>
<td>I</td>
<td>139</td>
<td>32</td>
<td>86 - 205</td>
</tr>
<tr>
<td></td>
<td>BOUGHT</td>
<td>143</td>
<td>24</td>
<td>86 - 192</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>161</td>
<td>43</td>
<td>79 - 242</td>
</tr>
<tr>
<td></td>
<td>THERE</td>
<td>133</td>
<td>23</td>
<td>74 - 186</td>
</tr>
<tr>
<td>C-E Male</td>
<td>I</td>
<td>141</td>
<td>26</td>
<td>96 - 195</td>
</tr>
<tr>
<td></td>
<td>BOUGHT</td>
<td>161</td>
<td>27</td>
<td>108 - 212</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>153</td>
<td>26</td>
<td>101 - 210</td>
</tr>
<tr>
<td></td>
<td>THERE</td>
<td>127</td>
<td>25</td>
<td>78 - 179</td>
</tr>
</tbody>
</table>
**F0 measurement**

*Female speakers.* Significant interaction was found between language group and stressed word \[F(3,126) = 9.39, p < 0.01\]. To investigate the possible language group difference, independent sample t-test was used. It was found that there was a significant difference \(p < 0.01\) in F0 of the stressed words produced by the C-E females and A-E females. C-E females produced the stressed words with a higher mean F0 (261.78 Hz) than A-E females (227.81 Hz). Moreover, one-way ANOVA was used to determine the influence of stressed words on F0. Results indicated a significant main effect \(p < 0.01\). Results of subsequent pairwise multiple comparisons revealed that F0 values were significantly different among all of the stressed words except between *BOUGHT* and *CAT* for C-E females. For A-E female speakers, F0 value was significantly lower for the stressed word *THERE* than the other three stressed words \(p < 0.01\).

*Male speakers.* A significant interaction effect was found between language group and stressed word \[F(3, 102) = 5.82, p < 0.01\]. This was followed by an independent sample t-test, which revealed no significant difference \(p > 0.01\) in F0 of the stressed words produced by the C-E males and A-E males. One-way ANOVA combined with pairwise multiple comparisons revealed that the C-E males produced stressed *I* at a significantly lower F0 than *BOUGHT* and *CAT* \(p < 0.01\) while produced stressed *BOUGHT* with a significantly higher F0 than the other 3 stressed words \(p < 0.01\). For A-E male speakers, there was significant difference only between the comparisons of stressed *I* and stressed *CAT* as well as stressed *THERE* and *CAT* \(p < 0.01\).

**Duration measurement**

*Female speakers.* There was no significant interaction between language group and stressed word \[F(3, 126) = 3.81, p > 0.01\]. Results of a two-way ANOVA revealed no
significant difference between the two language groups \([F(1,42) = 2.72, p > 0.01]\). However, significant main effects for stressed word \([F(3, 126) = 127.00, p < 0.01]\) were found. Results of pairwise multiple comparisons showed that each of the four stressed words was significantly different from the others regarding the duration \((p < 0.01)\) except between \textit{I} and \textit{THERE} \((p > 0.01)\). Stressed word \textit{I} and \textit{THERE} were produced with a significantly longer duration than other stressed words \((p < 0.01)\) while \textit{CAT} was produced with a significantly shorter duration than other stressed words \((p < 0.01)\) in both language groups.

\textit{Male speakers}. Significant interaction was found between language group and the duration of stressed words \([F(3, 102) = 11.33, p < 0.01]\), which was followed by an independent sample t-test. Results indicated no significant difference \((p > 0.01)\) in duration of the stressed words produced by the C-E males and A-E males. However, results of one-way ANOVA and pairwise multiple comparisons revealed significant difference in the duration of each stressed words except between \textit{BOUGHT} and \textit{CAT} in C-E males \((p < 0.01)\). \textit{I} and \textit{THERE} were found to be produced at a significantly longer duration than other stressed words \((p < 0.01)\) while \textit{CAT} was produced at a significantly shorter duration than other stressed words \((p < 0.01)\). For A-E male speakers, there was significant difference only between \textit{I} and \textit{CAT} as well as \textit{CAT} and \textit{THERE} \((p < 0.01)\).

\textit{Intensity measurement}

\textit{Female speakers}. Significant interaction was found between language group and stressed words \([F(3, 63) = 11.91, p < 0.01]\). The subsequent independent sample t-test showed no significant difference \((p > 0.01)\) in intensity of the stressed words produced by the C-E females and A-E females. However, one-way ANOVA and pairwise comparisons revealed that the stressed word \textit{BOUGHT} was produced with significantly greater intensity than other three stressed words \((p < 0.01)\) in C-E females but there is no significant difference in the stressed
Male speakers. From the two-way ANOVA, there was no significant interaction effect shown by the ANOVA. Moreover, it was found that there was no significant main effect between the male language groups \([F(1, 34) = 1.20, p > 0.01]\). However, there was a significant main effect for stressed words \([F(3, 102) = 12.9, p < 0.01]\), with the stressed words \(I\) being produced at a significantly lower intensity and \(BOUGHT\) being produced at a significantly greater intensity than the remaining three stressed words for both language groups.

Across sentence stress

\(F0\) measurement

Figure 1 shows the comparison of across-sentence \(F0\) between A-E and C-E speakers. Both language groups demonstrated the use of higher \(F0\) in differentiating the stressed words from the unstressed words.

Female speakers. The two-way ANOVA revealed that there was no significant interaction effect between language group and stressed words \([F(3, 126) = 0.09, p > 0.01]\). For female speakers, there was no significant main effect between language groups \([F(1, 42) = 0.258, p > 0.01]\). However, results of the two-way ANOVA indicated a significant main effect in across-sentence \(F0\) among stressed words \([F(3, 126) = 9.41, p < 0.01]\). Pairwise multiple comparisons revealed that the stressed word \(CAT\) was differentiated with a significantly higher \(F0\) than the other three words \((p < 0.01)\).

Male speakers. The two-way ANOVA showed that there was no significant interaction between language group and stressed words \([F(3, 102) = 3.09, p > 0.01]\). There was no significant main effect for the language group \([F(1, 34) = 0.62, p < 0.01]\). However, there was a significant main effect for the stressed words produced by male speakers \([F(3, 102) = 9.42,\]
Results of the pairwise multiple comparisons indicated that the stressed word CAT was significantly differentiated with higher F0 than other three words ($p < 0.01$).

Figure 1. Comparison of across-sentence F0 between A-E and C-E speakers. The values represent the F0 difference between stressed and unstressed words across sentences.

**Duration measurement**

Figure 2 depicts the average difference in vowel duration between the stressed and unstressed words. It shows that both language groups differentiated the stressed word from unstressed words by a longer duration.

*Female speakers.* The two-way ANOVA revealed no interaction effect between language group and stressed words [$F(3, 126) = 0.521, p > 0.01$]. In addition, no significant main effect was found in across-sentence duration between the language groups [$F(1, 42) = 0.003, p > 0.01$]. However, there was a significant main effect for stressed words [$F(3, 126) = 92.59, p < 0.01$]. Pairwise multiple comparisons showed that all stressed words were significantly different from the others ($p < 0.01$). The durational difference of *I* was the largest, and that of *CAT* was the smallest.

*Male speakers.* The two-way ANOVA revealed no interaction between language group and stressed words [$F(3, 102) = 3.40, p > 0.01$]. There was no significant main effect was found in duration between language groups [$F(1, 34) = 3.85, p > 0.01$]. However, there was a significant main effect for the stressed words [$F(3, 102) = 62.87, p < 0.01$]. Similar to female speakers, pairwise multiple comparisons showed that all stressed words were significantly
different from the others \( (p < 0.01) \), with the duration difference of \( I \) being the largest and that of \( CAT \) being the smallest.

**Figure 2.** Comparison of across-sentence duration between A-E and C-E speakers. The values represent the duration difference between stressed and unstressed words across sentences.

**Intensity measurement**

Figure 3 depicts the average intensity difference between the stressed and unstressed words. Speakers of both language groups were able to differentiate the stressed word from unstressed words by an elevated intensity.

**Female speakers.** The two-way ANOVA presented that there was no significant interaction between language group and stressed words was found \( [F(3, 126) = 1.04, p > 0.01] \). Moreover, no significant main effect was found in intensity difference between the language group \( [F(1, 42) = 2.57, p > 0.01] \). A significant main effect for the stressed words was observed \( [F(3, 126) = 8.49, p < 0.01] \). The pairwise multiple comparisons revealed that the stressed word \( THERE \) was significantly louder than the other three words.

**Male speakers.** The two-way ANOVA revealed no significant interaction between language group and stressed words \( [F(3, 126) = 1.04, p > 0.01] \). Moreover, no significant main effect was found in across-sentence intensity between the language groups \( [F(1, 34) = 188, p > 0.01] \). However, there was a significant main effect for the stressed words \( [F(3, 102) = 16.5, p < 0.01] \). Pairwise multiple comparisons revealed that the stressed word \( THERE \)
was significantly louder than the other three words. No significant interaction was found between language group and stressed words \[F(3, 126) = 1.04, p > 0.01\].

![Bar chart](image)

Figure 3. Comparison of across-sentence intensity between A-E and C-E speakers. The values represent the intensity difference between stressed and unstressed words across sentences.

**Reliability measurement**

For intra-rater reliability, average absolute errors for F0, intensity, and vowel duration were 2.18 Hz, 3.64 dB and 8.06 ms, respectively. Pearson correlation coefficients for F0, intensity, and vowel duration between the first and second measurement were 0.97, 0.76 and 0.71 \((p < 0.01)\), respectively. For inter-rater reliability, average absolute errors for F0, intensity, and vowel duration for inter-rater reliability were 2.01 Hz, 5.03 dB and 1.74 ms, respectively. Pearson correlation coefficients for F0, intensity, and vowel duration between the first and second measurement were 0.96, 0.72 and 0.91 \((p < 0.01)\), respectively.

**Perceptual experiment**

**Placement of stress**

Average percent correct identification of stress placement obtained from raters who were native speakers of American English is depicted in Table 3. Results showed that the stressed word *BOUGHT* had the greatest accuracy of identifying stress placement (90.6%), followed by *THERE* (90.3%) and then by *CAT* (82.0%). The stressed word *I* showed the lowest
Table 3. Average percent identification of stress placement by native A-E raters.

<table>
<thead>
<tr>
<th>Stressed word Perceived</th>
<th>Percent Identification of Stress Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stressed word Presented</td>
</tr>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>77.81</td>
</tr>
<tr>
<td>BOUGHT</td>
<td>15.31</td>
</tr>
<tr>
<td>CAT</td>
<td>2.19</td>
</tr>
<tr>
<td>THERE</td>
<td>4.69</td>
</tr>
</tbody>
</table>

To correlate the above results of placement of stress with the production of C-E speakers, the average F0, duration and intensity measurements of C-E speakers were analyzed by ANOVA. For F0, it was found that the C-E females produced the stressed I with a significantly lower F0 than stressed BOUGHT and CAT ($p < 0.01$) while produced stressed BOUGHT with a significantly higher F0 than stressed I and THERE ($p < 0.01$). The C-E males produced stressed I at a significantly lower F0 than BOUGHT and CAT ($p < 0.01$) while produced stressed BOUGHT with a significantly higher F0 than all other 3 stressed words ($p < 0.01$). Regarding the vowel duration, the C-E females produced the stressed I with a significant longer duration than stressed BOUGHT and CAT ($p < 0.01$) while produced stressed BOTUGH with a significantly shorter duration than stressed I and THERE ($p < 0.01$). Similar pattern was shown for males. Regarding intensity measurement, both the C-E female and male speakers produced the stressed BOUGHT with greater intensity than other three stressed words ($p < 0.01$).

Degree of stress

Table 4 shows the average rating of the degree of stress obtained from the perceptual
experiment. Results showed that the raters perceived the highest degree of stress when the word *THERE* (average rating = 6.48) was being stressed, followed by *CAT* (average rating = 6.18), then by *BOUGHT* (average rating = 6.05). The stressed word *I* (average rating = 5.87) was rated as having the lowest degree of stress compared to other three words.

Table 4. Percentage and average rating in degree of stress by native A-E raters, with 10 and 1 representing the most stressed and the least stressed respectively.

<table>
<thead>
<tr>
<th>Stressed word</th>
<th>Percentage Ratings (%)</th>
<th>Degree of stress</th>
<th>Average Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>I</em></td>
<td>1  7  7  10  14  20  21  13  7  1</td>
<td>5.87</td>
<td></td>
</tr>
<tr>
<td><em>BOUGHT</em></td>
<td>0  4  8  8  16  19  21  16  5  3</td>
<td>6.05</td>
<td></td>
</tr>
<tr>
<td><em>CAT</em></td>
<td>0  5  6  10  14  17  20  15  9  4</td>
<td>6.18</td>
<td></td>
</tr>
<tr>
<td><em>THERE</em></td>
<td>1  4  6  8  11  15  20  18  12  6</td>
<td>6.48</td>
<td></td>
</tr>
</tbody>
</table>

Naturalness of stress

Naturalness of stress was also rated by the raters and the average ratings of naturalness of stress are shown in Table 5. Results indicated that the raters perceived the naturalness to be the highest when the primary stress words was *I* (average rating = 5.01), followed by *BOUGHT* (average rating = 4.61), then by *CAT* (average rating = 4.44). The least natural stress perceived was on *THERE* (average rating = 4.36).

Table 5. Percentage and average rating in naturalness of stress by native American raters, with 10 and 1 representing the most natural and the least natural respectively.

<table>
<thead>
<tr>
<th>Stressed word</th>
<th>Percentage Ratings (%)</th>
<th>Naturalness of stress</th>
<th>Average Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>I</em></td>
<td>2  10  14  15  15  11  12  8  8  5</td>
<td></td>
<td>5.01</td>
</tr>
<tr>
<td><em>BOUGHT</em></td>
<td>4  13  18  18  18  9  10  5  4  2</td>
<td></td>
<td>4.61</td>
</tr>
<tr>
<td><em>CAT</em></td>
<td>6  14  18  17  17  9  9  4  5  2</td>
<td></td>
<td>4.44</td>
</tr>
<tr>
<td><em>THERE</em></td>
<td>4  15  15  16  18  8  9  7  6  2</td>
<td></td>
<td>4.36</td>
</tr>
</tbody>
</table>
Post-hoc test was performed to investigate the relationship between the degree and naturalness of stress. Pearson correlation coefficient between them was -0.895 indicating a strong negative correlation between the degree of stress and the naturalness of stress.

*Reliability measurement*

Pearson correlation coefficients for placement of stress, degree of stress, and naturalness of stress between the first and second time of the perception were 0.89, 0.66 and 0.57 ($p < 0.01$) respectively.
Discussion

Sentence stress according to F0

Results of the present study indicated that C-E bilingual speakers (L2) could use F0 to differentiate the stressed words from the unstressed words in English sentence production, in a way similar to Mandarin-English bilingual speakers and A-E monolingual speakers reported by Chen et al. (2001). C-E bilingual speakers were able to produce higher F0 of the stressed words to achieve the differentiation. This can be justified by the positive difference in F0 across sentence stress calculation. However, it was found that C-E female speakers tended to use a significantly higher average and across F0 than A-E female speakers to indicate the stressed words while they were not significantly different between the C-E and A-E male speakers.

There are two possible reasons accounting for the significant female language group difference. Firstly, the higher F0 in the stressed words produced by the C-E females than those by A-E females may be due to possible structural differences of their larynges. Yang (1996) correlated vowel F0 and formant frequencies produced by Caucasian and Korean and their laryngeal structures, and found that the F0 and formant frequencies associated with Korean speakers were higher than those with Caucasian speakers. Yang proposed that the difference was due to overall smaller larynx in Korean speakers; Korean speakers were found to have shorter average vocal tract length and back cavity than the Caucasian speakers, which was inversely correlated with formant frequencies (Johnson, 1997). Similarly, Adrianopoulos, Darrow, & Chen (2001) also found that Mandarin speakers produced vowels with a higher fundamental frequency than their counterparts of Caucasian, African American speakers of standard American English and native Hindi Indian speakers. Adrianopoulos et al. (2001) suggested that F0 is strongly affected by the length of vocal folds and the density of vocal fold tissues so that these structural differences were suggested to be the factors accounting for the
higher fundamental frequencies in Mandarin speakers. Therefore, the F0 difference between female speakers of language groups in the present study may also be contributed by the overall smaller vocal tract in C-E speakers when compared with A-E speakers. Apparently, future study should focus on the possible anatomical difference between C-E and A-E female speakers, in relation to the F0 and formants differences.

Another possible explanation is the influence of L1 on L2 in the C-E speakers; namely Cantonese tone may have influenced C-E speakers’ production of English sentences. As reported by Chen et al. (2001), similar deviation of a higher F0 produced by Mandarin than native A-E speakers was attributed by the significantly greater F0 fluctuations at syllable-level than the continuous speech of English suggested by Eady (1982). As both Cantonese and Mandarin are tonal languages in which each syllable possesses its own tonal contour (Eady, 1982), it may exert an effect on English stress pattern produced by C-E speakers. Xu (1999) suggested that lexical tone in each syllable is the main component determining the height and shape of the F0 contour of that syllable. At the same time, it equally affects the adjacent syllables. Xu (1999) also suggested that sentence stress solidly alters the overall F0 contour of the entire sentence. This suggestion was supported by the study of Yujia et al. (2004) in which surface F0 of a sentence was a combination of the tone-related local components and phrase-level long term variation. Moreover, the interplay between sentence intonation and lexical tones proposed by Shen (1990) may also be an account for the F0 deviation. Shen (1990) believed that such interplay is the cause of greater pitch fluctuations and increased overall pitch level in French, which is a non-tonal language, produced by native Mandarin speakers. Such interplay also appeared in Cantonese. The investigation of Cantonese by Ma et al. (2004) concluded that both tone contour and tone level are modified by intonation within a sentence in tonal language. Therefore, both the influence of tonal fluctuations at syllable level as well as the interplay between lexical tone and sentence intonation in Cantonese can account
for the raise in F0 in stressed words in C-E speakers than A-E speakers.

Sentence stress according to duration

Results from the present study indicated that both female and male C-E speakers were able to produce stressed words with a longer duration than unstressed words, in a way similar to Mandarin-English bilingual speakers and A-E monolingual speakers as reported by Chen et al. (2001). This is shown by the positive differences in the across sentence stress duration calculation. The lack of significant difference in duration between C-E and A-E speakers for both female and male speakers implies that, for both females and males, C-E and A-E speakers are having comparable ability in using duration to correctly contrast English sentence stress.

In the present study, C-E bilingual speakers were able to use durational differentials to signal stress in a way similar to A-E speakers. This is contradictory to the results reported by Chen et al. (2001) in which Mandarin-English bilingual speakers were different in differentiating stress by using duration compared with A-E speakers. According to Chen et al. (2001), such difference was attributed to the syllable-timed nature of Mandarin. Since Mandarin is a syllable-timed language, the interval between stressed and unstressed syllables are relatively regular (Roach, 1991) with a constant syllable duration across a sentence (Chun, 1982). It increases the difficulty for speakers of a syllable-time language to adjust for correct English sentence stress. However, Cantonese is also syllable-timed. Yet, the present data did not indicate the difficulty for C-E speakers to produce English stress pattern correctly. This is probably indicative of the high proficiency of English in the C-E speakers participated in the present study. As suggested by Guion et al. (2000), proficiency of speaking English as a L2 had a common feature of temporal variability. The degree of temporal variability is affected by a number of factors such as age of arrival to the English-speaking country, age of
acquisition, length of stay and the amount of speaking L1 and L2. The C-E speakers may have started acquiring English at an earlier age, probably since kindergarten. This may have reduced the effect of syllable-timed nature of Cantonese in C-E speakers.

Sentence stress according to intensity

Results from the present study confirmed that C-E bilingual (L2) speakers could use intensity to differentiate stressed words and unstressed words in English sentence. C-E speakers produced the stressed words with greater intensity than the unstressed words. The positive difference in across-sentence stress intensity achieved by the C-E speakers is similar to A-E speakers and Mandarin-English bilinguals as reported in Chen et al’s study (2001). Moreover, for both females and males, the lack of significant difference in the intensity differentials between C-E and A-E speakers implies that C-E bilingual speakers were able to use intensity to achieve sentences stress in a way similar to A-E speakers, and Mandarin-English bilingual speakers, as reported by Chen et al. (2001).

It can be concluded that C-E speakers had a higher proficiency in achieving English sentence stress by the use of intensity and durational difference than Mandarin-English bilingual speakers reported in Chen et al. (2001), as the C-E speakers appeared to have a more comparable intensity and durational pattern to A-E speakers than Mandarin-English speakers. Such difference in the ability in distinguishing English stress may be due to the different selection criteria for the bilingual speakers implemented in the present study. C-E speakers must have obtained at least a grade C in the HKCEE or D in the HKAL or A in the speaking section in the University Entrance Test for Speaker of other Languages. Moreover, in Hong Kong, most of the C-E speakers have been learning English since kindergarten. The long period of learning English might have contributed to the higher proficiency in sentence stress.
pattern observed in C-E speakers.

Perception of sentence stress produced by C-E bilingual speakers

The present results indicate that the A-E rater had the lowest accuracy in judging the placement of the stressed word *I* while the highest accuracy for *BOUGHT*. This matches with the acoustical measurements of average F0 and intensity. It appeared that the stressed word *I* was produced in a relatively lower F0 and less intense manner while the stressed word *BOUGHT* was produced in a higher F0 and greater intensity compared to other stressed words. With the combined effects of F0 and intensity, it is not surprising that the A-E raters could achieve the lowest perceptual accuracy in judging the placement of the stressed word *I* but with the highest perceptual accuracy in that of *BOUGHT*. This result also agrees with the finding of degree of stress which the listeners rated stressed word *I* (5.87) to be relatively low while stressed word *BOUGHT* (6.05) to be relatively high. The reason for the most prominent use of F0 and intensity to indicate stressed word *BOUGHT* is proposed to be due to its plosive nature. Such unique plosive nature with a burst of energy and voicing may have facilitated the elevation of F0 and intensity during signaling the stress.

According to Fry (1955), both duration and intensity are physical cues for the perception of linguistic stress patterns and duration ratio is a more effective cue. However, the results of average stress of duration in the present study were contradictory to the perceptual accuracy. Such discrepancy may suggest that the A-E raters did not solely rely on duration in perception of stress. It is proposed that the combined effects of frequency and intensity may be greater than the effect of duration alone in perceiving sentence stress in English produced by C-E speakers. Further study is needed to investigate the contribution of each acoustic cue on the perception of sentence stress in English. The present study also revealed a strong negative correlation between the degree of stress and the naturalness of stress. It follows that the more
the C-E speakers stressed on the stressed words, the more unnatural it would be from the point of view of listeners who were native speakers of A-E.

Conclusion

The present study revealed that C-E speakers were able to use F0, vowel duration and intensity to differentiate different stress patterns. Their use of vowel duration and intensity to signal the stress is similar to A-E speakers. However, the use of F0 in the stressed words by C-E female speakers was different from A-E female speakers. Moreover, F0 and intensity were found to correlate closely with perceptual judgment and the degree of stress with the naturalness of stress.

Being an international language, English is important in daily life communication among C-E bilingual speakers in Hong Kong. It has been suspected that C-E bilingual speakers would encounter difficulties in correctly speaking English as an L2. Chan and Li (2000) argued that achieving suprasegmental features of English in a way similar to native speakers of A-E is one of the pronunciation difficulties for C-E bilingual speakers. Based on the present results, C-E speakers are recommended to improve English stress production by reducing the elevation in F0 in the stressed words especially for female speakers. This might help the C-E bilingual speakers in reducing the accent in English sentence stress in order to achieve more effective communication in English.
References


Appendix A Chinese version of informed consent form for adult speakers

參與實驗同意書

雙語使用者(廣東話及英語)的英文句子重音模式研究

現誠意邀請閣下參與由香港大學言語及聽覺科學系四年級學生譚詩雅主理的研究調查。是項研究利用及涉及聲學數據之分析，來研究語言學與句子重音的關係，藉此研究中英文雙語者英文句子重音模式的自然和準確程度，以及對其作出改善建議。

實驗會在香港大學菲臘牙科醫院五樓的隔音室進行，聲學樣本將會通過一枝放於距離約10厘米的麥克風錄取。閣下只要根據指示，以正常速度及聲線分別說出12句英文句子“I bought a cat there”，每句句子的重音位置會以底線表示，例如“I bought a cat there”的重音位置是“bought”，如此類推。閣下可以在錄音開始前練習，而閣下的發聲將會被錄音作研究用途。閣下有權要求聽回個人的錄音樣本，也有權要求銷毀所有或部分樣本。

紀錄過程中閣下可能會有少許不適或疲倦，所以在有需要時可以隨時稍作休息。

是次參與純屬自願性質，閣下可隨時終止參與是項行動，有關決定將不會引致任何不良後果。

所收集的資料只作研究用途並於最遲五年後銷毀，個人資料將絕對保密。

是次研究並不為閣下提供個人利益，但所搜集數據將對研究語言學與句子重音的關係提供寶貴的資料。如閣下對是項研究有任何問題，請現在提出。

如日後閣下對是項研究有任何查詢，請與譚詩雅聯絡(電話號碼: 64361538)。如你想知道更多有關研究參與者的權益，請聯絡香港大學非臨床研究操守委員會(22415267)。

如閣下明白以上內容，並願意參與是項研究，請在下方簽署。在此多謝閣下的參與。

__________________________                _______________________
參與研究者姓名

__________________________                _______________________
參與研究者簽署

__________________________                _______________________
研究人員姓名

__________________________                _______________________
研究人員簽署

日期

日期
Appendix B English version of informed consent form for adult speakers

Informed Consent Form for Adult
Sentence Stress Pattern in English Produced By Cantonese-English Bilingual Speakers

You are invited to participate in a research study conducted by a year 4 student Tam Sze Nga, Emily in the Division of Speech and Hearing Sciences at the University of Hong Kong.

PURPOSE OF THE STUDY
This study will investigate the difference in the stress pattern associated with English sentences produced by Cantonese-English speakers and native American speakers through acoustic analysis.

PROCEDURES
For speakers:
You are invited to produce the 12 sentence “I bought a cat there” with different placement of stress in front of a microphone which placed approximately 10 cm away from your mouth. The primary stress is indicated by underline word. (1) I bought a cat there; (2) I bought a cat there; (3) I bought a cat there; (4) I bought a cat there. Each sentence will be produced 3 times and the order of presentation will be randomized.
Audio recording of the productions will be carried out and you are allowed to practice the speech materials as often as you wish prior to the actual recording. The whole procedure would take about 30 minutes in the sound-proof booth on 5/F in Prince Philip Dental Hospital.

POTENTIAL RISKS / DISCOMFORTS AND THEIR MINIMIZATION
You may experience some mild fatigue and discomforts during the procedure. Such fatigue and/or discomforts will be kept to a minimum because the tasks are self-paced and you are free to take short breaks.

COMPENSATION FOR PARTICIPATION
There will be no specific compensation for participation.

POTENTIAL BENEFITS
There will be no direct benefits to you but the results will be beneficial for current and future research studies.

CONFIDENTIALITY
Any information obtained in this study will remain very strictly confidential, will be known to no-one, and will be used for research purposes only. Codes, not names, are used on all test instruments to protect confidentiality. You can review the audio/video-recording of the procedure. We will erase the entire audio tape or parts of it if you want us to do so.

PARTICIPATION AND WITHDRAWAL
Your participation is voluntary. This means that you can choose to stop at any time without negative consequences. Once you have withdrawn, all your information will be obliterated.
STORAGE OF DATA
For research purposes, your participation will be audio/video-taped for further data checking. The record will be disposed of 5 years after publication of the relevant research results.

QUESTIONS AND CONCERNS
If you have any questions or concerns about the research, please feel free to contact Tam Sze Nga, Emily at HKU, at 64361538 or emilyeastlife@gmail.com. If you have questions about your rights as a research participant, contact the Human Research Ethics Committee for Non-Clinical Faculties, HKU (22415267).

SIGNATURE
I _________________________________ (Name of Participant)
understand the procedures described above and agree to participate in this study.

________________________________________
Signature of Participant
Date: __________________________________
Date of Preparation: _______________________ 
HRECNCF Approval Expiration date: ______
Appendix C Informed consent form for adult listeners

Informed Consent Form for Adult
Sentence Stress Pattern in English Produced By Cantonese-English Bilingual Speakers

You are invited to participate in a research study conducted by a year 4 student Tam Sze Nga, Emily, in the Division of Speech and Hearing Sciences at the University of Hong Kong.

PURPOSE OF THE STUDY
This study will investigate the difference in the stress pattern associated with English sentences produced by Cantonese-English speakers and Native American speakers through acoustic analysis.

PROCEDURES
You are invited to listen to some speech samples of the sentence “I bought a cat there” at a random order. For each sentence, the primary stress will be different. You are invited to judge for the placement of stress, degree of stress and naturalness of the stressed syllable using a 10-point equal-interval scale. Answer sheets on which the rating scales are printed will be provided to you. The whole procedure would take about one hour and a half inside the sound-treated room.

POTENTIAL RISKS / DISCOMFORTS AND THEIR MINIMIZATION
You may experience some mild fatigue and discomforts during the procedure. Such fatigue and/or discomforts will be kept to a minimum because the tasks are self-paced and you are free to take short breaks.

COMPENSATION FOR PARTICIPATION
There will be no specific compensation for participation.

POTENTIAL BENEFITS
There will be no direct benefits to you but the results will be beneficial for current and future research studies.

CONFIDENTIALITY
Any information obtained in this study will remain very strictly confidential, will be known to no-one, and will be used for research purposes only. Codes, not names, are used on all test instruments to protect confidentiality.

PARTICIPATION AND WITHDRAWAL
Your participation is voluntary. This means that you can choose to stop at any time without negative consequences. Once you have withdrawn, all your information will be obliterated.

STORAGE OF DATA
For research purposes, your participation will be audio/video-taped for further data checking. The record will be disposed of 5 years after publication of the relevant research results.

QUESTIONS AND CONCERNS
If you have any questions or concerns about the research, please feel free to contact Tam Sze Nga, Emily at HKU, at 64361538 or emilyeastlife@gmail.com. If you have questions about your rights as a research participant, contact the Human Research Ethics Committee for
Non-Clinical Faculties, HKU (2241-5267).

SIGNATURE

I _________________________________ (Name of Participant)
understand the procedures described above and agree to participate in this study.

________________________________________
Signature of Participant
Date:_____________________________________

Personal Information

Name:__________________________
Gender:_________________________
Age:____________________________

1. Are you a native speaker of American English?
□ Yes □ No

2. Do you have any prior training in Cantonese?
□ Yes □ No

3. Do you have any history of hearing problems?
□ Yes □ No

4. Do you have college-level education?
□ Yes □ No

Thank you very much for your kind participation!
Acknowledgements

I would like to express my wholehearted gratitude to my supervisor, Dr. Lawerence Ng, for giving me guidance, support and encouragement throughout the development of this study. Special thanks to all subjects who helped in this research. Thanks are also given to all staffs in Division of Speech and Hearing Sciences, The University of Hong Kong. Deepest thanks are given to my beloved family, friends, classmates, clinical supervisors and Mr. Ricky Fung for their psychological support and help.