

The primary importance of onsets: Timing and prediction in speech segmentation

Laurence White¹, Silvia Benavides-Varela², Katalin Mády³ & Sven Mattys⁴

¹Newcastle University; ²University of Padua;

³Hungarian Academy of Sciences, Budapest; ⁴University of York

Speech sounds run one into the next, even across word boundaries. Whilst word recognition often suffices for speech segmentation when hearing our native language(s), structural ambiguity or lack of context may require listeners to exploit the available range of non-lexical cues to word boundaries [1]. Such cues are particularly important in difficult listening conditions and for first language (L1) or second language (L2) learners. Paradoxically for acquisition, however, some non-lexical cues – such as boundary-related phonotactic regularity – rely on language-specific knowledge, thus can only be incrementally exploited by new learners. Cross-linguistic consistency of prosodic segmentation cues may resolve this paradox: certain timing and intonational features may be language-independent and thus potentially available for use by L1 and L2 learners early in acquisition [2].

In the first series of experiments, we used an artificial language (AL) stream of trisyllabic nonwords (e.g., *nudopa*, *subako*, etc.) to compare the segmentation power of three durational cues across native English, Hungarian and Italian speakers. Word- or phrase-final vowel lengthening (here **Syl3-V**, e.g., *nudopa*) is widely observed across languages [3] and has been held to be a universal segmentation cue [4]. If so, lengthening of the first-syllable vowel in polysyllabic words (**Syl1-V**, *nudopa*) should mislead listeners into perceiving a mid-word prosodic boundary, at least in the absence of intonational cues. Finally, word-onset consonant lengthening (**Syl1-C**, *nudopa*) is seen in prosodically-diverse languages [5], although its importance for segmentation has been relatively little explored [e.g., 6 - Dutch, 7 - English] and not within a single experiment using the same materials for multiple languages.

Listeners were exposed to the AL stream for six minutes, then completing a 24-item two-alternative forced-choice task to test how their nonword recognition was affected by the three lengthening conditions (**Syl1-C**, **Syl1-V**, **Syl3-V**) relative to an evenly-timed baseline (**Flat**). (In this between-subjects design, each participant only heard one AL timing condition.) As Figure 1 shows, participants across languages showed higher nonword recognition in the **Syl1-C** condition than the **Flat** baseline (English $p = .010$, Hungarian $p < .0001$, Italian $p = .012$), indicating they all benefited, during AL exposure, from the segmentation support provided by word-initial consonant lengthening. This accords with the critical importance of onsets for lexical processing, as they initiate activation of word hypotheses [8]. By contrast, **Syl3-V** improved recognition over **Flat** only for English listeners ($p = .002$), contradicting the hypothesis that final-vowel lengthening is a universal segmentation cue, but in line with recent findings for Italian [9]. For all languages, **Syl1-V** was equivalent to or worse than **Flat**, reinforcing the conclusion that it is *localised* word-initial consonant lengthening that is key.

A second series of segmentation experiments, with native English listeners, examined how timing cues are processed, using a method designed to enable within-subject comparison of different cues. On each trial, participants heard a 12-syllable nonsense utterance (e.g., *pabikugolatudaropi*) followed by a nonword target (e.g., *golatu*) and had to decide if the target had been in the preceding utterance. We used the same duration conditions – **Flat**, **Syl1-C**, **Syl1-V**, **Syl3-V** – but varied the position of the target within the preceding carrier utterance: early, medial or late. Results reinforced the primary importance of word onsets: nonword targets were detected better than the **Flat** baseline only when onset consonants were lengthened, and then only when the **Syl1-C** targets were late in the carrier ($p < .05$ vs all other conditions). This pattern strongly suggests the use of a predictive timing mechanism [10], with listeners exploiting foregoing speech rate within utterances to gradually build up durational expectations and thus detect lengthened onset consonants as word boundary cues.

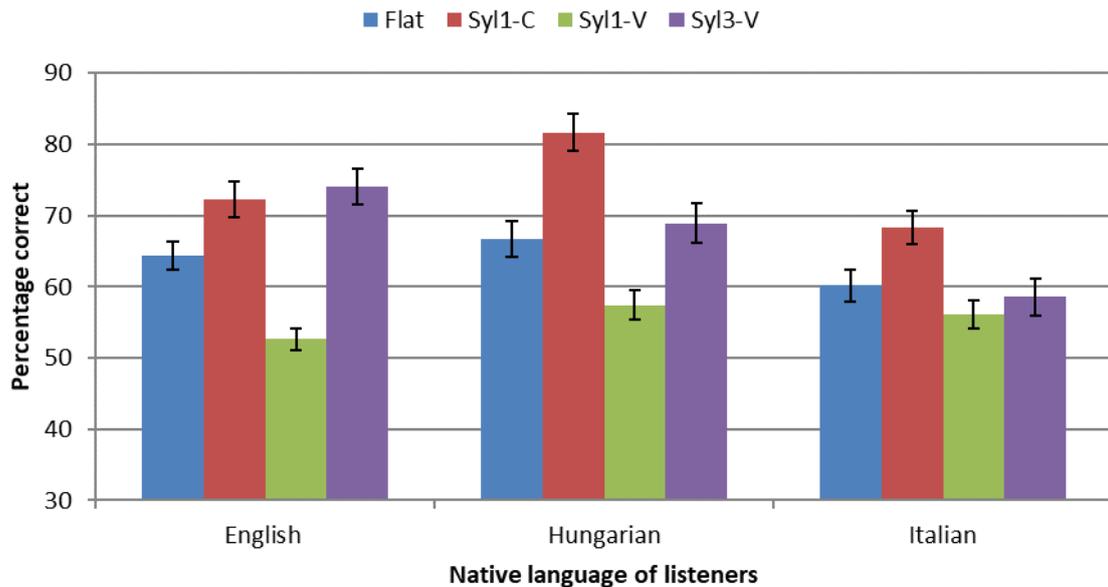


Figure 1. Mean correct recognition of nonwords after six minutes' exposure to an artificial language comprised of four trisyllables: *nudopa*, *subako*, *bitusa*, *ripolu*. See abstract (paragraph 2) for details of the timing conditions: **Flat**, **Syl1-C**, **Syl1-V**, **Syl3-V**.

- [1] S. L. Mattys, L. White, and J. F. Melhorn, 'Integration of multiple speech segmentation cues: a hierarchical framework.', *J. Exp. Psychol. Gen.*, vol. 134, no. 4, pp. 477–500, 2005.
- [2] A. Fernald and G. McRoberts, 'Prosodic bootstrapping: A critical analysis of the argument and the evidence', *Signal Syntax Bootstrapping Speech Gramm. Early Acquis.*, pp. 365–388, 1996.
- [3] J. Fletcher, 'The prosody of speech: Timing and rhythm', *Handb. Phon. Sci. Second Ed.*, pp. 521–602, 2010.
- [4] M. D. Tyler and A. Cutler, 'Cross-language differences in cue use for speech segmentation', *J. Acoust. Soc. Am.*, vol. 126, no. 1, pp. 367–376, 2009.
- [5] P. Keating, T. Cho, C. Fougeron, and C.-S. Hsu, 'Domain-initial articulatory strengthening in four languages', in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, J. Local, R. Ogden, and R. Temple, Eds. Cambridge: Cambridge University Press, 2004, pp. 143–161.
- [6] H. Quené, 'Segment durations and accent as cues to word segmentation in Dutch', *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 2027–2035, 1993.
- [7] L. White, S. L. Mattys, L. Stefansdottir, and V. Jones, 'Beating the bounds: Localized timing cues to word segmentation', *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 1214–1220, 2015.
- [8] A. Content, R. K. Kearns, and U. H. Frauenfelder, 'Boundaries versus onsets in syllabic segmentation', *J. Mem. Lang.*, vol. 45, no. 2, pp. 177–199, 2001.
- [9] M. Ordin, L. Polyanskaya, I. Laka, and M. Nespors, 'Cross-linguistic differences in the use of durational cues for the segmentation of a novel language', *Mem. Cognit.*, pp. 1–14, 2017.
- [10] E. Reinisch, A. Jesse, and J. M. McQueen, 'Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue', *Lang. Speech*, vol. 54, no. 2, pp. 147–165, 2011.