The production of English stress by L1 and L2 speakers: Beyond binary stress levels

Magdalena Ivok[1], H. Henny Yeung[1]
[1]*Simon Fraser University (Canada)*

Proficiency in English lexical stress, involving acoustic correlates such as pitch, duration, intensity, and vowel quality, is crucial for effective communication in English as a second language (L2). English learners often face challenges in producing and perceiving suprasegmental features, including lexical stress, particularly those learners with first languages (L1s) characterized by different speech rhythms [1, 2, 3].While previous work has mainly focused on primary-stressed (PS) and unstressed-reduced (UR) syllables in L1 and L2 speakers' productions, this study also examines an intermediate stress level (U) between PS and UR syllables, which includes unstressed- unreduced (UU) syllables, and similar secondary-stressed (SS) syllables [4]. This study specifically explores the acoustic cues for these stress levels (PS, U, and UR) in highly proficient L1-Mandarin speakers of L2-English and L1-English speakers.

Prior research into L1-English and L1-Mandarin speakers' English stress production [5, 3] has shown that acoustic contrasts for stress levels manifest primarily in duration, pitch, and vowel quality, but are weighted more towards pitch and duration in L1-Mandarin speakers (i.e., due to their familiarity with F0 as the primary cue for Mandarin tones). Less is known about how these speakers use acoustic cues across PS and U syllables, where vowel reduction is non-critical, and U and UR syllables, where vowel reduction is critical. We predicted that L1-Mandarin speakers would primarily use F0 to differentiate all English stress contrasts.

We recruited 20 L1-English (13 F, 4 M, 3 NB; *Range* = 19 years; *M* = 24 years) and 20 L1-Mandarin speakers (13 F, 7 M; *Range* = 7 years; *M* = 23 years) living in Anglophone Canada. All L1 Mandarin speakers lived in Canada for 1-5 years (*M* = 3.1), with an average of 13.6 years learning English. Participants produced 21 disyllabic and multisyllabic words containing the target vowel /u/ carrying three stress levels (e.g., PS, U, and UR in m**u**sic, ref**u**gee, and in**s**tr**u**ment, respectively) in both isolation and in a sentence context. Acoustic measurements implicated in English stress [4], mean F0, vowel duration, intensity, F1, F2, in addition to peak F0 location [3], were analyzed in different linear mixed-effects models, with fixed effects of stress level, language, and their interaction, presentation context, and random effects of word and subject. To account for variability from speech rate and gender, duration, F0 and intensity were normalized. Results, detailed in Table 1, indicate that L1-Mandarin and L1-English speakers employed similar cues, with duration crucial for distinguishing PS from U and UR syllables ($p < .001$, $p = .001$), but not UR and U syllables ($p = .252$). F0 played a role in differentiating PS and U syllables from UR syllables ($p < .001$, $p = .033$), but not PS and U syllables ($p = .153$). L1-Mandarin speakers exhibited higher intensity for UR syllables compared to L1-English speakers ($p < .001$), and later peak F0 in PS than UR syllables ($p = .002$) compared to L1-English speakers ($p = .740$). Vowel quality was not a significant correlate of stress.

These findings are significant in two ways. First, (highly proficient) L1-Mandarin speakers can produce 3 English stress levels, mirroring their L1-English counterparts. Second, for both types of speakers, duration was key to distinguishing PS from other stress levels, whereas spectral cues, like pitch, are crucial for distinguishing UR syllables from other stress levels. These findings challenge binary stress models and provide novel insights into the use of acoustic cues across levels of English stress. Furthermore, the study sheds light on L1-Mandarin speakers' English stress production, demonstrating slight differences in their use of intensity and F0 peak compared to L1 English speakers, potentially due to L1 influences. Future work will further analyze other languages, like Russian.

Tables and Figures

Table 1. Summary of Interaction, Language, Stress, and Context Effects on Acoustic Cues.

Acoustic Cue	Interaction of Language & Stress Type	Language	Stress Type	Context
F0 Mean	.265	.158	.001**	< .001***
Duration	.400	.041*	< .001***	< .001***
Intensity	.203	.961	.452	.901
F1	.239	.428	.225	.463
F2	.339	< .001***	.405	.004**
Peak F0 location	.003**	.092	.015*	< .001***

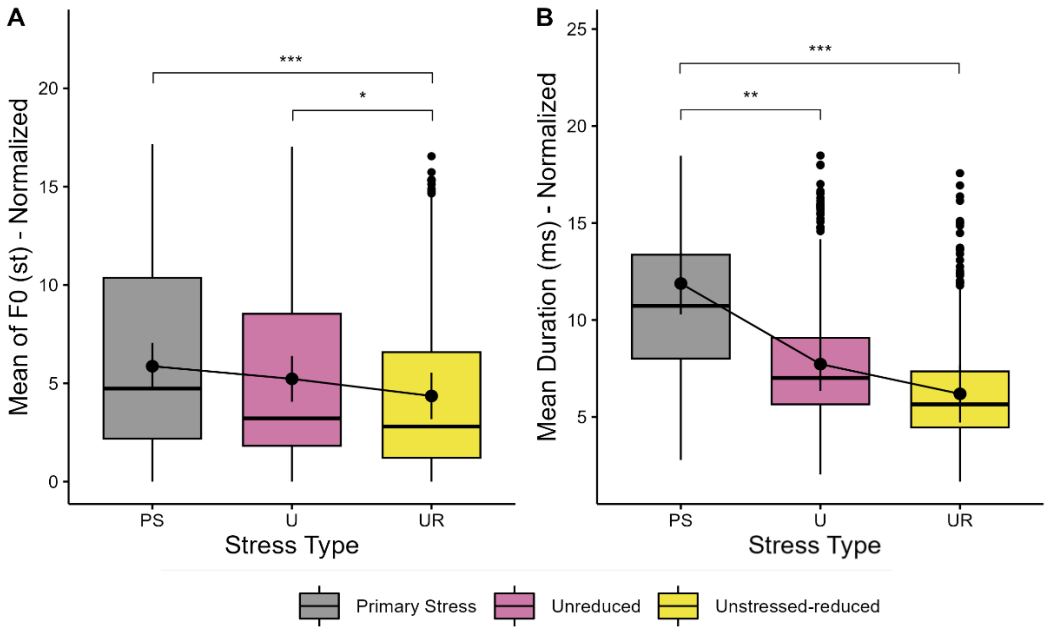

Fig. 1. Plot A displays Mean Fundamental Frequency (F0) and Plot B displays Duration across PS, U, and UR Stress Types (*$p < .05$, **$p < .01$, ***$p < .001$).

References

[1] Banzina, E., Dilley, C. L., Hewitt, E. L. (2016). The Role of Secondary-Stressed and Unstressed–Unreduced Syllables in Word Recognition: Acoustic and Perceptual Studies with Russian Learners of English. *J Psycholinguist Res*, 45, 813–831. doi:10.1007/s10936-015-9377-z

[2] Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, 38(3), 422-430. doi:10.1016/j.wocn.2010.04.001

[3] Zhang, Y., Nissen, S. L., Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America*, 123(6), 4498-4513. doi:10.1121/1.2902165

[4] Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97(3), 1893–1904. doi:10.1121/1.412063

[5] Mattys, S.L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, 62, 253–265. doi:10.3758/BF03205547