Levels of Lexical Stress Contrast in English and their Realization by L1 and L2 Speakers

Chiu-yu Tseng and Chao-yu Su
Phonetics Lab, Institute of Linguistics,
Academia Sinica, Taipei, Taiwan
cytling@sinica.edu.tw

Tanya Visceglia
Education Center for Humanities and Social Sciences,
National Yang Ming University, Taipei, Taiwan

Abstract—Following our previous extended paradigm of Pairwise Variability Index (PVI) to analyze both F0 and duration, we analyze and compare primary, secondary and tertiary lexical stress in English as it is realized by L1 English, L1 Taiwan Mandarin and L1 Beijing Mandarin speakers. Chao's tone letter system was also adapted to describe differences in degree of stress contrast produced by L1 and L2 speakers. 20 multisyllabic English words of varying stress patterns produced by 32 speakers were selected for analysis. Results indicate that: 1) the most acoustically robust stress contrast in L1 English is found between primary stress and secondary/tertiary stress; the distinction between secondary and tertiary stress is not strong enough to be categorical; 2) a substantial difference in degree of contrast realization between L1 and L2 speakers exists with respect to F0, but not duration; 3) words whose stress patterns were consistently transcribed across dictionaries require a five-point scale to distinguish native from non-native speakers' realization of English lexical stress, whereas for inconsistently described words, a three-point scale is sufficient. The results suggest that for L1 speakers, robust contrast of lexical stress/no-stress differentiation is critical, a 2-way contrast represented by a simple three-point scale; whereas more fine-grained secondary/tertiary differentiation, which requires more distinct 3-way contrasts as reflected in a five-point scale is less crucial. L2 speakers' failure to achieve the same differentiation patterns, as exemplified by their production data, is predicted to have a strong effect on the naturalness and intelligibility of L2 speech.

Index Terms: L2 English, lexical stress, tone letter scale, PVI, L2 prosody

I. Introduction

Previous studies of L2 pronunciation have focused primarily on segmental variation between L1 and L2; however, recent studies shown that prosodic errors often produce as much of an effect on the intelligibility, comprehensibility and perceived accentedness of L2 speech as segmental errors do [1, 2, 3]. Research comparing the discourse prosodic organization and speech planning of L1 and L2 English speakers has also demonstrated patterns of difference: L2 speakers use more intermediate chunking units and fewer larger-scale planning units in their prosodic discourse organization, and L1 speakers are able to plan on a larger scale than L2 speakers at every prosodic layer [4]. These studies show that prosody plays an important role in shaping what we perceive to be L2 accent.

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. Research has shown that rightward stress shift and stress shift unaccompanied by a change in vowel quality were both found to have a strong effect on intelligibility for both native and non-native groups [5]. In addition to misplacement of lexical stress, our recent study found that even when placement of lexical stress is correct, contrast underdifferentiation of lexical stress is one of the key features that distinguish L1 and L2 speakers' production of English lexical prosody [6]. Using PVI as an indicator of rhythmic contrast between adjacent phonological units, recent research has revealed L2 underdifferentiation of duration between stressed and unstressed syllables [7].

Performing cross-feature analysis with the same indicator, we adapted and extended PVI to measure F0 and intensity as well. Results show a general pattern of contrast underdifferentiation in L2 speakers with respect to F0, duration and intensity, with the most striking difference found in F0 [8].

The following study presents F0 and duration analysis of lexical stress as it is realized by L1 English and L1 Taiwan-Mandarin English speakers, using speech data extracted from the AESOP-ILAS corpus (Asian English Speech cOrpus Project, Institute of Linguistics, Academia Sinica). It focuses on differences between L1 and L2 realization of phonological contrast and categorization by examining F0 and duration within 3 levels of stress (primary, secondary and tertiary), with the aim of investigating (1) how phonological contrast and stress level categorization are reflected in the acoustic features produced by L1 and L2 speakers (2) whether a degree difference between L1 and L2 speakers exists with respect to stress contrast and categorization (3) what the simplest scale is that we can use to distinguish between L1 and L2, while still maintaining the phonological contrasts and categorizations of L1. In order to determine the simplest scale, Chao's tone letter scale [9] was adapted and extended to describe stress levels and the degree to which an acoustic feature is realized.

II. METHOD

2.1 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 multi-L1 speech corpora to represent the varieties of English spoken in Asia [10]. This experiment uses materials from 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. Examining the phonetic transcription of those words in dictionaries, we found that 6 target words out of 20 are transcribed with different stress patterns across dictionaries. Based on this observation, we classified the 20 words into 2 groups: words with varying stress transcriptions across dictionaries and words with consistent stress transcriptions. The 6 inconsistent words are listed below:

- **TO - MORR – OW [132, 131, 132, 131]**
- **VI - DE – O [312, 311, 312]**
- **HOS - PI - TAL [321, 311, 321 , 321]**
- **VI - ET - NA – MESE [1213, 213, 2113, 213, 1113]**
- **MI - SUN - DER – STAND [2213, 1213, 2113, 1213]**
- **O- VER – NIGHT [213, 313, 313]**

Primary/secondary/tertiary stresses are represented by the numbers 3/2/1. Green, blue and black indicate stress patterns as transcribed in the online Merriam-Webster, CMU, and Cambridge dictionaries, respectively. Preliminary observation of the 6 words with various stress transcription revealed that almost all cases of inconsistency involve ambiguity between placement of secondary and tertiary stress. Ambiguity between primary and secondary stress was found only in the stress shift represented by the adjectival and adverbial uses of **OVERNIGHT.**

2.2 Participants and procedure

A total of 32 speakers: 9 L1 North American English speakers (4 male and 5 female), 20 Taiwan L2 speakers (10 male and 10 female) and 3 Beijing L2 speakers (1 male and 2 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.

2.3 Data Analysis

All data were preprocessed automatically for segmental alignment using the HTK Toolkit, which was then manually spot-checked by trained transcribers for accuracy. F0 values were extracted and measured using a semitone scale. Duration extraction values were normalized to remove the effect of inherent segmental duration and boundary effects using the multi-layered normalization algorithm that appears below [11], 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 εi represents all other unpredictable values.

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

III. RESULTS

3.1 L1/L2 comparison by contrast degree

The following analyses reveal discrimination patterns across stress levels and speaker groups. Figure 1 below shows prosodic patterns by acoustic correlate, stress level and speaker group. In L1 speakers, primary stress exhibits a distinctive contrast across 3 stress levels for both F0 and duration; however, the contrast between secondary and tertiary stress is not nearly as clear. For L2 speakers, F0 shows a lack of contrast degree, but duration patterns are similar to those produced by L1speakers.
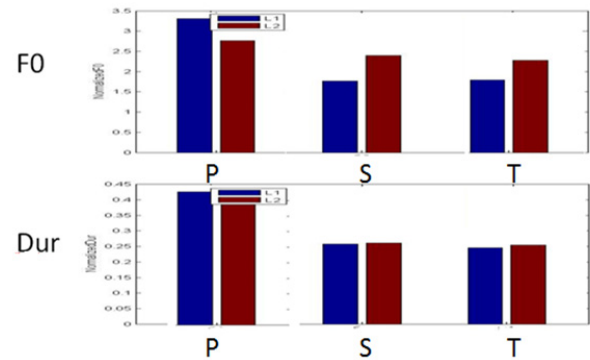

Figure 1: Prosodic patterns by acoustic correlate, stress level (Primary, Secondary and Tertiary) and speaker group.

Based on the data reported above, it appears that phonological category and stress-level contrasts are reflected in F0 and duration differences. In L1 English, the most obvious stress contrast is found between primary stress

and secondary/tertiary stress; the distinction between secondary and tertiary is not strong enough to be categorical. These results are consistent with the observation made in section 2.1: stress transcription ambiguity occurs most often in labeling secondary and tertiary stress. Comparing the degree of contrast made by L1 and L2 speakers between primary stress and secondary/tertiary stress, we found F0, but not duration, to exhibit a degree difference. So it seems as if L2 speakers are using duration to distinguish stress in a native-like manner, but are unable to do the same with F0.

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

Using PVI as a measure of contrast, these data show detailed stress contrasts between adjacent phonological units (syllables, in this case). PVI cross-feature comparison demonstrates that F0 exhibits a larger L1/L2 ratio than duration or intensity; thus, F0 is more significant and stable indicator than duration and intensity for differentiating L1 and Taiwan L2 speech across lexical items [8].
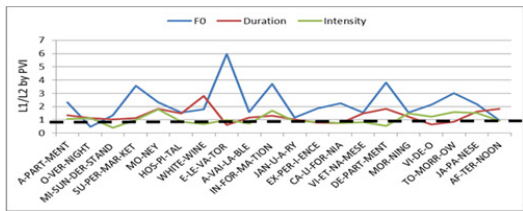

Figure2. L1/L2 ratio of PVI by acoustic parameter and word

Results show that underdifferentiation of lexical stress contrasts across adjacent phonological units in Taiwan L2 speakers is primarily reflected in L2 speakers' failure to produce an F0 height contrast between stressed and unstressed syllables.

3.3 Feature categorization by multiple scales-words with consistent stress transcriptions

This section examines the 14 words with consistent stress transcriptions using an adapted tone-letter scale to determine the simplest scale which can sufficiently distinguish between L1 and L2 production while still maintaining L1 phonological contrasts and categories. F0 and duration values are categorized using numerals from 5 to 10 to represent the relative relationship among primary, secondary and tertiary stress. For F0, L1 speakers produce a high degree of contrast across stress levels, in contrast with both L2 (T) and L2 (B). The F0 register of L2 (B) is also higher than that of L1 and L2 (T) speakers. As for duration, degree of contrast among the three groups is similar, which does not allow for clear discrimination among speaker groups.
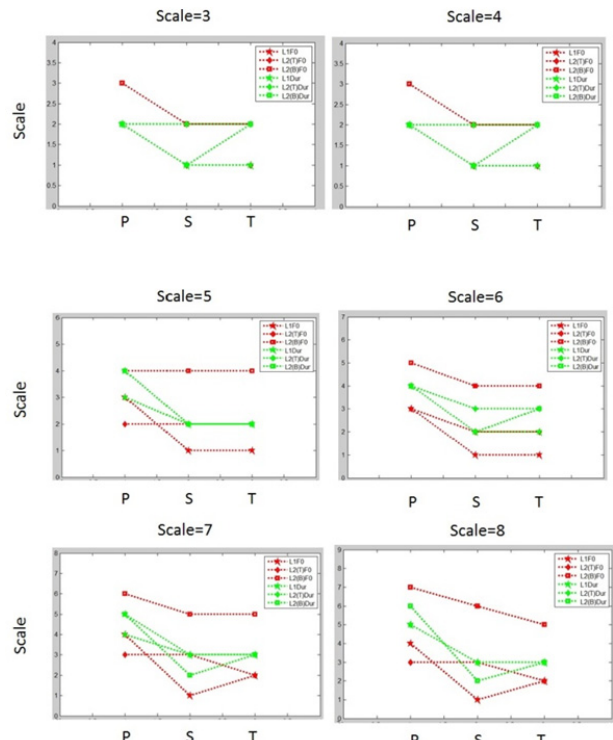

Figure 3: F0/duration categorization using multiple scales and stress levels for words with consistent stress transcriptions.

Figure 3 shows that although scales with more than five points provide more detailed information a finer grain of analysis does not improve discrimination accuracy. Scales of fewer than five points, in contrast, fail to tease apart differences between L1 and L2 production. Thus, it appears that a five-point scale is sufficient to discriminate between native and non-native speakers' realization of English lexical stress.

3.4 Feature categorization by multiple scales-words with inconsistent stress trascrptions

This section examines the 6 words with inconsistent secondary/tertiary stress transcriptions across dictionaries. With respect to F0, L1 speakers exhibit a greater degree of contrast across stress levels. Both L2 (T) and L2 (B) fail to produce such a contrast; for some words, values across the three levels are equivalent. The F0 register of the L2 (B) speaker is higher than that of the L1 and L2 (T) speakers.
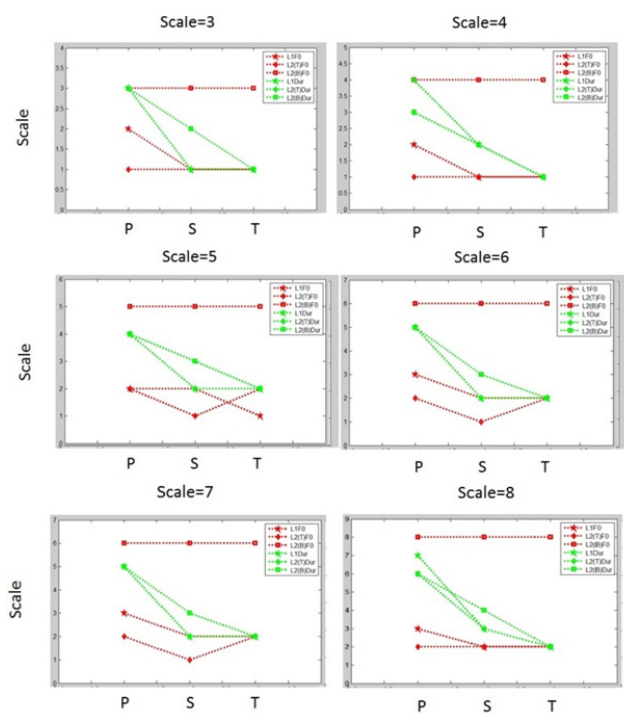
Figure 4: F0/duration categorization by multiple scales and stress levels for words with inconsistent stress transcriptions.

It can be seen in Figure 4 that no phonological contrast by stress level category is found in many L2 English tokens of inconsistently described words. As L1 English speakers do produce a contrast in these cases, a 3-level system is sufficient to discriminate between native and non-native speakers.

IV. DISCUSSION

In a previous study of the interaction between lexical and sentence in Taiwan L2 English, we have shown that interaction with higher levels of prosody diffuses lexical-level contrasts for both L1 and L2 speakers. Production of the prosodic cues to mark lexical stress (F0, duration and amplitude) becomes much more difficult for L2 speakers when disyllabic and multisyllabic words are embedded in higher-level prosodic contexts. F0 and duration contrasts clearly maintained by L2 speakers in carrier sentences were almost completely washed out when those words were overlaid with prosodic boundaries or contrastive focus cues [6]. The data from the present study further indicate that the most robustly realized stress contrast in L1 English is found between primary stress and secondary/tertiary stress, and that the distinction between secondary and tertiary stress is not strong enough to be considered categorical. As for realization of the contrast between primary and secondary/tertiary stress, we found a difference in degree between L1 and L2 speakers with respect to F0, but not

duration. Thus, F0 is a more reliable cue than duration to discriminate the lexical stress contrasts produced by L1 and L2 speakers, and the underdifferentiation of lexical stress contrasts across adjacent phonological units observed in Taiwan L2 speakers is more strongly reflected in their production of less distinct F0 contrast, compared with duration and intensity. The present results differ from a comparative study of Australian English produced by Vietnamese learners; the Vietnamese learners produce higher F0 contrast and longer duration patterns across the board [7]. These differences demonstrate how prosody underdifferention in L2 English is language-specific and why foreign accents are distinct.

With respect to the two groups of lexical items, the following patterns were found for words with consistent stress transcriptions: (1) L1 speakers produce a larger degree of contrast across stress levels. (2) Both L2 (T) and L2 (B) show lack of contrast degree. (3) The F0 register of L2 (B) is higher than that of L1 and L2 (T) speakers. As for the 6 words with inconsistent transcriptions, in some cases, L2 English speakers produced no stress contrast at all, which may be attributable to insufficiently clear modeling by native speakers. As for developing a scale to capture differences in degree of contrast across speaker groups, for consistently transcribed words, a five-point scale is necessary to discriminate between native and non-native speakers' realization of English lexical stress, whereas for inconsistently described words, a three-point scale suffices.

V. CONCLUSION

These results suggest that for L1 speakers, categorical contrast of lexical stress/no-stress differentiation is critical, while the less categorical secondary/tertiary differentiation leaves room for a wider range of variation in pronunciation. To maintain the 3-way primary/secondary/tertiary contrast, more fine-grained differentiation pattern is required. Our L2 data suggest that robust realization of differentiation is no easy task for L2 speakers. Less robust realization of necessary stress contrasts is consistent with the use of more intermediate chunking units and fewer larger-scale planning units in their prosodic discourse organization. Collectively, these features of L2 speech suggest that one of its major features is production of less distinct levels of contrast, which results in underdifferentiation. Highly varied underdifferentiation is therefore a major feature which contributes to the naturalness and, in some cases, the intelligibility of L2 English.

VI. REFERENCES

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

[2] Munro, M. J., & Derwing, T. M. (1994). "Evaluations of foreign accent in extemporaneous and read material", Language Testing, 11, 253–266

[3] Baker, W., & Trofimovich, P. (2006). "Perceptual paths to accurate production of L2 vowels: The role of individual differences", International Review of Applied Linguistics in Language Teaching (IRAL), 44, 231-250.

[4] Visceglia, Tanya, Tseng, Chiu-yu, Su, Zhao-yu and Huang, Chi-Feng (2010). "Discourse Prosody Planning in L1 and L2 English." Oriental COCOSDA 2010, (Nov 24-25, 2010), Kathmandu, Nepal. 6 pages.

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

[6] Visceglia, T., Tseng, C, Su, Z. and Huang, C. (2010) "Interaction of Lexical and Sentence Prosody in Taiwan L2 English" SLaTE Workshop, Interspeech 2010. Tokyo.

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

[8] Tseng, C, Su, C-Y and Visceglia, T. (2013). "Underdifferentiation of English Lexical Stress Contrasts by L2 Taiwan Speakers", SLATE 2013, Grenoble France.

[9] Chao, Yuen-Ren (1930), "A system of tone-letters", Le Maître Phonétique 45: 24–27

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

[11] Tseng, C, and Su, C-Y (2013). "Temporal characteristics of emphasis in continuous speech", Speech prosody 2012, Shanghai, China