

The Role of Durational Reduction on Speech Segmentation: Evidence from Artificial Language Learning

Shu-chen Ou¹, Hung-lin Wu

¹ Department of Foreign languages and literature, National Sun Yat-Sen University, Taiwan

Background: The present study examines how durational reduction influences speech segmentation using an artificial language learning (ALL) experiment. In listening to continuous speech, listeners must identify discrete units such as words, using various distributional cues, such as transitional probability (TPs) [1]. Previous research has also shown that salient prosodic cues, such as final vowel lengthening or onset consonant strengthening of a syllable, facilitate segmentation across languages. However, much less is known about the role of reduced prosodic cues, despite their systematic occurrence in natural speech.

Prior studies suggest that phonetic reductions, such as durational shortening, often indicate continuity or a single unit [2–4]. This pattern is also observed in Mandarin Chinese (e.g., huǒ chē zhàn [xwǒ tʂʰɿ tʂan] “train station”, in which the middle syllable is frequently shortened and reduced) [5, 6]. Therefore, shortening the middle syllable may function as a cue to holistic units and facilitates segmentation. The present study examines whether durational reduction, a form of reduced prosody, influences speech segmentation.

Method: Using an artificial language learning paradigm, we designed six “words,” each comprising a meaningless CVCVCV trisyllabic sequence (e.g., [ba.nu.me], [bi.mo.na]). Each syllable used to construct these words was recorded separately and manipulated to have the same duration (300ms, with equal duration for the consonant and vowel) and a flat F0 contour (i.e., 130 Hz). TPs between syllables served as the sole cues for word definition.

Three additional experimental conditions were then created in which TPs were held constant while durational reduction was manipulated. In these conditions, the duration of one target syllable was reduced by 50% (i.e., from 300ms to 150 ms). In the Middle Shortening (MS) condition, reduction was applied to the middle syllable, whereas in the Initial Shortening (IS) and Final Shortening (FS) conditions, reduction was applied to the first and third syllable respectively. The phonetic manipulation of an example was shown in Figure 1.

The experiment was conducted using E-Prime 3.0. Native Mandarin listeners (N = 120) were recruited and randomly assigned to one of the four conditions (mean age = 26.78; SD = 10.65). During the learning phase, participants initially heard 150 repetitions of each of the six words, which were pseudo-randomly concatenated without pauses. They subsequently took a two-alternative forced-choice test, in which each trial presented one actual word from the artificial language and one part-word composed of syllables drawn from two artificial words (e.g., [me.bi.mo] from [ba.nu.me] and [bi.mo.na]), and participants were asked to identify the word belonging to the language.

Results: Response accuracy, as shown in Figure 2, showed 55.56% in the TP-only condition, 55.46% in the IS condition, 65.28% in the MS condition and 52.87% in the FS condition. A mixed-effects logistic regression model was fitted to the binary responses (correct vs. incorrect), with Condition included as the fixed factor. Random effects included a random intercept for the words of the artificial languages. The results revealed a significant effect of shortening position. The MS condition significantly facilitated segmentation compared to the TP-only baseline ($\beta = 0.307$, $SE = 0.13$, $z = 2.29$, $p = 0.0217$). In contrast, The FS condition significantly hindered performance ($\beta = -0.348$, $SE = 0.13$, $z = -2.64$, $p = 0.0084$), suggesting that durational reduction at word boundaries disrupts speech perception. No significant effect was found for the IS condition ($\beta = -0.105$, $SE = 0.13$, $z = -0.79$, $p = 0.428$).

Discussion: The results demonstrate a robust effect of shortening position. Shortening the middle syllable significantly facilitated speech segmentation relative to the TP-only baseline, indicating that

durational reduction at highly predictable, word-internal positions enhances unit identification. In contrast, shortening the final syllable significantly impaired performance, suggesting that durational reduction at word boundaries disrupts speech segmentation, consistent with previous studies. Taken together, these findings suggest that durational reduction functions as a prosodic cue to internal unit structure rather than as a boundary marker. Reduced prosody therefore plays an active role in speech segmentation, but only when aligned with distributional predictability (e.g., TPs).

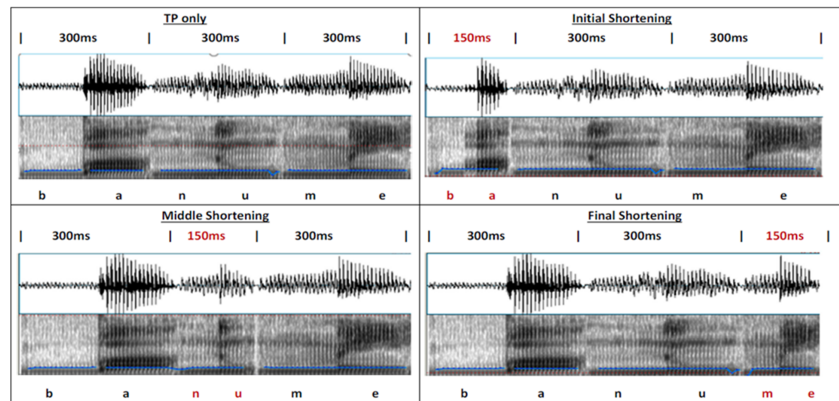


Figure 1: Durational manipulation of the example word [banume] under four experimental conditions. Panels show (upper left) TP-only baseline, (lower left) Middle Shortening (MS), (upper right) Initial Shortening (IS), and (lower right) Final Shortening (FS).

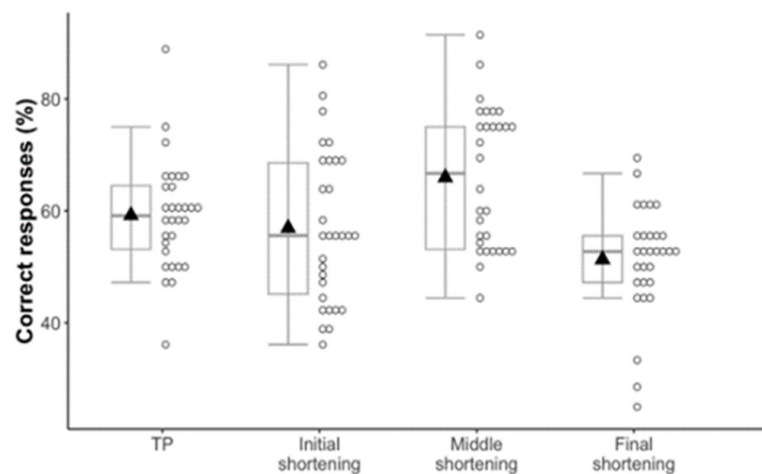


Figure 2: Percentages of correct responses of individual participants (empty circles) and the means of different experimental conditions (filled triangles).

References

- [1] Saffran, J., Newport, E., & Aslin, R. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, 35(4), 606–621. <https://doi.org/10.1006/jmla.1996.0032>
- [2] Arnon, I., & Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3), 349–371. <https://doi.org/10.1177/0023830913484891>
- [3] Arnon, I., & Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, 9(3), 377–400.
- [4] Tremblay, A., & Tucker, B. (2011). The effects of transitional probabilities on the articulation of French liaison. *Journal of Memory and Language*, 64(3), 259–269.
- [5] Chen, A.-H. (2021). Durational Patterns of Recurrent Multiword Combinations in Mandarin Spontaneous Speech Production. *Language and Speech*, 64(3), 742–767. <https://doi.org/10.1177/0023830920966010>
- [6] Chen, A. (2023). *F0-based pairwise variability index: A prosodic metric for holistic language processing*. Guarant International.