

Real-time Integration of Acoustic Cues of Lexical Pitch Accents: An Eye Tracking Study

Dušan Nikolić¹, Kiana Gardner², Angeliki Athanasopoulou²

¹*Austrian Academy of Sciences*, ²*University of Calgary*

The mapping of continuous and variable acoustic information onto segmental and prosodic categories involves multiple acoustic dimensions [1]. These acoustic cues to a single category are often asynchronous. There are two major models that describe the integration of acoustic cues. First, the Buffered Integration model claims that the cues are stored in a memory buffer, and once they are all available in the signal, listeners make a lexical decision [2] [3]. Second, the Continuous Cascaded model claims that listeners make a lexical decision as soon as the first cue is available, and then as the other cues arrive, the listeners update this decision [1] [4] [5]. Our understanding of the time course of cue integration is primarily based on segmental contrasts in words. Only a handful of studies explore lexical stress and tone cue integration [6] [7], all of which support the continuous cascaded model. Yet there are no studies on the integration of acoustic cues of lexical pitch accents. These could uniquely test the models as the synchronicity and the weight of the acoustic cues signalling lexical pitch accents varies.

In particular, we investigate how Serbian (native) and English (naïve, non-native) speakers process Serbian lexical pitch accent cues. Serbian has a typologically rare four-way accent contrast signalled through duration (long or short vowels) and pitch (falling or rising tone). Based on these two cues, there are three important contrast types: 1) vowel length, 2) tone, 3) accent (or both length and pitch). Duration of the accented syllable largely determines the contrast in length, while the pitch movement from the end of the accented to the beginning of the post-accented syllable differentiates the type of tone. According to the Buffered Integration model, we hypothesize that the integration of cues on tone, length, and accent will be delayed. In contrast, according to the Continuous Cascaded model, the integration of cues will be immediate.

To test these hypotheses, (so far) we tested 10 Canadian-English and 2 Serbian listeners. The stimuli were 16 trisyllabic nonwords synthesized in Praat. The words created minimal pairs based on length, tone or accent or based on vowel quality for fillers in the first syllable. Each word was linked to a nonsense object [8]. Following the visual world paradigm [9], participants completed a learning phase (320 trials in 7 blocks) where they learned all the word-object pairs (Serbian_{ACC} = 79%, English_{ACC} = 80%). Then, they were asked to find the object named among four possible objects on the screen, while their eye movements were recorded (48 trials in 2 blocks). To collect the eye-tracking data, we used the Tobii Pro Fusion 120Hz eye-tracker, and the experiment was designed and run in E-Prime. We analyzed the data by calculating the proportion of fixations towards the target, competitor, and 2 distractors in 100ms time bins. The predicted probability of fixations and group comparisons were carried out using GAMMs [10].

The results show that English listeners processed tone faster than length (Fig. 1), while Serbian listeners processed length faster than tone (Fig. 2), which indicates greater weighting of length in comparison to tone for native listeners, as previously shown [11]. This suggests that language experience modulates the weighting of acoustic cues without preventing non-native listeners from forming stable lexical representations. The results show that English listeners appear to be more sensitive to lexical pitch modulations than duration, and they process pitch as it arrives, while Serbian listeners process duration immediately. The findings thus support the Continuous Cascaded model in which listeners begin lexical processing as soon as the first acoustic cue is encountered and refine their interpretation as subsequent cues arrive.

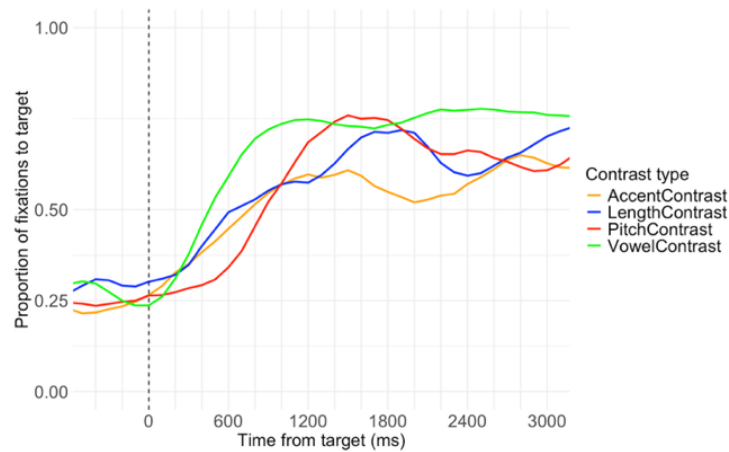


Fig. 1. English listeners' proportion of fixations to target

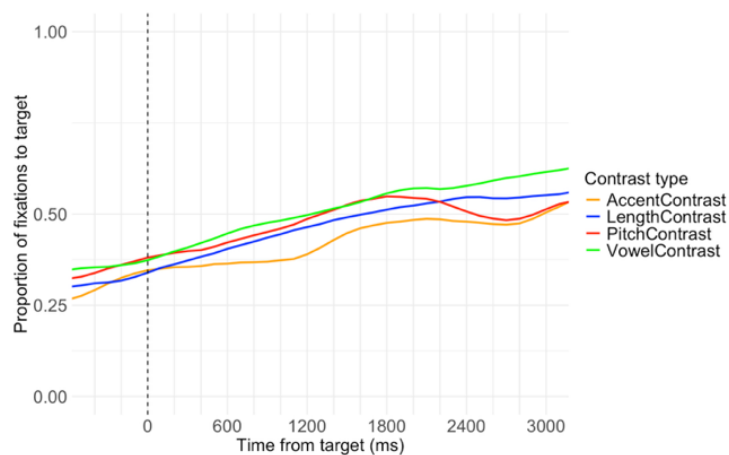


Fig. 2. Serbian listeners' proportion of fixations to target

References

- [1] Kim, H., Tremblay, A., & Cho, T. (2024). Perceptual Cue Weighting Matters in Real-Time Integration of Acoustic Information During Spoken Word Recognition. *Cognitive Science*, 48(12), e70026.
- [2] Galle, M. E., Klein-Packard, J., Schreiber, K., & McMurray, B. (2019). What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cognitive science*, 43(1), e12700.
- [3] Schreiber, K. E., & McMurray, B. (2019). Listeners can anticipate future segments before they identify the current one. *Attention, Perception, & Psychophysics*, 81(4), 1147-1166.
- [4] McMurray, B., Galle, M., Farris-Trimble, A., & Seedorff, M. (2018). What the/f/? We're not done with fricatives yet. Integrating across time and frequency bands. *The Journal of the Acoustical Society of America*, 144(3_Supplement), 1716-1716.
- [5] Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of phonetics*, 41(2), 101-116.
- [6] Jesse, A., Poellmann, K., & Kong, Y. Y. (2017). English listeners use suprasegmental cues to lexical stress early during spoken-word recognition. *Journal of Speech, Language, and Hearing Research*, 60(1), 190-198.
- [7] Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics*, 87, 101059.
- [8] Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods*, 48(4), 1393-1409.
- [9] Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 55(7), 3461-3493.
- [10] Porretta, V., Kyröläinen, A. J., van Rij, J., & Järvikivi, J. (2015). A step forward in the analysis of visual world eye-tracking data. *New developments in the quantitative study of languages*, 64.
- [11] Nikolić, D., & Winters, S. (2022). Are Serbian and English listeners insensitive to lexical pitch accents in Serbian?. *Phonetica*, 79(4), 397-423.