DO YOU *SEE* WHAT I MEAN?
CUE CLASHING IN L2 SPANISH LEXICAL STRESS PERCEPTION

Sebastian Leal-Arenas, University of Pittsburgh
Amanda Huensch, University of Pittsburgh

The present study explored the effects of conflicting aural and visual cues on the perception of lexical stress in minimal pairs of conjugated Spanish verbs (e.g., *canto* 'I sing' vs. *cantó* 's/he sang'). English-speaking adults (n = 60) enrolled in Spanish language courses participated in the perception task. The visual cue, i.e., eyebrow raising, was depicted via a *Memoji*, with constant mouth movement. The findings indicated that when eyebrow raising coincided with the stressed syllable, participants were more accurate and faster in their perception of lexical stress. The implications of these results are explored, specifically as they relate to the potential of gestural cues to enhance perception processing. Pedagogical implications are discussed in relation to speech perception training and the use of technology to create pronunciation materials.

INTRODUCTION

Effective communication often relies on the integration of linguistic and visual elements. The combination of these factors can facilitate the conveyance of an intended meaning, or, conversely, detract from it. For instance, saying 'I am angry' with a furious facial expression or typing 'I am sad' to a friend and concluding the message with a sad *emoji* can enhance the intended message. In contrast, stating 'I am angry' with a grin or typing 'I AM SAD' in capital letters could obfuscate the intended meaning. Harnessing these cues in language learning materials is not common in the second language (L2) classroom, but given prior evidence of the role visual information plays in pronunciation perception (e.g., Bosker et al., 2020; Jaekl et al., 2015), doing so has multiple potential pedagogical advantages. In this context, our study explores the impact of combining aural and visual cues on Spanish lexical stress perception using a *Memoji* to depict face movements.

Speech Perception and Animated Avatars

Oral language production can be visually connected with the movements of the lips, jaw, and tongue. This means that visual cues related to these articulators make it easier for people engaged in a conversation to detect, among other things, emphasized syllables. For instance, a more open mouth and more pronounced lip movements facilitate the perception of stressed syllables (see e.g., Beckman & Edwards, 1994; Cho, 2005, 2006; Dohen & Lœvenbruck, 2005; Erickson, 2002; Scarborough et al., 2009). Prosodic information, particularly in terms of emphasis, has also been associated with head movement (Hadar et al., 1983) and eyebrow raising (Cavé et al., 1996; Gast, 2023). The latter has been found to signal emphasis alongside the aural stimulus, and to be

Leal-Arenas, S & Huensch, A. (2024). Do you *See* What I Mean? Cue Clashing in L2 Spanish Lexical Stress Perception. In D. J. Olson, J. L. Sturm, O. Dmitrieva, & J. M. Levis (eds). *Proceedings of 14th Pronunciation in Second Language Learning and Teaching Conference*, (pp. 1-11). Purdue University, September 2023. https://doi.org/10.31274/psllt.17053

a more consistent visual cue than head movement for non-native speakers (Granström et al., 1999).

Research on the effects of facial movements on speech perception has been conducted using animated avatars (e.g., House et al., 2001). These talking heads facilitate the isolation of distinct facial motions, permitting their synchronization with specific acoustic signals. Visual cues can be manipulated to be in alignment with aural cues or to conflict with those cues. Krahmer et al. (2002) investigated the relative impact of eyebrow movement and pitch accent on the perception of stress by presenting six different alignment combinations of visual and aural stimuli in the two-word Dutch phrase *blauw vierkant* 'blue square.' In four of the conditions, they crossed pitch accents with eyebrow movement. In two of the conditions, both words carried pitch accents with eyebrow raising either on the first or the second word. They found that pitch perception can be influenced by both auditory and visual stimuli, but eyebrow movement had a comparatively weaker impact than auditory cues did. Nevertheless, other studies have shown that in situations where auditory cues fail to provide conclusive information, visual cues tend to play a more significant role in accurate perception. Prieto et al. (2015) presented conflicting audiovisual information in the perception of focus statements alongside reaction times. Their results showed that participants were more accurate and faster in their perception of focus when both stimuli were combined in the same word. However, they noted that facial movements were more influential than aural information in stress perception when facial gestures were more pronounced. Likewise, the effects of prosody were greater when gestures were subtler. While prior research has focused on stress perception among speakers in their first language (L1), little is known about the interaction of visual and aural cues in L2 lexical stress perception. The current study integrates eyebrow raising as a visual indicator of emphasis, alongside the use of auditory stimuli to investigate their effects on L2 Spanish lexical stress perception in matched and mismatched conditions.

Spanish Lexical Stress and the English L2 Learner

Stress in Spanish, not always orthographically associated with an acute accent mark on the most prominent vowel, tends to predominantly fall on the penultimate syllable of words ending in a vowel and on the final syllable of consonant-ending words (Navarro-Tomás, 1957). However, stress placement for verb conjugations in the first- and third-person singular forms of the preterit is morphologically marked by an irregular stress pattern where the final vowel is accented. Various minimal pairs across specific grammatical persons and tenses stem from this phenomenon. For instance, the present, first-person singular conjugation of *cantar* 'to sing' is *<u>can</u>to* /ˈkan.to/ 'I sing', while the preterit, third-person singular conjugation of the same verb is *cant<u>ó</u>* /kan.ˈto/ 's/he sang'.
It is perceptually challenging for English speakers learning Spanish to differentiate words that differ only in lexical stress (Kim, 2020; Ortega-Llebaria et al., 2013; Ortín & Simonet, 2022; Romanelli & Menegotto, 2015; Saalfeld, 2012), because vowel quality, an important cue in English, is comparable in stressed and unstressed syllables in Spanish (Harris, 1989; Ortega-Llebaria et al., 2007; Ortín & Simonet, 2023). Thus, exploring the effect of cues outside the acoustic domain seems promising to help learners with stress perception.

Research Questions

In speech perception, listeners determine which dimensions are most important in auditory processing. A single dimension is rarely sufficient for accurate acoustic discernment. However, perception is mainly determined by a primary cue, followed by a secondary cue whose role is still detectable (Abramson & Lisker, 1985). The experiment addressed the following research questions:

RQ1: To what extent do conflicting eyebrow raising and aural prominence affect accurate lexical stress perception?

RQ2: How does accuracy in perception relate to reaction times?

METHODS

Participants

Data were obtained from 60 Spanish language learners attending a large, public university in the northeast of the United States. Participants were recruited from different Spanish language courses, ranging from first to third-year courses in the sequence of the program. English was their first language (L1), and they were aged 18 to 21 years (M = 19.3, SD = 0.86). There were 42 women, 14 men, and 4 participants who identified as non-binary.

Language Surveys

An adapted version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) and the Language History Questionnaire (LHQ 2.0; Li et al., 2014) were employed. Additionally, as a proficiency measure, participants completed a shortened version of the Cervantes Language Test, which included 20 multiple-choice items.

Linguistic Stimuli

Twenty-four stress minimal pairs, adapted from Kim (2015), differing in verbal inflection were used. They consisted of disyllabic regular *–ar* verbs in the first-person singular present tense conjugation and the third-person singular preterit tense. The verbs were embedded in meaningful sentences with the same structure. Target words were at the end of each sentence and were preceded by four syllables (see Figure 1).

Audio-visual Stimuli

Aural stimuli were technologically generated using *Narakeet*'s (https://www.narakeet.com/) text-to-audio feature, with standard volume (-12 dB), normal speed (4 syllables/second) and .wav format as settings. Visual stimuli were created using Apple's *Memoji*. A *Memoji* is a custom, 3D avatar which can mimic a person's facial expression by using a device's camera. Mouth

movements were maintained constant to not interfere with the isolation of the effects of eyebrow raising in the perception of prominence. The stimuli were merged using Adobe Premier Pro. Eyebrow raising was present in two different conditions for each sentence. The first sentence had eyebrow raising aligned with the stressed syllable. A second sentence presented eyebrow raising in unstressed syllable position. For example, in the sentence *por el jardín fumo* 'through the garden, I smoke,' the first audio-visual stimulus had eyebrow raising on *fu-*, whereas the second stimulus of the same sentence had it on *–mo*. Additionally, eyebrow movement was matched to the beginning of the syllable in which it was present. Figure 1 shows the *Memoji* and a sample item. Materials are available via the Open Science Framework (https://osf.io/n9u4w/).

- Por el jardín fumo.
- Por el jardín fumó.

Figure 1. Stimuli example

Procedure

Participants were sent a link to the experiment, which was conducted online on *Qualtrics*. The landing page provided a general overview of the study and procedures. Participants first completed the Linguistic Background Survey, followed by the Language Proficiency Test and the perception experiments. Participants watched and listened to an animated avatar uttering 24 sentences and were asked to select the utterances they heard. They were only provided with two options: one sentence in the present and one in the preterit. Reaction times were recorded using a JavaScript (Dohen & Lœvenbruck, 2009; Prieto et al., 2015). Code and step-by-step instructions can be found on GitHub (https://github.com/SebastianLealArenas/JavaScriptQualtrics). Reaction times were measured as the temporal duration from the onset of the video reproduction until the participant selected a response. Participants were instructed to watch and listen to the video once.

Data Analysis

Participants' answers were labelled as 1 for accurate responses and 0 for inaccurate responses. The answers were then averaged to obtain the Mean Accuracy Score for each variant. Thus, mean accuracy values approaching 1 indicated more accurate responses. Reaction Time (in milliseconds) for accurate answers was also aggregated to calculate the Mean Reaction Time for each factor. Subsequently, the mean difference effect sizes and 95% confidence intervals (CIs) were calculated and interpreted following benchmarks suggested by Plonsky and Oswald (2014).

Finally, two models were constructed. Because accuracy is a categorical variable, the first model was a mixed-effects regression with accuracy as the dependent variable, eyebrow raising and stress as fixed effects, and participant and item as random effects. As reaction time is a continuous variable, the second model was a linear regression with mean reaction time as the dependent variable and the fixed and random effects of previously mentioned regression. Statistical significance was determined by *p*-values below 0.05 alongside CIs that did not cross 0.

RESULTS

This study investigated the extent to which accurate perception is affected by presenting competing gestural (i.e., eyebrow movement) and aural cues and their effect on reaction times in accurate responses. Results (Table 1) showed that matched cues presented higher mean accuracy scores (M = 0.783, SD = 0.412) and faster mean reaction times in accurate answers (M = 2123, SD = 1063) compared to the mean accuracy (M= 0.549, SD = 0.498) and accurate responses' reaction time (M = 3086, SD = 1091) of mismatched cues. The effect size of the difference in mean accuracy in matched and mismatched cues is not meaningful (d = -0.51 [-0.87, -0.15]). However, the effect size of mean reaction times in accurate responses is small but reliable (d = 0.89 [0.51, 1.26]). When considering eyebrow movement and stress placement, the highest mean accuracy was achieved when eyebrow raising and stress were both present in the first syllable (M = 0.825, SD = 0.380). Cohen's d effect sizes and 95% CIs indicated a small but reliable effect (d = -0.83 [-1.19, -0.45]) in accuracy when both variables coincided. The second-highest mean accuracy score corresponded to both eyebrow raising and stress taking place in the second syllable (M = 0.742, SD = 0.438). However, compared to eyebrow raising occurring in the second syllable and stress in the first, the difference is not meaningful (d = 0.22 [-0.14, 0.58]). Based on Cohen's d effect size and 95% CIs, there is a small effect for reaction time in accurate responses when eyebrow raising occurs in the first syllable (d = 0.88 [0.50, 1.25]), with mean reaction times in accurate responses being slower when the second syllable is stressed (M = 3023, SD = 987) in comparison to the first (M = 2150, SD = 1002). A small effect for reaction time in accurate responses is observed when eyebrow raising occurs in the second syllable (d = -0.89 [-1.27, -0.52]), with slower mean reaction times when the first syllable is stressed (M = 3125, SD = 1106) as opposed to the second (M = 2094, SD = 1128).

Table 1

Mean Accuracy and Reaction Time Values for Accurate Responses in Independent Variables

Eyebrow Raising	Stress	Mean Accuracy Score	Effect Size Cohen's *d* [95% CI]	Mean Reaction Time (ms)	Effect Size Cohen's *d* [95% CI]
1st syllable	1st syllable	0.825 (*SD* 0.380)	-0.83 [-1.19, -0.45]	2150 (*SD* 1002)	0.88 [0.50, 1.25]
	2nd syllable	0.458 (*SD* 0.499)		3023 (*SD* 987)	
2nd syllable	1st syllable	0.639 (*SD* 0.481)	0.22 [-0.14, 0.58]	3125 (*SD* 1161)	-0.89 [-1.27, -0.52]
	2nd syllable	0.742 (*SD* 0.438)		2094 (*SD* 1128)	
	Matched cues	0.783 (*SD* 0.412)	-0.51 [-0.87, -0.15]	2123 (*SD* 1063)	0.89 [0.51, 1.26]
	Mismatched cues	0.549 (*SD* 0.498)		3086 (*SD* 1091)	

The mixed-effects regression (Table 2) predicted accurate lexical stress perception. Eyebrow raising did not statistically nor independently predict accurate performance ($\beta = 0.12$, $p = 0.310$). Stress had a significant effect ($\beta = -0.62$, $p < .001$), with answers being less accurate when the second syllable was stressed, based on the negative estimate and z-value. The interaction between eyebrow raising and syllable was significant ($\beta = 2.21$, $p < .001$), indicating that when eyebrow raising and stress placement coincided on the second syllable, performance improved, mitigating the negative effect observed for stress on the second syllable. The linear mixed-model for reaction time in accurate responses (Table 3) did not yield statistical significance for eyebrow raising ($F_{1, 959} = 0.11$, $p = 0.737$) nor stress ($F_{1, 959} = 1.03$, $p = 0.320$). However, the interaction between eyebrow raising and stress was significant ($F_{1, 959} = 127.08$, $p = < .001$), predicting lower reaction times in accurate answers, indicated by the negative estimate, when eyebrow raising and stress were aligned on the second syllable.

Table 2

Mixed-effects Regression for Accurate Responses in the Perception of Lexical Stress with Mis/matched Cues (Reference Level: Inaccurate response)

Variable	Estimate	95% CI	Std. error	*Z*-value	*p*-value
(Intercept)	0.75	[0.63, 0.87]	0.06	12.31	< .001
Eyebrow Raising					
1st syllable	*reference*				
2nd syllable	0.12	[-0.11, 0.36]	0.11	1.01	0.310
***Stress**					
1st syllable	*reference*				
2nd syllable	-0.62	[-0.85, -0.38]	0.11	-5.17	< .001
***Interactions**					
1st ER: 1st stressed syllable	*reference*				
2nd ER: 2nd stressed syllable	2.21	[1.74, 2.68]	0.24	9.21	< .001
Random Effects	Variance	*SD*	*N*		
Participant	0.00	0.09	60		
Item	0.00	0.00	24		

N = 1440; Log. Likelihood: -856.33; AIC = 1724.66, $R^2_{marginal} = 0.10$, $R^2_{conditional} = 0.11$; * = significant in the best model. Shaded factors are significant.

Table 3

Linear Regression of Reaction Times for Accurate Responses in the Perception of Lexical Stress with Mis/matched Cues

Variable	Estimate	95% CI	Std. error	t	p-value
(Intercept)	2664.1	[2484, 2739.1]	65.1	40.08	< .001
Eyebrow Raising					
1st syllable	*reference*				
2nd syllable	28.6	[-136, 193.4]	84.1	0.34	0.737
Stress					
1st syllable	*reference*				
2nd syllable	-85.5	[-250, 79.1]	84	-1.01	0.320
***Interactions**					
1st ER: 1st stressed syllable	*reference*				

2nd ER: 2nd stressed syllable	-1891	[-2220, -1562.7]	167.8	-11.27	< .001
Random Effects	Variance	*SD*	*N*		
Participant	149136	368	60		
Item	15670	125	24		

N = 959; Log. Likelihood: -8000.420; AIC = 16058, $R^2_{marginal}$ = 0.161, $R^2_{conditional}$ = 0.280; * = significant in the best model. Shaded factors are significant.

Regardless of the syllable on which the stress falls, the matching of aural and visual stimuli presented higher accuracy scores and faster reaction times in accurate answers compared to mismatched cues.

DISCUSSION

The present study assessed the relevance of auditory and gestural cues in the perception of lexical stress in contexts where eyebrow movement was presented in prominent and non-prominent syllables. The findings indicated that participants exhibited higher accuracy and quicker responses when presented with matched cues. This result is in line with previous studies with incongruent cues. Prieto et al. (2015) found that participants experienced more uncertainty when presented with inconsistent matches, as shown by a decrease in accuracy and an increase in reaction times. The effects of matched cues on perception in Prieto et al. (2015) were greater in accuracy than in reaction times, whereas in ours, the effect is stronger on reaction times. A potential explanation for this difference could stem from the type of stress studied. Prieto et al. (2015) explored phrasal stress, whereas this study investigated lexical stress. This could indicate that the impact of visual cues on stress perception might vary depending on the stress type. For instance, consider that lexical stress affects the interpretation of individual words and phrasal stress impacts the meaning conveyed in the sentence. Given this, we can hypothesise that visual cues may facilitate faster perception of literal meaning, as shown by a decrease in reaction times in the present study, whereas the accurate perception of implied meaning may be enhanced by visual cues, as shown by an increase in accuracy in Prieto et al. (2015). Exploring this relationship would be a fruitful avenue for future research.

In the mismatched cue condition of the present study, participants were less accurate and slower in lexical stress perception. Terken and Nooteboom (1987) observed that reaction times were slower when given information was accented or when new information was deaccented. The fact that accuracy scores decreased and reaction times increased when conflicting cues were presented in both studies could indicate that participants perceived these cue combinations as contradictory. Traditionally, gestures in oral languages tend to be considered redundant. If gestural cues were redundant or optional, results associated with accuracy and reaction times would not be enhanced by eyebrow raising, as noted in Leal-Arenas and Huensch (under review), nor hindered by mismatched cues, as shown in this study.

Additionally, participants were more accurate when matched cues were present in the first syllable of the target word. This could be due to two main reasons: common lexical stress placement in the target language and learner bias. Lexical stress typically falls on the second-to-last syllable in Spanish (Navarro-Tomás, 1957). As a result, participants may have been more accustomed to this stress pattern. Beginning and intermediate learners are more exposed to the present than the preterit (Kim, 2015; Saalfeld, 2012); thus, suggesting a bias towards the present

tense stress pattern, which is their default. An anonymous reviewer asked about the role the L1 (English) plays in the results. A robust influence of the L1 would predict bias towards the second syllable since in minimal pairs, the stress falls on the last syllable in the verb (e.g., increase / /ɪŋˈkriːs/) and on the first syllable in the noun (e.g., increase /ˈɪŋkriːs/).

A limitation of the study pertains to this specific point. The experiment solely featured verbs, making it challenging to tease apart whether higher mean accuracy and reaction times resulted from the participants' familiarity with the present or due to the prevalent position of lexical stress in Spanish. Future studies could incorporate words belonging to different word classes (*camino* 'a path,' *camino* 'I walk,' *caminó* 's/he walked') to address this limitation. In addition, an audio-only condition could be incorporated to better assess the impact of visual elements. An anonymous reviewer also pointed out that language proficiency may play a role. In fact, the current study did include three different measures of language proficiency (grammar test, years of L2 study, and course). While expert L2 users were slightly more accurate in lexical stress perception than novice L2 users, proficiency variables did not reach statistical significance and were therefore not included in the final models.

The findings provide valuable insight into the multimodal nature of language processing and their implications for the language classroom. Educators should consider incorporating perception-training materials accompanied by visual cues. We used a *Memoji* in the study due to its accessibility and easy-to-use interface in comparison to other alternatives. *Memojis* mimic the facial movements a person makes; hence, the creation of teaching materials is as easy as recording a video with an iPhone's front-facing camera. By using *Memojis* to illustrate the visual aspects of speech, which go beyond lexical stress, instructors could make the learning process more interactive and immersive. Not only would students hear language, but they could also see subtle differences in segmental and suprasegmental production of speech.

ABOUT THE AUTHORS

Sebastian Leal-Arenas is a doctoral student at the University of Pittsburgh. His pedagogical experience and interest in prosody inform his research on L2 speech perception and production. His recent work addresses the importance of perception, the expression of focus and theticity, the use of available technology, and the intricate relationship between AI, L2ers and language teaching.
Contact information:
Department of Linguistics, University of Pittsburgh
4200 Fifth Ave
Pittsburgh, PA 15213
E-mail: sebastianleal@pitt.edu

Dr. Amanda Huensch is Assistant Professor in the Department of Linguistics at the University of Pittsburgh. Her research examines second language speech development in and outside of the classroom. Her most recent work has been published in *Language Learning* and *Studies in Second Language Acquisition.* She is currently Associate Editor for Open Science of *Applied Psycholinguistics*.
Contact information:

Department of Linguistics, University of Pittsburgh
4200 Fifth Ave
Pittsburgh, PA 15213
E-mail: amanda.huensch@pitt.edu

REFERENCES

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). New York: Academic.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological Structure and Phonetic Form*: *Papers in laboratory phonology III* (pp. 7–33). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511659461

Bosker, H. R., Peeters, D., & Holler, J. (2020). How visual cues to speech rate influence speech perception. *Quarterly Journal of Experimental Psychology*, *73*(10), 1523–1536. https://doi.org/10.1177/1747021820914

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996, October). About the relationship between eyebrow movements and F0 variations. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 4, pp. 2175–2178). IEEE. https://doi.org/10.1109/ICSLP.1996.607235

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *The Journal of the Acoustical Society of America, 117*(6), 3867–3878. https://doi.org/10.1121/1.1861893

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory Phonology, 8*, 519–548. https://doi.org/10.1515/9783110197211.3.519

Dohen, M., & Lœvenbruck, H. (2005). Audiovisual production and perception of contrastive focus in French: A multispeaker study. In *Interspeech/Eurospeech 2005* (pp. 2413–2416). ISCA.

Dohen, M., & Lœvenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, *52*(2-3), 177–206. https://doi.org/10.1177/0023830909103166

Erickson, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica, 59*(2-3), 134–149. https://doi.org/10.1159/000066067

Gast, V. (2023). The temporal alignment of speech-accompanying eyebrow movement and voice pitch: A study based on late night show interviews. *Behavioral Sciences*, *13*(1), 52. https://doi.org/10.3390/bs13010052

Granström, B., House, D., & Lundeberg, M. (1999). Prosodic cues in multimodal speech perception. In Proceedings of the *International Congress of Phonetic Sciences (ICPhS99)* (pp. 655–658).

Hadar, U., Steiner, T. J., Grant, E. C., & Clifford Rose, F. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech, 26*(2), 117–129. https://doi.org/10.1177/002383098302600202

Harris, J. (1989). How different is verb stress in Spanish?. *Probus, 1*(3), 241–258. https://doi.org/10.1515/prbs.1989.1.3.241

House, D., Beskow, J., & Granström, B. (2001). Timing and interaction of visual cues for prominence in audiovisual speech perception. In *Seventh European Conference on Speech Communication and Technology (Eurospeech 2001) (pp. 387–390)*. https://doi.org/10.21437/Eurospeech.2001-61

Jaekl, P., Pesquita, A., Alsius, A., Munhall, K., & Soto-Faraco, S. (2015). The contribution of dynamic visual cues to audiovisual speech perception. *Neuropsychologia*, *75*, 402–410. https://doi.org/10.1016/j.neuropsychologia.2015.06.025

Kim, J. Y. (2015). Perception and Production of Spanish Lexical Stress by Spanish Heritage Speakers and English L2 Learners of Spanish. In E. Willis, P. M. Butragueño, & E. Herrera Zendejas (Eds.), Selected proceedings of the *6th Conference of Laboratory Approaches to Romance Phonology (pp. 106–128)*. Somerville, MA: Cascadilla Proceedings Project.

Kim, J. Y. (2020). Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition, 23*, 233–250. https://doi.org/10.1017/S1366728918001220

Krahmer, E., Ruttkay, Z., Swerts, M., & Wesselink, W. (2002). Pitch, eyebrows and the perception of focus. In B. Bel, & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference, Aix-en-Provence, France, April 11-13, 2002* (pp. 443-446). Laboratoire Parole et Language.

Leal-Arenas, S., & Huensch, A. (under review). Using *Memoji*'s facial gestures to enhance L2 word stress perception. In E. Tergujeff & A. Kirkova-Naskova (Eds.), *Achievements in Second Language Pronunciation*. Cambridge University Press.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680. https://doi.org/10.1017/S1366728913000606

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Navarro-Tomás, T. (1957). *Manual de pronunciación española*. Hafner.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics, 41*, 186–197. https://doi.org/10.1016/j.wocn.2013.01.006

Ortega-Llebaria, M., Prieto, P., & Vanrell, M. D. M. (2007). Perceptual evidence for direct acoustic correlates of stress in Spanish. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress on Phonetic Sciences, Saarbrücken* (pp. 1121–1124).

Ortín, R., & Simonet, M. (2022). Phonological processing of stress by native English speakers learning Spanish as a second language. *Studies in Second Language Acquisition, 44*, 460–482. https://doi.org/10.1017/S0272263121000309

Ortín, R., & Simonet, M. (2023). Perceptual sensitivity to stress in native English speakers learning Spanish as a second language. *Laboratory Phonology 14*(1). https://doi.org/10.16995/labphon.7978

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics, 49*, 41–54. https://doi.org/10.1016/j.wocn.2014.10.005

Romanelli, S., & Menegotto, A. C. (2015). English speakers learning Spanish: Perception issues regarding vowels and stress. *Journal of Language Teaching and Research, 6*, 30–42. https://doi.org/10.17507/jltr.0601.04

Saalfeld, A. K. (2012). Teaching L2 Spanish stress. *Foreign Language Annals, 45*, 283–303. https://doi.org/10.1111/j.1944-9720.2012.01191.x

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech, 52*(2-3), 135–175. https://doi.org/10.1177/0023830909103165

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*(3-4), 145–163. https://doi.org/10.1080/01690968708406928