Amplitude envelopes as a means to quantify vowel length contrasts

Friederike Hohl, Hendrik Behrens-Zemek, Bettina Braun
University of Konstanz, Department of Linguistics, Germany

The analysis of amplitude envelopes has become a wide-spread method in the speech sciences, language acquisition and neurolinguistics (Assaneo et al., 2019; Frota et al., 2022; Goswami, 2019; Gross et al., 2013; Leong & Goswami, 2015; Poeppel, 2014). Amplitude envelopes track the amplitude distribution over an utterance and are interesting because they track the part of the signal that is relevant to convey rhythm. Furthermore, the method is easy to apply without demanding manual annotation. Despite the increasingly wide-spread usage across disciplines, there is little research on which aspects of the speech signal influence the amplitude envelopes in what way. Cross-linguistic research has shown that stress-timed German led to higher power between 7 and 10 Hz than more syllable-timed Brazilian Portuguese (Frota et al., 2022). In this paper we compare amplitude envelopes in disyllabic words in German differing in vowel length. We hypothesize that target words with short vowels have an amplitude distribution more similar to syllable-timing than those with long vowels and hence have higher power between 7 and 10 Hz. In carrier sentences, the effect of vowel length is expected to be reduced because longer stretches of speech convey the general rhythmic structure of the language, rather than segmental length.

Nine German native speakers each recorded 20 disyllabic word pairs that differed in the length of the stressed vowel (e.g. Mitte [ˈmɪtʰə] ‘centre’ vs. Miete [ˈmiːtʰə] ‘rent’) in a carrier sentence (*Das nächste Wort war XXX*. ‘The next word was XXX’). The long vowels were nearly twice as long as the short vowels: 155.0 vs. 80.9 ms). The target word was cut out of the carrier sentence. Following Gross et al. (2013), the signal (target word and carrier sentence) was first filtered into nine frequency bands in the range from 100–10,000 Hz, which are equidistant on the cochlear map. The cutoff frequencies were 100.5Hz, 250.7Hz, 458.6Hz, 748.8Hz, 1159.0Hz, 1449.0Hz, 2619.8Hz, 3954.2Hz, 6121.8Hz and 10000.8Hz. To remove high-frequency components, the amplitude envelopes were low-pass filtered using a Hann Band pass filter from 0 to 10 Hz in praat (He & Dellwo, 2016). The resulting narrowband envelopes were added to compute the wideband amplitude envelope, which were spectrally analyzed (Fast Fourier transform) in 100 0.1-Hz steps. The effect of vowel length on the amplitude modulation spectra was investigated using general additive mixed models (Porretta et al., 2016; Wood, 2006; Zahner et al., 2019; Zahner-Ritter et al., 2022), which allow comparison of power across frequency bands. For model fitting and plotting, we employed the R packages *mgcv* (Wood, 2015) and *itsadug* (van Rij et al., 2015). The response variable was log-normalized power. The initial model included *length* as parametric effect (fixed effect), along with a factor smooth for the modulation of *length* over frequency bands, s(*fband*, by = *length*). Smooths for *speakers* and *items* (random intercept and over frequency bands) were also included. The model was corrected for auto-correlation in the data using a correlation parameter, determined by the acf_resid() function and used the scat linking function to normalize the otherwise skewed distribution of residuals. The results showed a significant effect of length, see Figure 1 for the target words, but not for the sentences (not shown for space reasons). The target words with a long vowel had a lower power in a small frequency band just below 2Hz and, importantly, between 5 and 8.5 Hz.

Amplitude envelopes can capture the length contrast in German word pairs, but the difference disappears when the same contrast is analysed in longer utterances. For the target words, the energy differences mirror those for stressed- vs. syllable-timed languages. For sentences, the pervasive stressed-timed rhythm (Das NÄCHste WORT war MIEte, stressed syllables are capitalized) seems to reduces the effect of vowel length of a single word.

doi:10.11576/pundp2022-1008

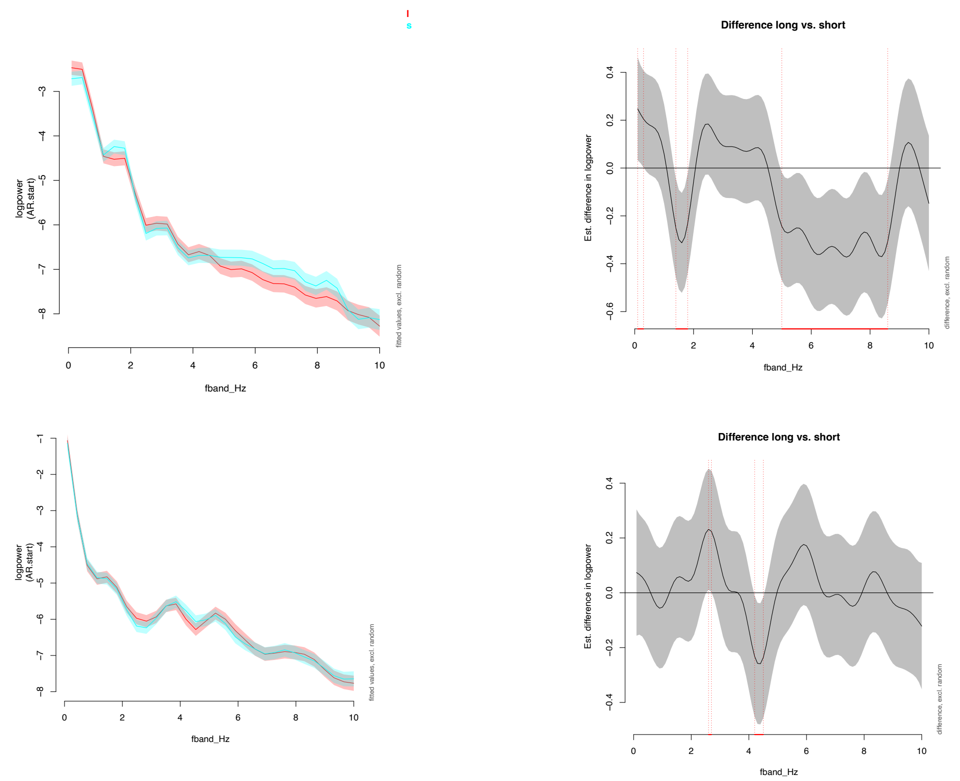

Figure 1. Factor smooths (left panels, long vowels in red, short vowels in blue) and the difference plots (long vs. short vowels). Top: target word, bottom: whole utterance.

References

Assaneo, M. F., Ripolles, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat Neurosci*, *22*(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Frota, S., Vigário, M., Cruz, M., Hohl, F., & Braun, B. (2022). *Amplitude envelope modulations across languages reflect prosody*. Proceedings of Speech Prosody, Lisbon, Portugal. https://doi.org/10.21437/SpeechProsody.2022-140

Goswami, U. (2019). A neural oscillations perspective on phonological development and phonological processing in developmental dyslexia. *Language and Linguistics Compass*, *13*(5), e12328.

Gross, J., Hoogenboom, N., Thut, G., Schys, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLOS Biology*.

He, L., & Dellwo, V. (2016). *A Praat-based algorithm to extract the amplitude envelope and temporal fine structure using the Hilbert transform*. 530–534.

Leong, V., & Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS One*, *10*(12), e0144411.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Curr Opin Neurobiol*, *28*, 142–149. https://doi.org/10.1016/j.conb.2014.07.005

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1–21.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, D. (2015). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. https://cran.r-project.org/package=itsadug

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman & Hall/CRC Press.

Wood, S. N. (2015). *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*.

Zahner, K., Kutscheid, S., & Braun, B. (2019). Alignment of f0 peak in different pitch accent types affects perception of metrical stress. *Journal of Phonetics*, *74*, 75–95.

Zahner-Ritter, K., et al. (2022). Three Kinds of Rising-Falling Contours in German wh-Questions: Evidence From Form and Function. *Frontiers in Communication*, *7*.

doi:10.11576/pundp2022-1008