Distal rhythmic structure and speech rate in perception: exploring prosodic and rate-dependent influences

Jeremy Steffman

University of California, Los Angeles (USA)
jsteffman@g.ucla.edu

Background: The role of prosody in modulating listeners' perception of phonetic cues has been a recent topic of interest in the literature. Work investigating these questions has focused on adjacent, or proximal prosodic effects (e.g. boundary marking) [1,2]. The present study explores how *distal*, non-adjacent rhythmic patterns influence listeners' processing of phonetic (durational) cues in speech, as informed by the documented importance of rhythmic/metrical structure in word segmentation/lexical processing [3,4].

Listeners are argued to project a prosodic structure in segmenting speech based on the rhythmic properties of a distal preceding context, grouping ambiguous strings of sounds into words based on patterns of alternating duration/pitch [4]. These findings are couched in the *perceptual grouping hypothesis* (PGH), which predicts that perceptual grouping of alternating patterns will affect listeners' expectations about incoming information in the speech signal. The present study extends the PGH to test if grouping influences perception of vowel duration as a cue to coda voicing [5]. Predictions based on rhythmic grouping are compared with expected durational contrast effects [6] (outlined below). Three experiments tested rhythmic effects in perception by manipulating the rhythmic structure and rate of a series of CV syllables preceding a target sound from a "coat" ~ "code" continuum which varied only in vowel duration (90-150ms, 15ms steps). The target was categorized in a 2AFC task. Participants were native speakers of American English.

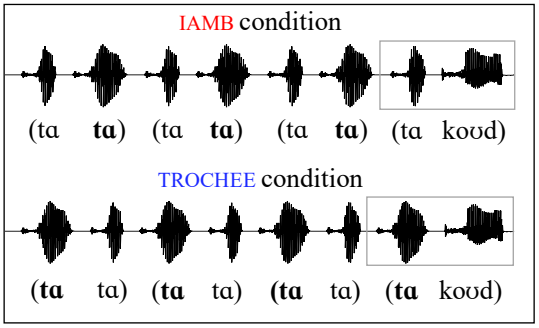
Experiments 1 & 2: In Experiment 1 (n = 32), the crucial manipulation was whether the precursor formed a sequence of durational trochees (long-short), or iambs (short-long). Preceding syllables were CV ([t^hɑ]), and were identical except for duration of the vowel. Three trochaic or iambic feet preceded a final foot in which the target was grouped with either a long syllable, forming a potential trochee (the TROCHEE condition), or a short syllable, forming a potential iamb (the IAMB condition; Fig. 1). Following the PGH, if listeners group the target as the second syllable in a foot based on preceding durational alternations, expectations about the duration of the target vowel might change based on whether it was the implied second syllable in a trochee (where it would be expected to be shorter), or an iamb (where it would be expected to be longer). An expectation of relative lengthening in the IAMB condition predicts that "code" responses *decrease* in the *IAMB condition*, reflecting the influence of distal rhythmic patterns. Proximal durational contrast effects [6] predict an opposite shift in categorization. Given that a relatively *longer* syllable precedes the target in the TROCHEE condition, categorization would be expected to shift to *longer* required vowel durations for a "code" response (*decreasing* "code" responses in the *TROCHEE condition*). As these two effects make opposite predictions, this design directly tests whether distal rhythmic structure or proximal duration will influence perception. Results are assessed by mixed-effects logistic regression (by-participant random intercepts, maximal random slopes). **Results:** A main effect of rhythm was found whereby the IAMB condition significantly *decreased* "code" responses ($p < 0.01$; Fig. 2). This result supports predictions based on the PGH, suggesting that listener *expectations* about duration based on rhythmic grouping influence the perception of vowel duration as a cue to voicing, in defiance of proximal contrast effects. These results thus provide novel insight into how rhythmic/prosodic context affects listeners' processing of durational cues. To test if proximal contrast effects are indeed present, Experiment 2 (n = 32) used *only the immediately preceding syllable* from the stimuli in Experiment 1 (boxed region in Fig. 1) and blocked stimuli by condition; it was otherwise identical. Results show that a preceding longer precursor (the TROCHEE condition in Experiment 1) significantly *decreased "*code" responses ($p < 0.05$; Fig. 3), confirming proximal contrast effects obtain in the absence of distal context.

Experiment 3: Experiment 3 (n = 32) explored how changing the *rate of repetition* of rhythmic structure influenced perception. Stimuli from Experiment 1 were manipulated to create a FAST condition (full precursor linearly compressed by 20%) and a SLOW condition (precursor linearly expanded by 20%); resulting in four conditions in a 2x2 design (FAST TROCHEE, SLOW TROCHEE, etc.). The experiment was otherwise identical to Experiment 1. **Results:** A main effect of rate was found whereby FAST rate increased

"code" responses ($p < 0.05$; not shown). This is expected based on previous findings in the speech rate literature (e.g. [7,8]). A main effect of rhythm was also found, whereby the TROCHEE condition significantly *decreased* "code" responses ($p < 0.01$; see Fig. 4). Perhaps surprisingly, this result matches the proximal contrast effects observed in Experiment 2. Because the *entire* precursor was temporally expanded, these rate manipulations include expansion and compression of the syllable precedingly the target and therefore also exaggerate *proximal* differences in duration. This result therefore highlights that rhythmic effects compete with proximal contrast effects such that they disappear when proximal differences are larger in magnitude, and rate varies. An interaction between rhythm and rate was also found ($p < 0.05$). Post hoc testing of the interaction finds a significant effect of rate only in the TROCHEE condition ($p < 0.01$), and *no* effect in the IAMB condition (cf. Figs. 5 & 6). Given that all participants were speakers of American English, this interaction raises an intriguing possibility: that listeners more readily entrain to changes in rate when the rhythmic structure of speech more closely matches language patterns, given that English is largely trochaic, and listeners have been shown to be sensitive to this sort of metrical structure in other tasks [3]. This prediction is currently being tested in a follow-up experiment. Crucially, Exp. 3 shows that in the presence of increased variability in rate, the rhythmic grouping effects seen in Exp. 1 are not observed.

Summary: These experiments find an intricate interplay between durational contrast effects and rhythmic context in perception. Extending these results will therefore better our understanding of distal prosodic cues in perception and their interface with other perceptual processes.

Figure 1: Waveforms are bracketed according to listeners' hypothesized grouping. The longer precursor syllable is bolded. The gray box highlights proximal context.



***In Figures 2-6*:** Categorization is plotted split by condition. Target vowel duration is on the *x* axis, the *y* axis shows the proportion of "code" responses. Parenthetical (short)/(long) refer to the syllable immediately preceding the target.

Figure 2 - Exp. 1

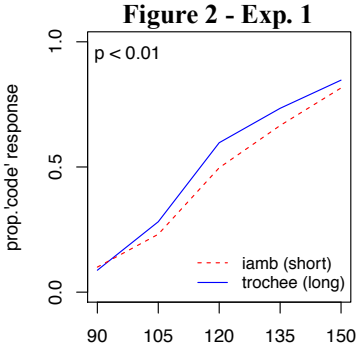


Figure 3 - Exp. 2

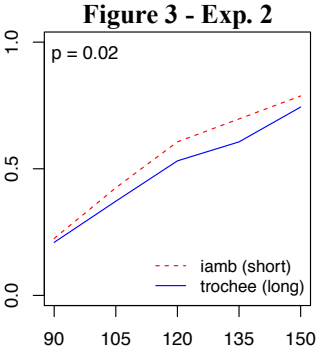


Figure 4 - Exp. 3

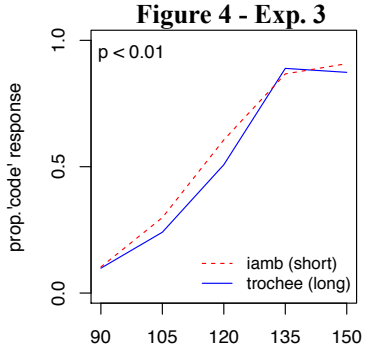


Figure 5 - Exp. 3

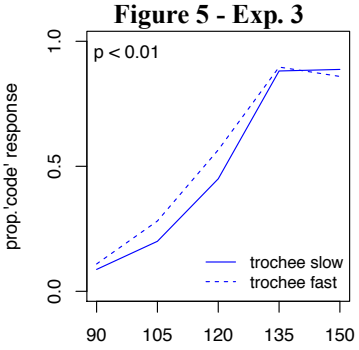
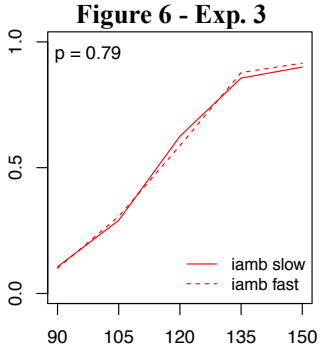


Figure 6 - Exp. 3



References

[1] Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *JPhon*, *54*, 68–79.

[2] Steffman, J. (2019). Intonational structure mediates speech rate normalization in the perception of segmental categories. *Jphon*, *74*, 114–129.

[3] Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psych: Human percept. and performance*, *14*(1), 113.

[4] Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *JML.*, *59*(3), 294–311.

[5] Raphael, L. J. (1972). Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English. *JASA*, *51*(4B), 1296–1303.

[6] Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *JASA*, *85*(5), 2154–2164.

[7] Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attn., Percept. & Psychophysics*, *79*(1), 333–343.

[8] Reinisch, E. (2016). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attn., Percept. & Psychophysics*, *78*(4), 1203–1217.