Underdifferentiation of English Lexical Stress Contrasts by L2 Taiwan Speakers

Chiu-yu Tseng[1], Chao-yu Su[1] and Tanya Visceglia[2]

[1]Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan
[2]Education Center for Humanities and Social Sciences, National Yang Ming University, Taipei, Taiwan

cytling@sinica.edu.tw

Abstract

Learning the stress patterns of English words presents a challenge for L1 speakers from syllable-timed and/or tone languages. Realization of stress contrasts in previous studies has been measured in a variety of ways. This study adapts and extends Pairwise Variability Index (PVI), a method generally used to measure duration as a property of speech rhythm, to compare F0 and amplitude contrasts across L1 and L2 production of stressed and unstressed syllables in English multisyllabic words. L1 North American English and L1 Taiwan-Mandarin English speech data were extracted from the AESOP-ILAS corpus. Results of acoustic analysis show that overall, stress contrasts were realized most robustly by L1 English speakers. A general pattern of contrast underdifferentiation was found in L2 speakers with respect to F0, duration and intensity, with the most striking difference found in F0. These results corroborate our earlier findings on L1 Mandarin speakers' production of on-focus/post-focus contrasts in their realization of English narrow focus. Taken together, these results demonstrate that underdifferentiation of prosodic contrasts at both the lexical and phrase levels is a major prosodic feature of Taiwan English; future research will determine whether it can also be found in the L2 English of other syllable-timed or tone language speakers.

Index Terms: L2 English, lexical stress, Taiwan Mandarin, underdifferentiation

1. Introduction

While past studies of non-native accent have focused primarily on segmental variation between L1 and L2, a growing body of research has shown that differences between L1 and L2 production of prosody also make a significant contribution to the perception of non-native accent [1, 2]. One of the factors which have been demonstrated to affect intelligibility across a range of listener groups is misplacement or non-target realization of lexical stress. Field [3] required groups of native and non-native listeners to transcribe recorded material in which lexical stress had been acoustically manipulated. For both native and non-native groups, rightward stress shift and stress shift unaccompanied by a change in vowel quality were found to have the strongest effect on intelligibility. Tajima et al. [4] re-synthesized two-word utterances in Mandarin-accented English to match temporal characteristics of the same utterances recorded by native English speakers and temporally distorted the same utterances recorded by native English speakers to match the temporal characteristics of Mandarin-accented ones. Intelligibility of unmodified L1 English stimuli declined after temporal distortion from 94% to 83%. Intelligibility of unmodified L1 Mandarin English phrases was 39%, which increased to 58% after temporal correction.

Other specific features found in both Taiwan and Beijing L2 English include underdifferentiation of narrow-focus and on-focus/no focus contrasts [5, 6]. Similar underdifferentiation patterns were obtained in PVI analyses of Vietnamese Australian L2 English speech rhythm [7]. Both Taiwan Mandarin and Vietnamese are syllable-timed tone languages; thus, it is possible that such prosodic patterns are found in the L2 English of many other syllable-timed or tone language speakers.

The following study presents acoustic analysis of L1 English and L1 Taiwan-Mandarin English speech data extracted from the AESOP-ILAS corpus (Asian English Speech cOrpus Project, Institute of Linguistics, Academia Sinica) for the purpose of investigating differences in the realization of English lexical stress by L1 speakers of North American English and Taiwan Mandarin. This study focuses on contrast insufficiency at the lexical level and uses PVI to measure the production of stress contrasts in F0, duration and intensity. Traditionally, PVI has been used to measure differences in duration as a component of speech rhythm [8], but we have adapted and extended this method to measure F0 and intensity. PVI measures average difference in acoustic features between adjacent phonological units such as vowels, consonantal intervals or syllables. In present study, the syllable is adopted as the unit of feature extraction for representing stress-related contrasts. Our purpose in performing these analyses was to compare L1 English and Taiwan Mandarin speakers' realization of English lexical stress contrasts and to determine whether F0 and intensity patterns similar to those found in our comparison of L1 and L2 narrow focus would be obtained [5, 6]. Similar patterns would suggest that similar planning strategies are employed by L2 speakers at both the lexical and phrase levels. If different patterns emerge, our focus would shift to determining which acoustic correlates represent the most substantial source of difference between the L1 and L2 speaker groups, and discussing the implications of those differences.

2. Method

2.1. Recording Materials

The materials used in this study represent a subset of the core phonetic experimental tasks developed by AESOP (Asian English Speech cOrpus Project), a multinational collaboration established with the goal of building speech corpora to represent the varieties of English spoken in Asia [9] using the same recording set-up. This experiment uses Task 1, in which 1, 2-, 3- and 4-syllable target words of all possible stress patterns were embedded in a fixed, sentence-medial position; a total of 20 target words were selected (money, morning, white wine, hospital, apartment, department, tomorrow, video,

overnight, January, supermarket, elevator, available, Japanese, afternoon, misunderstand, information, experience, California, Vietnamese). Each of the experimental sentences contains one target word appearing in a broad-focused position two syllables removed from any phrase boundary.

2.2. Procedure

A total of 14 speakers: 7 L1 speakers (2 male and 5 female) and 7 Taiwan L2 speakers (3 male and 4 female) were recorded by trained proctors in quiet rooms directly into a laptop computer. Proctors used a recording platform developed specifically for the AESOP project with pre-loaded experimental sentences, each appearing individually on a computer screen. Participants wore head-mounted Sennheiser PC155 microphones positioned 2 cm away from their mouths; they were instructed to speak naturally at a normal rate and volume. All data were preprocessed automatically for segmental alignments using the HTK Toolkit then manually spot-checked by trained transcribers for accuracy of segmental alignment. Subsequent manual checking of F0 and intensity values was also performed to ensure extraction accuracy.

2.3. Data Analysis

The PVI index, i.e. the average difference in duration between adjacent phonological units such as vowels, consonantal intervals or syllables is among the most accepted methods to compare and represent rhythmic differences among languages [10]. Stress-timed languages are reported to exhibit higher PVI values than syllable-timed languages [8, 11]. Analyses of Japanese (mora-timed) and Estonian L2 English have shown that PVI is also a useful detector of non-native speech rhythm [11]. We began our analysis with the acoustic correlate duration and chose the syllable as the phonological unit of PVI analysis to more accurately reflect Mandarin speech rhythm. The syllable is also the phonological unit of Mandarin tone; thus, this choice also facilitates the inclusion of tone in future prosodic analyses. We then applied the same rationale to analyze average difference in the acoustic correlates F0 and intensity.

The data analysis procedure includes 2 steps: (1) calculating the difference between the current interval and the one that follows in terms of a particular acoustic parameter (2) computing the average of all differences. The PVI extraction equation appears below for duration d, in which k=syllable index and m= number of syllables in the target word:

$$\mathrm{PVI} = \sum_{m-1}^{k-1} (|d_k - d_{k-1}| / (\mathrm{m} - 1))$$

To facilitate comparison across speakers, duration values were subjected to Z-score normalization. F0 and intensity were normalized using the maximum and minimum values in each sentence. In addition, duration extraction was further refined to remove the effect of inherent segmental duration and boundary effects using the multi-layered normalization method that appears below [12], in which *factor1* represents information at the segmental level, *factor2* represents respective syllable position within the word (to remove word-final boundary lengthening effects), and ε_i represents all other unpredictable values. Extracted values μ_i thus represent duration values which have been normalized for inherent segmental duration and boundary effect:

$$x_i = \mu_i + factor_1 + factor_2 + ... + \varepsilon_i$$

3. Results

3.1. L1/L2 Production of Lexical Stress Contrasts

Average values for twenty English words across two speaker groups and three acoustic parameters are given in Table 1. Overall, results show between-group differences in all three acoustic parameters measured, with the most obvious difference appearing in F0, for which the degree of contrast produced by L1 English speakers is twice that of L2 speakers. In terms of duration and intensity, the degree of contrast produced by L1 speakers is only slightly higher, with a L1/L2 ratio of 1.281 and 1.003, respectively.

Table 1: *Average stress contrast for 20 English words by speaker group and acoustic parameter*

Speaker group / Prosodic attributes	L1	L2	L1/L2
F0	0.031	0.015	2.036
Duration	0.161	0.134	1.207
Intensity	0.1999	0.1994	1.003

3.2. PVI distribution by lexical item, speaker group and acoustic parameter

Figure 1 shows distribution of the L1/L2 PVI ratio across words and acoustic parameters, which was calculated in order to determine the most stable indicator across lexical items for distinguishing L1 and L2 speech. The dotted line in Figure1 represents equal L1/L2 PVI. Values above the dotted line indicate a higher level of stability as an indicator to distinguish L1 and L2. For F0, 19 out of 20 words exhibit higher than equal values; for intensity and duration, only 10 words exhibit higher than equal values.

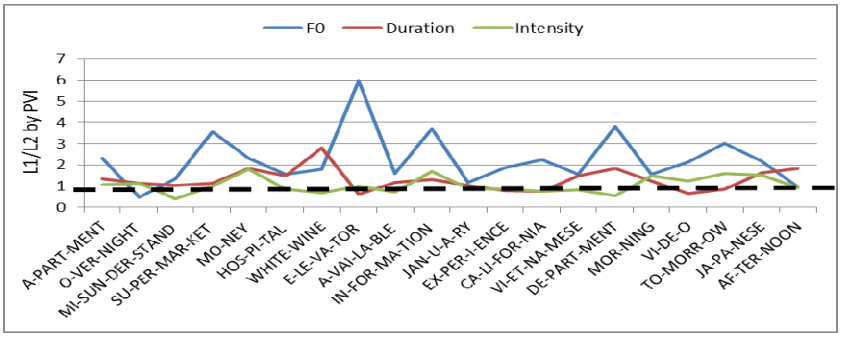


Figure1. L1/L2 ratio of PVI by acoustic parameter and word

3.3. Prosodic patterns by syllabicity, acoustic parameter and speaker group

In this section, we observe prosodic patterns occurring in 2-, 3- and 4-syllable words in order to determine whether production of stress contrasts could be related to the number of syllables or placement of stress in different lexical items. This analysis is illustrated in the figures below, containing the items "money" (2-syllable initial stress), "tomorrow" (3-syllable penultimate stress) and "California" (4-syllable penultimate primary stress, initial secondary stress) respectively. Each figure contains 3 sub-figures representing normalized F0, duration and intensity. Each sub-figure compares L1 and L2 English by individual acoustic parameter. The word "money", seen in Figure2, shows similar F0 and intensity patterns for both L1 and L2 English. F0 and intensity in the first syllable are higher than second; however, for both F0 and intensity a higher degree of contrast between syllables is found in L1 English. In terms of duration, the degree of contrast for both L1 and L2 English is minimal.

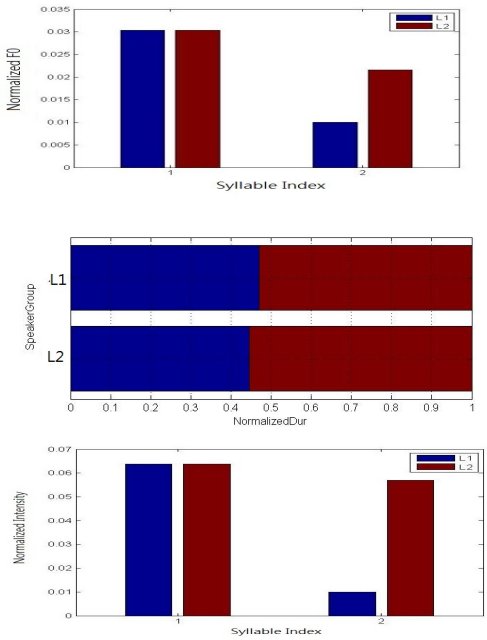


Figure2: *Acoustic patterns of "money" by acoustic parameter and speaker group*

The F0 and intensity pattern of "tomorrow", illustrated in Figure3, shows that L1 English speakers consistently produce the highest F0 and intensity on the second syllable, whereas L2 English speakers realize the same stress contrast using duration only.

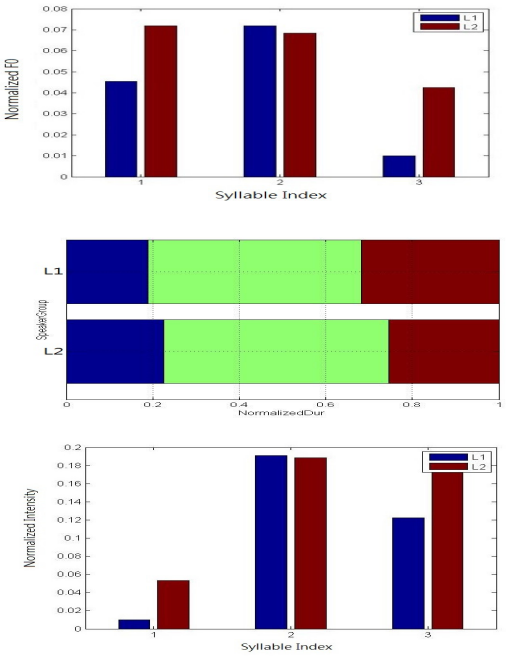
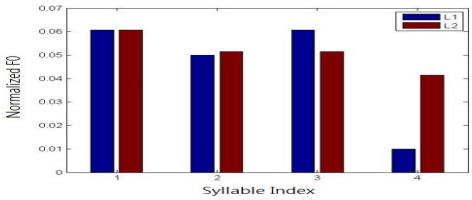


Figure 3: *Acoustic patterns of word "Tomorrow" by feature and speaker group*

In "California", L2 speakers produce a smaller intensity contrast than L1 speakers do. As for F0, only L1 English corresponds with the canonical stress pattern; for L2 speakers, the highest F0 value occurs in first syllable rather than the third. No clear stress contrast patterns in duration were found for either speaker group.



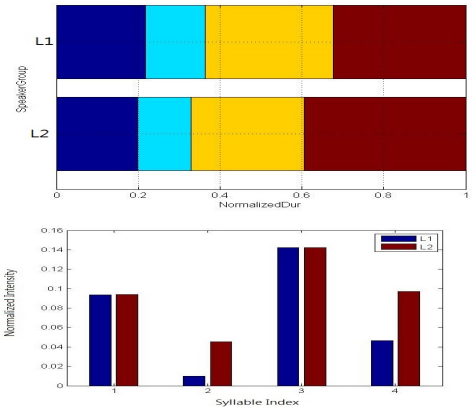


Figure 4: *Acoustic patterns of word "California" by feature and speaker group*

4. Discussion

Overall, results show between-group differences in all three acoustic parameters measured, with the most obvious difference appearing in F0, for which the degree of contrast produced by L1 English speakers is twice that of L2 speakers. Item-based analyses also confirm that F0 is a more stable indicator than duration and intensity to distinguish L1 from L2 speech across lexical items. Thus, underdifferentiation of F0 contrast in realization of lexical stress seems to make a more substantial contribution to Taiwan-accented English than either duration or intensity. Analyses based on syllabicity found that in L1 English, the highest F0 and intensity are always realized on primary-stress syllables, but the difference between primary and secondary stress syllables is not very distinct (e.g. the distinction between the first (secondary-stress) and the third (primary-stress) syllables in the word "California"). In contrast, L2 English patterns of F0 and intensity do not always correspond to canonical stress patterns, and in cases in which they do correspond, the degree of contrast produced is lower than that of L1 English. It is interesting to note that no clear duration patterns were observed for either group in this analysis.

The present study has obtained results similar to those of previous studies [7] in which the L2 English of syllable-timed L1 speakers exhibits substantially less duration contrast in realization of lexical stress, suggesting rhythmic difference is a major prosodic feature. However, PVI analysis of F0, duration and intensity has revealed that stress contrast is realized more robustly by means of F0 than by duration or intensity. Moreover, for the words recorded by the L1 speakers in our experiment, the highest F0 and intensity were always realized on stressed syllables, whereas duration either often exhibited no clear pattern or played a relatively smaller role. In contrast, the Taiwan English speakers' production of F0 and intensity did not always correspond to canonical stress patterns; when they did, however, the L2 speakers realized the contrast less robustly than native speakers did. Interestingly, this speaker group also exhibited no clear patterns with respect to duration.

5. Conclusion

Based on these results, it appears that L1 English speakers produce lexical stress contrasts more robustly than L2 Taiwan English speakers do. Moreover, their pattern of contrast underdifferentiation is realized in terms of F0 and intensity, echoing the pattern found in our study of narrow focus, which suggests that insufficient contrast is a feature of L2 prosody at both the lexical and phrase levels. Using PVI to measure F0 and intensity, in addition to duration, further revealed that F0 and intensity appear to play a larger role than duration in marking English stressed syllables. However, we must note here that that inconsistent stress assignment were found across three different dictionaries for 6 of the 20 words in our task (tomorrow, hospital, video, overnight, misunderstand and Vietnamese). Subsequent studies will investigate whether and how the inconsistencies are realized by both L1 and L2 speakers. Since F0 appears to be the most salient cue of underdifferentiation, future studies will include more refined, syllable-internal analysis of the same words embedded in a variety of intonation contexts in order to examine the effect of layering higher levels of prosodic information on their production. Future research will also investigate the question of whether similar patterns can be found in the L2 English of other syllable-timed and tone language speakers.

6. References

[1] Magen, H.S., "The perception of foreign-accented speech", Journal of Phonetics, vol. 26, 381-400, 1998.

[2] Anderson-Hisieh, J., Johnson, R. and Koehler, K. "The relationship between native speakers judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure", Language Learning 42: 4 529-555, 1992.

[3] Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399– 423.

[4] Tajima, K., Port, R., and Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. Journal of Phonetics, 25, 1-24.

[5] Visceglia, T., Tseng, C. Y., Su, Z. Y. and Huang, C. F. "Realization of English Narrow Focus by L1 English and L1 Taiwan Mandarin Speakers", The 7th International Congress of Phonetic Sciences. Hong Kong, China, 2011.

[6] Visceglia, T., Su, C. Y. and Tseng, C. Y. "Comparison of English Narrow Focus Production by L1 English, Beijing and Taiwan Mandarin Speakers", Oriental COCOSDA 2012 47-51. Macau, China, 2012

[7] Mixdorff, H. and Ingram, J. "Prosodic analysis of foreign-accented English", Proc. Interspeech 2009, 6-10 Sep. Brighton UK, 2009.

[8] Grabe, E. & Low, E. L. "Durational variability in speech and the rhythm class hypothesis ", In Gussenhoven, C. & Warner, N. (eds.) Papers in Laboratory Phonology 7, Berlin, Mouton de Gruyter, 515-546, 2002.

[9] Visceglia, T., Tseng, C. Y., Kondo, M., Meng, H and Sagisaki, Y. "Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project)", Oriental COCOSDA 2009. Beijing, China, 2009.

[10] Asu, E.L. & Nolan, F. "Estonian and English rhythm: a two-dimensional quantification based on syllables and feet", In Proceedings of Speech Prosody 2006, Dresden, Germany, OS1-5_0229, 2006.

[11] Kinoshita, N. and Sheppard, C., "Validating acoustic measures of speech rhythm for second language acquisition", ICPhS XVII, Hong Kong, 1086-1089, 2011.

[12] Tseng, C. Y. Su, Zhao-yu. "Dynamic Discourse Speech Tempo and Phonological Timing", The 7th International Congress of Phonetic Sciences. Hong Kong, China, 2011