10 Linguistic Factors Affecting Amplitude Modulation Spectra

BETTINA BRAUN

10.1 Introduction

The analysis of amplitude envelopes has become a widespread method in the speech sciences (Gibbon, 2021; Mermelstein, 1975; Rosen, 1992), language acquisition research (Goswami, 2018, 2019), and neurolinguistics (Assaneo et al., 2019; Gross et al., 2013; Obleser et al., 2012; Poeppel, 2014; Poeppel and Assaneo, 2020). Loosely speaking, amplitude envelopes capture the amplitude distribution over the waveform of an utterance; a common metaphor is that their shape is like a blanket put over a waveform such that the individual positive glottal spikes "carry" the blanket. High-energy sounds (vowels, but also some sonorants and sibilants) lead to bumps in the amplitude envelopes; low-energy sounds (non-sibilant fricatives, stops) lead to troughs, as visualized in Figure 10.1. These amplitude envelopes present a low-frequency time-varying signal.

Like for any other signal, one can extract the spectrum of the low-frequency amplitude envelope, which gives an indication of the frequencies of modulation that have the most direct association with the concept of speech rhythm. Three reasons make the analysis of amplitude envelopes and modulation frequencies (i.e., modulation spectra) interesting. First, envelopes capture the part of the signal that is relevant to convey rhythm, which makes them a prime candidate for a fresh look at acoustic research on speech rhythm (Arvaniti, 2009, 2012; Barry, 2007; Ramus et al., 1999), beyond the analysis of durational cues (Dauer, 1983) and acoustic isochrony. Second, the method is easy to apply to speech, without manual annotation. Third, it has been proposed that neural oscillations at different frequency bands track linguistic structure in speech, such as phonemes or syllables (Assaneo et al., 2019; Gross et al., 2013; Poeppel, 2014).

Despite the increasingly widespread use across disciplines, there is little research on which aspects of speech influence the amplitude envelopes in what ways. Cross-linguistic comparisons suggest that a language's rhythm affects the modulation spectra in particular (see Section 10.3.1). The aim of this chapter is twofold: It first provides an acoustic and statistical overview of previous studies (Section 10.2); then it reviews some literature on modulation spectra for

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

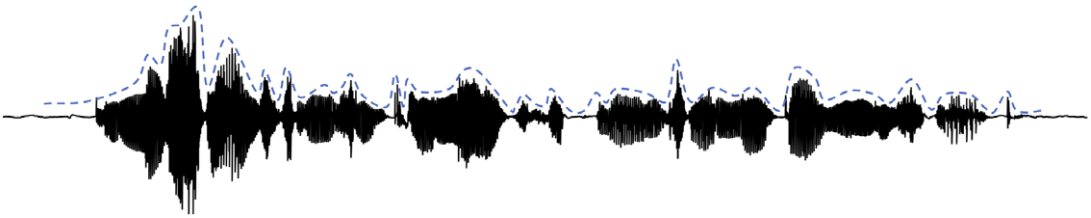


Figure 10.1 Schematic display of amplitude envelope.
Waveform (black) with a stylized amplitude envelope (dashed line) shifted upwards to increase visibility.

linguistic contrasts at the phrase level (Section 10.3: cross-linguistic differences, speaking styles, and illocution type) and more locally at the word level (Section 10.4: the role of segmental length and pitch accent type). Where possible, we provide data of our own, analyzed with the same workflow to allow for comparison of the magnitude of the differences (as a measure of effect size) and the frequency bands at which linguistic contrasts exert their differences (to allow for generalizations and the generation of hypotheses for future studies). Of the five linguistic contrasts included in this chapter, pitch accent type is the only one that does not alter the general rhythmic structure of an utterance (irrespective of accent type, the pitch accent is associated with the very same word; pitch accents primarily consist of frequency modulations, not amplitude modulations; see Gibbon, 2021). Pitch accent type is hence not expected to change the amplitude envelopes and serves as a perfect control condition to probe the specificity of the procedure for capturing rhythmic aspects. The chapter closes with an overview of the effect sizes and frequency domains in which effects are located for the own data (presented in Sections 10.3.1, 10.3.3, 10.4.1, and 10.4.2), discusses potential generalizations, and postulates desiderata for future research (Section 10.5).

This contribution focuses on the interpretation of amplitude envelopes in terms of phonetic and linguistic factors. Adhering to the book's scope, it primarily covers research that investigates aspects that pertain to speech rhythm. Hence, it does not cover the neurolinguistic literature that is concerned with testing links between brain oscillations and modulation frequencies of the amplitude envelope (see Chapters 3 and 5). Amplitude envelopes share some overlap with research on perceptual centers (p-centers; see Marcus, 1981; Rathcke et al., 2021) and amplitude rise times, which are related to vowel onsets (Chapter 11). This literature is, however, not covered here.

10.2 Acoustic and Statistical Background

Low-pass-filtered amplitude envelopes capture the slow-varying energy distributions over the course of an utterance. Technically, modulations can be extracted in different ways (for a comparison between extraction techniques and their relation to vowel onset, see MacIntyre et al., 2022). A common procedure is to filter the spectrum into a range of bands (between 70 Hz and 10,000 Hz). The cutoff points for filtering are either spaced logarithmically or equidistantly on the cochlear map (Chandrasekaran et al., 2009; Gross et al., 2013; Todd and Brown,

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

1994; Varnet et al., 2017). The initial decomposition of the signal into bands has the advantage of representing amplitude information from different frequency bands, information that has been shown to affect intelligibility (see Chapter 8) because the different frequency bands provide information on an utterance's linguistic structure (Assaneo et al., 2019; Gross et al., 2013; Poeppel, 2014). The envelope itself can be extracted using half-wave rectification (making the signal positive) followed by low-pass filtering (Caetano and Rodet, 2011; Gibbon, 2021; Kolly and Dellwo, 2014; Loizou et al., 1999) or by using the Hilbert transform (Braiman et al., 2018; MacIntyre et al., 2022; O'Sullivan et al., 2015). These two methods are said to achieve largely similar results (Ding et al., 2017, p. 182). The wideband envelope is derived by the sum of the narrowband envelopes (Chandrasekaran et al., 2009) or their average (Varnet et al., 2017). Other authors do not initially decompose the signal into frequency bands but first band-pass-filter the signal (e.g., between 400 and 4,000 Hz) and then low-pass-filter it (<10 Hz), for example, by using Butterworth filters (Tilsen and Johnson, 2008).

In most cases, researchers analyze amplitude envelopes spectrally, that is, in the frequency domain. To that end, modulation frequencies are extracted by discrete Fourier transform (Gibbon, 2021; Tilsen and Johnson, 2008). The result is a spectrum, that is, a continuous power curve across frequency bands. The signal typically follows a 1/f trend; that is, it has the highest energy in low-frequency bands (note that this 1/f trend is removed in some approaches). The different frequencies relate to linguistic units of different sizes (phrasal rate: 0.6–1.3 Hz; word/stress rate: 1.8–3 Hz; syllable rate: 4–5 Hz; phone rate: 8–10 Hz [Keitel et al., 2018]). One metric that is employed is the frequency of the highest peak (when the 1/f trend is removed), which is consistent across languages (Ding et al., 2017; Varnet et al., 2017) and related to the syllable rate (but see Zhang et al., 2023). Other researchers analyze the envelope in the time domain using empirical mode decomposition of the amplitude modulation (Tilsen and Arvaniti, 2013) and extract different parameters from that signal. Some authors combine measures from both the frequency and the time domains (Lau et al., 2022). It is clear that the variety of signal-processing methods makes it hard to directly compare results and effect sizes across studies. For the analysis of our own data presented here (specifically in Sections 10.3.1, 10.3.3, 10.4.1, and 10.4.2), we used nine gammatone frequency bands between 100 and 10,000 Hz, spaced equidistantly on the cochlear map, followed by low-pass filtering of the rectified narrowband signals. These narrowband envelopes were summed to derive a wideband envelope, which was then submitted to Fourier analysis (see Einfeldt and Braun, 2023 and Frota et al., 2022, for more details on these methods). Neither the sound files nor the amplitude envelope modulation spectra were normalized (i.e., the 1/f trend was not removed).

There are different kinds of analyses of amplitude envelopes in the literature. One approach has been to correlate aspects of the amplitude modulation spectra with visual phonetic information (Chandrasekaran et al., 2009, with mouth-opening data), acoustic information (MacIntyre et al., 2022, with acoustic vowel onsets; Varnet et al., 2017, with temporal rhythm metrics such as pairwise variability indices), or phonological information (Gibbon, 2021, with morphophonological information; Leong and Goswami, 2015, with feet, syllables, and onset–rime units; Todd and Brown, 1994, with the phonological stress hierarchy of a word). It has also been tested how useful the amplitude envelope information is for prediction (Ludusan et al., 2011: prominent syllables; MacIntyre et al., 2022: vowel onsets) and clustering (Gibbon,

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

2021). In studies that investigate effects of particular factors on the amplitude modulation spectra, mixed-effects models (Baayen et al., 2008) or general additive mixed models (GAMMs) have been employed (Wood, 2006). For our own data presented here (see Sections 10.3.1, 10.3.3, 10.4.1, and 10.4.2), we used GAMMs as they allow the researcher to model differences in power across frequency bands in a continuous manner, while accounting for autocorrelation (see Frota et al., 2022, for details of the statistical modeling). GAMMs establish the frequency bands in which significant differences occur across conditions and can be used for both exploratory and hypothesis-testing research questions. They allow for a statistic comparison of effect sizes (differences in log power, mathematically equivalent to ratios of raw power) and frequency bands across conditions.

10.3 Cross-linguistic and Cross-stylistic Comparisons

10.3.1 Cross-linguistic Comparisons

In linguistics, languages have been grouped into different rhythm classes (stress-timed, syllable-timed, mora-timed) based initially on perceptual impression. A common assumption is that there is a roughly equal distance between the respective linguistic units. Since such isochrony (typically established by annotation of segments) has been difficult to establish (e.g., Chapter 14; Dauer, 1983), other composite metrics have been suggested (see Grabe and Low, 2002; Ramus et al., 1999), all relying on manual annotation. However, these measures are also not without difficulties (Arvaniti, 2009, 2012), and the analysis of low-frequency amplitude envelopes is expected to support cross-linguistic differences in speech rhythm. Interestingly, there are a number of studies that suggest that amplitude envelopes have globally similar shapes across languages, including peaks in power for low-modulation frequencies followed by a decrease in amplitude (Ding et al., 2017; Poeppel and Assaneo, 2020; Varnet et al., 2017). One of the most comprehensive studies was conducted by Ding et al. (2017). They analyzed amplitude modulations from nine different languages (American English, British English, Chinese, Dutch, Danish, French, German, Norwegian, and Swedish). Their analyses show a peak in the (1/f-corrected) speech modulation spectrum around 4 Hz, independent of (the rhythm of the) languages. Varnet et al. (2017) analyzed semi-spontaneous productions from 10 languages (Basque, Dutch, English, French, Japanese, Marathi, Polish, Spanish, Turkish, and Zulu). They showed that the amplitude envelopes differed between stress-timed and syllable-timed languages (p. 1981). Frota et al. (2022) compared amplitude envelopes in three rhythmically different languages: stress-timed German, and more syllable-timed Brazilian and European Portuguese. They showed that rhythmic differences across languages are reflected in modulation spectra: stress-timed German had higher power than more syllable-timed Brazilian and European Portuguese in the delta (1–2 Hz) and theta bands (6–8 Hz). European and Brazilian Portuguese also differed, but only in the delta band (1–2 Hz), with similar mean differences as between German and the two Portuguese varieties. The difference between the two Portuguese varieties may be due to macro-rhythmic difference (see Jun, 2012): There are more frequent pitch accents in Brazilian than European Portuguese, affecting the energy distribution over the utterance, which is then picked up in the low-frequency amplitude

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

envelope. Tilsen and Arvaniti (2013) compared six languages (English, German, Greek, Italian, Korean, and Spanish) (see Section 10.3.2 for more detail). Their cross-linguistic analysis showed that English differed from the other languages, but no other comparisons were significant. The special characteristics of English were explained in terms of the "relatively high degree of supra-syllabic periodicity in English and minimal differences among the other languages in the corpus" (p. 635). In sum, Varnet et al. (2017) and Frota et al. (2022) reported differences between stress-timed and syllable-timed languages, while Ding et al. (2017) and Tilsen and Arvaniti (2013) did not. Possible reasons for discrepancies in outcomes are differences in the type and amount of materials (read sentences versus longer texts and more spontaneous productions; see Section 10.3.2) or differences in the type of signal processing.

10.3.2 Speaking Styles

The most rhythmic style analyzed in the literature is probably English nursery rhymes spoken in synchrony to a metronome in infant-directed speech (Leong and Goswami, 2015). But also, quite generally, infant-directed speech showed a higher amplitude in the delta band (around 2 Hz) than adult-directed speech (Leong et al., 2017), suggesting a clearer foot-based rhythm. In their cross-linguistic study introduced in Section 10.3.1, Tilsen and Arvaniti (2013) further tested for effects of speaking styles (spontaneous speech, read sentences, read passages). They showed an effect of speaking style on time domain aspects of the speech envelopes, distinguishing spontaneous speech from the two read speech recordings (p. 635).

10.3.3 Illocution Type

Hohl and Braun (2023) tested the influence of illocution type, in particular whether a question was meant as an information-seeking or a rhetorical one. This contrast can be signaled lexically – see (1) versus (2) – but also by means of prosody.

(1) What time is it? information-seeking question

(2) Who likes paying taxes? rhetorical question

Prior analyses have shown that rhetorical questions are realized with longer constituent durations than information-seeking questions (Braun et al., 2019; Dehé and Braun, 2019, 2020). Furthermore, they were more often realized with a breathy voice quality (in particular in German) and differed in global intonation. In English polar questions, moreover, participants often shifted the accent from the default accent on the final noun in information-seeking questions (e.g., Does anyone like LEMON?) to the indefinite in rhetorical questions (e.g., Does ANYONE like lemon? [small capitals signal accent location]). The analyses of amplitude envelopes showed differences across illocution types in all languages. The strongest difference (0.5 log power) was observed for English between 3.1 and 4.7 Hz (with larger power for rhetorical than for information-seeking questions). Further smaller differences were present between 1.4 and 1.9 Hz. Icelandic showed differences in slightly lower frequency bands (0.5–1.2 Hz and 3.4–3.8 Hz), but of a smaller magnitude than English (0.2 log power with large variability). The magnitude of the differences for German lay in between (0.4 log power), but occurred in a

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

higher frequency band (7.3–9.1 Hz). The authors argued that the differential effects of illocution type in the three languages cannot be explained by differences in segment durations alone (which were overall largest in Icelandic and smaller in German and English) but must be caused by other differences. Likely predictors are accent placement (for English) and the use of a breathy voice in initial segments in rhetorical questions (in German). We will return to this in the General Discussion (Section 10.5).

10.4 Local Differences in Segmental Length and Pitch Accent Type

10.4.1 Segmental Length

In languages with segmental length contrasts, the length of segments (vowels or consonants) distinguishes lexical candidates. It is expected that the different durations influence the "local" rhythmic structure of target words and hence the modulation spectra. Hohl et al. (2022) extracted spectra for German word pairs differing in vowel length, which were matched for lexical frequency (e.g., *Mitte* [ˈmɪtʰə] "center" versus *Miete* [ˈmiːtʰə] "rent"). Note that length in German is often, but not always, accompanied by vowel quality. The results showed that words with a long vowel had a lower power in a small frequency band just below 2 Hz and, importantly, between 5 and 8.5 Hz (mean difference of 0.4 in log power). Interestingly, the segmental length effect disappeared when the target words were embedded in carrier phrases (*Das nächste Wort war XXX*, "The next word was XXX"), suggesting that the effect of vowel length on the global rhythmic profile of utterances is negligible.

The effect of consonantal length in native speakers of Italian was studied in Einfeldt and Braun (2023) (see top panel of Figure 10.2). In their first experiment, they extracted modulation spectra from six Italian speakers from 16 minimal pairs only differing in phonemic consonant length (e.g., *fato* [ˈfato] "fate" and *fatto* [ˈfatːo] "fact, done"). The results showed that the Italian consonantal length contrast is evident in a large frequency band in the modulation spectra. Compared to the German vowel length contrast, the Italian consonantal length contrast has power differences in a larger frequency interval (2–7.4 Hz compared to 5–8 Hz) and is larger in magnitude (up to 0.5 larger log power in geminates than in singletons). The most likely reason for these cross-linguistic differences in magnitude and frequency band between Italian consonantal and German vocalic length is that the Italian length contrast is distributed over a longer temporal interval, with adjustments to not only the duration of the consonant but with complementary temporal adjustment in the preceding vowel (Payne, 1995) and even the word-initial consonant (Turco and Braun, 2014).

10.4.2 Pitch Accent Type

So far, there has been little information in the literature as to whether/how pitch accent type (i.e., rising or falling) directly affects the amplitude envelope. The data used for testing this factor was originally collected in Zahner-Ritter et al. (2022). The study tested three conditions that differed in the alignment of the low and high tonal targets with respect to the stressed syllable in rising–falling accents. In L+H*, the low (L) tone preceded the stressed syllable while

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

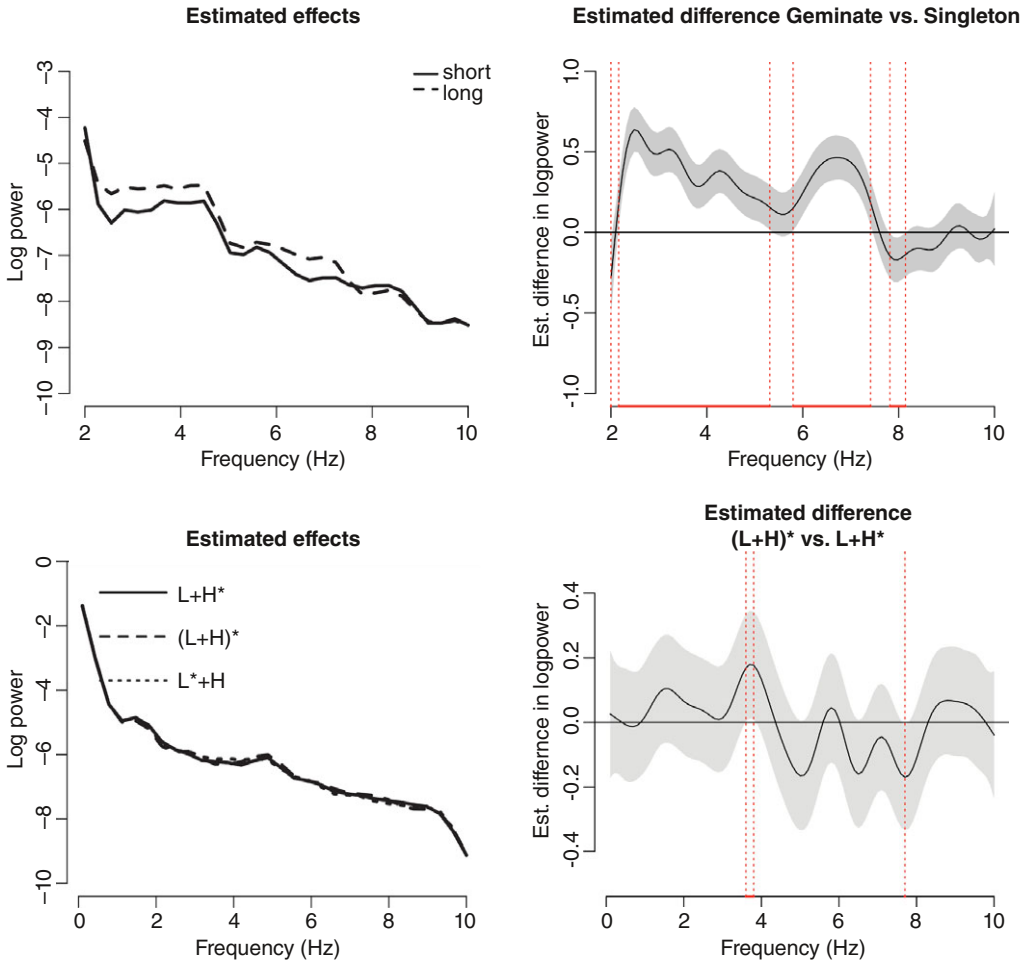


Figure 10.2 Differences in amplitude envelope spectra for two linguistic contrasts.
Estimated effects of consonantal length in Italian (top) and pitch accent type in German (bottom), as predicted by the GAMM model (left panel) and estimated differences (right panel). The gray band shows the 95% CI of the difference.

the high (H) tone was aligned with the stressed syllable; in (L+H)*, both the low and the high tonal targets were aligned with the stressed syllable; and in L*+H, the low tone was aligned with the stressed syllable and the high tone in the following syllable. Since a change in accent shape (on the same word) does not have a strong influence on an utterance's rhythm, we do not expect an effect of intonation condition on the amplitude modulation spectra. The data hence serve as a control condition to test the specificity of the modulation spectra for rhythmic contrasts.

For the purpose of this chapter, we analyzed the modulation spectra of imitation data from native German speakers who imitated the three-way accent contrast in rising–falling accents on words in wh-questions (e.g., *Wer malt denn* MANDALAS? "Who draws PRT mandalas"). The results of the

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

GAMM analysis showed an improvement in the modal fit when adding a random smooth for accent over a model without that random smooth term ($\chi 2(6) = 6903.6$, $p < 0.0001$). However, the estimated effects of the three accents overlap to a large degree (see Figure 10.2, left panel). Pairwise comparisons between the three kinds of accentual realizations showed only minimal differences in power for one of the accent contrasts: There was slightly higher power in (L+H)* than L+H* in the frequency range between 3.6 and 3.8 Hz. The mean difference in power was 0.2, with a lot of variability (95% CI [confidence interval] of the difference: [0.01;0.38]; see bottom panel of Figure 10.2). It is possible that the model picked up on the higher energy exerted by producing a salient rise in the stressed syllable compared to shallower rises in L+H* and L* +H accents. In any case, the effects of fundamental frequency (f0) contour are minimal.

10.5 General Discussion

This chapter provided an overview of the extraction and analysis methods of amplitude envelope modulations in the literature. It then focused on the role of some linguistic factors that may affect the rhythmic structure of utterances: rhythm class, speech style and illocution, as well as segmental length. As a control condition, to test the specificity of the method for capturing rhythmic contrasts, we added an analysis of the effect of pitch accent type on amplitude envelopes. As expected, this factor only marginally affected the amplitude envelope modulation spectra (in a very small frequency band and with little difference in power), supporting the view that amplitude envelopes carry first and foremost rhythmic information.

The different aspects examined here had different effects on the amplitude envelopes, both in terms of the magnitude of the (log) power differences[1] and of the frequency bands that were affected: The largest mean differences across conditions were observed for different rhythm classes (macro rhythm: Brazilian versus European Portuguese; stress timing versus syllable timing: German versus Portuguese), followed by consonantal length in minimal pairs in Italian, and rhetorical versus polar questions in English (mean differences > 0.5). The other linguistic factors had smaller effects (rhetorical versus information-seeking questions in German and Icelandic, differences in vowel length in German) (see Table 10.1). Pitch accent type, the control condition, showed very small effects with a large spread.

Differences in the *lowest frequency bands* (0.5–2.5 Hz) are observed for rhythm class (macro rhythm, stress timing versus syllable timing) as well as English and Icelandic rhetorical versus information-seeking questions (roughly in the range between), that is, aspects that concern the whole utterance. However, differences in the *highest frequency bands* (> 7.3 Hz) are also caused by phrasal aspects (stress timing versus syllable timing, German rhetorical versus information-seeking questions). This is surprising given that these frequency bands capture phone-sized units (Keitel et al., 2018). We conclude that rhythm class and illocution type exert influence on a segmental level as well, which is then tracked in these higher frequency bands. The more local length contrasts had different effects: vowel length in the vicinity of the syllable rate (around 4 Hz),

[1] The mean differences and the lower and upper bounds of the 95% CI of the differences were read off from difference plots of the GAMM analysis.

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

Table 10.1 Comparison of amplitude envelope differences across linguistic contrasts

Overview of differences in amplitude modulation spectra (frequency bands, maximum frequency range of significant differences Δf) and maximum differences in log power (including the 95% CI of the difference) for selected linguistic contrasts, analyzed with the same method, ordered by decreasing maximum mean difference in log power. (EP: European Portuguese; BP: Brazilian Portuguese).

Contrast	Frequency bands of significant differences (Hz)	Max Δf (Hz)	Max. mean difference	95% CI of difference
Macro rhythm (EP vs. BP)	1.3–2.5	1.2	0.7	[0.5;1.0]
Stress timing (German) vs. syllable timing	G vs. EP: 1.4–1.8; 6.7–9.3 G vs. BP: 1.6–2.4; 7–7.9; 9–10	2.6 1.0	0.6 0.6	[0.5;1.0] [0.4;1.0]
Consonant length	2–7.4	5.4	0.6	[0.5;0.7]
English rhetorical vs. information-seeking questions	1.4–1.9; 3.1–4.7	0.5	0.5	[0.35;0.65]
Vocalic length	2.1–3.8	1.7	0.4	[0.1;0.6]
German rhetorical vs. information-seeking questions	7.3–9.1	1.8	0.35	[0.15;0.55]
Icelandic rhetorical vs. information-seeking questions	0.5–1.2; 3.4–3.8	0.7	0.3	[0.15;0.45]
Pitch accent type	(L+H)* vs. L+H*: 3.6–3.8	0.2	0.2	[0.01; 0.4]

consonant length in a very broad frequency band. This is consistent with the observation that the consonantal length contrast is not limited to the duration of the respective consonant but is distributed more widely. The effect of the control condition, pitch accent type, was minimal.

In terms of the size of affected frequency intervals, Italian consonantal length exerted effects in the largest frequency range (Δ = 5.4 Hz; see Table 10.1). Pitch accent type, on the other hand, had differences in the smallest frequency range (Δ = 0.2 Hz). This latter finding strengthens the conclusion that differences in pitch accent type alone do not alter the amplitude envelope (or the rhythmic structure).

For future research, it would be desirable to provide access to off-the-shelf, open-source programs, such that results across languages and factors become more comparable across labs. In terms of the signal processing involved, application-oriented researchers may benefit from mathematically easier methods (e.g., low-pass filtering compared to Hilbert transform) because they can be more readily understood and results may find a broader audience.[2]

Carefully controlled materials provide a useful first step to target the linguistic aspects that influence low-frequency amplitude envelopes. This will ideally lead to a model from which we

[2] Publicly available code is often in MATLAB – for example, for extracting amplitude envelopes – https://github.com/alexisdmacintyre/AcousticLandmarks and https://osf.io/yp4k3/ – or for Spectral-Amplitude Modulation Phase Hierarchy – www.cne.psychol.cam.ac.uk/methods-spectral-amplitude-modulation-phase-hierarchy. Python code for extracting and visualizing rhythm formants can be found here: https://github.com/dafyddg/RFAGUI.

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

can derive testable predictions for future research and for analyzing more heterogeneous speech materials. To date, it clearly seems that (macro-)rhythmic differences across languages as well as consonantal length have a robust influence in the low-frequency range (around 2 Hz). Studies from more languages are needed to test the generalizability of this finding.

10.6 Acknowledgements

I thank Friederike Hohl, Hendrik Behrens-Zemek, and Marieke Einfeldt for recording participants and pre-processing data. Thanks are further due to Volker Dellwo for providing the script to extract the narrowband envelopes, to Sonia Frota and Antje Strauß for general discussion about low-frequency amplitude envelopes and cross-linguistic differences, to Anna Verebély for help with figure quality, and to Dafydd Gibbon for valuable comments on an earlier version.

Box 10.1 Chapter Overview

Summary

After an introduction to amplitude envelopes, the chapter compared several linguistic contrasts (rhythm class, speaking style, consonantal and vocalic length, pitch accent type) with regards to their respective effects on amplitude envelope modulation spectra. Results showed the largest power differences for rhythm class and the smallest differences for pitch accent type.

Implications

Amplitude envelopes provide a holistic measure that efficiently captures rhythmic differences. The comparison of controlled data allowed us to isolate different factors and work towards a model of factors that influence amplitude envelopes. Future research will benefit from more languages and the provision of transparent open-source scripts to ease comparison.

Gains

Neurobiology has long thought of links between amplitude envelopes in the speech signal and brain oscillations. The current research allows us to better understand the links between linguistic factors and the amplitude envelope, ultimately leading to a model that describes the connections between linguistic units and brain oscillations.

REFERENCES

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46–63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

Assaneo, M. F., Ripolles, P., Orpella, J., et al. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, 22(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

Barry, W. J. (2007). Rhythm as an L2 problem: How prosodic is it? In J. Trouvain and U. Gut (Eds.), *Non-Native Prosody: Phonetic Description and Teaching Practice* (pp. 97–120). Mouton de Gruyter.

Braiman, C., Fridman, E. A., Conte, M. M., et al. (2018). Cortical response to the natural speech envelope correlates with neuroimaging evidence of cognition in severe brain injury. *Current Biology*, 28(23), 3833–3839. https://doi.org/10.1016/j.cub.2018.10.057

Braun, B., Dehé, N., Neitsch, J., Wochner, D., and Zahner, K. (2019). The prosody of rhetorical and information-seeking questions in German. *Language and Speech*, 62(4), 779–807. https://doi.org/10.1177/0023830918816351

Caetano, M., and Rodet, X. (2011). Improved estimation of the amplitude envelope of time-domain signals using true envelope cepstral smoothing. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4244–4247. https://doi.org/10.1109/ICASSP.2011.5947290

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLOS Computational Biology*, 5(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.

Dehé, N., and Braun, B. (2019). The prosody of rhetorical questions in English. *English Language and Linguistics*, 24(4), 607–635. https://doi.org/10.1017/S1360674319000157

Dehé, N., and Braun, B. (2020). The intonation of information-seeking and rhetorical questions in Icelandic. *Journal of Germanic Linguistics*, 32(1), 1–42.

Ding, N., Patel, A. D., Chen, L., et al. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, 81, 181–187.

Einfeldt, M., and Braun, B. (2023). Amplitude envelopes as a means to study length contrasts: Evidence from L1 and L2 Italian. In R. Skarnitzl and J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences 2023* (pp. 664–668). International Phonetic Association.

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

Gibbon, D. (2021). The rhythms of rhythm. *Journal of the International Phonetic Association*, 53(1), 233–265. https://doi.org/10.1017/S0025100321000086

Goswami, U. (2018). A neural basis for phonological awareness? An oscillatory temporal-sampling perspective. *Current Directions in Psychological Science*, 27(1), 56–63.

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.

Grabe, E., and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology* (pp. 515–546). Mouton de Gruyter.

Gross, J., Hoogenboom, N., Thut, G., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLOS Biology*, 11(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Hohl, F., and Braun, B. (2023). Using amplitude envelope modulation spectra to capture differences between rhetorical and information-seeking questions. Language Technology Conference (LTC), Poznań, Poland.

Hohl, F., Behrens-Zemek, H., and Braun, B. (2022). Amplitude envelopes as a means to quantify vowel length contrasts. Phonetik und Phonologie im deutschsprachigen Raum: P&P 18. https://doi.org/10.11576/pundp2022-1008

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

Jun, S.-A. (2012). Prosodic typology revisited: Adding macro-rhythm. Proceedings of Speech Prosody 2012, Shanghai, China, pp. 535–538. https://doi.org/10.21437/SpeechProsody.2012-135

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, 16(3), e2004473. https://doi.org/10.1371/journal.pbio.2004473

Kolly, M.-J., and Dellwo, V. (2014). Cues to linguistic origin: The contribution of speech temporal information to foreign accent recognition. *Journal of Phonetics*, 42, 12–23. https://doi.org/10.1016/j.wocn.2013.11.004

Lau, J. C. Y., Patel, S., Kang, X., et al. (2022). Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study. *PLoS ONE*, 17(6), e0269637. https://doi.org/10.1371/journal.pone.0269637

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

Leong, V., Kalashnikova, M., Burnham, D., and Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind*, 1(2), 78–90. https://doi.org/10.1162/OPMI_a_00008

Loizou, P. C., Dorman, M., and Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America*, 106(4), 2097–2103. https://doi.org/10.1121/1.427954

Ludusan, B., Origlia, A., and Cutugno, F. (2011). On the use of the rhythmogram for automatic syllabic prominence detection. Proceeding of Interspeech 2011, pp. 2413–2416. https://doi.org/10.21437/Interspeech.2011-628

MacIntyre, A. D., Cai, C. Q., and Scott, S. K. (2022). Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques. *Journal of the Acoustical Society of America*, 151(3), 2002–2026. https://doi.org/10.1121/10.0009844

Marcus, S. M. (1981). Acoustic determinants of perceptual center (p-center) location. *Perception & Psychophysics*, 30(3), 247–256.

Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. *Journal of the Acoustical Society of America*, 58(4), 880–883.

Obleser, J., Herrmann, B., and Henry, M. (2012). Neural oscillations in speech: Don't be enslaved by the envelope. *Frontiers in Human Neuroscience*, 6, 00250. www.frontiersin.org/articles/10.3389/fnhum.2012.00250

O'Sullivan, J. A., Power, A. J., Mesgarani, N., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, 25(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

Payne, E. M. (1995). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association*, 35(2), 153–181.

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

Poeppel, D., and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21, 322–334.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.

Rathcke, T., Lin, C.-Y., Falk, S., and Dalla Bella, S. (2021). Tapping into linguistic rhythm. *Laboratory Phonology*, 12(1), 11. https://doi.org/10.5334/labphon.248

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences*, 336(1278), 367–373.

Tilsen, S., and Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, 134(1), 628–639. https://doi.org/10.1121/1.4807565

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press

Tilsen, S., and Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124(2), EL34–EL39. https://doi.org/10.1121/1.2947626

Todd, N. P. M., and Brown, G. J. (1994). A computational model of prosody perception. Third International Conference on Spoken Language Processing (ICSLP 1994), pp. 127–130. https://doi.org/10.21437/ICSLP.1994-35

Turco, G., and Braun, B. (2014). Initial consonant strengthening in Italian geminate words. 10. Tagung Phonetik und Phonologie im deutschsprachigen Raum. https://doi.org/10.13140/2.1.2212.2242

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *Journal of the Acoustical Society of America* , 142(4), 1976. https://doi.org/10.1121/1.5006179

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Press.

Zahner-Ritter, K., Einfeldt, M., Wochner, D., et al. (2022). Three kinds of rising-falling contours in German wh-questions: Evidence from form and function. *Frontiers in Communication*, 7, 838955. www.frontiersin.org/articles/10.3389/fcomm.2022.838955

Zhang, Y., Zou, J., and Ding, N. (2023). Acoustic correlates of the syllabic rhythm of speech: Modulation spectrum or local features of the temporal envelope. *Neuroscience & Biobehavioral Reviews*, 147, 105111. https://doi.org/10.1016/j.neubiorev.2023.105111

https://doi.org/10.1017/9781009295888.013 Published online by Cambridge University Press