PERCEPTION OF ENGLISH LEXICAL STRESS IN DIFFERENT INTONATIONS BY MANDARIN LISTENERS

Tzu-Hsuan Yang[1], Annie Tremblay[2]

University of Kansas[1], University of Texas at El Paso[2]
thyang@ku.edu[1], actremblay@utep.edu[2]

ABSTRACT

The present study seeks to elucidate the nature of first language transfer effects on the perception of English lexical stress in different intonational contexts. We examine whether Mandarin learners of English assume a one-to-one relationship between pitch and lexical stress in English by testing their stress perception in four different intonations: H*L–L%, L*H–H%, H*H–H%, and L*L–L%, where F0 cues to stress were realized differently in each intonation. The results of a stress identification task showed that Mandarin listeners falsely associated higher F0 with stress and lower F0 with the absence of stress, and their performance did not improve as they became more proficient in English. The findings provide corroborating evidence for the Cue-Weighting Transfer Hypothesis by showing that the use of a suprasegmental cue can transfer from one phonological category (tone) to another (stress).

Keywords: Second language acquisition, cue-weighting transfer, lexical stress, lexical tone

1. INTRODUCTION

Languages with lexical stress differ in how acoustic cues signal stress. One approach that seeks to explain how differences in cue weighting affect second-language (L2) speech perception and processing is the Cue-Weighting Transfer Hypothesis [1]–[5]. This hypothesis predicts that listeners can transfer the weighting of an acoustic cue (e.g., F0) from one phonological phenomenon in the first (i.e., native) language (L1; e.g., lexical tones) to another in the L2 (e.g., lexical stress). To illustrate, in English, the stressed syllable is produced with a full vowel, longer duration, and greater intensity [6]–[10]. Crucially, pitch (F0) can only serve as a *surface* cue to lexical stress: Different pitch accents (e.g., H*, L*) can be aligned with the stressed syllable, and the pitch contour that is realized over the word is created by the combination of a pitch accent, a phrase accent (e.g., H– or L–, and a boundary tone (e.g., H% or L%) [11]–[13]. Unlike English, Mandarin is a tonal language in which lexical tones, conveyed primarily by F0 cues, serve to distinguish word meanings [14], [15]. Given the great functional load of pitch for signaling lexical contrasts in Mandarin, Mandarin listeners are expected to rely on F0 cues when distinguishing lexical stress contrasts in English as well.

This prediction was largely confirmed in previous studies. For example, Qin et al. [3] compared learners of English whose L1 was Standard Mandarin (SM), which has lexical stress contrasts signaled primarily by duration cues, or Taiwan Mandarin (TM), which does not have stress contrasts. The authors found that both groups of Mandarin listeners could use F0 to perceive lexical stress in English, but only SM listeners could use duration cues. This provides clear evidence that listeners can transfer the use of acoustic cues from one phonological phenomenon (lexical tones) to another (lexical stress) in speech perception.

Crucially, though, these studies all examined the perception or processing of stress in the canonical intonational context—that is, when the stressed syllable had a H* pitch accent and thus was realized with higher pitch relative to the unstressed syllable. To investigate whether and how L2 learners' processing of English lexical stress is affected by different sentence intonations, Ou [16] tested L1 Taiwan Mandarin listeners' perception of English stress in two intonations: a falling intonational contour where the stressed syllable is realized with H* (H*L–L%) and a rising intonational contour where the stressed syllable is realized with L* (L*H–H%). The results of the discrimination tasks revealed that overall, English native listeners showed significantly greater sensitivity to stress than L2 learners. Crucially, the difference between native and non-native listeners was greater in the rising contour than in the falling contour, indicating that L2 learners had more difficulty perceiving stress in this context.

Adopting a similar design as [16], Choi [17] investigated the perception of English lexical stress by L1 Cantonese L2 learners of English in the falling contour and the rising contour. The results of the discrimination task showed that when lexical stress contrasts were not cued by vowel quality changes, Cantonese listeners outperformed English listeners in the falling intonational context. Nonetheless, L2 learners' discrimination ability significantly worsened in the rising contour, and an advantage over native listeners was no longer found.

However, one important limitation arises in the interpretation of [16]–[17] due to the authors' adoption of discrimination tasks: The use of a discrimination task may have encouraged listeners to rely on the acoustic realization of the words at a surface level, leaving open the question of whether listeners can correctly assign nouns and verbs to different categories despite variability in their intonational realizations. In addition, [16]–[17] did not investigate whether L2 learners can rely on non-tonal cues to English lexical stress in contexts where the intonation is relatively flat.

Adopting an identification task, the current study seeks to extend [16]–[17] by addressing the research questions of whether Mandarin listeners assume a one-to-one relationship between pitch and lexical stress in English, and whether they can rely on non-tonal (suprasegmental) cues when the pitch cues heard over the target word are ineffective.

2. METHODS

2.1. Participants

Participants were 30 Mandarin-speaking L2 learners of English from Taiwan (14 female, 15 male, 1 preferred not to answer, aged 18–31, *M* = 25.3) and 29 native speakers of American English (20 female, 9 male, aged 18–29, *M* = 20.6). The English proficiency of the L2 listeners was assessed using the LexTALE test [18]. Mandarin listeners' averaged LexTALE score was 65 out of 100 (*SD* = 14), which roughly amounts to an intermediate level of proficiency.

2.2. Materials

A two-alternative forced-choice identification task was conducted to examine Taiwan Mandarin listeners' perception of lexical stress. The target words included ten disyllabic minimal pairs with stress on either the first (noun) or the second (verb) syllable (Table 1). As the focus of the present study was to examine how Mandarin listeners utilize suprasegmental cues to process lexical stress, stimuli were selected so that the unstressed syllable of these disyllabic words would not undergo vowel reduction.

Trochaic (noun)	Iambic (verb)
PERmit	perMIT
IMpact	imPACT
IMport	imPORT
DEcrease	deCREASE
INcrease	inCREASE
INsult	inSULT
DIScharge	disCHARGE

Table 1: Stress minimal pairs used in the stress identification task.

The target words were elicited in four intonations: H*L–L% (i.e., a H pitch accent followed by a L phrase accent then by a L boundary tone), L*H–H%, H*H–H%, and L*L–L% (or HLL, LHH, HHH, and LLL hereafter). The auditory stimuli were produced and then extracted from two carrier sentences: *Mary said _____ before* or *Mary say _____ before?*, with the former used to elicit the HLL and LLL contours and the latter to elicit LHH and HHH. Note that the target words were elicited in non-sentence-final position to decrease the amount of word- and sentence-final lengthening. All four contours were naturally occurring intonations in English [13], [19]. The pitch contours of the target words in each intonation are shown in Figure 1.

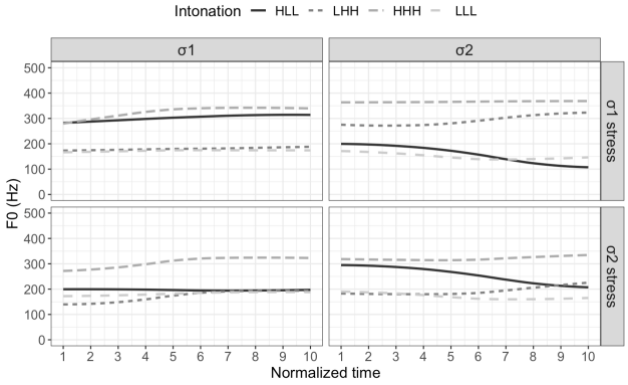

Figure 1: Pitch contours of the stimuli (sample word: *insult*).

These intonations were chosen so that the F0 cues to stress would be realized differently in each intonation, and to allow for an examination of the use of non-tonal prosodic cues to English lexical stress (duration and intensity cues, which were not teased apart in this study) by systematically manipulating the pitch contour. In the HLL condition, a H* was aligned with the stressed syllable for both trochaic and iambic words. The stressed syllable was thus always realized with higher F0, which was consistent with the canonical association expected from Mandarin L2 learners of English. However, in the LHH condition, a L* was aligned with the stressed syllable, and H phrase and boundary tones surfaced on the second syllable for both stress patterns. In other words, the second syllable was higher in pitch for both stress patterns, possibly preventing Mandarin listeners from resorting to their canonical association to perceive lexical stress. On the other hand, the HHH and LLL conditions contained relatively neutral F0 cues, as the pitch contours that were heard over the target word were relatively flat. Crucially, in these two conditions, words with stress on the first or second syllable were realized with very similar pitch contours, and therefore F0 cues were ineffective for

participants to distinguish between the stress minimal pairs. Table 2 summarizes the acoustic information of the stimuli in each intonation condition.

The task included a total of 160 stimuli (10 pairs × 2 stress positions × 4 intonations × 2 talkers). All stimuli were randomized.

Intonation	Stressed syllable	Duration ratio of σ1 to σ2	Mean F0 ratio of σ1 to σ2
HLL	First	0.47	1.62
	Second	0.35	0.78
LHH	First	0.49	0.63
	Second	0.35	0.87
HHH	First	0.53	0.94
	Second	0.35	0.94
LLL	First	0.5	1.08
	Second	0.36	1.07

Table 2: Acoustic measurements of the stimuli. Values were averaged across all tokens in each intonation condition.

2.3. Procedure

The experiment was conducted online via Gorilla Experiment Builder [20]. Participants were told that they would hear English words that can either be a noun or a verb depending on which syllable is stressed. They were instructed to press the left arrow key if they perceived the stress to be on the first syllable (if they heard a noun) and the right arrow key if on the second syllable (if they heard a verb). The experiment took approximately 25 minutes.

3. RESULTS

The raw accuracy data were converted into d-prime scores [21]. Figure 2 presents L1 and L2 listeners' d-prime scores in each intonation. Mixed-effects linear regression models were fitted to participants' d-prime scores in R [22]. Table 4 summarizes the results of the model with the best fit. We also correlated L2 learners' z-transformed LexTALE scores with their d-prime scores to determine whether L2 listeners' sensitivity to lexical stress was modulated by their proficiency in English.

The results showed that both Mandarin and English listeners were significantly more sensitive to stress in the HLL intonation than in the other intonations (demonstrated by simple effects of Intonation with negative coefficients, which were also found in the releveled model with English as the baseline of L1). The model also revealed a significant simple effect of L1. The negative coefficient indicates that in the HLL intonation, Mandarin listeners' sensitivity to lexical stress was in fact higher than that of native English listeners, suggesting that Mandarin listeners were better able to use the higher F0 to facilitate their perception of stress in the falling intonation. Additionally, we found a significant interaction between Intonation (LHH) and L1, suggesting that the difference between HLL and LHH was larger for Mandarin listeners than for English listeners. Collectively, these findings indicate that Mandarin listeners made a stronger association between higher F0 and lexical stress than English listeners did.

To examine listeners' performance when F0 cues were not as effective (relatively flat), we releveled the Intonation variable with HHH as the baseline. The releveled model reported a non-significant intercept (β = 0.217, SE = 0.151, t = 1.437, p > .1). As a d-prime score of 0 signals no sensitivity, this finding showed that Mandarin listeners' performance in the HHH condition was at chance level. There was also no significant simple effect of Intonation (LLL) (β = 0.032, SE = 0.185, t = <|1|, p >.1), suggesting that L2 listeners' performance was similar across HHH and LLL conditions. Taken together, these results imply that Mandarin listeners were not able to utilize non-tonal cues (i.e., duration and intensity) to distinguish words with stress on the first or second syllable when F0 cues were not informative. The lack of simple effect of L1 (β = –0.174, SE = 0.216, t = <|1|, p > .1) and Intonation (LLL)-by-L1 interaction (β = 0.301, SE = 0.264, t = 1.14, p > .1) indicated that English listeners patterned similarly to Mandarin listeners.

Finally, Figure 3 shows the relationship between learners' d-prime scores and their L2 proficiency (measured by LexTALE). The results demonstrate that when Mandarin listeners encountered difficulty in identifying stress, that is, when their association between higher F0 and lexical stress was violated (i.e., LHH, HHH, and LLL), their performance did not improve with increasing proficiency in English.

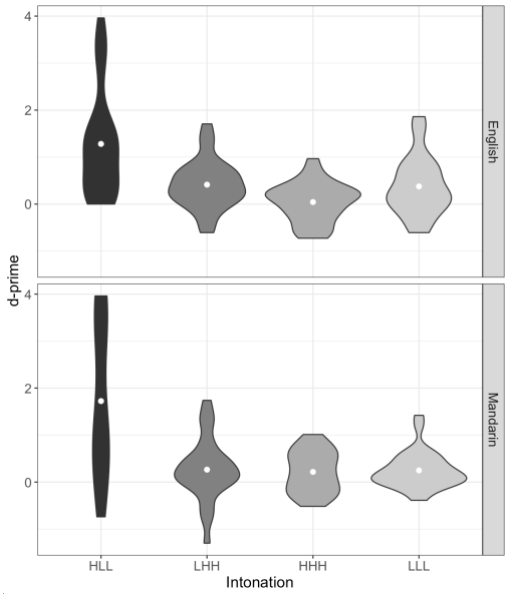

Figure 2: Participants' d-prime scores in each intonation (the white dots represent the mean).

	β	SE	*t*	*p*
(Intercept)	1.724	0.151	11.394	< .001
Intonation (LHH)	–1.459	0.185	–7.888	< .001
Intonation (HHH)	–1.507	0.185	–8.147	< .001
Intonation (LLL)	–1.475	0.185	–7.976	< .001
L1 (English)	0.444	0.216	–2.054	< .05
Int. (LHH) × L1 (Eng)	0.592	0.264	2.243	< .05
Int. (HHH) × L1 (Eng)	0.269	0.264	1.02	0.309
Int. (LLL) × L1 (Eng)	0.569	0.264	2.16	< .05

Table 3: Summary of fixed-effects for the model on participants' d-prime scores. Formula: dprime ~ intonation*language + (1|Participant), with HLL being the baseline for Intonation and Mandarin being the baseline for L1.

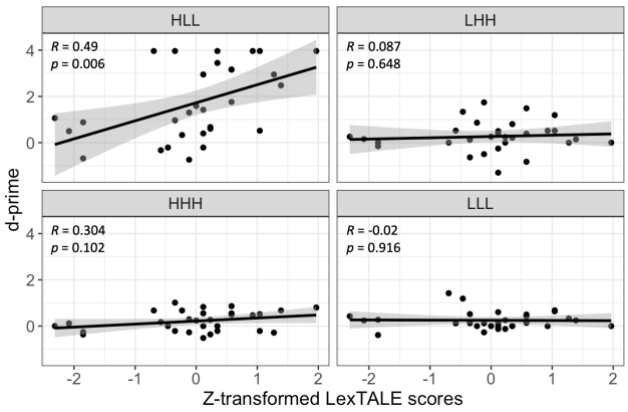

Figure 3: Relationship between L2 learners' d-prime scores and their English proficiency.

4. DISCUSSION

The results of the stress identification task provided clear evidence for the hypothesis that Mandarin listeners assume a one-to-one relationship between pitch and lexical stress in English. Mandarin listeners showed the greatest sensitivity to stress in the HLL condition, where the stressed syllable was always realized with a higher F0. In contrast, their performance worsened in the LHH contour where this association was violated, regardless of their L2 proficiency. These results were consistent with the findings of previous studies [16], [17] and support the proposal of the Cue-Weighting Transfer Hypothesis [1]–[5] that L2 learners can transfer the use of prosodic information (here, F0) from a different phonological category in the L1 (here, lexical tones) to lexical stress in the L2, even if their L1 does not contain lexical stress contrasts.

An interesting observation arises from the results: In the falling contour, L2 learners outperformed native listeners of English. The low accuracy of L1 listeners is attributed to the lack of vowel quality cues in the stimuli. Given that English listeners rely on vowel quality among all acoustic cues to stress [1], [5], it might have been difficult for them to distinguish stress contrasts using suprasegmental cues only [2]. Furthermore, in recent years there has been emerging evidence showing that in the absence of vowel quality cues, native English listeners show less sensitivity to stress compared to L2 learners of English whose L1 uses pitch to signal lexical contrasts [2], [17]. Mandarin listeners' advantage over native English listeners has been ascribed to their positive transfer of F0 processing skills: As F0 is the primary acoustic correlate of lexical tones, it carries greater functional weight in Mandarin, and thus Mandarin listeners are more sensitive to it than English listeners.

To examine how Mandarin listeners weighted non-tonal cues when F0 was an ineffective cue to stress, we examined L2 learners' performance in the HHH and LLL conditions. In these intonations, the words with stress on the first and second syllables were realized with very similar contours, and therefore participants had to rely on duration and intensity cues to distinguish stress contrasts. The results revealed that L2 learners had almost zero sensitivity to stress placement in these conditions, suggesting that they were not able to use the duration and intensity cues to identify the position of stress, even as they became increasingly proficient in English. These results corroborated [3], who reported that Taiwan Mandarin listeners were unable to use duration as a cue to stress due to it not being a primary cue to any lexical contrasts in their L1.

Note that English listeners' sensitivity was very low in the HHH and LLL conditions as well. This, again, stems from their reliance on the vowel quality cues. The results were consistent with the findings of [23] that without vowel quality cues, L1 and L2 learners of English (whose L1 is Cantonese) performed equally bad in a discrimination task when the F0 information was neutralized and only duration and intensity cues were available.

5. CONCLUSION

The present study examined the perception of English lexical stress by L1 Taiwan Mandarin L2 learners of English. The results of the stress identification task showed that Mandarin listeners associated higher F0 with the presence of stress and lower F0 with the absence of it, and they failed to make use of non-tonal suprasegmental cues to process stress in the L2 when F0 cues were ineffective. The findings provide robust evidence for the Cue-weighting Transfer Hypothesis by showing that the use of a suprasegmental cue can transfer from one phonological category to another and that how L2 learners use this cue to process lexical stress is affected by how much weight it carries in their L1.

6. REFERENCES

[1] A. Chrabaszcz, M. Winn, C. Y. Lin, and W. J. Idsardi, "Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 4, pp. 1468–1479, 2014.

[2] H. Kim and A. Tremblay, "Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach," *J Phon*, vol. 87, p. 101059, 2021.

[3] Z. Qin, Y.-F. Chien, and A. Tremblay, "Processing of word-level stress by Mandarin-speaking second language learners of English," *Appl Psycholinguist*, vol. 38, no. 3, pp. 541–570, 2017.

[4] A. Tremblay, M. Broersma, Y. Zeng, H. Kim, J. Lee, and S. Shin, "Dutch listeners' perception of English lexical stress: A cue-weighting approach," *J Acoust Soc Am*, vol. 149, no. 6, pp. 3703–3714, 2021.

[5] Y. Zhang and A. Francis, "The weighting of vowel quality in native and non-native listeners' perception of English lexical stress," *J Phon*, vol. 38, no. 2, pp. 260–271, 2010.

[6] M. E. Beckman and J. Edwards, "Articulatory evidence for differentiating stress categories," *Phonological structure and phonetic form: Papers in laboratory phonology III*, pp. 7–33, 1994.

[7] T. Gay, "Physiological and acoustic correlates of perceived stress," *Lang Speech*, vol. 21, no. 4, pp. 347–353, 1978.

[8] B. Lindblom, "Spectrographic study of vowel reduction," *J Acoust Soc Am*, vol. 35, no. 11, pp. 1773–1781, 1963.

[9] D. B. Fry, "Experiments in the perception of stress," *Lang Speech*, vol. 1, no. 2, pp. 126–152, 1958.

[10] M. Liberman, *The Intonational System of English*. Cambridge: MIT Press, 1975.

[11] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology*, vol. 3, pp. 255–309, 1986.

[12] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[13] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," Doctoral dissertation, Massachusetts Institute of Technology, 1980.

[14] Y. R. Chao, *A grammar of spoken Chinese*. Univ of California Press., 1965.

[15] S. Duanmu, *The phonology of standard Chinese*. OUP Oxford, 2007.

[16] S. Ou, "Perception of English lexical stress with a marked pitch accent by native speakers of Mandarin," *Taiwan Journal of Linguistics*, vol. 14, no. 2, pp. 1–31, 2016.

[17] W. Choi, "Theorizing positive transfer in cross-linguistic speech perception: The Acoustic-Attentional-Contextual hypothesis," *J Phon*, vol. 91, p. 101135, 2022.

[18] K. Lemhöfer and M. Broersma, "Introducing LexTALE: A quick and valid lexical test for advanced learners of English," *Behav Res Methods*, vol. 44, no. 2, pp. 325–343, 2012.

[19] M. E. Beckman and G. Ayers, "Guidelines for ToBI labelling," *The OSU Research Foundation*, vol. 3, no. 30, pp. 255–309, 1997.

[20] A. L. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed, "Gorilla in our midst: An online behavioral experiment builder," *Behav Res Methods*, vol. 52, no. 1, pp. 388–407, 2020.

[21] N. A. Macmillan and C. D. Creelman, "1991. Detection theory: a user's guide," *Cambridge University PressMacmillanDetection theory: a user's guide1991*, 2005.

[22] R Core Team, "R: A language and environment for statistical computing; 2018." 2018.

[23] W. Choi, X. Tong, and A. G. Samuel, "Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners," *Cognition*, vol. 189, pp. 188–192, 2019.