Amplitude envelopes as a means to study length contrasts: evidence from L1 and L2 Italian

Marieke Einfeldt, Bettina Braun

Department of Linguistics, University of Konstanz, Germany
marieke.einfeldt@uni-konstanz.de, bettina.braun@uni-konstanz.de

ABSTRACT

Previous research has shown that amplitude envelopes can distinguish between rhythmically different languages and between words differing in phonemic vowel length. We use this method to study the realization of consonantal length contrasts for native Italian and advanced German learners of Italian. Consonant length is phonemic in Italian. Phonetically, geminates are longer than singletons and the vowels preceding geminates are shorter than those preceding singletons. German does not have a phonemic consonantal length contrast. We extracted amplitude envelopes from Italian word pairs differing in consonant length for natives and learners. The results of generalized additive modelling showed higher power between 2.3 and 7.4Hz for geminates vs. singletons for both groups. However, native speakers realized the contrast with larger power differences between 2.3 and 3.1Hz while the learners showed larger differences between 7.9 and 8.4Hz. A comparison to German vowel length data suggests that these differences are due to cross-linguistic influence.

Keywords: Amplitude envelopes, consonantal length, cross-linguistic influence, Italian, GAMMs.

1. INTRODUCTION

The analysis of amplitude envelopes has become a widely used method in the speech sciences, language acquisition and neurolinguistics [1-6]. Amplitude envelopes track the amplitude distribution over an utterance. They hence represent the part of the signal that is relevant to convey speech rhythm [7]. 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 a stress-timed language (German) led to lower power around 2Hz and between 7 and 10Hz than a more syllable-timed language (Brazilian Portuguese, cf. [1]). Here we compare amplitude envelope differences across two consonantal length conditions in Italian disyllabic words produced by Italian native speakers (Experiment 1) and advanced German learners of Italian (Experiment 2).

For the analysis of amplitude envelopes, the extraction of the modulation frequencies from the speech signal is necessary. This can be done with several procedures [8]: One approach is to filter the sound into a number of frequency bands; the filtering is spaced either in equal steps on the cochlea or logarithm-based. The frequency range usually chosen is between 100 and 8,000 or 10,000Hz respectively. In a next step, the high-frequency components are removed (low-pass filtering up to ~10Hz or Hilbert transform), so that the low-frequency amplitude envelopes remain (so-called narrowband envelopes). These narrowband envelopes are summed up and the modulation frequencies are derived by Fourier analysis. With this method, the final outcome received are spectra, which represent power distribution across frequency.

There are several ways on how to relate amplitude envelope modulation frequencies to linguistic units, such as the syllable rate [e.g., 8 for a review, 9] or to word or phrasal prosody. [4]'s study showed that there are the following clusters of energy: (1) word stress/stressed syllable rate (~2Hz), (2) and the syllable rate (~5Hz). Also [10] found timescales that could be related to certain clusters of energy: (1) the phrasal (0.6-1.3Hz), (2) the word (1.8-3Hz), (3) the syllable (2.8-4.8Hz), and the phoneme (>8Hz) scale.

In this paper we build on data from German vowel length contrast [11]. Their results showed that target words with a short vowel (e.g., [ˈmɪtʰə]) had a higher power in a small frequency band just below 2Hz and between 5 and 8.5Hz than words with a long vowel (e.g., [ˈmiːtʰə]). Here, we test the Italian consonantal length contrast in L1 speakers of Italian and in proficient German learners of Italian (L2).

2. EXPERIMENT 1

Italian differentiates between short consonants, so-called singletons, and long consonants, so-called geminates. This length contrast can result in minimal pairs such as *fato* /fato/ 'fate' and *fatto* /fat:o/ 'fact', which phonologically only differ in the length of the medial consonant [12]. Vowels preceding singletons have phonetically longer duration than vowels preceding geminates, but this vowel duration contrast

is not phonemic. [6] have further shown that the initial consonants are lengthened in words with an upcoming geminate. Given the vowel length contrast data in [11] and the allophonic decrease in vowel duration before geminates (creating a similarity to the short vowel condition in the German data), we predict higher energy in geminates than singletons in the range between 5 and 8Hz for native speakers of Italian, similar to the German vowel length contrast. Since duration differences in the Italian consonantal length contrast span a wider range (including the preceding vowel), we further hypothesize that the Italian consonantal length contrast results in power differences in a larger frequency range than the German vocalic length contrast.

2.1. Methods

The experiment was a self-paced reading task.

2.1.1 Participants

Six L1 speakers of Italian participated in the experiment (age range: 30-45 years). All participants lived in Germany during the time of testing and had not acquired a foreign language before 6 years of age.

2.1.2 Materials

We selected 16 disyllabic word pairs that differed in the length of the medial consonant. Eleven were minimal pairs and five pairs had the same consonant and preceding vowel but differed in other segments (e.g., /pinːa/ vs. /mina/, /ramo/ vs. /gamːa/). The target consonants were /p/ – /pː/, /t/ – /tː/, /l/ – /lː/, /n/ – /nː/, /m/ – /mː/, /z/ – /sː/. All target words were trochaic. The frequency of the words was on average similar for long and short consonants (SUBTLEX [13] mean-log-frequency score for "geminate words" was 6.6 (SD = 2.8) vs. 6.6 (SD = 3.5) for "singleton words"). We further selected 10 tri- and disyllabic words as distractors. The target words were embedded in a carrier sentence, such that the target word was accented: *Era "palla" che ho detto*, (English: It was "ball" that I said). The written sentence was accompanied by a picture of the item.

2.1.3 Procedure

Recording. The participants sat in front of a computer screen with a PowerPoint-presentation. The test sentences appeared one-by-one. Participants were asked to read the test sentences out loud at normal speed. The recordings were done in a quiet room at the University of Konstanz or at the participant's home. The recording device (Olympus Linear PCM Recorder LS-11/LS-5, 44.1 kHz/16 bit) was placed on the table next to the participant.

Extraction of the amplitude envelope modulation spectra. The start and end of target words were segmented using standard segmentation criteria [14] and the words were saved as separate wav-files, using *Praat* [15]. We analyzed the wideband amplitude envelopes of the productions, following the procedure in [9]. First, we extracted the narrowband amplitude envelopes. For this analysis we used a script developed by He and Dellwo [16, 17]. The speech signal was first down-sampled to 22050Hz and then filtered into nine frequency bands in the range from 100–10,000Hz, which are equidistant on the cochlear map. The cutoff frequencies were 100.5Hz, 250.7Hz, 458.6Hz, 748.8Hz, 1,159.0Hz, 1,449.0Hz, 2,619.8Hz, 3,954.2Hz, 6,121.8Hz and 10,000.8Hz. To remove high-frequency components, the amplitude envelopes were low-pass filtered (< 10Hz). These narrowband amplitude envelopes were then added to compute the wideband amplitude envelope, which were spectrally analyzed in 100 0.1-Hz steps, resulting in 19200 data points (6 speakers x 2 length conditions x 16 items x 100 frequency bands).

Statistical modeling. To model the effect of length across frequency bands, we used generalized additive mixed models, GAMMs [6, 18-22]. They are well-suited to pinpoint in which frequency bands differences occur, taking into account non-linear relationships and auto-correlation [23, 24]. The response variable was log-normalized power. We modelled non-linear dependencies of *length* over frequency bands using smooth functions s(*fband_Hz*, by = *length,* bs='tp', k = 20*)*). These smooth functions include a pre-specified number of base functions of different shapes, e.g., linear and parabolic functions of different complexity [e.g., 24]. *Length* was further added as fixed (parametric) effect. Smooths for *speakers* and *items* (random intercept and over frequency bands) were also included (e.g., s(speaker, fband_Hz, by='re')). For model fitting, we employed the R package *mgcv* [20, 25]; the package *itsadug* was used to plot the model results [26]. The model was corrected for auto-correlation in the data using a correlation parameter, determined by the acf_resid() function (package *itsadug* [26]). We use the function gam.check() to check whether the number of smooth functions (k) and the smoother (thin plate regression, 'tp') were adequate and adjusted if necessary. The final model included the scat-linking function.

2.2. Results

Figure 1 shows the averaged estimated effects of the factor *length*. It suggests that the long consonants have higher power between 2 and 8Hz. To determine whether these effects are significant, we analysed the differences across conditions.

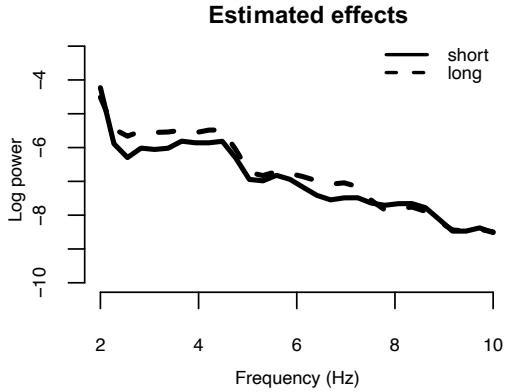

Figure 1: Estimated effects of the factor *length*.

Figure 2 shows the differences between geminates and singletons for the Italian native speakers (L1). Positive values indicate higher power for geminates than singletons; the difference is significant when the grey band (95% confidence interval) excludes 0. Given that the words were very short (on average 401 ms, SD = 80 ms), we only included the interval between 2 and 10Hz in the modelling. Figure 2 indicates that words with geminates had a considerably higher power between 2.2 and 7.4Hz with a short non-significant stretch between 5.3 to 5.8Hz and a slightly lower power in the small range between 7.8 and 8.1Hz.

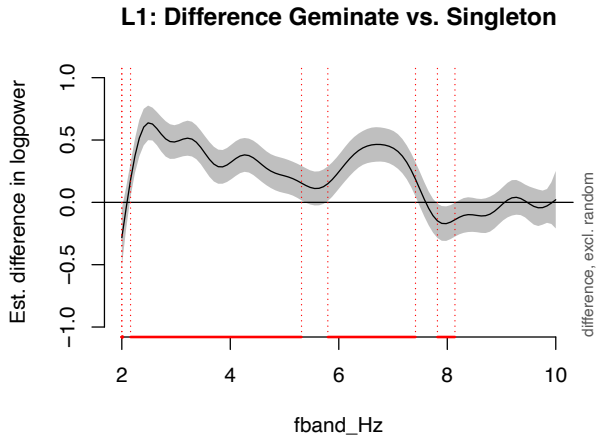

Figure 2: Effect of length for native speakers.

The power differences span a longer frequency interval for the Italian consonantal length contrast than for the German vowel length contrast (between 5 and 8.5Hz, cf. [11]). For direct statistical comparison, we combined the datasets and included a smooth term for the interaction between *language* and *length* (for details on modelling see section 3.2). This interaction model showed significant cross-linguistic power differences between 2.3 and 3.6Hz: In this frequency band the Italian power difference was significantly larger than the German one.

2.3. Discussion

The Italian consonantal length contrast is evident in a large frequency band in the amplitude envelope modulation spectra. Compared to the German vowel length contrast, the Italian consonantal length contrast resulted in significantly larger power differences between 2.3 and 3.6Hz. The most likely reason for these cross-linguistic differences is that the Italian length contrast is distributed over a longer temporal interval, including the preceding vowel and the consonant. Another, albeit less likely, option is that the contrast is larger in Italian due to the use of quasi-minimal pairs. We consider this explanation less likely because the number of quasi minimal pairs was small (fewer than 1/3 of the word pairs).

3. EXPERIMENT 2

Learners of Italian have been shown to produce the consonantal length contrast less clearly/robustly than native speakers [27, 28]. Further, the L2 acquisition of other length related aspects, such as general rhythm, is difficult and prone to L1 transfer [29]. Experiment 2 tested whether amplitude modulation spectra can bring out potentially smaller differences between groups, e.g., deviation from L1 due to transfer in the L2 productions. We predict that the difference between geminates and singletons is less pronounced in German L2 learners of Italian, i.e., differences in power are expected to be smaller and possibly limited to a narrower frequency band. If L2 speakers fail to adequately realize the Italian length contrast (consonantal and vowel duration difference), we particularly predict differences between L1 speakers and learners in the frequency band from 2.3 to 3.6Hz (in which differences between the Italian consonantal length contrast and the German vowel length contrast occurred).

3.1 Methods

Ten German L2 speakers of Italian (20–35 years) were recorded. All L2 speakers studied Italian at university. Their self-rated proficiency on a 6-point-scale (1 = beginner to 6 = native-like) was on average 4. All participants lived in Germany during the time of testing. Materials and procedure were identical to Experiment 1. The L2 speakers additionally completed a background questionnaire including information on proficiency and language use.

3.2 Results

Figure 3 shows the power difference between geminates and singletons for the learners. Geminates resulted in a significantly higher power in the frequency range between 2.3 and 8.7Hz with more

non-significant interruptions (grey confidence interval including 0) than the Italian natives (cf. Figure 2).

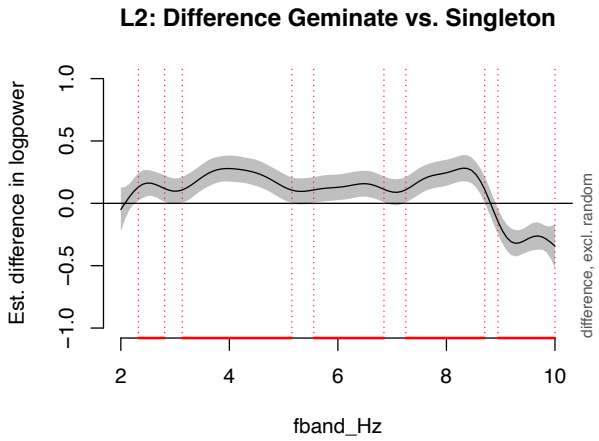

Figure 3: Effect of length contrast for L2 learners.

Furthermore, geminates had significantly lower power in the range from 8.9 and 10Hz, resulting in a biphasic pattern that was absent in this frequency area for the L1 speakers.

To test whether the power differences across language groups are significant, we combined the data from L1 (Experiment 1) and L2 speakers (Experiment 2) and tested for an interaction between *language group* and *length*. To this end we included a factor smooth for the interaction of *language group and length* over frequency bands, s(*fband*, by = *language_length,* bs ='tp', k = 20) and the interaction as parametric (fixed) effect. Otherwise, the model was fitted as described above. The model including the interaction smooth was subsequently compared to a simpler model with separate smooth terms for *language group* and *length,* but without the interaction smooth, using the function CompareML() [26]. The comparison confirmed that the inclusion of the smooth term significantly improved the fit of the model in terms of Maximum Likelihood ($p < 0.003$). To investigate the interaction in detail, we fitted binary smooths (see [24, 30]). The results confirmed the visual comparison of the difference curves between Figures 2 and 3: L1 speakers have significantly higher power in low frequency bands, in particular between 2.3 and 3.1Hz. Furthermore, the L2 speakers produced a larger power difference for the contrast than the L1 speakers at 7.9-8.4Hz.

3.3 Discussion

The results indicate that the L2 speakers produce the consonantal length contrast similar to L1 speakers, with higher power for words containing geminates across most frequency bands from 2.3 to 7.4Hz (intersection between L1 and L2 frequency bands that showed significant differences or differences that approached significance). However, the L1 speakers have higher power difference between 2.3 and 3.1Hz than the L2 speakers, while the L2 speakers show larger power differences between 7.9 and 8.4Hz compared to the L1 speakers. The differences in the higher frequency band are similar to the power differences of the German vowel length contrast [11] (where words with a short vowel had higher power between 5 and 8.5Hz) and could indicate cross-linguistic influence.

4. GENERAL DISCUSSION

Given that the extraction of amplitude envelopes does not necessitate manual segmentation and that amplitude envelopes capture more than just the duration of segments, they provide an efficient means to operationalize the length contrast. This is particularly appealing for phonological contrasts that are distributed over a number of segments, such as Italian gemination (with cues in consonant duration and duration in preceding segments). Our data further show that amplitude envelopes can capture similarities and dissimilarities between L1 and L2 speakers: In terms of similarities, learners produced the length contrast in a similar fashion as the L1 speakers. There was a larger power for geminates than for singletons in the frequency range from 2.3 to 7.4Hz. In terms of dissimilarities, the power differences were smaller for L2 than for L1 speakers in the frequency band from 2.3 to 3.1Hz. On the contrary, learners produced a larger contrast than L1 speakers from 7.9 to 8.4Hz. Interestingly, the smaller effect in the lower frequency band and the larger effect in the higher frequency band mirrors what German speakers do when producing the German vowel length contrast (cf. [11]). Therefore, this difference might be explained in terms of transfer from the L1 (German) to the L2 (Italian).

5. CONCLUSION

Amplitude envelopes seem a reliable and efficient method to compare the production of length contrasts in L1 and L2. They track fine-grained differences between advanced learners and native speakers and are susceptible to transfer effects.

6. ACKNOWLEDGEMENTS

We thank Friederike Hohl for preparing the data, and Tanja Kupisch and Mechtild Tronnier for help with constructing the stimuli. We thank Cesko Voeten for discussion on the function gam.check(), linking functions and smoothing options.

7. REFERENCES

[1] Frota, S., Vigário, M., Cruz, M., Hohl, F., Braun, B. 2022. Amplitude envelope modulations across languages reflect prosody. *Proc. Speech Prosody*, Lisboa, 688-692.

[2] Poeppel, D. 2014. The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology* 8, 142-149.

[3] Goswami, U. 2019. Speech rhythm and language acquisition: an amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences* 1453, 1, 67-78.

[4] Leong, V., Goswami, U. 2015. Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PLoS ONE* 10, 12, e0144411.

[5] Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., Poeppel, D. 2019. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience* 22, 4, 627-632.

[6] Gross, J. Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S. 2013. Speech Rhythms and Multiplexed Oscillatory SensoryCoding in the Human Brain. *PLoS Biologogy* 11, 12, e1001752.

[7] Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66, 46-63.

[8] Poeppel, D., Assaneo, M. F. 2020. Speech rhythms and their neural foundations. *Nature Review Neuroscience* 21, 322–334.

[9] Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., Ghazanfar, A. A. 2009. The natural statistics of audiovisual speech. *PLOS Computational Biology* 5, 7, e1000436.

[10] Keitel, A., Gross, J., Kayser, C. 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology* 16, 3, e2004473.

[11] Hohl, F., Behrens-Zemek, H., Braun, B. 2022. Amplitude envelopes as a means to quantify vowel length contrasts. Presented at: *Phonetik und Phonologie im deutschsprachigen Raum (P&P)*, Bielefeld, Germany.

[12] Payne, E. 2005. Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association* 35, 2, 153-181.

[13] Crepaldi, D., Mandera, P., Keuleers, E., Brysbaert, M. *Subtlex-it: A new frequency list for Italian based on movie subtitles*. [Online]. Available: http://crr.ugent.be/subtlex-it/.

[14] Turk, A. E., Nakai, S., Sugahara, M. 2006. Acoustic segment durations in prosodic research: A practical guide. In: S. Sudhoff, D. Lenertová,, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, J. Schließer (eds), *Methods in Empirical Prosody Research*. Walter de Gruyter, 2006.

[15] Boersma, P., Weenink, D. 2016. *Praat: doing phonetics by computer*.[Online]. Available: https://www.fon.hum.uva.nl/praat/

[16] He, L., Dellwo, V. 2017. Amplitude envelope kinematics of speech signal: parameter extraction and applications. In: J. Trouvain, I. Steiner, and B. Möbius (eds), *Proc. Elektronische Sprachsignalverarbeitung 2017*, Dresden.

[17] He, L., Dellwo, V. 2016. A Praat-based algorithm to extract the amplitude envelope and temporal fine structure using the Hilbert transform. *Proc. Interspeech 2016*, San Francisco, 530-534.

[18] Wood, S. N., Saefken, B. 2016. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* 111, 1548-1575.

[19] Wood, S. N. 2017. *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*.

[20] Wood, S. N. 2006. *Generalized additive models: an introduction with R*. Chapman & Hall/CRC Press.

[21] Wieling, M., Margaretha, E., Nerbonne, J. 2012. Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics* 40, 2, 307-314.

[22] 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.

[23] Baayen, R. H., van Rij, J., de Cat, C., Wood, S. N., Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In: D. Speelman, K. Heylen, and D. Geeraerts (eds), *Mixed effects regression models in linguistics*, Springer, 2018, pp. 49-69.

[24] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86-116.

[25] Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, 1, 3-36.

[26] van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H. 2017. *itsadug: Interpreting time series and autocorrelated data using GAMMs*.

[27] Harada, T. 2006. The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition* 28, 4, 601-632.

[28] Kabak, B., Reckziegel, T., Braun, B. 2011. Timing of second language singletons and geminates. *Proc. International Congress of Phonetic Sciences*, Hong Kong, 994-997.

[29] Li, A., Post, B. 2014. L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German Learners of English. *Studies in Second Language Acquisition* 36, 2, 257-281.

[30] Zahner-Ritter, K., Einfeldt, M., Wochner, D., James, A., Dehé, N., Braun, B. 2022. Three Kinds of Rising-Falling Contours in German wh-Questions: Evidence From Form and Function. *Frontiers in Communication* 7, 838955.