Acoustic cues to stress perception in Spanish – a mismatch negativity study

Karolina Broś

University of Warsaw
k.bros@uw.edu.pl

Abstract

In this paper we investigate the cues governing stress perception in Spanish – an issue that has been subject to debate and remains largely unresolved. While there is general agreement as to the ability of Spanish listeners to detect and reliably produce stress contrasts, there is some disagreement on the roles played by individual stress cues. In this study, we focus on early stress processing in a passive oddball paradigm aimed at eliciting a mismatch negativity response to changes in stress. Individual features (spectral tilt and f0) rather than feature bundles are used to induce stress shift. A vowel change is used as a control condition. The results show that while both spectral tilt and f0 manipulations result in mismatch negativity, the latter evokes a stronger response that equals the effect of a change in the quality of the stressed vowel. The results are in line with previous studies on stress correlates in other languages, pointing to a possible cross-linguistic pattern.

Index Terms: stress perception, MMN, mismatch negativity, Spanish, spectral tilt

1. Introduction

The Spanish language is characterised by variable stress with an uneven distribution of penultimate, antepenultimate and final-stressed words. Over 64% of all Spanish words are stressed on the penultimate syllable while antepenults constitute at most 8% [1, 2] and can be therefore considered exceptional. Further evidence for this comes from studies that look at disyllabic and trisyllabic words separately (70% of them are penults, [3, 4]). This frequency pattern, supported by grammatical predictability of certain forms over others, suggests that there is a default penult stress pattern in Spanish and other types of stress, especially antepenults, are lexical exceptions and have to be learned (see e.g. [5]). At the same time, given variable stress and the existence of minimal pairs, we expect that Spanish speakers are sensitive to stress differences in perception and successfully identify stressed syllables (they are not 'stress-deaf', [6]). It is, however, not entirely clear which syllable characteristics are the most salient and guide stress marking in production and perception. Although many studies have been conducted on the phonetics of Spanish stress to date, there is no general agreement concerning the actual cues that govern it.

The aim of this paper is to take on this debate and provide direct, objective evidence for the role the most important cues pointed out in behavioral studies play in the perception of Spanish stress. To this end, we conducted a passive oddball study aimed at eliciting mismatch negativity [7], an ERP component illustrating perceptual sensitivity to salient changes in the presented stimuli. To look at the relative importance of particular stress cues, we compared listeners' neurophysiological responses to stress shift driven by f0 vs intensity in the higher regions of the spectrum, i.e. spectral tilt, both previously reported as robust perceptual stress cues in many languages (see section 2.2). Thus, we look at the neurophysiological response to both f0 and spectral tilt changes that affect stress perception. Changes of stressed vowel quality were used as a control condition checking the participants' general sensitivity to melodic shifts in the acoustic signal.

2. Theoretical background

2.1. Phonetic studies on Spanish stress cues

As mentioned in the Introduction, the literature is not in agreement as to what constitutes the principal cue to stress perception in Spanish. Cues to lexical stress were first discussed in Contreras [8] and Navarro [9], according to whom variations in pitch can be associated with intonation, while duration and intensity are the domain of general lexical stress. In addition, Navarro called lexical stress an 'intensity accent' that is different from the 'pitch accent' marking intonation. This is supported by the fact that many lexical items in Spanish tend to differ solely in stress and no other phonetic property. Such triplets as *límite-limite-limité* 'limit' (noun), 'limit' (1st / 3rd p. sg. present subjunctive), 'limited' (1st p. sg. past tense) are often cited to exemplify this. According to Navarro, the only phonetic difference between these variants lies in intensity, which means that Spanish people should be particularly sensitive to this cue ([9], pp. 177). At the same time, there are studies that go against this claim. Llisterii and colleagues [10], for instance, state that intensity alone is not sufficient for the identification of the stressed syllable. The same applies to duration. F0, on the other hand, when combined with intensity and/or duration, is a relevant acoustic cue. To add further doubt about what is important for stress marking in the Spanish language, Prieto and Ortega-Llebaria [11] claim that syllable duration, vowel quality and spectral tilt (i.e. intensity in the higher regions of the spectrum) are reliable acoustic correlates of lexical stress, while overall intensity marks accentual differences.

One of the reasons why it is so difficult to determine which acoustic cues are the most robust and play a role in perception is that stressed syllables tend to coincide with pitch accents marking intonational contours in Spanish. A study focused on separating the two contexts revealed that the intonational phrase boundary is a determining factor in the patterning of pitch and duration, but not intensity [12]. Apparently, duration is sensitive to all phonological categories, i.e. accented, unaccented, stressed and unstressed syllables, but pitch changes (f0) are correlated with accent only. The results for intensity were unfortunately not definitive due to interspeaker variation. The latter factor was later found to be a determinant of stress in perception [13, 14], however. According to [13, pp. 19], stressed syllables are possibly perceived as more prominent in Spanish "due to an increase in the intensity levels in the higher

[...] regions of the spectrum", as opposed to overall intensity. Following up on the perception issue, there is ample evidence that Spanish speakers are able to correctly identify stress in minimal pairs, relying either on duration or on intensity in unaccented positions. Spanish speakers may be contrasted with e.g. French speakers who have been deemed 'deaf' to stress contrast [6, 15-17]. Although perceptual recognition of stress was also confirmed by [18], the latter study suggests that in spite of duration and intensity being used as perceptual stress cues, stress pattern neutralisation may be at work in production. The authors of the study demonstrated that durational and intensity cues to stress are produced by speakers and used by listeners above chance level. However, there is a substantial amount of phonetic overlap between stress categories in production, and numerous errors in identification are made by listeners (up to 37%).

To conclude the discussion on stress production and perception, it may be said that Spanish native speakers should be rather apt at identifying stress and detecting stress shift in their own language as they are exposed to lexical contrasts driven by a series of prosodic differences between stressed and unstressed syllables. This assumption was borne out in an EEG study [19] which reported high accuracy rates in response to the shift of stress to a different syllable in an N400 paradigm. The question is, which cues are responsible for the correct perception of stress by speakers of Spanish. As implied by the above discussion, although certain acoustic correlates repeat across studies, their respective roles in perception as opposed to production remain unclear. While it is impossible to study all of them in detail in a single experiment, we decided to take on the task of looking at those two which have been repeatedly associated with lexical stress (intensity) and accent (f0).

2.2. Neural correlates of lexical stress

The problem with determining which acoustic cues are the most salient in perceptual terms goes beyond Spanish. As noted by [20] while f0 is usually considered the most important correlate, studies conducted so far differ in their conclusions concerning the role of intensity, especially spectral tilt [21-23]. Listeners' sensitivity to stress contrasts, including the processing of their acoustic correlates, e.g. f0 or intensity, has been studied quite successfully to date using the mismatch negativity (MMN) component in event-related potential (ERP) research. More specifically, it has been shown that MMN is not only an automatic response of the brain to purely acoustic changes in auditory input but also to the long-term representation of stress and other linguistic units [7, 24, 25]. Although no study was conducted to date that would specifically address these questions in Spanish, studies on other languages provide some insight on the neural correlates of stress perception. For instance, Zora and colleagues [20] investigated the link between f0 and intensity in stress perception with lexical access in English. They used a multi-deviant oddball paradigm in which changes in stress were made by manipulating f0, intensity or both parameters at the same time. The results show that both intensity and f0 elicit an MMN, although the response to intensity is less robust. In addition, it was shown that the scalp distribution of the component was broader in language-related processing compared to pseudowords, with centroparietal and not only frontal regions involved. Another study relevant for the present experiment explores the processing of spectral emphasis, duration and f0 in stress perception in Turkish [26]. Spectral tilt was used instead of overall intensity to better reflect the role of loudness in stress perception as, according to the literature, it is a more reliable stress correlate both in production and perception [23, 27-28]. The amplitude of the MMN evoked by changes in these features was to reflect their relative salience. As in Spanish, in Turkish, the role of f0 vis à vis intensity and duration in stress perception and lexical access is also unclear as previous studies produced divergent results. It was also suggested that only f0 can be treated as a reliable perceptual correlate. The results of the study showed an MMN response that was maximal at the frontal and central scalp locations around 230 ms after the onset of the stress change [26]. F0 was found to be the most prominent feature, while spectral tilt and duration are less robustly responded to. The authors suggest that f0 must be lexically specified in Turkish.

3. Method

3.1. Participants

16 native speakers of Spanish (7 males) aged 19-35 participated in the study after giving their informed consent. All of them were right-handed and received remuneration for participation.

3.2. Stimuli

One target word was selected to serve as a standard stimulus in a passive oddball paradigm. The existing word *fábula* 'fable' with antepenult stress was manipulated in Praat [29] and in Adobe Audition CS6 to create three deviants. The first deviant involved a lowering of the f0 in order to induce final stress without changing the acoustic properties of the originally stressed syllable. The second deviant involved a change in the spectral balance. To induce final stress, the intensity in the spectrum above the f0 was increased by 5 dB using the fast Fourier transform filter in Adobe Audition (the technique was based on [26]). The third deviant was created by changing the vowel [a] of the first (stressed) syllable to [u] by cutting out and splicing the (prosodically adjusted) second vowel. This was a control condition aimed at comparing the MMN elicited by the stress cue deviants to the one elicited by other acoustic changes. As a result, we had three deviants, two of which involved a change in stress. The manipulations made ensured that no reaction to the lack of initial stress is evoked since only the last syllable was manipulated in the *fabulá* deviants. Also, the word *fábula* was selected specifically to ensure that the stress shift is not associated with any other (melodic) change, such as vowel quality changes. To avoid boundary and other intonational effects, the original word was cut out from a longer sentence in which it was not final. The stimuli were presented in one block, with the probability of the standard *fábula* amounting to 74.5% (N=870) and the probability of each deviant (f0, spectral tilt, vowel) equal to 8.5% (N=100 each).

3.3. Procedure

The participants were sitting in front of a computer screen and watching a silent video presenting sand on glass art which lasted around 22 minutes. At the same time they were exposed to auditory stimuli (N=1170) presented with a variable 500-600 ms interstimulus interval via loudspeakers at a comfortable listening level (65dB). The stimuli were pseudo-randomized with the restriction that each deviant must be preceded by at least one standard stimulus. The participants were told not to pay special attention to the audio stimuli. The experiment was run using Presentation software (Neurobehavioural Systems, www.neurobs.com).

3.4. EEG data acquisition

EEG data were recorded using a BioSemi ActiveTwo EEG system (ActiveTwo, BioSemi B.V., Amsterdam, Netherlands), with 128 channels (Ag/AgCl electrodes). Two ocular electrodes were placed below the left and right canthi to record vertical eye movements. The sampling rate was 512 Hz. Impedances were kept equal to or below 20kΩ, and an online band pass filter of 0.1–100 Hz was applied. The data were preprocessed using EEGLAB [30] combined with ERPLAB [31]. The signal was bandpass-filtered (0.1-30 Hz), baseline-corrected and average-referenced. Blinks and saccades were corrected using independent component analysis (ICA) [32]. 1000 ms epochs were extracted with a baseline period of 200 ms based on triggers marking the onset of the syllable of interest (1^{st} syllable for the vowel deviant, and 3^{rd} syllable for the stress shift deviants). Data from all participants were viable for analysis, with an average of 7% of epochs rejected due to artifacts.

3.5. Statistical analysis

Three regions of interest (ROIs) were taken into account: frontal, central and parietal. Data from 24 electrodes per region were extracted via the ERPLAB measurement tool and then pooled. The measurement window was determined by visual inspection of grand average ERP waveforms. Amplitudes were computed as mean voltage over the selected time window. Statistical analyses were conducted in R [33] using the *lme4* package [34] Linear mixed effects models were built with mean amplitude as a dependent variable and condition (f0, spectral tilt, vowel), stimulus type (standard, deviant) and ROI (frontal, central, parietal) as fixed factors. Full random structure could not be used in most cases due to model singularity and problems with convergence. In models built for the complete electrode pool, a random slope of ROI was used together with the random intercept of subject. In models excluding parietal electrodes random slopes of ROI and stimulus type were used. The results are presented in the next section based on the summary table for the models, as well as *F*-testing of fixed effects, obtaining denominator degrees of freedom via Satterthwaite approximation (*anova()*). Simple effects were obtained using marginal means from the model (*emmeans* package).

4. Results

The grand averages show an MMN response in the 150-350 ms window after change onset (see Fig. 1). The MMN component is distributed fronto-centrally, as expected, with slight differences in latency depending on the condition (Fig. 2). It is earliest and longest in the vowel condition (100-330 ms), a bit later for f0 (150-300 ms) and the most delayed in spectral tilt (200-350 ms). Additionally, there seemed to be a positive component of the P3 type following the MMN in the vowel condition at the 330-430 ms time range. Activity was also detected in the parietal region but with a reverse polarity (Fig. 1). Separate models were first run for each of the tested acoustic correlates, followed by a joint analysis of the difference waves.

In the model for f0 changes, there was a significant effect of stimulus type, with the deviant stimuli having significantly lower amplitude compared to the standards (t=-15.875, p<0.001), and a significant effect of ROI, with parietal but not frontal electrodes showing a significantly higher amplitude compared to central electrodes (t=9.813, p<0.001). There was also a significant interaction between stimulus type and ROI, driven by the different polarity over the parietal regions of the scalp (t=26.07, p<0.001).

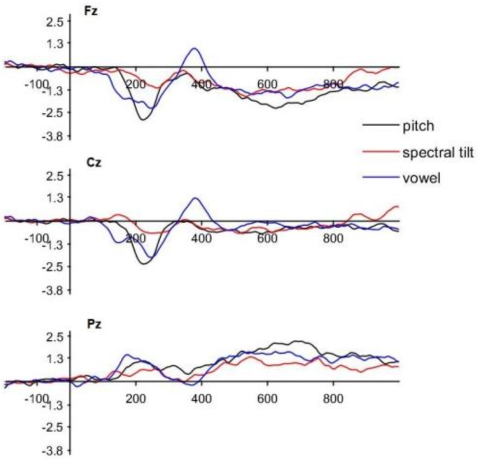

Figure 1: *MMN effect in the case of pitch (f0), spectral tilt and vowel changes. Positive values up.*

In the model constructed for the spectral tilt manipulation, there was a significant effect of stimulus type, with the deviant stimuli having significantly lower amplitude compared to the standards (t=-6.45, p<0.001), and a significant effect of ROI, with parietal but not frontal electrodes showing a significantly higher amplitude compared to central electrodes (t=11.5, p<0.001). There was also a significant interaction between stimulus type and ROI, driven by the different polarity over the parietal regions of the scalp (t=12.12, p<0.001).

The model estimating the effects of stimulus type and ROI in the vowel change condition showed a significant effect of stimulus type, with the deviant stimuli having a significantly lower amplitude compared to the standards (t=-19.21, p<0.001; overall effect as per the analysis of variance: F=80.59, p<0.001). The effect of ROI did not reach significance (F=3.62, p=0.052) but there was a significant interaction between the two factors (F=625.61, p<0.001), with deviants in both frontal and parietal electrodes differing from standards over the central electrode pool. Additionally, we tested the positive deflection of the P3 type following the MMN in the vowel condition. The effect was significant for stimulus type, with deviants showing higher mean amplitude than standards (F=7.9, p=0.005). ROI did not reach significance (F=2.08, p=0.15). There was also a significant interaction between the two, which showed the effect to be centroparietal, with frontal regions unaffected, i.e. no significant change in mean voltage from standard to deviant (F=53.31, p<0.001).

In all the above models testing the MMN effect, there was no significant difference between the central and frontal electrodes and no interaction between stimulus type and ROI once the parietal electrodes were excluded, which suggests the same level of activation over the frontocentral part of the scalp. The last model run on the data included all three conditions (f0, spectral tilt and vowel change) and all the three ROIs, with the mean amplitude calculated over the difference waves (i.e. the deviant-standard subtraction curves, as per [35, 36] and others), in order to provide a comparative analysis. The results show a significant effect of condition (F=28.85, p<0.001) and ROI (F=71.13, p<0.001), as well as a significant interaction (F=119.62, p<0.001). Simple effects of condition show that there is a significant difference between the results for f0 compared to spectral tilt (t=-6.8, p<0.001) and between the results for spectral tilt and vowel change (t=6.33, p<0.001) but no significant difference between the difference waves for vowel change and f0 (t=-0.47, p=0.89). Thus, the MMN found for spectral tilt is significantly smaller than the ones found for f0 and vowel changes. The marginal means calculated for ROI show significant differences in the MMN effect between both central (t=-9.8, p<0.001) and frontal sites (t= -6.7, p<0.001) compared to parietal sites but no significant difference between the central and frontal sites themselves (t=-0.87, p=0.67). The interaction term is best appreciated on the effects plot (Fig. 3).

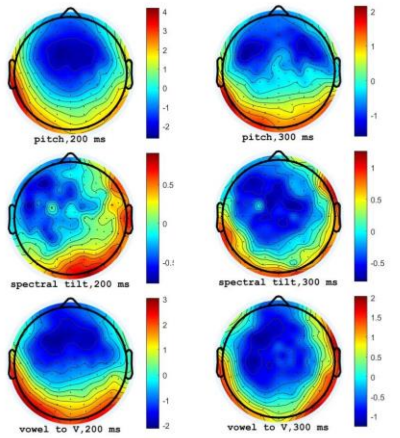

Figure 2: *Scalp topographies showing the MMN at 200 ms and 300 ms after change onset.*

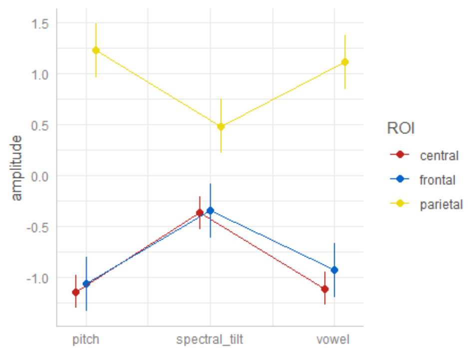

Figure 3: *Predicted values of amplitude per condition and ROI showing the interaction between the two.*

5. Discussion

The results of the study confirmed that Spanish listeners are sensitive to changes in two major stress correlates: f0 and spectral tilt. They also react to a change in the quality of the stressed vowel. There are three differences between the tested conditions. First, spectral tilt appears to be less robust than f0 while the latter patterns with vowel changes in terms of activation strength. Second, there is a slight latency shift in the MMN depending on the condition. Maximum activation was observed at around 200 ms from change onset in the case of f0 and vowel changes, and at 300 ms in the case of spectral tilt (Fig. 2). This is perhaps due to the fact that spectral tilt consists in the slope of intensity between the lower and the higher regions of the spectrum within the syllable. However, changes in f0 in languages such as Spanish (and as manipulated in the present study) are also distributed in time but are perhaps more salient at the outset. Finally, vowel changes but not f0 and spectral tilt-based stress shift induce a P3 component. This might be interpreted as P3b, which is usually maximal on parietal electrodes and has to do with context updating operations and memory storage [37]. Since a change in the vowel is more likely to cause a change in the lexical item and hence the meaning of the word, the memory trace of the standard stimulus is more likely to be updated quickly after a change in the sensory input. At the same time, it should be noted that stress shift was created on the last syllable of the word, when full recognition must have taken place and the meaning was decoded, while changing the vowel of the first syllable triggers item search in the mental lexicon. It should also be mentioned that in previous studies on stress correlates, f0 (but not intensity) changes in words as opposed to pseudowords induced an early P2 component [20, 26]. Although we used an existing word in the present experiment, no P2 was registered prior to the MMN wave. This may be due to the design of the study. In our case, we shifted the perceived stress by inducing it in a different syllable instead of destressing the originally stressed syllable as in e.g. [26]. Finally, our results resonate very well with those of previous studies in which both intensity (overall or spectral tilt) and f0 resulted in MMN, with f0 inducing a much stronger activation. This was interpreted as a strong indicator of the role played by f0 in lexical access and perhaps even its lexical storage [20, 26]. Additionally, it has been argued that different activations caused by f0 and intensity manipulations suggest that acoustic correlates of stress are processed separately, as individual features [26]. Moreover, it must be noted that the robust role played by f0 in stress perception has been reported for many languages of different language families, such as Swedish, Japanese, English, Turkish and Spanish, among others [12, 20, 21, 26, 37-39]. Thus, given the convergent results concerning the importance of this cue, it might be suggested that the prominence of this feature is cross-linguistic in nature and might be a generally lexicalized property of language across speakers from different backgrounds around the world. Research testing this hypothesis is needed to disentangle exposure-based acoustic sensitivity of the listener’s ear from a general phonological property of language acquisition and use.

One limitation of this study is that another important stress cue, i.e. duration, was not tested. We can therefore only speculate about its role in stress perception in Spanish. More importantly though, the study involved the use of context-free language, which is usually the case in EEG research, especially such designs as the oddball paradigm. Future research should focus on more naturalistic speech that consists in phrases or sentences that convey meaning and induce linguistic processing as in everyday communication. Hopefully, a crosslinguistic study of this type will help us decide how much of the linguistic reality is part of a universal speech processing system in our brain and how much belongs to the specificity of a given language.

6. Conclusions

In this study, we have shown that both f0 and spectral tilt are reliable stress perception correlates in Spanish. Additionally, f0 evokes a stronger brain activation, which coincides with MMN studies on other languages, suggesting a cross-linguistic listener sensitivity to this cue.

7. References

[1] A. Morales-Front, "El acento," in R. Nuñez-Cedeño, S. Colina, and T. Bradley, *Fonología generativa contemporánea de la Lengua española*, Washington, DC: Georgetown University Press, 2014, 2nd ed., pp. 235–265.

[2] A. Quilis, *Fonética acústica de la lengua española*, Madrid: Gredos, 1981.

[3] J. Alcina and J.M. Blecua, *Gramática española*, Barcelona: Editorial Ariel, 1975.

[4] N. Sebastián-Gallés, M.A. Martí, M. Carreiras, and F. Cuetos, *LEXESP: Léxico informatizado del español (online database)*, Barcelona: Ediciones Universitat de Barcelona, 2000.

[5] V. Martínez-Paricio and F. Torres-Tamarit, "Trisyllabic hypocoristics in Spanish and layered feet," Natural Language and Linguistic Theory, vol. 37, no. 2, pp. 659–691, May 2019, doi: 10.1007/s11049-018-9413-4.

[6] E. Dupoux, S. Peperkamp, and N. Sebastián-Gallés, "A robust method to study stress 'deafness,'" Journal of the Acoustical Society of America, vol. 110, no. 3, pp. 1606–1618, Aug. 2001, doi: 10.1121/1.1380437.

[7] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, "The mismatch negativity (MMN) in basic research of central auditory processing: A review," Clinical Neurophysiology, vol. 118, no. 12, pp. 2544–2590, Dec. 2007, doi: 10.1016/j.clinph.2007.04.026.

[8] H. Contreras, "Tiene el español un acento de intensidad," Boletín De Filología, no. 16, pp. 237–239, Jan. 1964, [Online]. Available: https://boletinfilologia.uchile.cl/index.php/BDF/article/download/49445/51906.

[9] T. Navarro Tomás, "La medida de la intensidad", Boletín del Instituto de Filología de la Universidad de Chile, vol. 16, pp 231–235, 1964.

[10] J. Llisterri, M. Machuca, C. de la Mota, M. Riera, and A. Ríos, "The perception of lexical stress in Spanish," in *Proceedings of ICPhS*, Barcelona, 2003.

[11] P. Prieto and M. Ortega-Llebaria, "Stress and accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance," in *Proceedings of Speech Prosody*, 2006, pp. 337–340.

[12] M. Ortega-Llebaria, "Phonetic Cues to Stress and Accent in Spanish," in M. Diaz-Campos, *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, Somerville, MA: Cascadilla, 2006, pp. 104–118.

[13] M. Ortega-Llebaria, P. Prieto, and M.M. Vanrell, "Perceptual evidence for direct acoustic correlates of stress in Spanish," in J. Trouvain and J. Barry, *Proceedings of the 16th international congress of phonetic sciences*, Saarbrücken, Germany, 6–10 August 2007, pp. 1121–1124.

[14] M. Ortega-Llebaria and P. Prieto, "The perception of stress in Castilian Spanish. The effects of sentence intonation and vowel type," in M. Vigário, S. Frota, and J. Freitas, *Interactions in phonetics and phonology*, 2007, Amsterdam/Philadelphia: John Benjamins, pp. 35–69.

[15] E. Dupoux, C. Pallier, N. Sebastián-Gallés, and J. Mehler, "A distressing 'deafness' in French?," Journal of Memory and Language, vol. 36, pp. 406–421, 1997.

[16] S. Peperkamp and E. Dupoux, "A typological study of stress "deafness"," in C. Gussenhoven and N. Warner, *Laboratory phonology*, 2002, Berlin: de Gruyter, pp. 203–240.

[17] S. Schwab and V. Dellwo, "Intonation and talker variability in the discrimination of Spanish lexical stress contrasts by Spanish, German and French listeners," Journal of the Acoustical Society of America, vol. 142, no. 4, pp. 2419–2429, Oct. 2017, doi: 10.1121/1.5008849.

[18] F. Torreira, M. Simonet, and J.I. Hualde, "Quasi-neutralization of stress contrasts in Spanish," in N. Campbell, D. Gibbon, and D. Hirst, *Proceedings of speech prosody 2014*, pp. 197–201.

[19] K. Broś, M. Meyer, M. Kliesch, and V. Dellwo, "Word stress processing integrates phonological abstraction with lexical access – An ERP study," Journal of Neurolinguistics, vol. 57, p. 100959, Feb. 2021, doi: 10.1016/j.jneuroling.2020.100959.

[20] H. Zora, I.-C. Schwarz, and M. Heldner, "Neural correlates of lexical stress," Neuroreport, vol. 26, no. 13, pp. 791–796, Sep. 2015, doi: 10.1097/wnr.0000000000000426.

[21] D. B. Fry, "Experiments in the Perception of Stress," Language and Speech, vol. 1, no. 2, pp. 126–152, Apr. 1958, doi: 10.1177/002383095800100207.

[22] P. Lieberman, "Some Acoustic Correlates of Word Stress in American English," Journal of the Acoustical Society of America, vol. 32, no. 4, pp. 451–454, Jun. 1959, doi: 10.1121/1.1908095.

[23] A. M. C. Sluijter and V. J. Van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," Journal of the Acoustical Society of America, vol. 100, no. 4, pp. 2471–2485, Oct. 1996, doi: 10.1121/1.417955.

[24] G. Dehaene-Lambertz, "Electrophysiological correlates of categorical phoneme perception in adults," Neuroreport, vol. 8, no. 4, pp. 919–924, Mar. 1997, doi: 10.1097/00001756-199703030-00021.

[25] Y. Shtyrov and F. Pulvermüller, "Neurophysiological evidence of memory traces for words in the human brain," Neuroreport, vol. 13, no. 4, pp. 521–525, Mar. 2002, doi: 10.1097/00001756-200203250-00033.

[26] H. Zora, M. Heldner, and I.-C. Schwarz, "Perceptual Correlates of Turkish Word Stress and Their Contribution to Automatic Lexical Access: Evidence from Early ERP Components," Frontiers in Neuroscience, Jan. 2016, doi: 10.3389/fnins.2016.00007.

[27] A. M. C. Sluijter, V. J. Van Heuven, and J. J. A. Pacilly, "Spectral balance as a cue in the perception of linguistic stress," Journal of the Acoustical Society of America, vol. 101, no. 1, pp. 503–513, Jan. 1997, doi: 10.1121/1.417994.

[28] M. Heldner, "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," Journal of Phonetics, vol. 31, no. 1, pp. 39–62, Jan. 2003, doi: 10.1016/s0095-4470(02)00071-2.

[29] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*, 2005 (Version 4.3.01).

[30] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21, Mar. 2004, doi: 10.1016/j.jneumeth.2003.10.009.

[31] J. Lopez-Calderon and S. J. Luck, "ERPLAB: an open-source toolbox for the analysis of event-related potentials," Frontiers in Human Neuroscience, vol. 8, Apr. 2014, doi: 10.3389/fnhum.2014.00213.

[32] T.-P. Jung et al., "Removing electroencephalographic artifacts by blind source separation," Psychophysiology, vol. 37, no. 2, pp. 163–178, Mar. 2000, doi: 10.1111/1469-8986.3720163.

[33] R Development Core Team. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2008, ISBN 3-900051-07-0. URL http://www.R-project.org.

[34] D. M. Bates, R. Kliegl, S. Vasishth, and R. H. Baayen, "Parsimonious Mixed Models," *arXiv (Cornell University)*, Jun. 2015, [Online]. Available: https://arxiv.org/pdf/1506.04967.pdf

[35] R. Näätänen and I. Winkler, "The concept of auditory stimulus representation in cognitive neuroscience.," Psychological Bulletin, vol. 125, no. 6, pp. 826–859, Nov. 1999, doi: 10.1037/0033-2909.125.6.826.

[36] S. J. Luck and E. S. Kappenman, "ERP components: the ups and downs of brainwave recordings," in S. J. Luck and E. S. Kappenman, *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press, 2013, pp. 3–30.

[37] J. Polich, "Updating P300: An integrative theory of P3a and P3b," Clinical Neurophysiology, vol. 118, no. 10, pp. 2128–2148, Oct. 2007, doi: 10.1016/j.clinph.2007.04.019.

[38] R. J. Zatorre, A. C. Evans, and E. Meyer, "Neural mechanisms underlying melodic perception and memory for pitch," The Journal of Neuroscience, vol. 14, no. 4, pp. 1908–1919, Apr. 1994, doi: 10.1523/jneurosci.14-04-01908.1994.

[39] A. Koso and H. Hagiwara, "Event-related potential evidence of processing lexical pitch-accent in auditory Japanese sentences," Neuroreport, Sep. 2009, doi: 10.1097/wnr.0b013e32833017af.