

To wait or not to wait? An eye-tracking account of the processing of acoustic cues of lexical pitch accents

Dušan Nikolić (dusan.nikolic@oeaw.ac.at)¹, Eva Reinisch¹, ¹Acoustics Research Institute,
Austrian Academy of Sciences

To accurately process speech, listeners must be able to recognize individual words. This presents a challenge as words unfold over time and listeners must integrate multiple acoustic cues that are often spread out over several hundred milliseconds [1]. Prior work contrasts two approaches: (1) a buffered integration approach, in which listeners store acoustic cues in their working memory and make a final lexical decision only once sufficient information is available [2]; and (2) a continuous cascaded approach, which shows that the earliest available cues influence lexical activation immediately [3]. In the present study, we explore the processing of acoustic cues to lexical pitch accents, a suprasegmental category spanning across at least two syllables, requiring the cue integration and spanning over time.

We compare native Serbian listeners, whose language has contrastive pitch accent, with German listeners with lexical stress and no prior exposure to Serbian. Serbian words exhibit a typologically rare four-way contrast expressed through length and pitch, yielding three key distinctions: length, pitch, and combined accent [4]. Participants were required to learn the names of 16 nonsense objects [5] in a picture-selection task: on each trial, they heard a nonce word and selected the corresponding object from a display of four, with eye movements to target versus competitor objects used to track the moment-by-moment resolution of lexical competition. The acoustic stimuli were synthesized in Praat [6] so that they contrasted in length (short vs. long vowel duration), pitch (high vs. low f_0 contour), accent, or a vowel. Participants completed five blocks of training with feedback and one block of testing (without feedback) for a total of 384 trials. We tracked eye movements during the final training block and the testing phase using a visual world paradigm [7]. Separate GAMMs [8] were fitted for Serbian and German listeners, with the target-over-competitor fixation advantage as the dependent variable, the contrasts (length, pitch, accent, and vowel) as the fixed factor, and random smooths for participants and trials.

First results show that both Serbian and German listeners process length cues faster than pitch cues, as evidenced by the difference in target-competitor disambiguation points in Figure 1. Accent contrasts lead to faster word recognition than contrasts coded by length and pitch individually, suggesting that initial pitch movement, when combined with length, is sufficient for listeners. The GAMMs reveal that the target-over-competitor advantage significantly differs across all conditions for both groups ($p < .001$), and that this advantage is weakest for the pitch contrast ($p < .001$). Taken together, these findings suggest that the processing of lexical prosody supports the continuous cascaded model, as the earliest cues are immediately used during word processing.

[1] McMurray B, Clayards MA, Tanenhaus MK, Aslin RN. Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic bulletin & review*. 2008. Dec; 15(6):1064-71.

[2] Kim H, Tremblay A, Cho T. Perceptual Cue Weighting Matters in Real-Time Integration of Acoustic Information During Spoken Word Recognition. *Cognitive Science*. 2024 Dec; 48(12):e70026.

[3] Galle ME, Klein-Packard J, Schreiber K, McMurray B. What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cognitive science*. 2019 Jan; 43(1):e12700.

[4] Subotić L, Sredojević D, Bjelaković I. *Fonetika i fonologija: ortoepska i ortografska norma standardnog srpskog jezika*. Novi Sad: Filozofski fakultet. 2012.

[5] Horst JS, Hout MC. The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods*. 2016 Dec; 48(4):1393-409.

[6] Boersma, Paul & Weenink, David. *Praat: doing phonetics by computer* [Computer program]. Version 6.4.65. 2026.

[7] Ito A, Knoeferle P. Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*. 2023 Oct; 55(7):3461-93.

[8] Porretta V, Kyröläinen AJ, van Rij J, Järvikivi J. A step forward in the analysis of visual world eye-tracking data. *New developments in the quantitative study of languages*. 2015.

Figures

Figure 1: German (left) and Serbian (right) listeners' proportion of fixations to target and competitor. (Currently actively recruiting participants).

