SUPRASEGMENTAL CUES TO PERCEPTION OF ENGLISH LEXICAL STRESS BY MANDARIN SPEAKERS

Mosi He, Ting Zhang, Bin Li,

City University of Hong Kong, Hong Kong SAR, China
mosihe2-c@my.cityu.edu.hk, tzhang343-c@my.cityu.edu.hk, binli2@cityu.edu.hk

ABSTRACT

Lexical stress in English is signaled by a combination of suprasegmental cues, including pitch, duration and intensity. Previous research reported that non-native listeners might not use all cues when perceiving English stress. The current study investigated how pitch and duration cues affected stress identification in resynthesized English disyllabic nonwords by native speakers of English and Mandarin speakers who learn English as a foreign language (NE and EFL speakers). Within each stimulus, the pitch and/or duration cue suggested either a strong-weak or weak-strong stress pattern. Results showed that both listener groups were able to exploit pitch and duration cues to perceive lexical stress. NE speakers used pitch and duration cues similarly for both stress patterns. EFL speakers, however, relied more on the pitch cue than the duration cue, especially for stimuli with a strong-weak pattern. The perceptual differences were attributed to influences of EFL speakers' first-language phonology.

Keywords: lexical stress, speech perception, EFL, Mandarin Chinese, phonetic cue

1. INTRODUCTION

Lexical stress in English is encoded by a combination of pitch, duration and intensity cues [1-3]. Previous research has shown that non-native listeners might not use all cues when perceiving lexical stress due to influences of their L1. Native speakers of non-tone languages, such as Russian speakers, used duration and intensity as cues to lexical stress in English and disregarded pitch cues [4]. In contrast, native speakers of tone languages, such as Vietnamese [5], Mandarin Chinese [6-8], and Cantonese [9], exhibited stronger pitch perception for lexical stress. It is unclear regarding the extent of the effect of L1 on their weighting of pitch and duration cues. The current study aims to investigate how tonal speakers use pitch and duration as cues to English lexical stress.

As a tone language, Mandarin Chinese uses pitch variations in tones to distinguish lexical meanings, while English uses pitch in word and sentence stresses. In addition to pitch, lexical stress in English is indexed by intensity and duration which only serve as secondary cues to tones in Mandarin [10, 11].

The predominant role of pitch in L1 influences perception of lexical stress by Mandarin speakers. In a stress identification task, Mandarin speakers relied on pitch differences for perception of stress in English disyllabic nonwords [6]. Similar findings were reported in the other studies [7, 8]. Comparison between Mandarin speakers' and English speakers' use of duration as cues to stress, on the other hand, showed mixed results. Specifically, Qin et al. [8] reported that speakers of Standard Mandarin used more duration cues than speakers of Taiwan Mandarin in a sequence recall task of English disyllabic nonwords, though both Mandarin groups used fewer duration cues than native speakers of English did. In contrast, Lai [12] found that beginning Mandarin L2 learners of English showed comparable use of duration cues to native speakers of English in identification of lexical stress.

The weighting of pitch and duration to stress perception may be affected by stress pattern. In English, the position of stress within a word indicates word class and syllable structure [13]. Pitch cues were found to facilitate perception of stress when they indicate a weak-strong pattern [4], whereas the findings in [6] showed that a word-initial syllable with higher pitch was more likely to be perceived as the stressed syllable.

To provide further insights into tonal speakers' perceptual cues to non-native lexical stress, the current study examines the effects of pitch and duration cues, and stress pattern on stress identification patterns of resynthesized English disyllabic nonwords by native speakers of English and Mandarin speakers who learn English as a foreign language (NE and EFL speakers). It intended to answer the following research questions: 1) What suprasegmental cues do EFL speakers employ to perceive English lexical stress? 2) Is the cue weighting to English lexical stress similar of different between EFL and NE speakers? 3) Do strong-weak and weak-strong stress patterns affect EFL speakers' perception of English lexical stress?

2. METHODS

2.1. Stimuli preparation

The stimuli were two minimal stress pairs of English disyllabic nonwords (e.g., /'tu.ku/ vs. /tu.'ku/) with a CV.CV syllable structure (C = consonant, V = vowel). The four nonwords were produced four times in a carrier sentence *I said ___ twice.* by a female native speaker and a male native speaker of American English. Two of the four repetitions for each nonword and each speaker, with natural and correct stress placement, were selected and manipulated in Praat [14] in two steps.

Firstly, intensity of all vowels was normalized to 70 dB to minimize the influence of intensity differences on perception of stress. Intensity was excluded from examination because the auditory treatment of the intensity, loudness, is dependent on duration factors [15]. Secondly, F0 and duration of vowels were manipulated so that the stressed-unstressed difference was co-indicated by both cues (higher F0 and longer duration, "F0+Dur" condition) or by a single cue (higher F0 or longer duration, "F0_only" condition or "Dur_only" condition). Specifically, the duration parameters in F0_only condition and the F0 parameter in Dur_only condition were normalized to the average values of all vowels. In addition, the ratios between vowels of the first and second syllables (V_1/V_2) were determined with reference to those in [4, 12]. A ratio larger and smaller than 1 indicates a strong-weak stress and weak-strong pattern, respectively.

Table 1 shows the acoustic values for the resynthesized stimuli across cue conditions and stress patterns from the recordings of the male speaker. The strong-weak (SW) pattern was represented by higher F0 and/or longer duration on the vowel of the first syllable, whereas the weak-strong (WS) pattern was indicated by higher F0 and/or longer duration on the vowel of the second syllable.

2.1. Participants

Thirty-five native speakers of Mandarin Chinese (mean age = 21.1, SD = 1.4) including 20 females and 15 males, participated in the study. They were born and raised in northern China. All were undergraduate students at a university in Guangzhou, China. Self-reported language proficiency and results of a cloze test showed that they were intermediate to advanced EFL speakers. Twenty-one native speakers of American English (mean age = 32.5, SD = 9.6), including 11 females and 10 males, were recruited as a control group. None of the participants reported any speech, hearing, or visual disorders.

Table 1: Pitch and duration parameters for the resynthesized stimuli across cue conditions and stress patterns from the recordings of the male speaker.

			Cue condition		
			F0+Dur	F0_only	Dur_only
F0 (Hz)	SW	V_1	212	212	165
		V_2	133	133	165
		ratio	**1.61**	**1.61**	**1**
	WS	V_1	146	146	165
		V_2	168	168	165
		ratio	**0.87**	**0.87**	**1**
Dur (ms)	SW	V_1	187	160	188
		V_2	127	160	123
		ratio	**1.47**	**1**	**1.53**
	WS	V_1	119	160	113
		V_2	221	160	217
		ratio	**0.54**	**1**	**0.52**

2.4. Procedure

An identification task was conducted online using the Psytoolkit platform [16]. To identify inattentive participants, four attention checks were embedded in the task. Participants took the task in an uninterrupted quiet place, using a laptop or a computer. Those who used devices other than laptops/computers were automatically blocked by the platform.

Before the formal test session, participants took a practice session, where pairs of English real words such as "IMport" /ˈɪm.pɔːrt/ and "imPORT" /ɪm.ˈpɔːrt/ were presented. They listened to one stimulus at a time and indicated whether the stress was on the first or the second syllable by pressing a corresponding button on the keyboard. Feedback was given immediately with the sign "correct" or "wrong". There was a 1500 ms interval between stimuli. If no response were made within 3000 ms, the next audio would be played after 1500 ms.

The same procedure was used in the formal test session, except that there was no feedback given in the test session. There were two blocks in the test session. All stimuli across three cue conditions were presented randomly in each block where the genders of the speakers were counterbalanced.

3. RESULTS

In total, 1344 responses were collected (56 participants × 4 stimuli × 3 cue conditions × 2 blocks). One EFL and two NE participants failed 50% of the attention checks and were excluded from further analysis.

Responses were marked as correct if they matched the stress patterns indicated by the cue conditions. For each participant, accuracy rates were calculated by dividing the number of correct responses by the total number of trials in each cue condition and each stress pattern. Figure 1 shows the average accuracy for the two listener groups across three cue conditions. Both listener groups had accuracy rates above 60% across three cue conditions. But the perceptual patterns varied between the two listener groups. For the EFL group, F0_only condition and F0+Dur condition yielded similar accuracy, whereas lower accuracy was found in Dur_only condition. The NE group, however, had the lowest accuracy in F0_only condition.

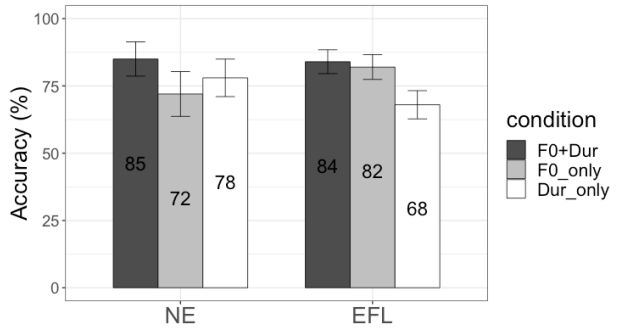

Figure 1: Mean accuracy rates for two listener groups by three conditions

A logistic regression mixed effects model (glmer) was fit to the data using the lme4 package [17] in R [18]. The dependent variable was accuracy, with correct responses coded as "1" and incorrect responses as "0". *Group* (NE vs. EFL), *cue condition* (F0+Dur, F0_only, Dur_only), *pattern* (SW vs. WS), and their interactions were entered as fixed factors. *Participants* and *items* were entered as random intercepts. All fixed factors were coded with sum coding. The Anova() function in the *car* package [19] was used to obtain significance of main effects and interactions. Post-hoc pairwise comparisons, wherever needed, were conducted with the *emmeans* package with Tukey adjustment [20].

Statistical results revealed that accuracy significantly differed across cue conditions [$\chi^2(2) = 12.8, p < .002$]. There was no significant difference in *group* [$\chi^2(1) = .31, p = .58$] or *pattern* [$\chi^2(1) = 2.30, p = .13$]. Two significant interactions were observed, *group* × *cue condition* [$\chi^2(2) = 7.20, p < .05$] and *pattern* × *cue condition* [$\chi^2(2) = 15.10, p < .001$], which will be elaborated in Sections 3.2 and 3.3 below. No other significant interaction was found.

3.1. Cue condition effect

Due to the significant difference among the three cue conditions, pairwise comparisons using Tukey adjustment were carried out. Results showed that F0+Dur condition brought about significantly higher accuracy than Dur_only condition ($b = .67, z = 3.58, p = .001$), while no significant difference was found between F0+Dur condition and F0_only condition, or F0_only condition and Dur_only condition (both bs $<.38$, zs < 1.99, ps $> .11$).

3.2. Group and condition interaction

The significant *group* × *cue condition* interaction indicated that the effect of cue condition varied between the NE and EFL groups. Post-hoc pairwise comparisons revealed that the EFL group had higher accuracy in conditions with F0 cue than Dur_only condition (both bs $> .75$, zs > 3.45, ps $< .002$), whereas the NE group had higher accuracy in conditions with duration cues than F0_only condition, as shown in Figure 1. But the differences did not reach statistical significance (all $|b|$s $< .57$, $|z|$s < 1.95, ps $> .13$).

Additionally, comparison within cue conditions showed that, in the F0_only condition, the EFL group had higher accuracy than the NE group, although this was not statitically significant ($b = .57, z = 1.79, p = .07$). No significant difference between the two groups was found in other cue conditions. Taken together, it indicated that the EFL group employed more F0 cues than the NE group did to encode stress in English nonwords.

3.3. Pattern and condition interaction

There was also a significant interaction between *pattern* and *condition*, suggesting that accuracy among cue conditions was influenced by stress patterns of stimuli in that condition. When averaged over *group*, for F0_only condition, the accuracy of SW nonwords was significantly higher than WS nonwords ($b = .80, z = 3.07, p < .01$). Contrary to F0_only condition, the accuracy in Dur_only condition was lower for SW nonwords than WS nonwords ($b = -.52, z = -2.15, p < .05$).

In addition, the effect of *cue condition* also varied within stress patterns. For SW nonwords, accuracy in conditions with F0 cue was significantly higher than Dur_only condition (both bs $> .95$, zs > 3.68, ps $< .001$). For WS nonwords, accuracy in F0+Dur condition was higher than F0_only condition, but it was not statistically significant ($b = .57, z = 2.27, p = .06$). There was no significant difference between F0_only and Dur_only conditions ($b = -.37, z = -1.49, p = .29$). Overall, listeners tended to identify a first

syllable with higher F0 as stressed and a second syllable with longer duration as stressed.

Although there was no significant three-way interaction among *pattern*, *condition*, and *group*, the two groups showed different perceptual patterns across stress patterns and cue conditions. As shown in Figure 2, the NE group used more F0 cues for SW nonwords and more duration cues for WS nonwords. The EFL group, on the other hand, constantly used more F0 cues than duration cues regardless of stress patterns, and more so for SW nonwords. In other words, EFL speakers showed similar use of F0 cues between the two stress patterns.

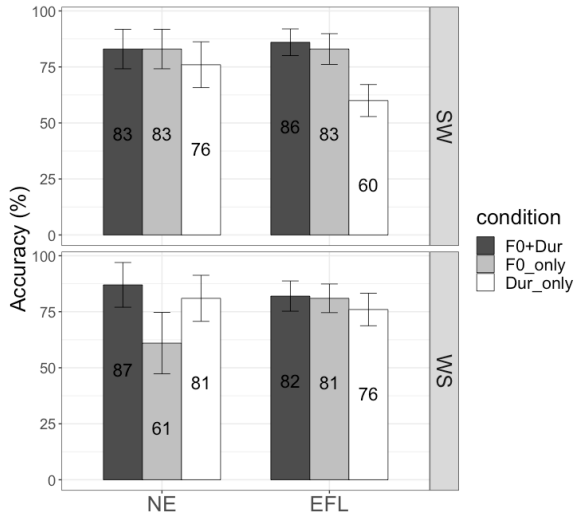

Figure 2: Mean accuracy rates for three conditions and two listener groups by two stress patterns

4. DISCUSSION

The current study examined stress identification in resynthesized English disyllabic nonwords by the NE and EFL speakers. Results showed that both listener groups were able to exploit pitch and duration cues to perceive lexical stress, which suggested that listeners did not attend exclusively to one cue but rather used a variety of cues at the same time [1]. The perceptual pattern, however, appears to vary between the two listener groups.

Firstly, the NE speakers used pitch and duration cues similarly, while the EFL speakers relied more on pitch cues than duration cues. These results are in line with those in [6] and [8]. EFL speakers' reliance on pitch cues to encode lexical stress in English could be explained by the influence of cue-weighting in their L1, where pitch variations in Mandarin contrasting lexical meanings. In addition, the NE speakers used more duration cues than the EFL speakers did, which suggested that lexical stress in English is encoded by the intergral of intensity and duration [15].

Secondly, the NE and EFL speakers employed duration cues to perceive stress in stimuli with the weak-strong pattern in a similar way, suggesting that duration cues also facilitated EFL speakers' stress perception when they indicated a weak-strong pattern. A possible explanation would be that the EFL speakers were influenced by the preference for a weak-strong pattern between syllables at the phonetic level in Mandarin [21, 22]. Specifically, a word-final syllable with sufficient duration and wide pitch range was perceived as more stressed [23]. The current findings contradict those in [8], which could be accounted for by different experimental paradigms. A sequence recall task was used in [8], which required a higher memory load.

Thirdly, the NE speakers alternated perceptual cues for different stress patterns. The EFL speakers, on the other hand, showed reliance on pitch cues irrespective of stress patterns. Similarly, Mandarin speakers in [12] were found to adopt a fixed and inflexible pattern between trochaic and iambic words, while NSE participants altered the weighting of cues accordingly. This could be attributed to a word-by-word learning strategy used by Mandarin speakers when processing stress patterns [24, 25]. Rather than associating stress patterns with grammatical functions, Mandarin speakers might associate an initial stress with the high-level tone and a final stress with the high-falling tone in Mandarin [6, 12]. Taken together, although both two groups used pitch cues to identify stress in English nonwords, the cross-language different weighting of pitch and duration across stress patterns indicated an overall influence of typological differences between tone languages and stress languages.

In conclusion, this study investigated the effects of pitch and duration, and stress pattern on perception of lexical stress in English by NE and EFL speakers. Results revealed the EFL speakers' reliance on pitch cues and inflexible use of cues between stress patterns, suggesting certain influences of their L1 tonal system. The findings contribute to L2 speech acquisition and studies of speech perception from a typological perspective. In future research, we will compare the perception of lexical stress by naïve native speakers of Mandarin to experienced EFL speakers.

5. ACKNOWLEDGEMENT

The study was supported by EDB(LE)/P&R/EL/175/12 funded by SCOLAR of Hong Kong and by TDG #6000807 funded by City University of Hong Kong.

6. REFERENCES

[1] Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech*, *1*, 126–152.

[2] Lieberman, P. 1960. Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, 32, 451–454.

[3] Lehiste, I. 1970. *Suprasegmentals*. MIT Press.

[4] Chrabaszcz, A., Winn, M., Lin, C. Y., Idsardia, W. J. 2014. Acoustic Cues to Perception of Word Stress by English, Mandarin, and Russian Speakers. *Journal of Speech, Language, and Hearing Research*, 57, 1468–1479.

[5] Nguyễn, T., Ingram, C. L. J. 2005. Vietnamese Acquisition of English Word Stress. *TESOL Quarterly*, 39(2), 309–319.

[6] Wang, Q. 2008. *Perception of English stress by Mandarin Chinese learners of English: An acoustic study*. Unpublished Ph.D. dissertation, University of Victoria.

[7] Yu, V. Y., Andruski, J. E. 2010. A Cross-Language Study of Perception of Lexical Stress in English. *Journal of Psycholinguistic Research*, 39(4), 323–344.

[8] Qin, Z., Chien, Y. F., Tremblay, A. 2017. Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics*, 38(3), 541–570.

[9] Lai, W. W. 2017. *The Production and Perception of English Lexical Stress by Hong Kong Cantonese Learners of English*. Unpublished Ph.D. dissertation, University of Hong Kong.

[10] Lin, T., Wang, L. J. 1992. *Course in Phonetics*. Beijing University Press.

[11] Liu, S., Samuel, A. G. 2004. Perception of mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47(2), 109–138.

[12] Lai, Y. W. 2009. *Acoustic Realization and Perception of English Lexical Stress by Mandarin Learners*. Unpublished Ph.D. dissertation, University of Kansas.

[13] Guion, S., Clark, J. J., Harada, T., Wayland, R. 2003. Factors Affecting Stress Placement for English Nonwords include Syllabic Structure, Lexical Class, and Stress Patterns of Phonologically Similar Words. *Language and Speech*, 46(4), 403–427.

[14] Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer. Version 6.1.08. http://www.praat.org/

[15] Turk, A. E., Sawusch, J. R. 1996. The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, 99 (6), 3782-3790.

[16] Stoet, G. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teaching of Psychology, 44(1), 24-31.

[17] Bates, D., Maechler, M., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

[18] R Core Team, R: A language and environment for statistical computing. Vienna, Austria. URL: http://www.R-project.org/, 2020.

[19] Fox, J., Weisberg, S. 2019. An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA.

[20] Lenth, R. V. 2018. *emmeans*: Estimated Marginal Means, aka Least- Squares Means. R package version 1.3.0. https://CRAN.R- project.org/package=emmeans.

[21] Chao, Y. R. 1968. A Grammar of Spoken Chinese. Reprinted in 2011. The Commercial Press.

[22] Yu, V. Y. 2021. Effects of Syllable Position, Fundamental Frequency, Duration and Amplitude on Word Stress in Mandarin Chinese. *Journal of Psycholinguistic Research*, 50(2), 293–312.

[23] Lin, M. C., Sun, G. H. 1984. Preliminary study on stress of disyllabic words in Beijing Mandarin. *Dialect*, (1), 57–73.

[24] Archibald, J. 1997. The acquisition of English stress by speakers of nonaccentual languages: lexical storage versus computation of stress. *Linguistics*, 35(1), 167–181.

[25] Wayland, R., Landfair, D., Li, B., Guion, S. G. 2006. Native Thai Speakers' Acquisition of English Word Stress Patterns. *Journal of Psycholinguistic Research*, 35(3), 285–304.