Amplitude envelope modulations across languages reflect prosody

Sonia Frota[1]*, Marina Vigário*[1]*, Marisa Cruz*[1]*, Friederike Hohl*[2]*, Bettina Braun*[2]

[1]Center of Linguistics, School of Arts and Humanities, Universidade de Lisboa, Portugal
[2]Department of Linguistics, University of Konstanz, Germany
{sfrota, mvigario, marisac}@edu.ulisboa.pt, {friederike.hohl, bettina.braun}@uni-konstanz.de

Abstract

The speech signal has been shown to contain a fine structure that consists of the fast changing spectral content (e.g., formant transitions, voicing, spectral energy distributions), together with amplitude modulations of the envelope with different timescales. These different modulation frequencies have been associated with linguistic units of different sizes, and neuronal oscillations seem to track this linguistic structure. As the amplitude envelope mostly captures suprasegmental information, the different modulation frequencies are natural candidates to convey prosodic information. In this paper we put these assumptions to the test by comparing effects of sentence length and language, focusing on languages with distinct prosodic profiles: Brazilian Portuguese (syllable-timed), European Portuguese (mixed rhythm, with syllable-timed and stress-timed properties), and German (stress-timed). There are further differences regarding the roles of the syllable, the foot, the prosodic word and the intonation phrase. We analyzed wideband amplitude envelopes using general additive mixed models and show that German differs from Brazilian Portuguese and European Portuguese in the delta (1-2Hz) and theta bands (6-8Hz). European and Brazilian Portuguese also differ, but only in the delta band (1-2Hz). The language differences in amplitude modulation are discussed in terms of speech rhythm and differences in prosodic structure across languages.

Index Terms: amplitude envelope, rhythm, prosodic structure, German, Brazilian Portuguese, European Portuguese, general additive mixed models

1. Introduction

Within the cortical oscillatory framework, it is proposed that neural oscillations at different frequency bands track linguistic structure in speech, such as phonemes, syllables, lexical words, syntactic phrases, or sentences [e.g., 1, 2, 3]. Some scholars have argued that neural tracking can be dissociated to some extent from the encoding of cues in the speech signal, and that oscillations at a specific rate, in particular at the theta band (around 4 Hz), are language independent [4-6]. Other studies have emphasized neural entrainment to speech at different timescales, matching the specific properties of the stimuli [e.g., 7, 8]. Irrespective of the precise mechanisms supporting neural entraining to speech, there is ample evidence for the key role played by the speech envelope in the process [9]. Strikingly, several researchers have argued that the amplitude envelope mostly captures suprasegmental information [10, 11] and thus the different modulation frequencies are natural candidates to convey prosodic information.

In other words, they may reflect prosodic structure (e.g., syllable, foot, prosodic word, and intonation phrase rates in speech), as well as key properties of prosody such as syllable structure, stress, or phrasal prominence. Consequently, prosody would play a central role in the neural tracking of speech.

Modulation frequencies can be extracted from the speech signal in a number of ways [9]. Most procedures first filter the sound into a number of frequency bands (spaced either logarithmically or such that they are equidistant on the cochlea), typically in the range between 100 and 8,0000 (or 10,000 Hz). These signals are then filtered to remove the high-frequency components, leaving frequencies in the range of 0 to approximately 10Hz. These narrowband envelopes are then summed and the modulation frequencies are derived by Fourier analysis. The result is a spectrum, i.e. power values across frequency.

Many studies have focused on a specific timescale, related to the syllable rate [e.g., 9 for a review, 12]. Some studies, however, highlighted other timescales in the speech signal which are related to phrasal prosody and word prosody. For example, [13] found clusters of energy at three different timescales that approximated the word stress/stressed syllable rate (~2Hz), the syllable rate (~5Hz), and the onset-rime rate (~20Hz). In [7], four different timescales are described: the phrasal scale (0.6-1.3Hz), the word scale (1.8-3Hz), the syllable (2.8-4.8Hz), and the phoneme (>8Hz) scale.

Very few studies have looked at the temporal regularities in the acoustic signal across languages, and attempted to compare them. Amplitude envelopes with globally similar shapes have been reported, including peaks or increase in power for low modulation frequencies followed by a decrease in amplitude [9, 14, 15]. Importantly, not enough attention has been given to variations in the amplitude modulations of the speech envelope across languages. However, amplitude modulations could differ across languages because languages differ in some of the features that may affect the amplitude envelope, in particular those related to syllable structure, speech rhythm, phrasal prosody and word prosody. Despite general convergent findings across speakers, speech materials and languages, variations in the frequency bands for each language (English and French) were noted in [12], which were not further explored. Similarly, [15] reported no differences in the amplitude modulation spectra of 9 languages (which were not grouped according to their prosodic properties, though). However, the spectra were normalized by their maximum amplitude values, precluding amplitude differences to emerge.

To the best of our knowledge, only one study [14] examined the effects of prosodic factors on amplitude modulations, by comparing 10 languages grouped according to speech rhythm type and phrasal prominence patterns. They found overall similar spectra, together with significant effects of both prosodic factors. In particular, there was an effect of rhythm on the type of amplitude modulation spectra that is closer to that computed in the current study, with higher amplitudes in the target window analyzed (2-8Hz) in stress-timed languages.

To what extent regularities in the amplitude envelope are relatively stable across languages, or differentiate between languages, and at what timescales, constitutes an open and timely empirical research question that the current study addresses.

2. Experiment

We compared the amplitude envelopes from three different languages with different prosodic properties, Brazilian Portuguese (BP), European Portuguese (EP) and German (G), from five speakers in each language, each producing sentences with different lengths (15-18 syllables). The role of sentence length has been hardly examined in prior work.

2.1 The languages

BP, EP and G were chosen due to their distinct prosodic profiles. The prosodic properties of BP approximate it to other Romance languages (e.g., Italian, Spanish). The prosodic properties of German are similar to those that characterize other Germanic languages (e.g., Dutch, English). EP, in turn, has an atypical prosodic profile within the Romance space, as it mixes prosodic properties more of the Germanic type with prosodic properties of the Romance type (e.g., [16-19]).

Independently of the different views on speech rhythm found in the literature [20-22], the finding that listeners perceive rhythmic differences across languages remains. G has been described as a stress-timed language, where the intervals between stresses, the trochaic patterning of feet, and the complexity of syllable structure play a major role. BP, by contrast, is a syllable-timed language, characterized by a simple syllable structure and tendency to a regular alternation of consonants and vowels. EP has been reported to display a mix of syllable- and stress-timed properties, with a simple phonological syllable structure combined with vowel reduction and vowel deletion phenomena. The prosodic word is the domain for resyllabification in G, whereas a high-level phrasal domain for resyllabification, the intonation phrase, is found in BP and EP [16, 18]. However, BP differs from EP in the features of prosodic phrases, with a lower phrase domain (or even the word) being marked with a pitch accent in BP, whereas only the intonation phrase has to be marked with a pitch accent in EP. Consequently, the distribution of accentual prominences is sparse in EP, and dense in BP [23]. A further difference concerns the direction of clitic attachment, which is proclitic in BP, and also in EP (but with the notable exception of postverbal pronominal clitics), and tends to be enclitic in G [16, 24]. Hence, the distribution of unstressed syllables relative to stress is expected to vary across the languages.

These patterns are manifested in the tonal, prominence and durational speech cues, and might as well be manifested in the rates of the various prosodic units. It is also expected that the language-specific prosodic profiles impact the amplitude modulation envelope yielding differences at the lower modulation rates (namely, delta and theta).

2.2 Methods

2.2.1 Participants

Five female speakers were recorded for each language. They were between 20 and 55 years old and took part voluntarily. The German speakers were from different areas of Germany (two from the north, two from the south, one from the middle), but they all read the sentences in Standard German. The EP speakers were all from the greater Lisbon area. The BP speakers were from the Southeastern areas of São Paulo and Espírito Santo. The speakers signed informed consent that the recordings could be used for a study on cross-linguistic comparisons of speech.

2.2.2 Materials

For each language, we constructed 20 sentences with four different sentence lengths, ranging from 15 to 18 syllables. The sentences did not demand contrastive accents, cf. (1-3) for an example sentence in G, EP and BP with 15 syllables (stressed syllables in bold, brackets marking theoretical prosodic words). We used the rhythm corpus within the Interactive Atlas of Portuguese Prosody project [25] as a baseline for constructing the sentences in German. The German sentences were adapted to arrive at the respective numbers of syllables.

(1) (**Ho**ffentlich) (**gibt** es im) (**näc**hsten) (**Som**mer) (**mehr**) (**Nie**der)(**schlä**ge). (G)

'Hopefully there will be more precipitation next summer.'

(2) (O me**ni**no) (levan**tou**-se) (**ce**do) (para **ver)** (o **sol)** (EP)

(3) (O me**ni**no) (se levan**tou)** (**ce**do) (para **ver)** (o **sol)** (BP)

'The boy got up early to see the sun'

2.2.3 Procedure

Recording. The German participants were recorded in a sound-proof booth at the University of Konstanz, using an MXL 990, condensor microphone, and were digitized onto a computer with 44.1 kHz, 16 Bit. The European Portuguese participants, were recorded at the Phonetics Lab of the University of Lisbon, using DPA microphones and a sampling frequency of 44.1kHz. The Brazilian Portuguese recordings were made at different sites and digitized onto a computer with 44.1kHz, 16 bit. In some cases, participants repeated sentences. We only included one rendition for each sentence. In total, we analyzed 60 sentences (20 recordings per speaker, five for each sentence length). The average durations of the sentences and the number of prosodic words across languages are summarized in Table 1.

Table 1. Average sentence duration and number of words across languages.

Sentence length	**Duration (s)**			**prosodic words (N)**		
language	G	BP	EP	G	BP	EP
15 syllables	2.9	2.2	2.0	9.2	4.8	4.8
16 syllables	2.8	2.5	2.2	9.2	5.4	5.2
17 syllables	3.1	2.7	2.4	9.0	6.0	5.8
18 syllables	3.1	2.8	2.6	8.2	6.0	6.0

Since average durations differed across languages, the durations were normalized across languages for each sentence length, using PSOLA resynthesis, as implemented in praat [26]. The resulting average durations of the sentences were 2.38s, 2.49s, 2.75s and 2.83s, respectively. The prosodic word rates differed across languages (cf. Table 2.), as well as the distribution of unstressed syllables (cf. Table 3).

Table 2. Average phonological syllable and prosodic word rates after duration normalization.

Sentence length	**Syllable rate per sec**	**prosodic word rate per sec**		
language	all languages	G	BP	EP
15 syllables	6.3	3.9	2.0	2.0
16 syllables	6.4	3.7	2.2	2.1
17 syllables	6.2	3.3	2.2	2.1
18 syllables	6.4	2.9	2.1	2.1

Table 3. Number of unstressed syllables.

Sentence length	**syllables after stress**			**syllables between stresses**		
language	G	BP	EP	G	BP	EP
15 syllables	1.1	0.3	0.9	1.2	2.3	2.4
16 syllables	0.9	0.7	0.8	1.2	2.2	2.4
17 syllables	1.2	0.8	0.8	1.3	1.9	1.9
18 syllables	1.0	0.7	0.7	1.2	1.8	1.8

Extraction of the amplitude envelope modulation spectra. We analyzed the wideband amplitude envelopes of the productions, following the procedure in [12]. The first step was to calculate the narrowband amplitude envelopes. For this analysis we used a script developed by Volker Dellwo and Lei He [27, 28].

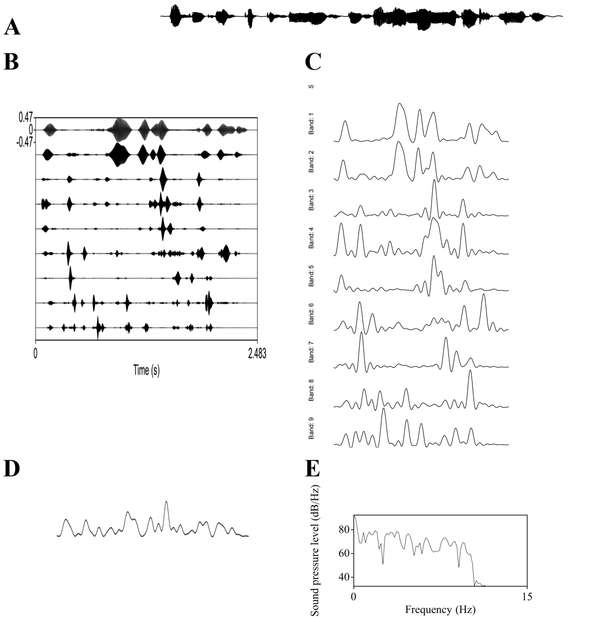

Figure 1: *Processing step: panel A: sound pressure wave; panel B: filtered signals from frequency bands (lowest frequency band at the top); panel C: narrowband amplitude envelopes; panel D: wideband amplitude envelope; panel E: spectrum of the wideband amplitude envelope.*

The speech signal (panel A in Fig. 1) was first downsampled to 22050 Hz and then filtered into nine frequency bands in the range from 100–10,000 Hz, which are equidistant on the cochlear map [2]. 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, see panel B in Fig. 1. To remove high-frequency components, the amplitude envelopes were computed using the Hilbert transform. The resulting narrowband envelopes are shown in panel C of Fig. 1. These narrowband amplitude envelopes were then added to compute the wideband amplitude envelope (panel D of Fig. 1), which were spectrally analyzed (panel E of Fig. 1) in 100 0.1-Hz steps, resulting in 30000 data points (3 languages x 5 speakers x 4 sentence lengths x 5 sentences x 100 frequency bands).

Statistical modeling. We initially calculated a **linear-mixed effect regression model** with log-power as dependent variable, language, frequency band and sentence length as fixed effects and participants and items as crossed random effects [29]. This model showed main effects for all fixed effects and interactions between frequency band and language and frequency band and sentence length. For further analysis of the interactions, we used **generalized additive mixed models**, GAMMs [30-34]. They are suited to pinpoint where differences occur; taking into account non-linear relationships and auto-correlation [35, 36]. We modelled non-linear dependencies of *language* and *sentence length* over the frequency bands using smooth functions. 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., 36]. *Language* and *sentence length* were further added as fixed effects. For model fitting, we employed the R package *mgcv* [37, 38]; the package *itsadug* was used to plot the model results [39]. The response variable was log-normalized power. The initial model included *language and sentence length* as parametric effects (fixed effects and in interaction), along with a factor smooth for the interaction of *language* over frequency bands, s(*fband*, by = *language*) and a factor smooth for the interaction of *sentence length* over frequency bands, s(*fband*, by = *length*). Smooths for *speakers* (random intercept and over frequency bands) were also included (s(speaker, fband, by='re'). The model was corrected for auto-correlation in the data using a correlation parameter, determined by the acf_resid() function (package *itsadug* [39]). 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 (we thank Cesko Voeten for discussion). The model including the smooth terms that captured the interactions with frequency bands was subsequently compared to a simpler model without the respective smooth term, using the function CompareML().This comparison tested whether the inclusion of the smooth term significantly improved the fit of the model in terms of Maximum Likelihood [see 40]. We removed non-significant term if this did not deteriorate the fit of the model.

2.3 Results

Figure 2 shows the average power across frequency bands split by language. The final GAMM included *language* and *sentence length* as parametric effects, along with factor smooths for the interaction of *language* over *frequency bands, sentence length* over *frequency bands,* as well as random smooths for *speakers* and *items*, varying over *frequency bands*. Model comparisons showed no significant interaction between the parametric effects *language* and *sentence length*. The model accounted for 40.6% of the variance. Fig. 3 shows the averaged difference between EP and BP (top panel), between EP and G (middle panel) and between BP and G (bottom panel) as a solid line. The grey shading displays the 95%CI (confidence interval) of the predicted mean difference. The difference is significant if zero is not included in the 95%CI. This is marked by the vertical red lines on the x-axis. The main significant differences were as follows: EP has lower spectral power than BP from 1.3 to 2.5 Hz (top panel), EP has lower spectral power than G in the band from 1.4 to 1.8Hz and lower spectral power from 6.7 to 9.3Hz (middle panel). BP has higher power than G in the band from 1.6 to 2.4Hz and lower power from 7.0 to 7.9Hz and from 8.99 to 10Hz.

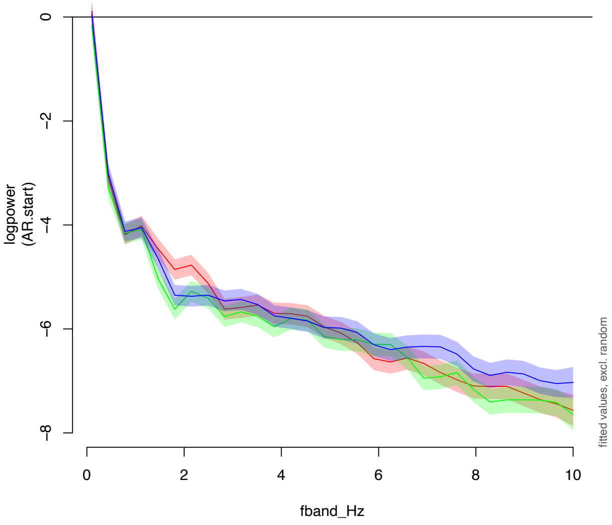

Figure 2: *Energy in frequency bands from 1 – 10Hz, split by language (red: BP, green: EP, blue: G)*

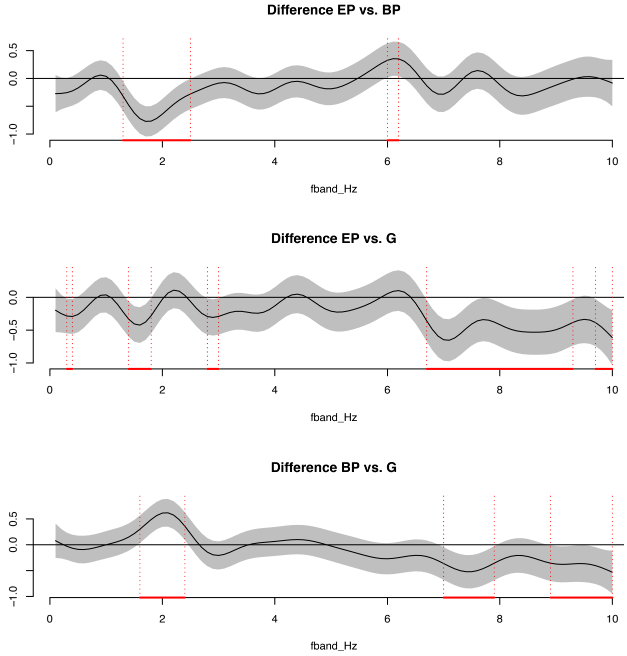

Figure 3: Predicted pairwise differences between languages. Red areas show significant differences.

3. Experiment 2

Exp. 2 is a control experiment to test whether the differences across languages can be explained by differences in recording conditions and resulting signal to noise ratios, instead of prosody. We compared the German utterances with 15 syllables to identical utterances to which white noise with 40dB was added (using [41]). The data were processed and analyzed as in Experiment 1. The difference plots did not reveal differences across noise conditions.

4. Discussion

In the current paper we examined whether languages with distinct prosodic profiles are distinguished on the basis of amplitude envelope modulations. We found that German differs from Brazilian Portuguese and European Portuguese in the delta (1-2Hz) and theta bands (6-8Hz). European and Brazilian Portuguese also differ, but only in the delta band (1-2Hz). In previous work [7,9,12,13], similar timescales have been related to the phrasal, word or word stress rates (delta band) and the syllable scale (theta band), respectively. Moreover, an effect of rhythm was previously found in [14] with stress-timed languages showing higher amplitudes than syllable-timed languages in a previously defined target window (2-8Hz).

The differences across languages were not influenced by sentence length. Shorter sentences had more energy than longer sentences, but, crucially, this did not interact with language. However, larger differences in length may be a stronger test case. The amplitude modulation differences between EP and BP in the current study match the contrast between the languages in the features of prosodic phrases, namely in the patterns of accentual prominences, that is expected to be reflected in the delta band. Interestingly, more accentual prominences, a feature of BP, seem to be reflected in higher spectral power in the delta band. No difference between EP and BP was found in the theta band, in agreement with the similarity in syllable structures and distribution of stressed and unstressed syllables (cf. Tab. 2 and 3). The mixed rhythmic nature of EP had no impact on the theta band, suggesting that it is not reflected in the modulation domain. These results are in line with the perceptual salience of syllable-timed properties over stress-timed properties in EP, and the key role played by intonation in distinguishing the two languages [42]. The modulation differences found between G, on the one hand, and EP and BP on the other, also match the prosodic contrasts between the languages. EP tends to have sparser accentual prominences than G, and BP to have more, and both EP and BP have lower word rates than G, factors that might influence the delta band. Additionally, G is a stress-timed language, unlike EP and BP, and displays a distribution of stressed and unstressed syllables different from both EP and BP (cf. Tab 2 and 3), properties that might influence the theta band. Remarkably, stress-timed G shows higher power than EP or BP in the theta band, along the lines of [14].

5. Conclusion

The findings strongly suggest that the amplitude modulation envelope reflects prosody, in particular rhythm and aspects of phrasal and word prosody. Sentence length did not interact with language, which strengthens the cross-linguistic differences. Beyond the overall similar shape of amplitude envelopes across languages, meaningful differences emerge in the delta and theta bands, which reflect prosodic differences across languages. These differences may have implications for the neural tracking of speech, speech processing and language acquisition.

6. Acknowledgements

We thank Volker Dellwo for providing the praat-scripts for extracting the narrowband amplitude envelopes and the Hilbert transform. Furthermore, we thank Antje Strauß for discussion on amplitude envelopes and Marieke Einfeldt for comments on an earlier draft. This research was partially funded by Fundação para a Ciência e Tecnologia, Portugal (Grant UIDB/00214/2020) awarded to the Center of Linguistics.

7. References

[1] M.F. Assaneo, et al., "*Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning". Nat* Neurosci, 22(4): p. 627-632. 2019.

[2] J. Gross, et al., "*Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain".* PLOS Biology. 2013.

[3] D. Poeppel, "*The neuroanatomic and neurophysiological infrastructure for speech and language".* Curr Opin Neurobiol, 28: p. 142-9. 2014.

[4] X. Teng, et al., "*Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process".* European Journal of Neuroscience, 48(8): p. 2770-2782. 2018.

[5] H. Getz, et al., "*Cortical tracking of constituent structure in language acquisition".* Cognition, 181: p. 135-140. 2018.

[6] N. Ding, et al., "*Cortical tracking of hierarchical linguistic structures in connected speech".* Nature Neuroscience, 19(1): p. 158. 2016.

[7] A. Keitel, J. Gross, and C. Kayser, "*Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features".* PLoS Biol, 16(3): p. e2004473. 2018.

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

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

[10] B.R. Myers, M.D. Lense, and R.L. Gordon, "*Pushing the Envelope: Developments in Neural Entrainment to Speech and the Biological Underpinnings of Prosody Perception".* Brain Sciences, 9(3). 2019.

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

[12] C. Chandrasekaran, et al., "*The Natural Statistics of Audiovisual Speech".* PLOS Computational Biology, 5(7): p. e1000436. 2009.

[13] V. Leong and U. Goswami, "*Acoustic-emergent phonology in the amplitude envelope of child-directed speech".* PloS one, 10(12): p. e0144411. 2015.

[14] L. Varnet, et al., "*A cross-linguistic study of speech modulation spectra".* J Acoust Soc Am, 142(4): p. 1976. 2017.

[15] N. Ding, et al., "*Temporal modulations in speech and music".* Neuroscience & Biobehavioral Reviews, 81: p. 181-187. 2017.

[16] M. Vigário, "*Phonetics and phonology of Portuguese"*, in *Manual of Romance phonetics and phonology*, C. Gabriel, R. Gess, and T. Meisenburg, [Eds], De Gruyter: Berlin/New York, 2021.

[17] S. Frota, "*The intonational phonology of European Portuguese"*, in *Prosodic typology II: The phonology of intonation and phrasing*, S.-A. Jun, [Ed], Oxford University Press: Oxford. p. 6-42, 2014.

[18] R. Wiese, "*The Phonology of German".* Oxford: Oxford University Press, 1996.

[19] S. Frota and M. Vigário, "*On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case".* Probus, 13(2): p. 247–275. 2001.

[20] L. White, S.L. Mattys, and L. Wiget, "*Language categorization by adults is based on sensitivity to durational cues, not rhythm class".* Journal of Memory and Language, 66(4): p. 665-679. 2012.

[21] A. Arvaniti, "*Rhythm, timing and the timing of rhythm".* Phonetica, 66: p. 46-63. 2009.

[22] M. Nespor, M. Shukla, and J. Mehler, "*Stress-timed vs. syllable timed languages"*, in *The Blackwell Companion to Phonology*, M. van Oostendorp, et al., [Eds], Wiley-Blackwell: Malden, MA. p. 1147-1159, 2011.

[23] S. Frota and J.A. de Moraes, "*Intonation in European and Brazilian Portuguese"*, in *The Handbook of Portuguese Linguistics*, W.L. Wetzels, J. Costa, and S. Menuzzi, [Eds], Wiley Blackwell: New Jersey. p. 141-166, 2016.

[24] A. Lahiri and F. Plank, "*Phonological phrasing in Germanic: The judgement of history, confirmed through experiment.".* Transactions of the Philological Society, 108(3): p. 370-398. 2011.

[25] S. Frota, "*Interactive Atlas of the Prosody of Portuguese Project"*, University of Lisbon: http://labfon.letras.ulisboa.pt/InAPoP, 2012-2015

[26] P. Boersma and D. Weenink, "*Praat: doing phonetics by computer"*: http://www.praat.org/, retrieved 11 May 2018, 2018

[27] L. He and V. Dellwo. "*Amplitude envelope kinematics of speech signal: parameter extraction and applications".* in *Elektronische Sprachsignalverarbeitung 2017*. Dresden, Germany, 2017.

[28] L. He and V. Dellwo. "*A Praat-based algorithm to extract the amplitude envelope and temporal fine structure using the Hilbert transform".* in *Interspeech 2016*. San Francisco, USA, 2016.

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

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

[31] S.N. Wood. *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. 2015 [cited 2017.

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

[33] M. Wieling, E. Margaretha, and J. Nerbonne, "*Inducing a measure of phonetic similarity from pronunciation variation".* Journal of Phonetics, 40(2): p. 307-314. 2012.

[34] K. Zahner, S. Kutscheid, and B. Braun, "*Alignment of f0 peak in different pitch accent types affects perception of metrical stress".* Journal of Phonetics, 74: p. 75-95. 2019.

[35] R.H. Baayen, et al., "*Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models"*, in *Mixed effects regression models in linguistics*, D. Speelman, K. Heylen, and D. Geeraerts, [Eds], Springer: Berlin. p. 49-69, 2018.

[36] M. Wieling, "*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: p. 86-116. 2018.

[37] S.N. Wood, "*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): p. 3-36. 2011.

[38] S.N. Wood, "*Generalized additive models: An introduction with R"*, 2nd ed, ed, n. edition. Boca Raton [u.a.]: CRC press, 2017.

[39] J. van Rij, et al., "*itsadug: Interpreting time series and autocorrelated data using GAMMs"*, 2017

[40] V. Porretta, B.V. Tucker, and J. Järvikivi, "*The influence of gradient foreign accentedness and listener experience on word recognition".* Journal of Phonetics, 58: p. 1-21. 2016.

[41] R. Corretge, "*Praat Vocal Tookit"*: http://www.praatvocaltoolkit.com, 2012-2021

[42] S. Frota, M. Vigário, and F. Martins. "*Language discrimination and rhythm class: evidence from Portuguese".* in *Proceedings of Speech Prosody* Aix-en-Provence: Laboratoire Parole et Langage, 2002.