

Prosodic Differences between Taiwanese L2 and North American L1 speakers— Under-differentiation of Lexical Stress

Chiu-yu Tseng & Chao-yu Su

Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan

cytling@sinica.edu.tw

Abstract

Assuming that categorical differentiation is major acoustic characteristics of English lexical stress through binary instead of more complex 3-way distinction, we investigated lexical stress in broad and narrow focus positions and found how binary distinction is achieved by the concomitancy of secondary stress defined by its position and distance in relation to primary stress. Similar results are found in broad (sentence initial) and narrow focus as well. These results suggest that binary categorical contrast is the optimal choice while differentiation is dependent on robust contrast patterns in the speech signal. Comparison between Taiwanese L2 and North American L1 speakers revealed how L2 speakers' realization of the binary opposition is of a lesser degree. The results explain why L2 speech is less differentiable and not as expressive.

Index Terms: English stress, primary, secondary, binary, contrast pattern, differentiation

1. Introduction

Earlier studies of L2 or foreign accents that concentrated mainly on segmental features [1, 2, 3], however, more recent shift to prosodic features has led to the discovery that prosodic deviations have as much an effect on the intelligibility and comprehensibility of L2 speech, and contribute significantly to perceived foreign accent as well. For example, Field [4] showed that shift of lexical stress has a strong effect on the intelligibility of native vs. non-native speech group. Mixdorff et al. [5] further showed how the speech rhythm in L2 Vietnamese (tone and syllable-timed) and Japanese (pitch accent and mora-timed) Australian English differed from L1 speakers; both Vietnamese and Japanese speakers produce longer and more equal syllable durations than Australian English speakers. However, we see instead the more even syllable duration of L2 speech not as a rate issue but as lack of the required long/short contrast for categorical stress differentiation. Our hypothesis is that patterns of robust contrast in the speech signal are directly correlated to linguistic categorical contrasts, while lack of or under-differentiation is a major feature of L2 speech.

In the case of English lexical stress, while it is necessary for L2 speakers to learn where the stressed syllable of a word is, it is as important for them also to learn how to maintain the contrast patterns between stressed/unstressed syllables. In other words, even when the correct syllable is stressed, insufficient contrast degree would still result in less differentiable perception and impair intelligibility. Based on the rationale of contrast robustness, we studied the contrast patterns of English lexical stress of all three acoustic correlates the F0, duration and amplitude between English stressed/unstressed syllables produced by L1 American vs. L2 Taiwan Mandarin (TM) speakers and found that in L1 English

the most significant contrast is in F0 (pitch contrast), not duration (rhythmic contrast). And as expected, contrast by lesser degree is found in both F0 and amplitude in TM L2 English. TM L2 speakers were able to maintain similar rhythmic contrast as L1 speakers do but still sounds flatter and foreign due to lack of pitch contrast [6]. Our results of TM L2 speakers differ considerably from Vietnamese and Japanese L2 speakers in [5], thus reduces possible generalization of how syllable-timed L1 may affect L2 English rhythm in general.

In a subsequent study of stress contrast, we further discovered that the 3-way primary/secondary/tertiary stress contrasts as English lexical stress is defined were not found in both L1 and L2 speech [7]. Instead, significant difference is only found in a 2-way contrast between stressed/unstressed syllables in both speaker groups, and again TM L2 speech exhibited less degree of contrasts. In addition, we found no significant contrast between secondary and tertiary stress across L1 and L2 speech and further discovered that 6 of the 20 tested words differ in where the secondary stress should be in three dictionaries consulted. Nevertheless, our results do suggest that 2-way contrast seem most optimal. This has lead us to further hypothesize that (1) in forming the optimal 2-way contrast the role of secondary stress is a concomitant one; its varied realization a surface phenomenon and should predictable. (2) The same rationale of maintaining the optimal 2-way contrast can also be applied to other prosodic categories such as broad sentential focus and narrow focus induced by context or syntactic structure. Prosodic contrasts of larger sized units are even coarser ones to facilitate long distance prediction, providing contextual and pragmatic information that distinguishes speech from text most significantly.

In the following study, we will analyze English secondary stress under the assumption that its concomitancy is dependent on two factors: (1) its linear order (before or after) the primary stress and its distance from it as well. Namely, if a secondary stress appears immediately BEFORE the primary one it is likely to be assimilated to the target primary stress. However, if it appears AFTER the primary stress then it is likely to be assimilated to the following tertiary stress and be reduced to a lower level in order to create the robust contrast patterns. (2) However, if there is more distance between the primary and secondary stress, such as BEFORE but intercepted by a tertiary stress, then the secondary stress stands more chance to be more differentiable from either the primary or the tertiary counterparts.

In the following analysis, we will compare English secondary stress in different positions in a word, in sentence/broad focus position and in narrow focus position in relation to categorical stress differentiation.

2. Method

2.1. Speech Materials

A subset of the AESOP-ILAS (Asian English Speech cOrpus Project—Institute of Linguistics Academia Sinica) corpus was

used for the present study. AESOP is a multinational collaboration whose aim is to build up English speech corpora across Asia that would represent the varieties of English spoken in that region while ILAS is part of the consortium that specifically collects L2 English of Mandarin L1 speakers in Taiwan. The materials used here are 20 frequently used words from 2-, 3- and 4-syllables categorized according to syllabicity and stress type: (1) 2-syllable initial stress 2, (2) 3-syllable initial stress, (3) 3-syllable medial stress, (4) 3-syllable final stress, (5) 4-syllable initial stress, (6) 4-syllable medial 1 stress, (7) 4-syllable medial 2 stress, (8) 4-syllable final stress, (9) left-headed compounds (e.g. orange juice), (10) right-headed compounds (e.g. afternoon). The chosen words are money, morning, white wine, hospital, apartment, department, tomorrow, video, overnight, January, supermarket, elevator, available, Japanese, afternoon, misunderstand, information, experience, California and Vietnamese. These words are then embedded in two conditions: (1) in a fixed sentence-medial broad-focus position two words removed from any phrase boundary, i.e., “I said OVERNIGHT five times.” for the purpose of baseline comparison as well as broad focus. (2) As elicited narrow focus to create phrasal and sentential prominence in broad and narrow focus positions. For example, *Context: Will 3-day delivery be fast enough? Reply: “No. We need OVERNIGHT delivery”* where the provided context requires the answer to disambiguate. As illustrated, the same target word in the previous broad focus position would now re-appear as narrow focus. At the same time, the sentence-initial word “we” may receive sentential prominence, thus providing both narrow focus and sentential prominence in the same sentence.

Speech data were recorded by trained proctors in quiet rooms directly into a laptop computer, using a recording platform developed specifically for AESOP. Experimental sentences and context were preloaded and appeared 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. The speech data of a total of 25 speakers were analyzed: 11 L1 North American English speakers (5 male and 6 female), 16 Taiwan L2 speakers (8 male and 8 female)

2.2. Data Analysis

Prosodic contrast is presented by F0, duration and intensity using Z-score normalization by each sentence first. In order to extract F0 due to lexical stress without intonation effect for subsequent analysis, a straight line with minimal distance (RMSE) to original F0 contour is derived to represent intonation and subtracted, the residual is regarded as F0 without intonation effect. In turn, duration extraction is also refined to remove the effect of inherent segmental duration and boundary lengthening using a multi-layered normalization method shown below[8], 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:

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

3. Results and Discussion

3.1. Contrast patterns--primary vs. secondary stress

This study examines prosodic contrast of secondary stress by linear order (before or after) to the primary stress and its distance from primary stress to test (1) if a secondary stress is assimilated to the target primary stress when it immediately precedes the primary stress; if a secondary stress immediately follows the primary stress it is reduced to tertiary stress in order to create more robust contrast patterns. (2) If there is more distance between the primary and secondary stress due to intermediate tertiary stress, then the secondary stress is more differentiable from either the primary or the tertiary counterparts. Furthermore, this study tests if these characteristics could help distinguish L1 and L2 English.

3.1.1. F0 without intonation effect—pitch contrast

Figure 2 shows F0 of secondary stress without intonation effect by L1/L2, linear order (preceding or following) regarding the primary stress and its distance from the primary stress. When the primary stress precedes the secondary stress, no significant difference between L1 and L2 English is found. However, the most distinct difference between L1 and L2 English is found when the secondary stress is immediately BEFORE the primary stress (1-syllable distance). The F0 of primary stress in L1 English is higher than secondary stress while L2 English lacks the same contrast pattern and highly varied. When secondary stress appears BEFORE primary stress (2-syllable distance), both L1 and L2 English show how secondary stress “stands out” and becomes more differentiable from the primary stress.

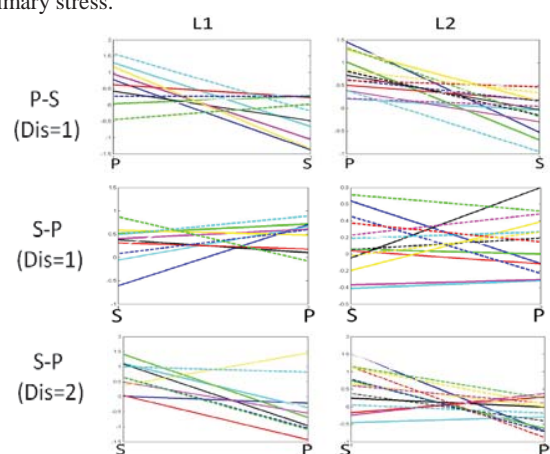


Figure 2: F0 of secondary stress without intonation effect by L1/L2, linear order (before or after) the primary stress and its distance from primary stress. Each color line denotes F0 patterns of one speaker.

The above results suggest that the F0 realization of secondary stress without intonation effect is context-dependent, concomitant but predictable. When secondary stress precedes the primary one, the F0 distinction does not always exist; the two categories are often under-differentiated. However, in reversed positions when secondary stress follows the primary one, it is lowered significantly, thus creating a sharper pitch contrast as shown in upper left panel. It is therefore no surprise why secondary stress is annotated differently in

different dictionaries. It is therefore also true that it is more difficult for L2 speakers to master.

3.1.2. Duration without segmental duration—rhythm contrast

Figure3 shows normalized duration of secondary stress by L1/L2, linear order (before or after) the primary stress and its distance from primary stress. The most distinct difference between L1 and L2 English is found when secondary stress appears immediately BEFORE primary stresses. By this context, the primary stress is distinctly longer than secondary stress while L2 English is again highly varied. For L1 English, primary stress is always longer than secondary stress thus the positions of primary stresses could be indicated by systematic rhythm/beat variation while L2 English exhibited no similar rhythm patterns.

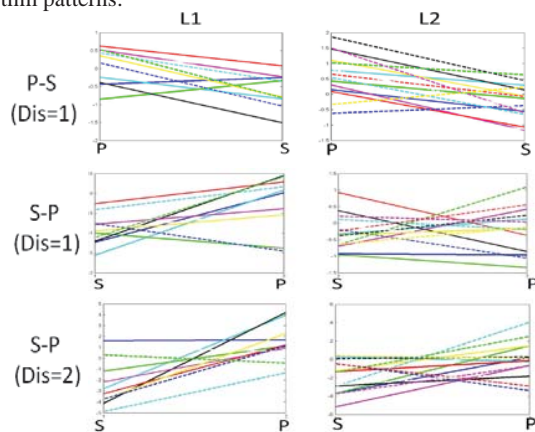


Figure 3: Normalized Duration of secondary stress by L1/L2 and primary/secondary stress context which is represented by order and distance between primary and secondary stress. Each color line denotes F0 patterns of one speaker.

The above results suggest the most robust feature to distinguish L1 and L2 English is normalized duration. For L1 English, primary stressed syllables are always longer than secondary stresses, creating systematic rhythmic patterns that imply distinctly the position of primary stresses in a word. However, L2 English lacks the same rhythmic contrast even some duration difference is exhibited. In other words, no categorical rhythm differentiation can be found in L2 English.

3.1.3. Intensity—loudness contrast

Figure4 shows intensity of secondary stress by L1/L2, linear order (before or after) to the primary stress and its distance from primary stress. For both L1 and L2 English, no significant difference is found between primary stress and secondary stress.

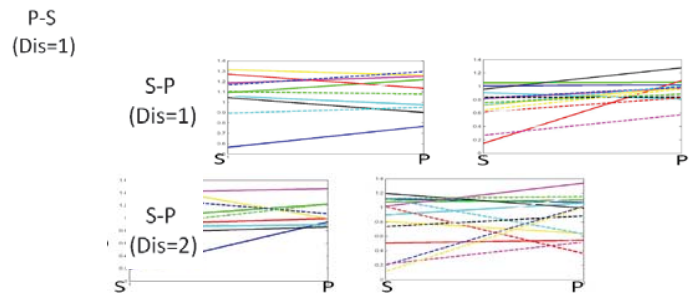
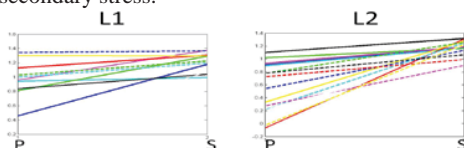


Figure 4: Normalized Duration of secondary stress by L1/L2, linear order (before or after) the primary stress and its distance from primary stress. Each color line denotes F0 patterns of one speaker.

3.2. Contrast patterns--narrow focus vs. sentential prominence

This study examines prosodic contrast of broad focus (sentential prominence) by linear order (before or after) to the narrow focus and its distance from narrow focus to test if (1) a sentence-initial broad focus is assimilated to the narrow focus when it is immediately BEFORE the narrow focus; or if the broad focus is reduced to a lower level in order to create more robust contrast patterns when it is AFTER the narrow focus. (2) However, if the same rationale can be applied to distance as well. Furthermore, this study tests if these characteristics could also help distinguish L1 and L2 English.

3.2.1. F0 without intonation effect—pitch contrast

Figure5 shows F0 of broad focus without intonation effect by L1/L2, linear order (before or after) to the narrow focus and its distance from narrow focus. When narrow focus precedes prominent word, the narrow focus is higher than broad focus for both L1 and L2 English. When broad focus precedes narrow focus, the contrast between narrow focus/broad focus is not clear for both L1 and L2 except for distance=8. It denotes broad focus “stand out” and is more differentiable from narrow focus when there is more distance between them. In addition, the contrast degree of L1 English is slightly higher than L2 English.

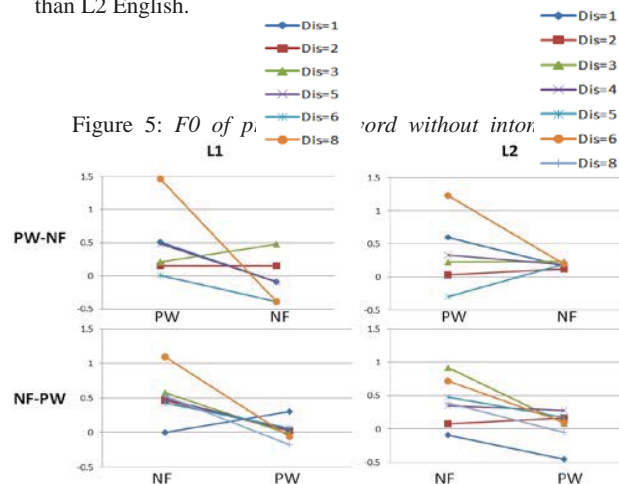


Figure 5: F0 of prominent word without intonation effect by L1/L2, linear order (before or after) the narrow focus and its distance from narrow focus PW, NF and Dis represent narrow focus, prominent word and distance by words.

The above results show how the contrast patterns of broad/narrow focus distinction is similar to patterns found between primary and secondary stress, as shown in 3.1. Broad focus is dependent on position and distance from narrow focus. The concomitant and more subtle differentiation again proves to be difficult for L2 speakers. L2 speech sounds flatter in melody.

3.2.2. Duration—Tempo contrast

Figure 6 shows duration patterns of broad focus by L1/L2, linear order (before or after) to the narrow focus and its distance from narrow focus. For narrow focus before or after broad focus, L1 English shows significant difference with L2 English, namely, a distinct contrast between narrow focus and broad focus. Narrow focus is always slower than broad focus thus the position of narrow focus could be indicated by systematic change of tempo. For L2 English, the tempo pattern between broad and narrow focus is more monotonous and exhibits no systematic tempo variations as found in L1.

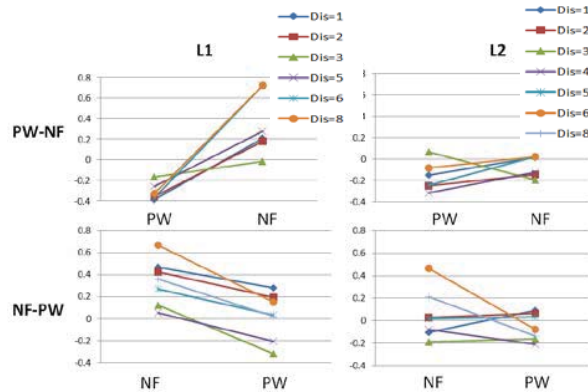


Figure 6: Tempo of prominent word by L1/L2, linear order (before or after) the narrow focus and its distance from narrow focus. PW, NF and Dis represent narrow focus, prominent word and distance by words.

The results above showed that the most significant difference of broad and narrow focus between L1 and L2 English is tempo patterns. While L1 speakers maintain distinct differentiating patterns as shown in left panels in Figure 6, L2 speakers could not realize the same tempo contrast patterns. As a result, L2 speech sounds more monotonous.

3.2.3. Intensity

Figure 7 shows intensity patterns of broad focus by L1/L2, linear order (before or after) to the narrow focus and its distance from narrow focus. Difference between L1 and L2 English is found when narrow focus precedes broad focus. By this context, a larger degree of contrast in L1 English is found than L2 English.

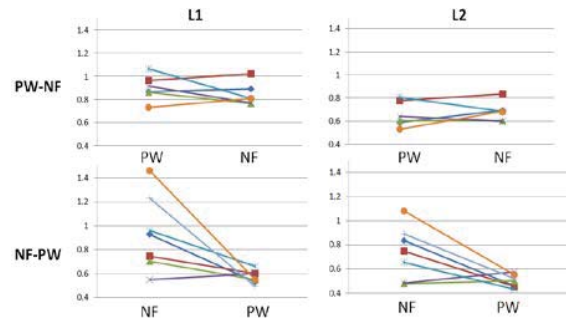
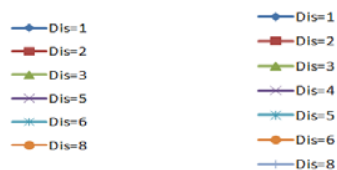


Figure 7: Intensity of prominent word by L1/L2, linear order (before or after) the narrow focus and its distance from narrow focus. PW, NF and Dis represent narrow focus, prominent word and distance by words.

The results above showed that the strong/weak contrast of broad/narrow focus is only differentiable for L1 when distance factor is bigger. L2 speech is less differentiable as expected.

4. General Discussion and Conclusion

Following our previous studies that showed the major acoustic characteristics of English lexical stress is F0 (pitch) contrast [6, 7], we further found in the present study that the major acoustic characteristic of primary vs. secondary stress is duration (rhythm) contrast, as shown in L1 speech. Though the same rationale is also found in the differentiation of broad vs. narrow focus, the patterns are additional contrast patterns on top of word level distinctions. The added results collectively suggest that binary contrast is the optimal choice of differentiating patterns, thus providing evidence of binary opposition, a crucial phonological concept, in the physical sense. Furthermore, these relative opposition patterns appeared to be quite difficult for L2 speakers to produce, suggesting on how perceptual sensitivity of such relative contrast patterns may be language dependent, and why L2 speech is less differentiable in production and flatter sounding in perception. Hence, under-differentiation of the necessary contrasts is a major feature of Taiwanese L2 English. We believe category related differentiation that require robust but sometime concomitant contrasts have not been properly addressed in language teaching in general and could be implemented to CALL technologies.

5. References

- [1] Magen, H.S., “The perception of foreign-accented speech”, *Journal of Phonetics*, vol. 26, 381-400, 1998.
- [2] Anderson-Hsieh, 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] Tajima, K., Port, R., and Dalby, J. “Effects of temporal correction on intelligibility of foreign-accented English”, *Journal of Phonetics*, 25, 1-24, 1997.
- [4] Field, J. “Intelligibility and the listener: The role of lexical stress”, *TESOL Quarterly*, 39(3), 399– 423, 2005.

- [5] Mixdorff, H. and Ingram, J. "Prosodic analysis of foreign-accented English", Proc. Interspeech 2009, 6-10 Sep. Brighton UK, 2009.
- [6] Tseng, C. Y. Su, Z. Y. and Visceglia, T. "Underdifferentiation of English Lexical Stress Contrasts by L2 Taiwan Speakers", *Slate* 2013 164-167. Grenoble, France, 2013.
- [7] Tseng, C. Y. Su, Z. Y. and Visceglia, T. "Levels of Lexical Stress Contrast in English and their Realization by L1 and L2 Speakers", KIIT Gurgaon, India, 2013.
- [8] Tseng, C. Y. and Su, Z. Y. "Dynamic Discourse Speech Tempo and Phonological Timing", The 7th International Congress of Phonetic Sciences. Hong Kong, China, 2011.