

The quality of prosodic quantity: Sonority is driving length illusions in perception

Fabian R. Eckert¹, Aviad Albert¹, Jeremy Steffman², T. Mark Ellison¹, Petra B. Schumacher¹, Martine Grice¹
¹Universität zu Köln, ²The University of Edinburgh

Prominence in the acoustic signal (see [1-3] for overviews) is typically associated with the acoustic dimensions of F0, duration, and power. The latter is usually represented by a log-transformed dB scale, also termed *intensity*. While providing an indication of the amount of energy present in the speech signal, intensity does not relate directly to auditory perception (e.g. to perceived loudness; see [4: pp. 19-21]). Investigating the power dimension in prosody-related studies thus requires a modification, such as the use of band-filtered signals to capture syllabic structures in the intensity curve (e.g. [5, 6]). The ProPer toolbox [7], employs an alternative method to capture acoustic power in prosody-related studies by extracting periodic energy. Rather than representing the power of selected frequencies in the signal, periodic energy represents the power of the periodic components: those which are best-suited to induce the sensation of pitch in the speech signal. These choices are grounded in the inherent link between pitch and periodicity [8] as well as in the suggestion that periodicity and sonority are linked in linguistic sound systems [4]. To that end, the following study uses systematic differences in the sonority of certain consonants as a proxy for corresponding differences in acoustic periodic energy.

The current study tests if periodicity differences in the speech signal lead to differences in auditory length perception. An analogy that served as inspiration for this endeavor is that of the well-known Müller-Lyer illusion, a visual illusion of length perception (as in Figure 1). Reproducing this in the auditory domain will further serve to test the underlying assumption that periodicity in the speech signal can be linked to prominence, through its effect on *syllable weight* [9]. The hypothesis of this study thus argues that spoken items with more sonorous material will be perceived to be longer than comparable items with less sonorous material, all other factors such as duration or loudness being held constant.

Method: We conducted an online study investigating the perceived relative length of word pairs, with controlled duration, loudness, and periodic energy. Two auditory stimuli were presented in sequence, and participants rated their relative length on a sliding scale by dragging towards either word to indicate greater perceived length. The experimental stimuli consisted of 12 monosyllabic (CVC) minimal triplets with an identical CV portion and an alveolar final consonant, which was either a voiceless sibilant (/s/), a voiced sibilant (/z/), or a nasal (/n/), e.g. *bus-buzz-bun*. All stimuli were based on recordings of a female speaker of Mainstream American English. We spliced the codas of each variant on the CV portion of the CVz variants (e.g. *buzz*) in the same triplet. This not only reduced segmental variation within the CV portions, but also kept the intonation contours largely constant within triplet. All stimuli were then modified to be fixed in duration at the across-triplet median of 630 ms, with longer (+25%, 788 ms) and shorter (-25%, 472 ms) variants being additionally created. Perceived loudness was controlled for by normalizing all targets to -23 LUFS (Loudness Unit relative to Full Scale; [10]). Programming and data collection of the experiment was carried out in PCIBex [11] and participants were recruited online using Prolific. 139 monolingual American English speakers took part in the study (120 trials per participant – 30 per trial type, see Figure 2).

Results: Figure 2 displays the distribution of length judgements per trial type, excluding responses in which participants did not move the slider, thus resulting in a default zero rating. Figure 3 displays the proportion of zero-ratings by individual response, including all data points. Importantly, the distributions of target trials (type C), show an effect of sonority, where positive responses represent an increased perception of length for the more sonorant word in a pair, *while no duration difference is present*. The results for trial type C resemble the distributions for trial type A, where actual duration differences drive the difference in attested length perception. Moreover, trial type C appears different from trial type B, in which identical items are compared. Note that the exclusion of default responses in Figure 2 reduced the number of data points in trial type B (where "same length"/zero reflects the correct response), as can be seen in Figure 3, which includes default ones.

Conclusion: This study provides preliminary evidence of links between sonorant material in the speech signal and perception of length, directly implicating prosodic prominence in terms of syllable weight.

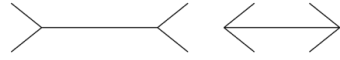


Figure 1. The Müller-Lyer illusion.

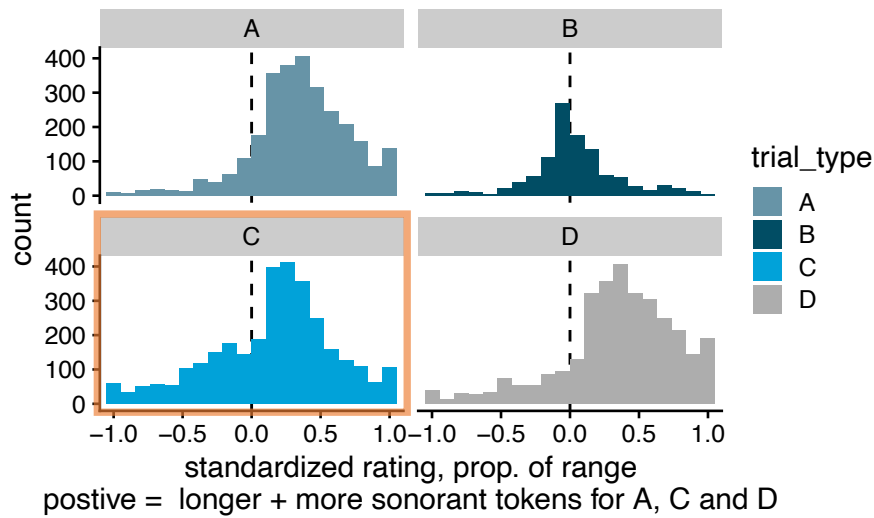


Figure 2. Distribution of results per trial type. **Trial type A:** same word, different duration (e.g. *bus^{med}* vs. *bus^{long}*); **Trial type B:** identical pairs (e.g. *bus^{med}* vs. *bus^{med}*); **Trial type C:** different word, same duration – the target comparisons (e.g. *bus^{med}* vs. *bun^{med}*); **Trial type D:** different word (including different CV), different duration – unanalyzed fillers (e.g. *bus^{med}* vs. *fuss^{long}*).

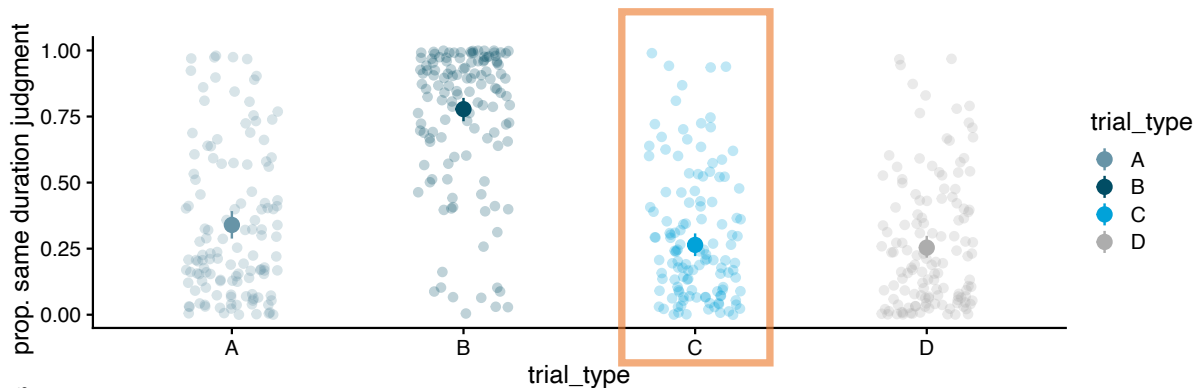


Figure 3. Proportion of zero-ratings, both active choice and default, per trial type (see trial types in Figure 2). Small points show individual responses; large points indicate the grand mean across trial types.

References

- [1] Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, 3(1).
- [2] Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, 70, 20–38.
- [3] Roessig, S., Winter, B., & Mücke, D. (2022). Tracing the Phonetic Space of Prosodic Focus Marking. *Frontiers in Artificial Intelligence*, 5.
- [4] Albert, A. (2023). *A model of sonority based on pitch intelligibility*. Language Science Press.
- [5] Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, 134(1), 628–639.
- [6] Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition*, 171, 130–150.
- [7] Albert, A., Cangemi, F., Silveira, G., Ellison, T. M., & Grice, M. (2025). *ProPer: PROsodic analysis with PERiodic energy* [computer software]. OSF. <https://doi.org/10.17605/OSF.IO/28EA5>
- [8] Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience*, 32(39), 13335–13338.
- [9] Gordon, M. (2004). Syllable weight. In B. Hayes, R. M. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (p. 277). Cambridge University Press.
- [10] ITU-R Recommendation BS. 1770-4. (2015). *Algorithms to measure audio programme loudness and true-peak audio level*. International Telecommunication Union.
- [11] Schwarz, F., & Zehr, J. (2021). Tutorial: Introduction to PCIBex—an open-science platform for online experiments: Design, data-collection and code-sharing. In *Proceedings of the annual meeting of the cognitive science society*, 43.