Interaction of Lexical and Sentence Prosody in Taiwan L2 English

Tanya Visceglia[1,] *Chiu-yu Tseng*[2], *Zhao-yu Su*[2] *and Chi-Feng Huang*[2]

1. Department of Applied English, Ming Chuan University
2. Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan

cytling@sinica.edu.tw

Abstract

This study investigates the effect of sentence-level prosody on production of English lexical stress, comparing L1 English and L1 Taiwan Mandarin speaker groups. 4 L1 North American English speakers and 9 L1 Taiwan Mandarin speakers were asked to produce a set of 20 disyllabic and multisyllabic words embedded in three different prosodic contexts: neutral broad focus, at a phrase/sentence boundary, and in narrow focus. Results suggest that 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.

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

1. Introduction

In considering questions of L2 pronunciation, the language teaching and learning community has shifted its focus away from accentedness, defined as how different a speaker's pronunciation is perceived to be from that of the L1 community, toward intelligibility, which has been broadly defined as "the extent to which a speaker's message is actually understood by a listener." [1] This shift has occurred for two reasons: first, many studies have demonstrated that L2 speech does not necessarily become less intelligible as a result of being different from native pronunciation. Many studies demonstrate no correlation between global accent ratings and level of overall intelligibility [2]. Second, the majority of English speakers in the world today are either ESL or EFL speakers engaged in communication with other ESL or EFL speakers, which suggests the need for an international and flexible set of phonological standards, rather than a single rigidly defined ENL norm [3]. Thus, investigations of ESL and EFL phonological variation would more productively focus on areas which can be demonstrated to influence intelligibility of individual words and speakers' intended meanings across a wide range of listener groups.

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 [4] 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. [5] 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.

Naïve and expert listeners have identified word stress, sentence stress and sentence intonation as three of the major factors affecting the overall comprehensibility of L1 Mandarin English [6] (Comprehensibility has been defined as the listener's perceived level of processing difficulty [7]). Yet a recent study of Mandarin speakers' production of lexical stress found that Mandarin speakers were able to approximate English-like patterns of duration, intensity and some F0 patterns [8]. The investigators suggested that the source of Mandarin speakers' greatest difficulty in production of lexical stress is target-like production of vowel reduction.

The experimental task in the aforementioned study involved use of disyllabic target words embedded in a fixed position in carrier sentences. No production study of Mandarin L2 English has yet used words representing a range of syllabicities and stress types uttered in neutral, carrier sentences and compared them with those same words embedded in a range of utterance-level prosodic contours such as narrow focus, utterance-final fall, continuation rise, or interrogative final rise. The experiment presented here investigates the effects of overlaying higher-level prosodic information, i.e. realization of boundary or narrow focus, on the production of lexical stress contrasts. Our results suggest that although L2 speakers can maintain the phonetic contrast between stressed and unstressed syllables when production of no additional prosodic events is required of them, production of lexical stress becomes much more difficult when disyllabic and multisyllabic words are embedded in higher-level prosodic contexts.

2. Procedure

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]. The materials, recording platform, and recording protocol manual (which includes guidelines for recording setup, hardware specifications, and a detailed set of recording instructions for both the proctor and the speaker) were developed in a collaborative effort by AESOP members in Taiwan, Japan and Hong Kong. In these materials, 2-, 3- and 4-syllable target words of all possible stress patterns were embedded in a fixed, sentence-medial position in carrier sentences and also in the following prosodic contexts: (1) at phrase boundaries in yes-no questions, wh-questions, and declarative sentences, (2) at minor phrase boundaries and (3) in narrow-focus positions. The purpose of this design was to investigate whether the competing demands of higher-level prosody, such as sentence-level illocution or production of narrow focus would affect production of the acoustic correlates of lexical stress, namely duration, F0 and intensity (Vowel quality/reduction will be considered separately in future research). Tokens of each word type were chosen from the CMU Electronic Dictionary [10] based on overall frequency of occurrence (database calculation) and level of familiarity to L2 speakers (piloted); two tokens of each stress type appear in each illocutionary condition (e.g. declarative fall, continuation rise). Two levels of stress are differentiated in our materials: primary stress and no stress. Syllables receiving secondary stress (such as the “for” in “information) were put into the same category as unstressed syllables. This decision was based on data from previous studies of running speech, which suggest that phonetic differences in stress, particularly differences in vowel duration and amplitude, are realized on primary stress and pitch-accented syllables when multisyllabic words are embedded in sentences [11,12].

In Task 1, each target word appears in a carrier sentence, two syllables from any phrase boundary (target words appear in boldface), e.g. “I said available five times.” Task 2 embeds target words in four prosodic boundary positions: 1) the final fall of a wh-question (ex. Where is the elevator?); 2) the final rise of a yes-no question (ex. Do you need any money?); 3) the continuation rise found in multiple-clause sentences and 4) the final fall in declarative sentences (e.g. When Sue left this evening for California, she said she would call me tomorrow). In Task 3, each target word appears as the subject of contrastive focus (e.g. “I said I want to go to the hospital, not the airport.”). Appendix A lists all 20 target words used in this study by number of syllables and stress pattern. Space considerations prevent us from including the materials used in Tasks 2 and 3.

Participants were recruited on university campuses in Taiwan. The four L1 speakers (2 male, 2 female) are instructors in the Department of Applied English at Ming Chuan University and native speakers of North American English. The nine L2 speakers (5 male, 4 female) are native speakers of Taiwan Mandarin and graduate students who have received at least ten years of English instruction. Most L2 speakers also have some knowledge of Taiwanese.

Speech data 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.

Data Analysis

55 English utterances from 4 L1 North American English and 9 L1 Taiwan Mandarin speakers were selected for analysis (total: 220 L1 North American and 495 L2 Taiwan English utterances). Speech tokens were sampled at a rate of 16kHz with a quantization of 16 bits. All data were pre-processed for segmental labeling using the phone sets from the CMU electronic dictionary [10] then manually spot-checked by trained transcribers for accuracy of segmental alignment.

Speech rate, syllable duration, average F0 and intensity of the target words in each of the three experimental conditions were derived for the purpose of comparison between L1 and L2 speaker groups. Two normalization methods were developed to remove features which we believed to be likely to interact with the features under observation: [1] position of the stressed syllable within the target word and within the phrase; [2] number of phones within the stressed syllable. Position of the stressed syllable is likely to interact with duration in the sense that utterance-final syllables tend to undergo a lengthening effect and likely to interact with pitch in the sense that utterance-final syllables will carry some form of illocutionary prosody. As for number of phones per syllable, this feature was found to strongly correlate with syllable duration for both L1 and L2 speakers. Therefore, normalization of those features prior to analysis was essential to obtaining accurate between-group comparisons. For a detailed description of the multi-layered normalization algorithms used in this paper, see [13]

3. Results

3.1. F0 height contrast between stressed and unstressed syllables

Figure 1 shows average F0 height on stressed and unstressed syllables across a range of three prosodic contexts and two speaker groups. In general, L1 speakers produce more pronounced F0 contrasts than L2 speakers do.

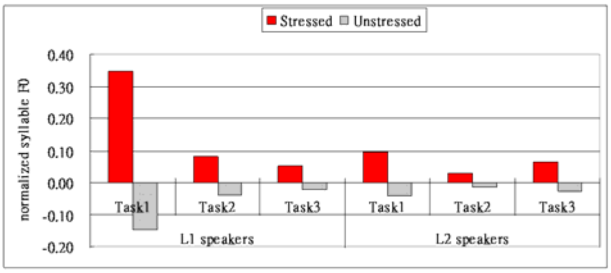

Figure 1 (the units of the y axis values represent normalized F0 ; they are expressed in terms of relative proportion)

We see that the contrast between stressed and unstressed syllables is most pronounced for both speaker groups in Task 1, in which words are produced in carrier sentences. L1 speakers realized this distinction much more clearly: the difference in F0 between stressed and unstressed syllables for the L1 group was 50%, whereas for the L2 group, it was only 14%. The contrast weakens considerably for both groups in Tasks 2 and 3, in which words are produced at phrase boundaries and in narrow focus contexts. Nevertheless, we see that the contrast is largely maintained by L1 speakers in Task 2, but not by L2 speakers. The difference in F0 between stressed and unstressed syllables for L1 speakers was 12%, whereas for L2 speakers, it was only 4%). In Task 3, in contrast, L2 speakers appear to produce a larger F0 distinction than L1 speakers do. Possible interpretations for this finding will be discussed in Section 5.

3.2. Duration contrast between stressed and unstressed syllables across speaker groups

Figure 2 provides comparisons of stressed and unstressed syllable duration across speaker groups and three prosodic positions.

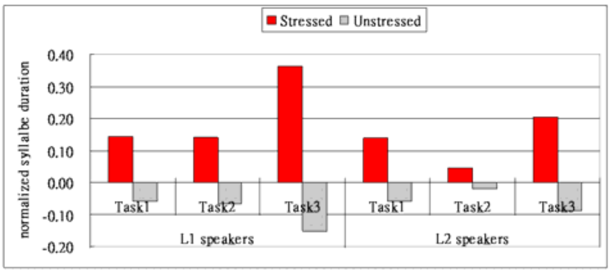

Figure 2 (the units of the y axis values represent normalized duration; they are expressed in terms of relative proportion)

For task 1, L1 and L2 speaker groups appear to make exactly the same level of contrast (20%). In Task 2, however, the contrast is preserved by L1 speakers, but neutralized by L2 speakers. The difference in duration between stressed and unstressed syllables for L1 speakers was 21%, whereas it was only 6% for L2 speakers. The between-group difference in syllable duration is most pronounced in Task 3, in which L1 speakers exhibited a difference of 51% and L2 speakers only 30%. Possible interpretations will be discussed in Section 5.

3.3. Intensity contrast between stressed and unstressed syllables

In Task 1, we see that although L1 and L2 speaker groups both realized intensity contrasts between stressed and unstressed syllables, L1 speakers' difference was twice that of L2 (L1 33% vs. L2 14%). In Task 2, however, L2 speakers exhibited a much larger contrast than L1 speakers did (L1 21%, L2 40%). In Task 3, both groups appeared to maintain approximately the same level of intensity contrast (L1 10%, L2 13%). Possible interpretations will be discussed in Section 5.

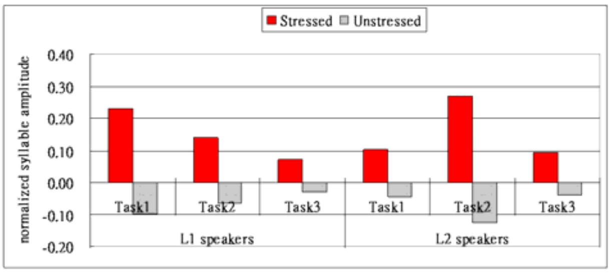

Figure 3(the units of the y axis values represent normalized amplitude [db.] ;they are expressed in terms of relative proportion)

4. Discussion

Overall, the results presented in Section 4 demonstrate that adding the processing demand of encoding higher-level prosodic information will influence both L1 and L2 speakers' realization of lexical stress. L1 speakers, however, maintain lexical stress contrasts in sentence context, whereas L2 speakers do not. Data obtained from target word tokens in Tasks 1, 2 and 3 will be discussed separately in Sections 5.1-5.3.

4.1. Task 1: target words in carrier sentences

F0 contrast is most pronounced for both speaker groups on target words embedded in carrier sentences, L1 speakers, however, demonstrate a substantially larger contrast than L2 speakers do. Duration contrasts are also maintained by both speaker groups, to approximately the same extent. Intensity contrasts are much more strongly realized by L1 speakers than L2 speakers in this context. Our results may appear to confirm previous [8] findings with respect to duration, but not intensity and F0. However, the differences between our results and those presented in [8] can largely be attributed to principled methodological differences in both measurement and comparison. We have chosen to measure each parameter using a logarithmic scale, and to present comparisons in ratio form rather than comparing absolute values in Hz or ms. in order to better capture the relative perceptual salience of each prosodic cue.

4.2. Task 2: target words at phrase boundaries

F0 contrast is largely preserved by L1 speakers in Task 2 but not by L2 speakers, which suggests that overlaying an utterance-level contour onto a phrase final word will make lexical prosody more difficult for L2 speakers to produce. Duration differences can also be observed between speaker groups, but these are weakened by the presence of a boundary, most likely as a result of phrase or utterance final lengthening [14]. Intensity differences were not found either within or across groups.

4.3. Task 3: target words in contrastive focus

In Task 3, syllable average F0 measurement suggests that both groups have neutralized the stressed/unstressed syllable contrast. In the case of L1 speakers', this could be attributed to their realization of the contrastive focus (scooped) L+H* pitch accent on the stressed syllable of narrow focus constituents [15], which was not observed in our L2 speakers' narrow-focus utterances. Since this pitch accent is realized as a fall followed by a rise for L1 English speakers, F0 average of rise and fall would yield a mid-range F0 value. Duration results for Task 3 provide support for the presence of a scooped accent: L1 speakers' stressed syllables were 51% longer than their unstressed counterparts. This may be due to the presence of a contour pitch accent to realize narrow focus, whose rise and fall is likely to take longer to realize than a level or directional accent. Contrasts in intensity values were not maintained by either speaker group in this condition.

In future research, we plan to investigate the syllable-internal pitch contours on narrow-focus and phrase-boundary tokens of target words in order to account for the duration and F0 anomalies described in Section 5. We also plan to perform comparisons of F0 range and number of pitch accents per utterance. Mixdorff and Ingram (2009) demonstrated that prosodic characteristics of L1 Vietnamese English include production of a wider F0 range than exhibited by L1 speakers and a larger number of pitch accents per utterance [16]. They also investigated the effect of syllable timing in Vietnamese on L2 English using a normalized pair-wise intervariability test, which compares consecutive syllable durations within an utterance. Less variability was found in the consecutive syllable duration of L1 Vietnamese utterances than L1 English utterances. Like Vietnamese, Taiwan Mandarin is a syllable-timed tone language, so we plan to replicate the analyses performed in [16] for the purpose of comparing L1 Vietnamese and L1 Taiwan Mandarin data.

These data will be used to conduct perception studies as well, which will test the relative intelligibility of tokens extracted from all three tasks by implementing them into transcription tests to determine the effect of stress neutralization on word identification, and to investigate how non-target realization of lexical stress may affect overall intelligibility of L2 English words. Separate tasks will include tokens appearing in the three contexts, both individually and embedded in sentences, in order to determine to what extent top-down information can be used to compensate for differences in pronunciation of individual words.

Plans for future resynthesis studies include systematic alteration of F0 and duration cues to determine their individual and cumulative effects on intelligibility. For that study, L1 Taiwan Mandarin speaker tokens extracted from all three conditions will be re-synthesized to match the F0 and duration characteristics of the same tokens recorded by L1 English speakers. Tokens recorded by L1 English speakers will be re-synthesized to match the F0 and duration characteristics of those produced by L1 Mandarin speakers for comparison.

5. Conclusion

The data presented here strongly suggest that interaction with higher levels of prosody diffuses lexical-level contrasts for both L1 and L2 speakers. In the case of L1 speakers, lexical stress contrasts are nevertheless maintained across prosodic contexts, whereas for L2 speakers, simultaneous production of lexical and sentence level prosodic cues appears to be much more challenging. 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. Intensity contrasts were not maintained in sentence prosody for either group, which confirms previous findings that in continuous speech, intensity is not often manipulated over units of speech as small as the syllable [16].

These data suggest that production of the prosodic cues to mark lexical stress (F0, duration and amplitude) presents more of a challenge to L2 speakers when disyllabic and multisyllabic words are embedded in higher-level prosodic contexts. Research has only recently begun to examine the effects on L2 speakers of the additional processing demands created by simultaneous production of lexical, utterance and discourse-level prosody in continuous speech. We believe that these interactions have an impact on L2 speakers' intelligibility and comprehensibility. Future work will include more detailed investigation of the interaction between lexical stress and higher levels of prosodic information in L2 speech.

6. Appendix

Target words by syllabicity and stress type

2-1. **mo**ney; **mor**ning
3-1. **vi**deo; **ho**spital **3**-2. a**part**ment; to**mo**rrow **3**-3. over**night**; Japa**nese**
4-1. **e**levator; **Ja**nuary **4**-2. a**vail**able; ex**pe**rience **4**-3. infor**ma**tion; Cali**for**nia **4**-4. misunder**stand**; Vietna**mese**
LH. **Su**permarket; de**part**ment
RH. White **wine**; after**noon**

7. References

[1] Munro, M. and Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73-97.

[2] Munro, M. J. (2008). Foreign accent and speech intelligibility. In Hansen Edwards, J. G. & Zampini, M. L. (Eds.). *Phonology and Second Language Acquisition* (pp. 193-218). Amsterdam: John Benjamins.

[3] Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83–103.

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

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

[6] Warren, P., Elgort, I., and Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. Language Learning & Technology, 13(3), 87-102.

[7] Derwing, T. M. and Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*, 1-16.

[8] Zhang, Y. Nissen, S. and Francis, A. (2008) Acoustic characteristics of English lexical stress produced by native Mandarin speakers J. Acoust. Soc. Am. Volume 123, Issue 6, pp. 4498-4513.

[9] Meng, H., Tseng, C., Kondo, M., Harrison, A. and Visceglia, T., "Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper" Interspeech 2009 1715-1718. Brighton.

[10] http://www.speech.cs.cmu.edu/cgi-bin/cmudict#about

[11] deJong, K. (2004) Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. Journal of Phonetics (32) 493-516.

[12] Wang, C. and Seneff, S. (2001) Lexical Stress Modeling for Improved Speech Recognition of Spontaneous Telephone Speech in the JUPITER Domain. EUROSPEECH 2001 September 2-7 2001, Aarlborg, Denmark.

[13] Tseng, C., Pin, S., Lee, Y., Wang, H. and Chen, C. 2005. Fluent speech prosody: Framework and modeling, Speech Communication (Special Issue on Quantitative Prosody Modeling for Natural Speech Description and Generation), Vol. 46: 3-4, pp. 284-309.

[14] Beckman, M.E, and Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston and M.E. Beckman (eds.) Papers in Laboratory Phonology I: Between the grammar and the physics of speech. Cambridge: Cambridge University Press.

[15] Silverman, K., Beckman M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). TOBI: A standard for labeling English prosody. In International Conference on Speech and Language Processing (ICSLP), volume 2, 867-870.

[16] Mixdorff, H. and Ingram, J. Prosodic analysis of foreign-accented English. Interspeech 2009 1715-1718. Brighton.