Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Categorical perception and prenuclear pitch peak alignment in Spanish

Germán Zárate-Sández

Department of Spanish, Western Michigan University, USA
german.zarate-sandez@wmich.edu

Abstract

Most dialects of Spanish seem to produce prenuclear pitch peaks displaced to the right of the stressed syllable in neutral declarative utterances. In Autosegmental-Metrical phonology, this delayed peak has usually been described as a L*+H pitch accent. Since evidence for this observation comes almost exclusively from production studies, the purpose of this paper was to investigate how Spanish speakers perceive prenuclear pitch alignment. Perception was tested using an imitation task aimed at capturing categorical effects (or lack thereof) in the perception of intonation. The stimuli consisted of the utterance "La nena lloraba" ["The girl was crying"], where the prenuclear pitch peak in "nena" was displaced 10 times in 25-millisecond increments. Seventeen native speakers of Spanish listened to the 10 resynthesized utterances and were asked to imitate each stimulus while being recorded. Resulting utterances were normalized for speech rate and analyzed acoustically for prenuclear pitch alignment. Data yielded a clear categorical perception effect, but did not necessarily lend support to a pitch accent with a delayed peak. The discussion addresses phonological representations of tonal events and the link between production and perception in prosody.

Index Terms: pitch alignment, speech perception, Spanish, prenuclear pitch accent, imitation task

1. Introduction

Phonetically, Spanish marks a distinction between nuclear and prenuclear pitch accents in neutral declarative utterances: the nuclear peak occurs within the boundaries of the stressed syllable, while for prenuclear pitch accents, the rise usually extends through the stressed syllable and reaches its peak on the posttonic syllable. Figure 1 provides examples of both situations in the sentence *Le dieron el número de vuelo* 'They gave him/her the flight number' [1]. This displacement was noted by scholars such as Navarro-Tomás [2] and investigated empirically in the 1990s [3, 4]. In a study of Peninsular Spanish, for example, [4] found that 79.5% of the peaks in paroxytone words appeared displaced to the right of the stressed syllable when they were not preceded by a pause, that is, in prenuclear position.

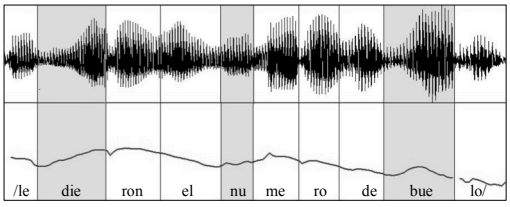

Figure 1: *Rendition of* Le dieron el número de vuelo *'They gave him/her the flight number.' Shaded areas represent stressed syllables*

Although the existence of this prenuclear delayed peak seems uncontroversial across most dialects of Spanish, the phonological representation for this peak in terms of Autosegmental-Metrical (AM) phonology [5, 6, 7] is still debated. At least two different positions exist in this regard [8]: (1) the peak is the trailing H tone of a bitonal L*+H pitch accent, where the starred L tone is associated with the tonic syllable [9, 1], and (2) the peak is the result of a monotonal H* associated with the tonic syllable [10, 11].

Face [9] provided evidence in support of Analysis (1). He tested peak alignment for prenuclear words in narrow focus. His findings supported the analysis whereby there is a non-focal pitch accent L*+H which contrasts with a focal pitch accent L+H*. He also presented evidence that the valley is constantly aligned with the stressed syllable in both cases (more or less at the beginning). The peak, on the other hand, tends to occur on a posttonic syllable in broad focus (with some variation depending on distance with the following stressed syllable) while it is aligned within the syllable boundaries in instances of narrow focus. Put simply, Face's rationale for proposing two different pitch accents went as follows: a clear phonetic distinction triggers a clear change in the meaning of the utterance (narrow vs. broad focus), thus the distinction must also be phonological, which, following AM tenets, is the result of different pitch accents.

Analysis (2), on the other hand, treats all pitch accents in the same way, and uses prosodic factors such as stress-clash avoidance and distance to a boundary to account for any differences in peak alignment. That is, peaks tend to displace to the right if they are relatively free to do so. Both Face [9] and Hualde [8], reported that, for instance, the prenuclear peak will tend to be timed within the syllable in oxytone words since the stressed syllable (and the associated pitch accent) occurs too close to the word boundary. Hualde [8] provided an example of this in the utterance *Le darán el número de vuelo* 'They will give him/her the flight number,' where the peak of *darán* should stay within the word instead of moving to the posttonic syllable.

It is quite surprising that most–if not all–of the evidence for this debate comes exclusively from production studies of Spanish intonation, while perception has been largely ignored. If we assume that perception and production of intonation function around a common system, it stands to reason that the study of perception will add to the picture of Spanish prenuclear alignment. As argued by the IPO (Institute for Perception Research, in Eindhoven) approach [12, 13], the study of listeners' perception and processing of intonation can provide us with information about the system of intonation. For the IPO approach, listeners process and make sense of pitch contours by *imposing* certain preexisting structures. This is usually called top-down processing. The major goal of perception research within this approach is to provide a

 doi: 10.21437/SpeechProsody.2016-136

description of the relevant acoustic properties that actually play a role in perceiving (and processing) f_0 as intonation. That is why this approach has been called a *listener's model* of intonation. With this in mind, the main goal of the present study was to bring perception data into the debate surrounding the phonological representation of prenuclear pitch peaks in Spanish and, in so doing, attempt to obtain a broader description of this particular feature of Spanish intonation.

2. Methodology

2.1. Participants

Seventeen monolingual speakers of Spanish (7 males, 10 females, mean age: 28.41, age range: 20–38) participated in this study. They represented the following dialectal areas: Southern Cone (n = 4), Andean (n = 4), Mexican (n = 3), Central American (n = 3), and Peninsular (n = 3). Within the Southern Cone, speakers of *rioplatense* (or *porteño*) Spanish were not included, since this dialect has been reported to lack the typical prenuclear delayed peak and its nuclear fall also differs substantially from most other varieties [14]. Nine speakers were born and had lived permanently in their respective dialectal areas, while eight had been living in the U.S. for less than three months and reported little to no knowledge or use of English. All participants reported normal hearing and normal or corrected vision.

2.2. Imitation task

2.2.1. Stimuli

The declarative sentence *La nena lloraba* 'The girl was crying' was used as the source utterance to create the stimuli. The following phonotactic and segmental features were taken into consideration, with the goal of maximizing sonority and improving pitch track detection: (1) all consonants are voiced, (2) only mid and low vowels (that is, [e, o, a]) are used (high vowels are known to produce small pitch changes by raising the larynx, and thus can potentially distort the pitch curve [15]), (3) syllables have a simple onset and no coda, (4) content words are paroxytone, which is the unmarked stress pattern in Spanish for words ending in a vowel. From a lexical point of view, the words in the stimulus are considered highly frequent among dialects of Spanish (Corpus del Español, http://www.corpusdelespanol.org/x.asp).

The sentence was recorded by the researcher (native speaker of western Argentinian Spanish), using a high-sensitivity microphone attached to a personal computer, at a sampling rate of 44 kHz. The file was resynthesized using the pitch-synchronous overlap and add (PSOLA) method included in Praat [16], which allows for stylization of the pitch track by reducing it to critical pitch points while keeping the overall shape of the curve.

Manipulation for prenuclear alignment consisted of the following steps (see Figure 2): (1) relevant pitch points were reduced to three: beginning of rise (A1), peak (A2), and end of fall (A3); (2) height (in Hz) of A2 was kept as in source utterance (138 Hz), while the timing (in ms) was displaced to make it equidistant from A1 and A3; (3) height and timing of A1 (118 Hz) were kept as in original, while height of A3 (114 Hz) was raised to match A1; (4) the three relevant points were displaced eight times to the left of the original and once to the right, in 25-ms increments, creating 10 stimuli spanning over 225 ms. Point A1 of the left-most stimulus coincided roughly with the beginning of the sound [a], while the peak (A1) matched the segmental boundary between [n] and [e]. Similar studies on pitch alignment have used diverse durations for the increments in stimuli: 15 ms [17], 20 ms [18], 25 ms [19], and 35 ms [20]. Unfortunately, seldom have authors provided a justification for their decisions. In the current study, since the segmental stretch under examination (see above) spanned over roughly 230 ms, the author considered that 10 stimuli in increments of 25 ms was a manageable number of stimuli and an increment duration that fell within parameters used in previous research.

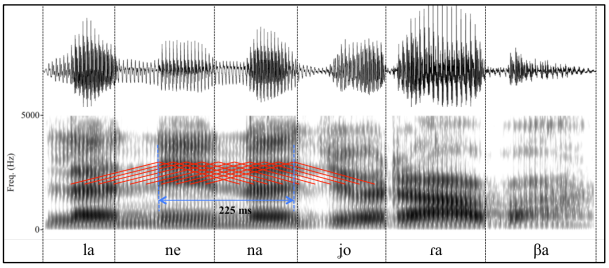

Figure 2: *Stimuli for prenuclear alignment.*

2.2.2. Procedure

Perception was examined via an imitation task. In this task, participants typically listen to one stimulus at a time and then reproduce it out loud. If there is a categorical distinction, participants will not reproduce the entire continuum they hear but, rather, utterances tend to group in a bimodal distribution. The technique has been employed to investigate pitch alignment [18, 19] and pitch height [21, 22]. The underlying assumption is that a purported perceptual system of intonation will restrict various renditions of intonation (the stimuli) to the unit or units that are linguistically meaningful. When hearers are asked to repeat the utterances, production thus reveals what these units—if any—are. If production yields a bimodal distribution [e.g., 18], results have been normally interpreted as indicative of categorical shifts in the perception of intonation. In regards to advantages of the imitation task over other perception tasks, it has been argued that, since subjects have to listen to only one stimulus before they are asked to repeat it, this task does not present the type of challenges for memory that more traditional tasks do (e.g., discrimination or identification, including ABX tasks [23]). Some scholars have actually argued that an imitation task is probably the best way to examine categoriality in intonation [e.g., 15, 24].

Participants were presented with the 10 prenuclear alignment stimuli twice, in two randomized blocks, for a total of 20 stimuli, and were instructed to reproduce the utterance they heard as faithfully as possible. They were asked to focus on the pronunciation of the sentence and encouraged to imitate it within a comfortable pitch range [19, 25, 26]. Participants first listened to a block of five practice utterances, which they could repeat until they felt comfortable with the task. Two of the five practice utterances were drawn from the stimuli, while the others were unrelated declarative utterances, similar in length [19]. For the trial blocks, participants had the possibility of saying the utterance again if they hesitated, paused, or considered that their output was not faithful to the stimulus they had heard. The stimuli were delivered over high-fidelity headphones and presented through E-prime software. Recordings were made in a sound-proof room, with a high-

sensitivity, head-mounted microphone attached to a personal computer and using Audacity software (Audacity Team, 2011, version 1.3.14), at a sampling rate of 32 kHz. Finally, participants completed a debriefing questionnaire where they were asked whether they thought the stimuli were different and, if that was the case, how they believed they differed from one another.

2.3. Data analysis

The quality of the 340 data points (20 utterances x 17 subjects) was both auditorily evaluated and visually inspected using spectrograms and intonational curves produced by Praat. Excessive creaky voice, hesitations, and uncommonly flat global pitch contours led to the exclusion of 54 data points. Twenty of these data points came from one subject who consistently struggled during the task and whose data were thus eliminated altogether (cf. [27], which argues that some participants are simply poor imitators). The other 34 data points excluded were uniformly distributed across stimuli. A trained phonetician, who was unfamiliar with the goals of the study, conducted this data-trimming phase.

Analysis of prenuclear peak alignment in the remaining 286 utterances was carried out by locating f_0 extrema in the segmental stretch comprised between the onset of [e] and the end of [a] in *nena*, since previous research would predict that most if not all peaks, both early- or late-aligned, should fall within these boundaries [1, 4, 9, 11]. The right edge of [a] was usually difficult to locate since the following sound—the initial sound in *lloraba*—was normally realized as the approximant [j]. Therefore, the right boundary for the stretch under analysis was redefined as the middle point of [j] (or [j̟]) and located acoustically at the point of minimum amplitude between the vowels [a] and [o] in [ajo]. These boundaries were manually marked by the same phonetician and annotated using a Praat text grid. Since location of boundaries was considered relatively objective, only one researcher performed this task. A special Praat script automatically extracted the time of f_0 maxima occurrence between these boundaries. In order to control for differences in speaking rate among participants, this value was normalized using the formula in (1) [19, 22].

$$T_N = \frac{t_1 - t_0}{d} \quad (1)$$

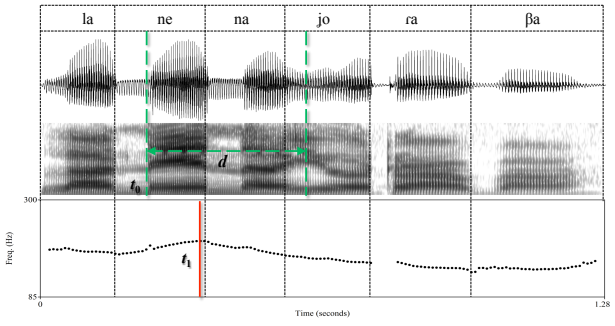

Figure 3: *Example of segmental and prosodic landmarks used in analysis.*

T_N is the normalized time for f_0 maximum, t_1 is the real time for f_0 maximum as extracted by the Praat script, t_0 is the onset of [e] in *nena*, and d is the total duration of the stretch under analysis, that is, from the onset of [e] to the midpoint of [j] (or [j̟]). Since all speakers produced t_1 within the boundaries of d, T_N in all cases consisted of a value in the range between 0 and 1.0. Figure 3 shows an example of analysis for a speaker with a t_1 aligned relatively early.

3. Results

Results for prenuclear alignment location (T_N) for all utterances are presented in Table 1, grouped by stimuli number. Results are also presented graphically in Figure 4. The X axis represents the 10 points in the stimuli, while the Y axis represents the normalized time scores for each group (T_N).

	Stimuli									
T_N	1	2	3	4	5	6	7	8	9	10
n	29	28	28	29	29	29	28	28	29	29
Mean	.323	.291	.333	.310	.788	.771	.759	.793	.811	.772
SD	.097	.088	.080	.089	.068	.082	.070	.074	.082	.076

Table 1: *Results of prenuclear peak alignment (T_N) in utterances produced as a response to imitation task.*

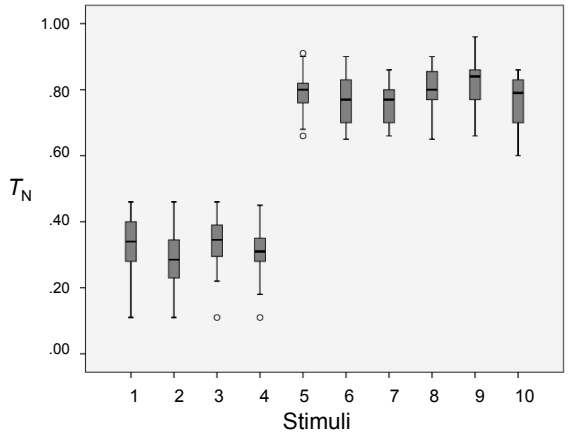

Figure 4: *Results of perception of prenuclear alignment.*

Figure 4 reveals a clear clustering of data points around two sets of stimuli, namely, 1-4 and 5-10, which is in principle consistent with a categorical effect in the perception of the tonal event under investigation. In order to verify the statistical significance of these results, a one-way omnibus ANOVA was run for the entire group, with Stimuli as the independent variable and T_N as the dependent variable. The test yielded significant results ($F(9, 276) = 257.66$, $p < .001$, partial $\eta^2 = .89$, observed power = 1.00). Post-hoc analyses (Tukey) were conducted in order to determine where significant differences actually lay. Results confirmed the patterns observed in the descriptive statistics: stimuli 1-4 are statistically homogeneous and so are stimuli 5-10 ($p > .05$ for all pair-wise comparisons). In turn, both clusters are statistically different from each other ($p < .001$ for all pair-wise comparisons).

Post-hoc results showed a clear categorical shift in perception and that the stimuli were perceived in two distinctive groups. Information from the debriefing questionnaire revealed that, as expected, participants perceived an unmarked, non-emphatic version and an emphatic (narrowly focused) version. They tended to describe the latter as 'said with more emphasis/accent on *nena*'.

4. Discussion and Conclusion

As shown in Figure 5, the shift in the perception of a Spanish declarative utterance as either emphatic (narrow focus) or non-emphatic (broad focus) seems to occur precisely at the end of the prenuclear stressed syllable (between points 4 and 5 of the stimuli used in the imitation task) for native speakers of Spanish. These findings may provide valuable data to the dilemma of how to treat phonologically prenuclear peaks in Spanish, which most production studies have found to appear in a delayed position, usually at the posttonic syllable. This is also represented graphically in Figure 5.

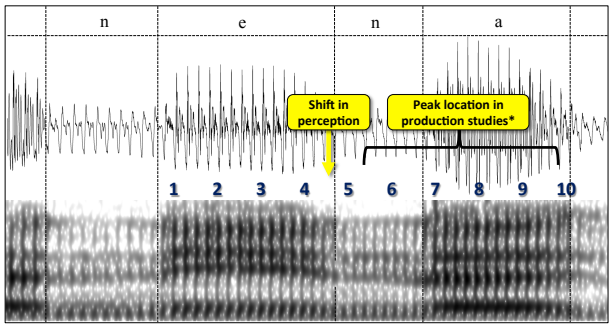

Figure 5: *Shift in perception in relation to segmental material as found in this study (left) and location of prenuclear delayed peaks as reported in previous literature on speech production (right, *). Location of (*) is not meant to be precise but given for comparison purposes only. Numbers indicate points where pitch peak for each stimulus occurred (see description of task).*

As reviewed above, even though this displacement is uncontroversial in most dialects of Spanish, its phonological representation within an AM model is still debated. Interestingly, the perception data in this study showed that shifts between narrow and broad focus occur inside the tonic syllable, just before the right boundary, and certainly quite far from the midpoint or right boundary of the posttonic syllable, where most production studies have located this delayed peak.

Alternative analyses to the L*+H tone have proposed that the prenuclear pitch accent is a monotonal H* [10, 11], a bitonal L+H* [28], or even an associated (L+H)* [8]. In all these proposals, the phonological representation for Spanish prenuclear pitch accents would consist of an H anchored within the stressed syllable, while the observed delay is a phonetic process, governed by factors related to timing, such as tonal crowding, avoidance of stress clash, or distance with upcoming prosodic boundaries.

Another potential shortcoming of the bitonal L*+H pitch accent is the treatment of tonal valleys as phonological (that is, as L tones). The presence of a valley before each peak is quite common in most Spanish utterances and arguably in many other languages. This can be observed, for example, in the melody curves of Figures 1 and 2, where a valley precedes every peak. We are confronted with the usual dilemma of whether to assign a phonological status to a phonetic regularity. In this regard, [1, 9, 29, 30] have suggested that valleys are actually considerably stable and tend to occur at or near the syllable onset. In these analyses, the presence of the L tone as part of the bitonal L*+H pitch accent would be warranted. On the other hand, [11] found that tonal valleys were not stable enough—in their scaling or timing—to be classified as L tones. In the author's analysis [11], valleys between H* would be the result of *dips,* a *sagging* transition between peaks. Pierrehumbert [6] noticed this feature in English and wrestled with the problem of how to treat it phonologically. In [28] the authors accepted that there is a difference in meaning between late and early peak alignment for prenuclear pitch accents, and that this difference translates into broad and narrow focus, respectively. However, they proposed L+H* as the pitch accent for both cases, the only difference being one of association: in narrow focus, the starred H tone has a secondary association with the syllable edge, represented as L+H*]σ, which forces it to remain within the syllable boundaries, while in broad focus the H* tone has a primary association only and does not need to align with a metrical unit.

Studies such as [10, 11, 28] propose H* (with or without a leading L tone) associated with the stressed syllable as the prenuclear pitch accent in Spanish. Any peak delay or displacement would be the result of the implementation at the phonetic level. This analysis simplifies the phonology of pitch accents in neutral declarative sentences in Spanish, as it treats all pitch accents the same. The present study provides strong evidence in favor of this proposal: the location of perceptual shifts between broad and narrow focus occurred within the limits of the stressed syllable. If we are to consider perception as another window into the intonational system of a language, in tandem with production [12, 13], these results can shed new light on processes of pitch alignment in Spanish.

Despite the relatively small scale of this study, restricted to one target utterance, it produced clear patterns of categorical perception. More studies are needed in order to determine if these findings hold in different, though comparable, research designs. The results in this paper also suggest that these future studies should prioritize the simultaneous examination of both perception and production of the tonal event under investigation.

5. Acknowledgements

I would like to thank Alfonso Morales-Front, Cristina Sanz, and Joan Carles Mora for their comments and support during the development of this study. The feedback from two anonymous reviewers also greatly improved the quality of this paper. All remaining errors are mine alone.

6. References

[1] J. M. Sosa, *La Entonación del Español: Su Estructura Fónica, Variabilidad y Dialectología*. Madrid: Cátedra, 1999.

[2] T. Navarro-Tomás, *Manual de Entonación Española*. New York: Hispanic Institute in the United States, 1944.

[3] J. M. Garrido, J. Llisterri, C. de la Mota, and A. Ríos, "Prosodic differences in reading style: Isolated vs. contextualized sentences," in *Eurospeech'93 - 3rd European Conference on Speech Communication and Technology, Berlin, Germany, Proceedings,* 1993, pp. 573–576.

[4] J. Llisterri, R. Marín, C. de la Mota, and A. Ríos, "Factors affecting F0 peak displacement in Spanish," in *Eurospeech'95 - 4th European Conference on Speech Communication and Technology, Madrid, Spain, Proceedings,* 1995, pp. 2061–2064.

[5] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook*, vol. 3, pp. 255–309, 1986.

[6] J. B. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, doctoral dissertation, Massachusetts Institute of Technology, 1980.
[7] J. Pierrehumbert, "Tonal elements and their alignment," in M. Horne (Ed.), *Prosody: Theory and Experiment: Studies Presented to Gösta Bruce,* pp. 11–36, Dordrecht: Kluwer Academic Publishers, 2000.
[8] J. I. Hualde, "Intonation in Spanish and the other Ibero-Romance languages: Overview and status quaestionis," in C. R. Wiltshire and J. Camps, J. (Eds.), *Romance Phonology and Variation: Selected Papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000,* pp. 101–115, Amsterdam: John Benjamins Pub, 2002.
[9] T. L. Face, "Focus and early peak alignment in Spanish intonation," *Probus*, vol. 13, pp. 223–246, 2001.
[10] H. J. Nibert, *Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation,* doctoral dissertation, University of Illinois, Champaign, 2000.
[11] P. Prieto, "The scaling of the L values in Spanish downstepping contours," *Journal of Phonetics*, vol. 26, no. 3, pp. 261–282, 1998.
[12] A. Cohen and J. 't Hart, "On the anatomy of intonation," *Lingua*, vol. 19, pp. 177–192, 1967.
[13] J. 't Hart, R. Collier, and A. Cohen, *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge, UK: Cambridge University Press, 1990.
[14] L. Colantoni and J. Gurlekian, "Convergence and intonation: Historical evidence from Buenos Aires Spanish," *Bilingualism: Language and Cognition*, vol. 7, pp. 107–119, 2004.
[15] C. Gussenhoven, *The Phonology of Tone and Intonation*. New York: Cambridge University Press, 2004.
[16] P. Boersma and D. Weenink, Praat: Doing phonetics by computer (Version 5.1.25) [Computer program], 2010. Retrieved from http://www.praat.org/
[17] M. D'Imperio, B. Gili Fivela, and O. Niebuhr, "Alignment perception of high intonational plateaux in Italian and German," in *International Conference on Speech Prosody, Chicago, USA*, 2010.
[18] J. Pierrehumbert and S. Steele, "Categories of tonal alignment in English," *Phonetica*, vol. 46, pp. 181–196, 1989.
[19] L. Redi, "Categorical effects in production of pitch contours in English," in *Proceedings of the 15th International Congress of the Phonetic Sciences, Barcelona,* 2003, pp. 2921–2924.
[20] M. D'Imperio and D. House, "Perception of questions and statements in Neapolitan Italian," in G. Kokkinakis, N. Fakotakis, and E. Dermatas (Eds.), *Proceedings of Eurospeech, Rhodes, Greece,* 1997.
[21] L. C. Dilley, *The Phonetics and phonology of Tonal Systems*, doctoral dissertation, Massachusetts Institute of Technology, 2005.
[22] L. C. Dilley and M. Brown, "Effects of pitch range variation on f0 extrema in an imitation task," *Journal of Phonetics*, vol. 35, pp. 523–551, 2007.
[23] C. Gussenhoven, "Experimental approaches to establishing discreteness in intonational contrasts," in S. Sudhoff et al. (Eds.), *Methods in empirical prosody research*, pp. 321–334, Berlin: Walter de Gruyter, 2006.
[24] C. Gussenhoven, "Discreteness and gradience in intonational contrasts," *Language & Speech*, vol. 42, no. 2, pp. 283–305, 1999.
[25] M. D'Imperio, R. Cavone, and C. Petrone, "Phonetic and phonological imitation of intonation in two varieties of Italian," *Frontiers in Psychology*, 5, Article 1226, 2014.
[26] J. German, "Dialect adaptation and two dimensions of tune," in *Proceedings of Speech Prosody, Shanghai*, 2012.
[27] L. C. Dilley, "Pitch range variation in English tonal contrasts: Continuous or categorical?," *Phonetica*, vol. 67, pp. 63–81, 2010.
[28] T. L. Face and P. Prieto, "Rising accents in Castilian Spanish: A revision of Sp_ToBI," *Journal of Portuguese Linguistics*, vol. 5, pp. 117–146, 2006.
[29] P. Prieto, J. van Santen, and J. Hirschberg, "Tonal alignment patterns in Spanish," *Journal of Phonetics,* vol. 23, no. 4, pp. 429–451, 1995.
[30] M. Simonet, "A contrastive study of Catalan and Spanish declarative intonation: Focus on Majorcan Spanish," *Probus*, vol. 22, pp. 117–148, 2010.