

Testing the Inverse Relationship Between Lexical Stress Strength and Macro-Rhythm Strength

Christine Prechtel

University of California, Los Angeles

In Jun's [1] model of prosodic typology, languages are classified based on how they mark word prominence. A word can be prominent by marking the head, such as with a pitch accent on the stressed syllable; the edge, such as with a boundary tone of a Prosodic Word (PW) or Accentual Phrase (AP); or both head and edge. She found that edge-prominence languages tend to have more rhythmic intonation patterns than head-prominence languages. These patterns of tonal rhythm are called macro-rhythm (MacR), and the typology categorized languages based on the degree of MacR strength. A language has strong MacR if its intonation regularly alternates between high and low tones over a word-sized prosodic unit in a phrase [1:522]. Jun further observed that languages with weaker MacR tend to have stronger acoustic correlates of stress, defined here as larger differences in duration and intensity between the stressed and unstressed syllables, while languages with stronger MacR tend to have weaker stress correlates. Therefore, she predicted an inverse correlation between the strength of lexical stress and the strength of MacR, reasoning that if word prominence is phonetically cued by pitch, then other cues such as duration and amplitude are less involved in cuing word prominence, and vice versa.

The current study tests this prediction in American English, Kolkata Bengali, and Uyghur. English is predicted to have the weakest MacR of the three languages because of its strong stress correlates [2] and its weaker MacR relative to other head-prominence languages [3, 4]. Bengali is predicted to have the strongest MacR because duration and intensity are not reliable cues for stress, which is fixed on the first syllable [e.g., 5]. Instead, stress is marked by F0 with a pitch accent on the stressed syllable (L*) followed by an AP edge tone (Ha), creating a rising tonal pattern [5, 6]. Uyghur is predicted to rank somewhere in between English and Bengali because lexical stress is acoustically marked by duration but not F0 [7, 8]. That is, Uyghur marks APs with a rising tonal pattern like Bengali but has no pitch accent on the stressed syllable [8], creating potentially contradictory prominence cues. A series of pilot experiments tested this predicted ranking in both production and perception, and the results of lexical stress production and MacR perception are reported here.

The lexical stress production experiment compared the duration ratios of stressed and unstressed vowels in 8 nonce disyllabic words in each language. Speakers from each language read short carrier sentences containing the target word in two prosodic conditions (accented and unaccented position). Duration ratios were log-normalized to account for speech rate [9]. Linear mixed effects models were run in R [10] with language and prosodic position as predictors and word and participant as random intercepts, and the results found that regardless of prosodic position, Uyghur had significantly smaller ratios than both English and Bengali, and there was no significant difference between English and Bengali ratios.

In the MacR perception experiment, 50 English participants listened to 30 phonetically manipulated sentences (10 utterances per language) two times in two conditions and rated how tonally rhythmic or "melodic" they sounded on a 5-point scale. A previous pilot study found that participants' responses differed depending on their musical background, so participants were asked at the start of the experiment about their music background, specifically whether they had formal music training (e.g., played an instrument or sang in a choir). The stimuli consisted of utterances taken from read speech corpora of each language [11, 12, 13] that were 13-20 syllables long and contained a single Intonational Phrase (IP). They were further

manipulated and presented in two conditions: Filtered and F0-only. In the Filtered condition, which was presented first, utterances were low-pass filtered to remove segmental information but still retained some information about syllable structure. In the F0-only condition, the same utterances were resynthesized as a continuous hum, retaining only the F0 of the original utterance. In each condition, linear mixed effects models were run with language and the participant’s music background as predictors, and the results found that English and Uyghur utterances were rated significantly less melodic than Bengali utterances but there was no significant difference between English and Uyghur, although English tended to have less melodic ratings than Uyghur in the F0-only condition (Figure 1). Rating responses were also affected by music background: participants who reported having formal music training rated English and Uyghur as significantly less melodic than Bengali, while participants with no music background did not have significantly different ratings between languages.

Taken together, these results suggest that the inverse relationship between lexical stress and MacR strength may be more nuanced in perception than in production. As expected, English had large duration ratios (corresponding to strong stress) and weaker perceived MacR, but while Bengali had the strongest perceived MacR, it also had larger than expected duration ratios. One possible explanation for this result is that participants may have treated nonce words differently than real words. Uyghur did not have stronger perceived MacR than English, which was surprising because, similarly to Bengali, it marks APs with a rising tonal pattern, so it was expected to sound more strongly macro-rhythmic than English. However, given that participants with no music background or training performed worse than those with musical training, this result may suggest that the stimuli were not ideal for testing this hypothesis. Another possible explanation for Uyghur data is that the Uyghur utterances had smaller and shallower rises on average than Bengali utterances, which could have contributed to the perception of Uyghur being less melodic despite having more consistent L/H tonal alternations than English. Therefore, the perception results may not fully reflect the MacR strength of each language. To further test the predicted inverse relationship, future work will quantify and compare MacR strength between the three languages in production data.

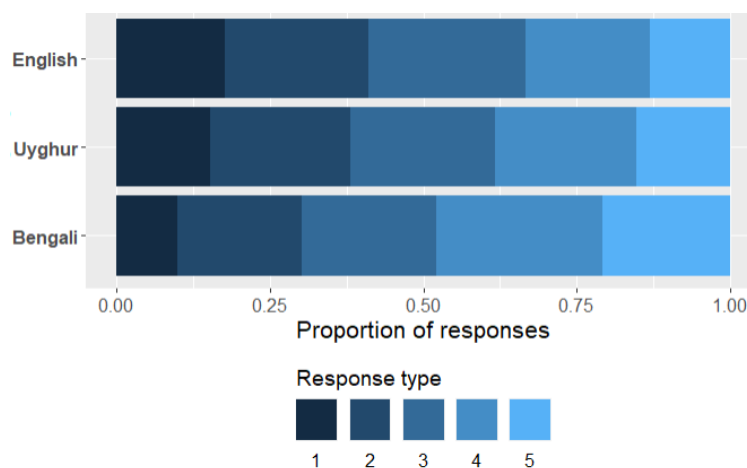


Figure 1. Proportion of rating responses of utterances in each language for the F0-only condition. 1 (darkest shade, left) = not melodic, 5 (lightest shade, right) = very melodic.

REFERENCES

- [1] Jun, S-A. (2014). Prosodic typology: by prominence type, word prosody, and macro-rhythm. In Jun, S-A (Ed.), *Prosodic Typology II*. Oxford: Oxford University Press, 520-539.
- [2] Mairano, P., Santiago, F., & Romano, A. (2015). Cross-linguistic differences between accented vs unaccented vowel durations. *Proceedings of the 18th International Congress of Phonetic Sciences* (Glasgow, Scotland).
- [3] Polyanskaya, L., Busà, M.G., & Ordin, M. (2019). Capturing cross-linguistic differences in macro-rhythm: the case of Italian and English. *Language and Speech*. doi: 10.1177/0023830919835849.
- [4] Prechtel, C. (2020). Quantifying Macro-rhythm in English and Spanish: A comparison of tonal rhythm strength. M.A. thesis. University of California, Los Angeles.
- [5] Hayes, B. & Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9, 47-96.
- [6] Khan, S.D. (2014). The intonational phonology of Bangladeshi Standard Bengali. In Jun, S-A (Ed.), *Prosodic Typology II*. Oxford: Oxford University Press, 81-117.
- [7] Yakup, M. & Sereno, J. A. (2016). Acoustic correlates of lexical stress in Uyghur. *Journal of the International Phonetic Association* 46.1: 61-77.
- [8] Major, T. & Mayer, C. (2018). Towards a phonological model of Uyghur intonation. *Proceedings of the 9th International Conference on Speech Prosody* (Poznań, Poland), 744-748.
- [9] Beckman, M. (1986). *Stress and Non-Stress Accent*. Dordrecht, The Netherlands: Foris Publications Holland.
- [10] R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- [11] Das, B., Mandal, S., & Mitra, P. (2011). "Bengali speech corpus for continuous automatic speech recognition system," *Proceedings of the 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA)*, 51-55.
- [12] Free ST American English Corpus, Surfing.ai. [Online]. Retrieved from <https://www.openslr.org/45/>.
- [13] Rouzi, A., Yin, S., Zhang, Z., Wang, D., Hamdulla, A., & Zheng, F. (2017). "THUYG-20: A free Uyghur speech database," *Journal of Tsinghua University*, vol. 57, 182-187. doi: 10.16511/j.cnki.qhdxxb.2017.22.012.