English Word Stress Production by Japanese and Mandarin ESL Learners: A Cross-linguistic Inquiry

Yizhou Lan

Abstract

The study intends to explore the role of L1 pitch configurations in Japanese and Chinese learners' realization of L2 English stress. In a production experiment, English tri-syllabic words with two stress patterns (words with the trochaic stress pattern, or Abc; and those with the iambic stress pattern, or aBc) and three vowel conditions (high front, low front, and low back) produced by Japanese and Mandarin Chinese speakers were examined for pitch height and range. The duration of the stressed syllable was also measured for each stimulus. Results echo previous findings that Japanese speakers used its pitch downfalls (HM) to represent English stress in narrative sentences whereas Mandarin speakers used high level tones or falling tones (HH or HL) for stress. However, findings also suggest tendencies not found in previous studies. On one hand, Japanese learners are more prone to use pre-set templates of the Japanese pitch-accent pattern globally across syllables since the pitch patterns for Abc and aBc words are not significantly differently configured. On the other hand, Mandarin speakers have shown more fixed local patterns of high level or falling tones for stressed syllables, with a clear difference of such in the said two types of words. Within each speaker group, the variation of duration for different vowel contexts has shown that only back vowels are correlated to longer durations. Also, we propose that gestural economy motivates the longer time lag for producing a back vowel coarticulation. Findings imply that the constructive pattern analogies between L1 and L2 pitch-patterns may be a learner strategy in both Japanese and Mandarin productions. Pedagogically, Japanese speakers are encouraged to focus on individual stresses whereas Mandarin speakers are reminded not to use pitch range as the acoustic cue for stress.

Key Words: second language acquisition, English stress, phonological transfer, speech production, Chinese tone, Japanese pitch accent.

1. Introduction

Languages rich in pitch variations are ones that either distinguish lexical meaning by pitch-accent (e.g., Japanese and Dutch) or by tone (e.g., Chinese and Thai). Speakers of such languages usually pronounce words in a second language with an accent featuring their first language pitch patterns (So and Best, 2014; Wang, 2008). Both pitch-accent and tone languages use pitch variations to contrast lexical meaning. For example, in the standard Tokyo dialect of Japanese, /aˈmeꜜ/ 'rain' and /aˈme˥/ 'candy' are different lexical entries despite their segmental resemblance. Similarly, in Mandarin Chinese, /ma˧˥/ 'mother' and /ma˥˩/ 'scold' also mean differently. Researchers found that first language (L1) pitch variations would influence the produc-

tion of English stress and intonation Japanese and Chinese L2 English speech, while these two L1s exhibit varying patterns of pitch assimilation (Teaman, 1992; Wang, 2008; Wee, 2008).

Such renderings of stress have been repeatedly discovered and recorded in classroom observations. In the ESL classroom, Japanese speakers tend to adopt only the high pitch to indicate stress in spoken English instead of using intensity and pitch rise-falls which conforms to the high-mid down-stepping (HM) pitch patterns. They are also known for using level pitch with high rise-falls on the nucleus of the English words, which occurs mostly on stressed English words (Kondo, 2009; Teaman, 1992). The Japanese realization of English stress can be seen as a conversational focus in the down-stepping intonational structure (HL-L%). In contrast, Mandarin speakers indicate English stresses by a high-falling (HL or H%L) tone (Wang, 2008). They are also found to replace stress with persisting high pitch (HH or HH%) in certain conditions (Lan and Oh, 2014). But the two realizations take place only in fixed intonational patterns in an intonational phrase. For example, the high-falling tone only exists in phrase-initial stressed syllables. A similar pattern was also found in Cantonese speakers of English, regardless of the position of stressed words in an intonational phrase.

Moreover, different L1 pitch configurations are employed in the production of multi-syllabic words under varying stress patterns. For example, Mandarin Chinese speakers produce stressed syllables with a higher pitch range than unstressed ones in tri-syllabic words (ibid.). Cantonese speakers produce stressed syllables with a higher pitch value in words stressed on the first syllables than in those on the second syllable (Wee, 2008). In detail, the Mandarin speakers either use persisting high pitch or high falling tone in English word production tasks (Lan and Oh, 2014; Wang, 2008). Apart from Japanese and Chinese, speakers of other tone languages such as Thai also transferr the L1 tone to L2 speech. For example, native Thai speakers of Mandarin Chinese will employ pitch height differences to replace the pitch slope in their production of Chinese tones (Wu, Munro, and Wang, 2014). However, many of the literature above (Lan and Oh, 2014; Teaman, 1992; Wang, 2008) feature stimuli of various syllable structures without comparing specific stress positions for words consisting of a fixed number of syllables, and the procedure is highly freestyle in that the recordings were done in multiple intonational phrases. This leaves a gap in future studies to control the intonational phrase of the stimuli sentences and only study the realization of different stress patterns in target words embedded into the same simple intonational phrase.

A few previous studies have also investigated the relationship between duration and stress in L2 Japanese stress production (Ueyama, 2000). The study compared durational variations for different stress patterns and found the same types of errors universal for everyone. However, the individual differences of each speaker were more evident. For some speakers, the duration of each stressed syllable was too long but for others, the duration was too short. The durational difference between stressed and unstressed syllables also varied (ibid.). The author attributed the variation of duration to the remaining and insertion of the mora structures. For tokens with shorter duration, the mora structure did not change at all; but for tokens with longer duration, an additional mora was created to accommodate the stressed syllable. The duration of stress may also be affected by the gestural complexity of the stressed segment itself. Gick (2003) investigated the correlation between gestural complexity due to different vowel environments and the duration of the stressed syllable. He found the more complex the gesture was, the longer the time lag wonld be required for the stressed syllable to be produced.

With the abundance of the previous studies, the present study aims to attain the following research objectives: to see the effect of L1-specific features on L2 acquisition, specifically the influence of L2 pitch variation on English stress in a controlled intonational phrase; to compare transfer of pitch variations by L1 Mandarin Chinese and Japanese speakers of English in search of generalizations about the differences between pitch-accent and tone languages; and finally, to investigate the reasons to the variation and provide pedagogical implications that can be derived from the results. To achieve these aims, a production experiment is designed in the present study. The experiment attempts to answer the following research questions: 1) What

types of L1 Japanese and Mandarin pitch variations are represented in English prosody? 2) Are those pitch representations transferred to L2 English speech? 3) What are these representations motivated from?

The experiment protocol of this study includes both impressionistic and statistical analysis to ensure well-roundedness of the data. Since pitch variations is the main concern of the present study, pitch curves are examined as an indicator of learners' prosodic cues (cf. Wang, 2008; Kondo, 2009), and the detailed measurements of the pitch curves are average pitch height and pitch range (cf. Braun, Galts, and Kabak, 2014; So and Best, 2014). According to the previous literature, we predict that the Japanese learners will pronounce English words with a down-stepping pattern for stressed syllables whereas the Mandarin speakers will produce them with a high-falling pattern, following this protocol.

2. Method

Pitch patterns of L1 Japanese and Mandarin speakers' English productions were examined and how these speakers map L1 tonal features onto L2 English stress situated in two different stress patterns were explored. In the experiment, 3 Japanese- and 3 Mandarin-speaking participants from Japan and Hong Kong read out randomized target stimuli words in a carrier sentence in a recording booth. All recorded sounds are analyzed for both impressionistic pitch patterns and statistical comparison for key indicators of pitch.

2.1 Participants

Participants are three Japanese speakers studying in a university in Japan and three Mandarin speakers studying in a university in Hong Kong SAR. The participants were recruited through convenient sampling. Their mean age was 22.5 years and 23 years for Japanese and Mandarin speakers respectively and all of them had over 10 years of English learning experience. For Japanese speakers, their TOEIC test scores were over 600 and for Mandarin speakers, their IELTS test scores were over 6.5 by the time of the experiment. So, they were identified as higher intermediate to advanced learners. The unity of linguistic backgrounds of participants was strictly controlled to ensure the reliability of the results. Specifically, speakers of both languages do not speak a third language, and they are all decedents of their first-language speaking parents. The dialectical backgrounds of the participants were also carefully screened. The Japanese speakers are all capable of speaking Japanese in the standard Tokyo accent and the Chinese participants are all northern China dwellers though they have physically been living in Hong Kong. All participants were reported of no speech disorders by the time of experiment.

2.2 Stimuli

Stimuli are 20 tri-syllabic words in narrative carrier sentences. Tri-syllabic words (CVCVCV, with the final V possibly a monophthong or a diphthong) are chosen for replicating the Japanese and Chinese-like prosody, both featuring a CV syllable structure. Another reason of choosing tri-syllabic words is the reliability for intercepting target syllables. For data processed by Praat, the last syllable in an intonational phrase is usually not clearly recorded. So, we deliberately added another unstressed syllable so that the first and the second syllables, which are to be target syllables in respective word types, will be more accurately analyzed. The stimuli are categorized in two types. Ten of these stimuli are trochaic, i.e., they have their stresses on the first syllable (e.g., SImilar). The other ten stimuli are iambic, i.e., with stresses on the second syllable (e.g., narRAtor). These two types of words will be labelled Abc words and aBc words hereinafter. To investigate the effect of vowel conditions on the duration of the stressed syllables, the vowels for the stressed syllables are varied in three conditions: high front (/i/), low front (/æ, ə/), and low back (/ɔ, o/). However, the number of tokens of each vowel category is not balanced due to the poverty of real word tokens from some vowel contexts. The complete list of stimuli in the two word types is shown in Table 1.

Table 1. List of Stimuli

Words with Abc stress			
similar	domino	camera	cinema
capita	colony	sanity	bakery
botany	family		
Words with aBc stress			
narrator	tomato	Malawi	bikini
potato	banana	tornado	saliva
colossi	pagoda		

2.3 Procedure

The experiment is performed in the following procedure. First, stimuli words were repeated for three times, put in random order and inserted into a carrier sentence "*Now I say _____ again.*" As a result, the total number of token was 6 speakers × 10 words × 2 stress patterns × 3 repetitions = 360. Among them, some 345 out of 360 tokens were valid and included in the analysis. However, 12 tokens with mispronounced vowels were excluded. Then, participants were asked to read out these sentences aloud in a sound-proof booth with recording equipment. all carrier sentences were recorded by a Shure SM 57 Microphone with the sampling rate of 44100Hz in mono channel. After the data-transfer of the recordings, the stimuli were segmented for impressionistic display of the pitch contour by semitone, and the raw pitch height and range was sent to statistical analysis. For each token in the statistical analysis, the pitch height and pitch range (Pmax-Pmin) of the nuclei of the stressed syllable were measured and compared by the independent variables of speaker group and word type. The duration values of these stressed syllables were also measured and compared by the independent variables of stress type and vowel context, separately for each speaker group. The ProsodyPro script in Praat (Xu, 2013) was used for measurement of duration, average pitch height and pitch range.

3. Results

3.1 Overall Pattern

For each stress pattern, the 90 sound samples of the sentences were all averaged and put together in an analytical frame by using the ensemble function from Praat where the pitch contours in speckles were drawn with the unit of semitone at 100Hz. Example productions with pitch plotted in the temporal frame are shown in Figures 1 and 2, with the target word segmented and labelled between dashed lines. The figures for each stress pattern contain examples from three speaker groups: Japanese, Mandarin and English. From the impressionistic patterns, we can see that generally the high pitch value (H) was associated with stressed syllables whereas the low pitch value (L) was with unstressed ones for both Japanese and Mandarin speakers. English speakers also show tendencies of using pitch for stress, but the correlation is not so apparent.

The patterns of transfer are clear language-specific for each speaker group compared with that of English baseline. For Japanese speakers, a high-mid falling down-step pattern (HM) was used across two word types. Compared with native English speakers' production, the pitch pattern of Japanese often show a group of down-stepping pitch freckles within the intonational phrase. We can also find that Japanese speakers replaced English stress with a high-mid falling pitch pattern that resembles Japanese one-focus (HL-L-L%) intonational phrase; and such pattern remains constant regardless of the stress patterns. It is interesting to see that Japanese speakers had shown a same level of pitch range as native English speakers, which is different from previous studies that claimed Japanese speakers do not use pitch in sentence intonation. In contrast, Mandarin speakers on one hand used falling tones (HL) for stress in the Abc words but on the other used persisting high pitch for the aBc words. For Abc words, the Mandarin production showed large fluctuations of pitch within

one single word in the intonational phrase, resulting in a large pitch range as shown in Figure 1. For aBc words, the persisting high pitch form of the target word occurd after an abrupt rising from the end of the previous word as shown in Figure 2. However, in native English productions, no such steep pitch fluctuations were shown, nor were distinct tone patterns such as HL or HH manifested.

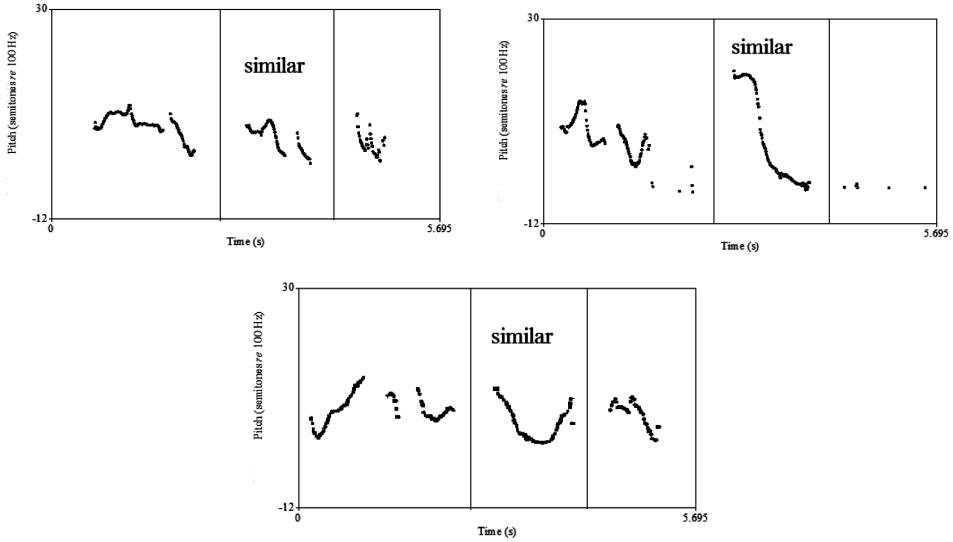

Figure 1. Pitch patterns for Japanese (upper left), Mandarin (upper right) and English (lower) productions of the Abc-carrier sentence "Now I say similar again."

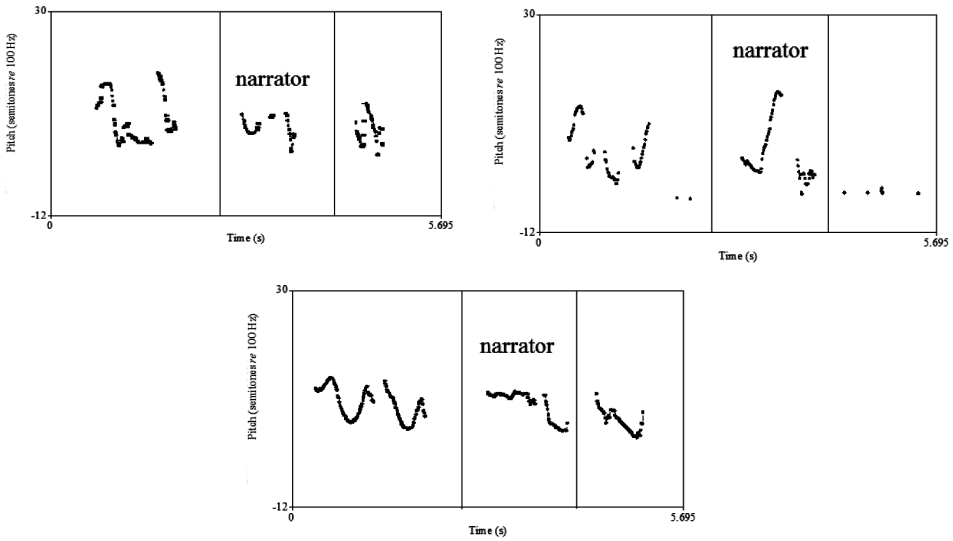

Figure 2. Pitch patterns for Japanese (upper left), Mandarin (upper right) and English (lower) productions of the aBc-carrier sentence "Now I say narrator again."

3.2 Statistical Comparison

The mean differences of pitch height and range were computed for four groups, namely Japanese Abc (JA), Mandarin Abc (MA), Japanese aBc (JB), and Mandarin aBc (MB). Following that order, the mean pitch height was averaged at 147.4, 160.4, 165.1 and 186.0 respectively and pitch range at 40.1, 57.9, 103.1, and 46.0 respectively. ANOVA comparisons between the four stress pattern groups also showed significance for both pitch value [$F(3, 336)=4.27$, $p<.016$] and pitch range [$F(3, 336)=3.58$, $p<.022$]. Seen form the Tukey's Post-hoc tests, Mandarin speakers' mean pitch value outnumbers that of Japanese speakers in both Abc [$t=57.37$, $md=8.05$, $p<.001$] and aBc [$t=55.12$, $md=8.05$, $p<.001$] stress patterns, indicating that Japanese speakers did not apply higher tone to English stress. In contrast, the variable of pitch range showed that Mandarin speakers' pitch was higher in Abc [$t=9.00$, $md=12.47$, $p<.05$] but not in aBc [$t=1.74$, $md=3.89$, $p=.376$] patterns. Moreover, significantly different patterns of pitch ranges were adopted by Mandarin speakers across stress patterns [$t=7.37$, $md=21.79$, $p<.001$] while Japanese speakers used almost the same pitch range for the two stress patterns [$t=0.45$, $md=4.82$, $p=.417$].

The duration of the four groups was also compared by two t-tests, one run by the independent variable of first language (Japanese and Mandarin) and the other by the stress types (Abc and aBc). Results show that the durational difference is not significant for the language factor, indicating that the Mandarin and Japanese usages of temporal cue are similar. The duration for stressed syllables in each stress type was not significant for Japanese speakers. However, the comparison was significant for Mandarin speakers. With the Abc showing a longer duration than the aBc patterns [$t=-4.57$, $df=337$, $p<.05$].

To find more details on which vowel environment has contributed to the significant differences of duration by stress type, mean duration was compared by vowel quality for stressed tops. Results in Table 1 below show that the duration was significantly influenced by vowel context of the word for both stress types for Japanese speakers. As shown in Table 2 below, Japanese speakers' productions of low back vowels were signifi-

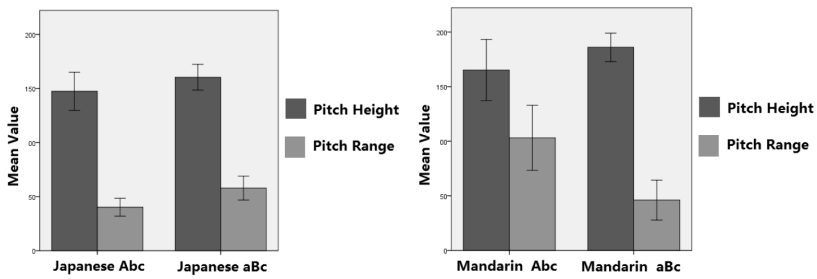

Figure 3. Average pitch value and pitch range for Japanese Abc, Japanese aBc, Mandarin Abc and Mandarin aBc stress-patterned words

Table 2. Mean duration for each vowel type in Abc stress type (upper panel) and aBc stress type (lower panel) for Japanese speakers

Vowel types	Mean duration (ms)	Significance
High Front	22.5	-
Low Front	27.3	-
Low Back	54.4	*** ($p<.001$)

Vowel types	Mean duration (ms)	Significance
High Front	17.5	-
Low Front	21.1	-
Low Back	43.6	* ($p<.05$)

cantly longer in duration than other vowel types. Moreover, the durations of stressed syllables had greater variation in the low back vowel conditions for the Abc [r=0.626, p<.001] than aBc types [r=0.423, p<.05]. However, the durational difference by vowel context was not significant for Mandarin speakers.

4. Discussions

Results generally indicate that both Japanese and Mandarin speakers showed a tendency of pitch-cueing regarding the realization of stress in English. However, such pitch patterns vary from the findings from speakers of two languages. On one hand, Japanese speakers are prone to replace English stress with a high-mid falling pitch pattern that resembles Japanese pitch-accent configurations; and such pattern remains constant regardless of the stress patterns. Vowel quality might be another reason for Japanese speakers' down-stepping pitch pattern. The findings suggest that, for diphthongs and back vowels, Japanese speakers are more prone to use duration as the acoustic cue. However, for front vowels, the use of duration is less prominent. This is possibly due to the greater time lag created from such diphthongs and back vowels for another mora to be inserted in the word (cf. Kondo, 2009). On the other hand, Mandarin speakers clearly differentiates the stress patterns by associating high-falling in Abc and High level in aBc patterns, which is a transfer by the Mandarin tone 4 and tone 1 (MA, MB), disregarding the duration as a whole. Both tone 4 and 1 are highly prominent in Chinese due to its high pitch ranges it spans across syllables. Therefore, we can conclude that Mandarin speakers resort to assigning specific prominent tones to each stressed syllable locally, and such local tone assignment is position-sensitive.

As mentioned above, the findings suggest that L1-specific pitch-usages in stress take place more holistically in Japanese, which supports that Japanese pitch-accents spans across syllables. Specifically, the Japanese speech resembles the one-intonational phrase structure in Japanese that the focus of the stressed word either takes an equal high pitch and drops down or it starts dropping down from the beginning of the intonational phrase. The Japanese pitch pattern in English, following either of these intonational phrase structures, also became highly global. That may explain why the Japanese learners' lack of stress cues in producing different stress patterns.

The durational results conform to that of Ueyama (2000), in that a high degree of variation was found for duration across participants and tokens, as has mentioned in the introduction. The current findings propose an explanation of gestural economy for such variations. The choice of remaining the same mora structure or adding an extra mora may be related to the quality of the nucleus of the syllable. Vowel quality apparently influences the realization of English stress by Japanese speakers. From the findings, Japanese speakers are more prone to allocate duration to stressed syllables when a sequence of two distant gestures (as in the context of a back vowel) occurs. Another evidence is that the down-stepping pitch pattern mentioned before that exists generally in Japanese productions was even more visible in such diphthongs and back vowels, probably due to the time lag created by the distant gesture for another mora to be inserted in the word. This can also be seen as supporting experimental evidence to the classroom observations proposed by Teaman (1992).

The current study, with a limited number of participants and measurements, only serves as a primer for a series of possible future studies. It was noted that pitch slope is another important factor for tone perception in Mandarin Chinese (Shen, 2016), and the effect of pitch slope on perception of Japanese and Mandarin English learners will be tested in future work that follows up this production study. Future studies will also include the testing of the English speakers' perception of the Mandarin and Japanese accented stress, in search for the common acoustic cue that will enhance the mutual intelligibility of stress for cross-linguistic perception from both native to non-native and the other way around.

Pedagogically, the study implies that even intermediate and advanced learners have not arrived at the mastery of L2 English stress. A clear reason shown from the results is the unawareness of L1 pitch transfers such

as from the Mandarin tone or the Japanese pitch accent. To tackle that, different pedagogical suggestions may be derived from the results. For Japanese learners, the instructors need to direct learners' attention to the acoustic prominence of stress to the precise syllables, rather than letting learners following a general intonation pattern. For example, instructors can make use of hand gestures to indicate intensity and pitch for individual stresses as in Krahmer and Swerts (2007). The study used a range of movements to signify the intra-phrasal and intra-lexical stress variations, which had been proven useful by pedagogical treatment results. However, for Mandarin speakers, the learners need to understand that pitch range is not a common indicator for English stress. The use of rising and falling tones will be considered as non-native renderings. Instructors are advised to direct the attention of the learners to the intensity and pitch value of the stressed syllables, instead of the pitch range.

In a nutshell, the study provides ESL teachers an experiment-based teaching method to help advanced learners notice the pronunciation gap between L1 and L2, and achieve the target pronunciation by self-driven awareness (Ellis and Shintani, 2013). The leaners' attention needs to be directed from global or local aspects of pitch to the interplay of intensity, pitch and duration in the English stress production. This way, learners' reminding themselves of the awareness will enhance the production through automatization of articulatory configurations (ibid.).

5. Conclusion

Tonal transfer in the present study can be seen as a form of assimilation (Best, 1995) clearly seen from the L1 global intonation pattern in Japanese and local tonal patterns in Mandarin. However, the L1 transfer is not the only factor for stress productions. It is also influenced by the cross-linguistic tendencies of gestural economy seen by a consistent correlation between stressed syllable duration and vowel quality. We can draw a conclusion from this study that L2 production is resulted from an interplay of L1 assimilation and universal grammar in articulatory phonology. The knowledge of such will sow seeds for the teachers and learners when facing the complex L2 pronunciation learning tasks and strive for the comprehensibility of L2 speech.

References

Best, C. T. (1995). A Direct-realist view of cross-language speech perception. In Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171-204.

Braun, B., Galts, T., and Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research,* 30(3), 323-350.

Ellis, R., and Shintani, N. (2013). *Exploring Language Pedagogy Through Second Language Acquisition Research. Routledge*.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 222-236.

Kondo, M. (2009). Is Acquisition of L2 Phonemes Difficult? *Production of English Stress by Japanese Speakers. Proceedings of the 10th GASLA Conference*, 105-112.

Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396-414.

Lan, Y., and Oh, S. (2014). The effect of Mandarin tone on the perception and production of English consonant clusters, *Proceedings of 4th Symposium of Tonal Aspects of Language*, 181-184.

Shen, R. (2016). Variations of F0 encoding in tone languages, *paper presented at the Interdisciplinary Workshop on Linguistics 2016*, Shanghai.

So, C. K., and Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Manda-

rin tones to native prosodic categories. *Studies in Second Language Acquisition, 36*(2), 195-221.

Teaman, B. D. (1992). Stress in Japanese English: evidence from native perceptual judgements. *Working Papers in Educational Linguistics, 8*(1), 69-83.

Ueyama, M. (2000). Prosodic *Transfer: an Acoustic Study of L2 English vs. L2 Japanese* (Doctoral Dissertation). Retrieved from https://www.semanticscholar.org/

Wang, Q. (2008). *Perception of English Stress by Mandarin Chinese Learners of English: An Acoustic Study* (Doctoral Dissertation). Retrieved from https://dspace.library.uvic.ca/

Wee, L. H. (2008). Phonological patterns in the Englishes of Singapore and Hong Kong. *World Englishes, 27* (3), 480-501.

Wu, X., Munro, M. J., and Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, 46, 86-100.

Xu, Y. (2013). ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France. 7-10.