Journal of Phonetics 102 (2024) 101294

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Dynamic multi-cue weighting in the perception of Spanish intonation: Differences between tonal and non-tonal language listeners

Peizhu Shang [a,*], Paolo Roseano [b,c], Wendy Elvira-García [d]

[a] Beijing Institute of Technology, 5, South Street, Haidian District, 100081 Beijing, China
[b] Universidad de Educación a Distancia, C. de Bravo Murillo 38, 28015 Madrid, Spain
[c] University of South Africa, Preller Street, Muckleneuk Ridge, 0002 Pretoria, South Africa
[d] Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain

ARTICLE INFO

Article history:
Received 3 January 2022
Received in revised form 15 December 2023
Accepted 29 December 2023

Keywords:
Intonation perception
Multiple cue weighting
Tonal and non-tonal languages
Perceptual trade-off
Stress pattern

ABSTRACT

This study investigates cue-weighting differences in intonation perception between tonal and non-tonal languages, specifically focusing on how native Spanish listeners and Mandarin learners of Spanish identify intonation categories using changes in multiple acoustic dimensions. Employing a relatively continuous response scale, we analyzed listener performance in two perceptual tests, in which stimuli were generated by manipulating the suprasegmental cues in sentence-final positions. The results of data analyses indicate that while f0 and duration cues are significant for intonation categorization in both Spanish and Mandarin listeners, intensity appears to be a redundant cue that exerts limited effect only on native Spanish listeners. Contrary to the general claim of a tonal language benefit in pitch perception, our two language groups showed similar sensitivities to f0 linear transitions perceived as sentence intonation. Moreover, Spanish natives used higher f0 contours for question recognition compared to Mandarin learners and relied more heavily on secondary cues in their auditory judgments. The study also demonstrates that perceptual weighting varies across acoustic conditions and stress patterns, suggesting that the dynamic mapping between acoustics and intonation is shaped by language background as well as specific acoustic and word-level suprasegmental contexts.

1. Introduction

Speech categories typically are defined along multiple acoustic dimensions, with continuous values along each dimension serving as cues for identifying abstract linguistic representations (Lisker, 1986; Holt & Lotto, 2006). Despite multiple cues available in listeners' estimation of category identity, their perceptual weighting is not equivalent (Holt et al., 2018). For example, fundamental frequency (f0) often emerges as a primary cue for recognizing prosodic contrasts across various languages, while duration and intensity are relegated to a secondary status (Fry, 1955; Ma et al., 2008; Peng et al., 2012). However, the extent to which duration and intensity cues play a "secondary" role in intonation recognition, and how they are integrated into the overall perceptual process remain inadequately understood. The highly redundant nature of speech signals (Carter, 2011) further amplifies this complexity and raises an important question: Do listeners truly rely on secondary cues for accurate perception when primary cues, such as f0, which are highly correlated with intonation, are already present in speech? Addressing this question is crucial to understanding how listeners prioritize and integrate multiple cues in intonation perception.

Perceptual weights are a function of the long-term statistics of linguistic input (Holt et al., 2018). Hence, listeners from different language backgrounds tend to develop distinct cue-weighting patterns for perceiving prosodic categories such as stress, tone, and intonation (Feng et al., 2019; Tremblay et al., 2021; Wiener, 2017). Moreover, the cue-weighting transfer hypothesis posits that the use of auditory cues can transfer from the first (L1) to the second language (L2) (Kim & Tremblay, 2020; Tremblay et al., 2018, 2021), even across different types of linguistic contrasts (Kim & Tremblay, 2021, 2022; Qin et al., 2017). According to this hypothesis, tonal language listeners, known for their acute sensitivity to f0 direction changes in tones (Deroche et al., 2019; Hallé et al., 2004; Xu

[*] Corresponding author at: Department of Spanish, School of Foreign Languages, Beijing Institute of Technology, 100081 Beijing, China.
E-mail address: shangpeizhu@bit.edu.cn (P. Shang).

https://doi.org/10.1016/j.wocn.2023.101294
0095-4470/© 2023 Elsevier Ltd. All rights reserved.

et al., 2006), should exhibit enhanced perception of f0 cues in all pitch-related events, including sentence intonation, compared to their non-tonal language counterparts. However, empirical evidence from neurobehavioral studies have yielded diverse outcomes, challenging this prediction. Determining whether and in what specific conditions tonal language listeners transfer their L1 perceptual strategies to L2 thus is a pivotal question in the study of mechanisms underlying cross-linguistic perception.

Additionally, extensive research indicates that listeners' perceptual weights can flexibly vary based on cue trade-off relations in dynamic acoustic environments (Kuang & Cui, 2018, Repp, 1982). Yet, fewer studies have explored how word-level suprasegmental elements impact the perception of sentence-level prosodic categories that share similar acoustic cues (e.g., Liu et al., 2022; Yuan, 2006, 2011; Ortega-Llebaria et al., 2019). In Mandarin, for instance, the presence of overlapping acoustic coding space between question intonation and the final Tone2 has been found to reduce the accuracy of intonation perception (Yuan, 2006). A parallel situation is hypothesized in Spanish, where stress and intonation are encoded using identical suprasegmental cues (f0, duration, and intensity) (Ortega-Llebaria & Prieto, 2011), whereas such cues are often rendered secondary or even redundant in English stress, due to the high functional weight of vowel reduction (Chrabaszcz, Winn, Lin, & Idsardi, 2014). Based on these insights, our study focuses on examining how native Spanish listeners and Mandarin learners of Spanish perceive intonation contrasts under diverse acoustic and stress conditions, seeking to elucidate the auditory mapping between acoustic details and intonation categories, and to enhance current understanding of the cue-weighting differences between tonal and non-tonal languages.

1.1. Multiple cue weighting in L1 and L2 intonation perception

Question-statement contrast can be encoded by several resources, other than intonation contour. In Peninsular Spanish, information-seeking yes/no questions differ from broad focus statements typically in word order: the former follow a verb-subject-object order, whereas the latter follow a subject-verb-object order (Haverkate, 2006).[1] In Mandarin Chinese, a key characteristic of yes/no question is the presence of the modal particle *ma* (Yuan, 2011). However, although syntactic and lexical mechanisms encode sentence types, intonation alone can achieve the same effect in both languages. Questions that differ from statements only in prosodic contours are known as "intonation questions," which are the focus of this study. Similar to other phonetic categories like vowels (Tillman et al., 2017) and stress (Chrabaszcz, Winn, Lin, & Idsardi, 2014), intonation is processed within a multidimensional acoustic space. Previous empirical research across various languages including English (Peng et al., 2012), Spanish (Romera Barrios et al., 2007), Cantonese (Ma et al., 2011), and Mandarin (G. Zhang et al., 2022; Yuan, 2006), has identified diverse degrees of acoustic contrast in sentence-final f0, duration, and intensity values between statements and yes/no questions. However, due to the selective nature of human attention (Holt et al., 2018), the relative contribution of these acoustic dimensions to listener perception varies. Not all acoustic cues serve as significant predictors for the identification of intonation contrasts.

The f0 dimension that reliably and saliently correlates with intonation categorization typically carries greatest perceptual weight in listeners' auditory decision-making. The prominence of f0 in intonation perception is recognized across various languages such as English (Peng et al., 2012), German (Niebuhr, 2007), and Cantonese (Ma et al., 2008; 2011). Bolinger's (1978) comprehensive analysis about 250 languages also illustrates the significant role of f0 in intonation, noting that roughly 70% of languages use a final rising pitch to signal questions, while others utilize pitch range variations. Spanish follows the former pattern, whereas Mandarin exhibits a complex use of f0 due to the tone-dependent nature of its intonation system (Liu et al., 2022). Specifically, in Mandarin, question intonation is characterized by an expanded overall f0 range (Wang et al., 2013) or a global upward f0 trend with a locally accelerating f0 rise towards the utterance end (Chen, 2022; Yuan, 2004). These cross-linguistic differences regarding f0 use raise a pertinent question: Do speakers with a tonal language background assign greater weight or show increased sensitivity to f0 cues in prosodic categories compared to those of non-tonal languages? Understanding this, within the framework of cue-weighting transfer theory, is crucial for predicting Mandarin listeners' perceptual weight of f0 cues in L2 Spanish intonation. Further discussion of this topic is presented in Section 1.2.

While f0 is a primary cue in intonation of most languages, secondary cues like duration and intensity can also influence listeners' judgments by conveying important, even if less reliable information (Peng et al., 2012). In Spanish, for example, it has been found that the perceptual weighting of intensity is relatively low and influenced by vowel types, which is in contrast with the more robust cueing provided by f0 and duration cues for stress perception (Ortega-Llebaria et al., 2007; Ortega-Llebaria & Prieto, 2011). However, the extent to which Spanish L1 listeners rely on non-f0 cues for recognizing intonation categories is not fully clear. Given the subtle and inconsistent nature of intensity contrasts in Spanish, particularly in differentiating statements from questions compared to more pronounced duration contrasts (Romera Barrios et al., 2007), it is hypothesized that Spanish listeners may place less weight on intensity than on f0 and duration cues in intonation perception.

On the other side, compared to statements, Yuan (2006) identified a distinctive acoustic profile for yes/no questions in Mandarin, marked by higher intensity across all final tone types and prolonged final syllables when ending in Tone3 and Tone4. Similar acoustic distinctions are also observed by G. Zhang et al. (2022), suggesting that Mandarin listeners may use differential duration and intensity patterns for more accurate intonation recognition. Yet, empirical research has yield mixed results for this prediction. For example, studies have shown that duration and intensity cues in sentence-final positions did not significantly impact intonation categorization for Cantonese listeners and Mandarin learners of English

[1] In other types of statements (e.g., narrow focus statements) and questions (e.g., echo questions and wh-questions), Peninsular Spanish displays different word orders. There are also differences across Spanish dialects. For a detailed discussion on this topic, please see Brown & Rivas, 2011; Escandell-Vidal, 2002; Shang et al., 2021; Bosch & Fernández-Soriano, 2013, and Zubizarreta, 1999.

(Feng et al., 2019; Ma et al., 2011), potentially due to the strong representation of f0 in phonated speech. In contrast, Heeren and van Heuven (2009) noted that, in whispered speech, duration can become an important cue for Dutch listeners' intonation perception. Therefore, the perceptual weighting of secondary cues appears to be context-sensitive, raising questions about how variations in duration and intensity affect intonation perception in cross-linguistic contexts with acoustically covarying cues.

A key aspect of this question involves the transferability of secondary cue-weighting from L1 suprasegmental contrasts to L2 intonation, an area where research has yielded mixed findings. For example, while English L1 listeners heavily rely on duration and intensity cues for differentiating questions from statements, this auditory sensitivity is not found in Mandarin learners of English (Feng et al., 2019). Conversely, Morrow and Liu (2013) reported that Mandarin learners of English were significantly influenced by the final word's intensity in their perception of English intonation, a finding not paralleled in native English listeners. These contrasting results reveal the variability of secondary cues in L1 and L2 intonation perception. Consequently, exploring how Mandarin listeners auditorily assign weights to duration and intensity cues in L2 Spanish intonation emerges a significant yet complex task. This complexity is further amplified by acoustic interplay between stress and intonation in Spanish, which is discussed in the subsequent Section 1.4.

1.2. Debate on f0 perception advantage and the scope of cue-weighting transfer

Given that Mandarin Chinese is a tonal language, its strong informational emphasis of f0 often leads to questions about whether its native listeners exhibit superior f0 processing abilities compared to listeners from non-tonal languages. Research in this area has yielded mixed results, broadly falling into three schools of thought. The first perspective aligns with the traditional claim that speaking a tonal language enhances listeners' sensitivity to acoustic variations in f0. Supporting this notion, Ortega-Llebaria et al. (2017) found that Chinese–English bilinguals detected f0 mismatches faster than speakers of non-tonal language. They also observed that Chinese listeners were more adept at utilizing the shape of f0 contours, especially falling f0 contours, to aid their perception of English words' intonation. Similar f0 advantages by tonal language speakers in comparison with their non-tonal language counterparts have been reported in lexical decision tasks with word-object pairs (Braun et al., 2014), as well as in the perception of contour (specifically Tone1 and Tone3, Chandrasekaran et al., 2007a) and level tones (Xu et al., 2006). Ortega-Llebaria et al. (2017) therefore proposed that extensive experience with a tonal language refines perception of f0 contours in general, irrespective of their association with tonal or intonational meanings.

Conversely, the second viewpoint argues that tonal language experience in one's L1 does not necessarily improve f0 perception in non-native tones and native intonations. So and Best (2010) and Tsukada et al. (2018) provide evidence for this position, showing that Hong Kong Cantonese and Burmese speakers, despite their long-term tonal experience, did not exhibit superior perceptual accuracy in non-native tone pairs compared to other language groups. Additionally, some researchers indicate that lexical-tone interference in tonal languages may be the primary factor in diminishing L1 listeners' sensitivity to f0 cues in sentence intonation. Liang and Heuven (2007), for example, found that Mandarin listeners were consistently slower in discerning questions and statements and less sensitive to f0 cues in intonation than non-tonal L2 learners. This phenomenon, according to the authors, arises from tonal language listeners primarily processing f0 at the lexical level, which could impede their efficiency and sensitivity in perceiving f0 cues at the sentence level.

The third perspective suggests that the f0 perception advantage in tonal language listeners is domain-specific and may not be applicable across all pitch dimensions. This view was long been reflected in Gandour' research (1983), which demonstrated that tonal and non-tonal language listeners relied on different pitch dimensions (e.g., height and direction) for tone perception. Listeners with a tonal language background were more sensitive to pitch directions in distinct tone pairs compare to non-tonal English listeners. Recent behavioral studies over the past decade (e.g., Bidelman, Hutka, & Moreno, 2013; Deroche et al., 2019; Krishnan, Gandour, & Bidelman, 2010) have also highlighted a distinct f0 advantage for tonal language speakers in perceptual tasks where f0 is processed in a manner akin to the auditory demands of their L1 pitch events, notably those involving specific curvatures and directional changes. Complementarily, neurobehavioral research (Chien, Friederici, Hartwigsen, & Sammler, 2020; Doherty, West, Dilley, Shattuck-Hufnagel, & Caplan, 2004; Gandour, 2009; Zatorre & Gandour, 2008) indicates that the neural mechanisms for processing various pitch events are not identical. Notably, f0 perceived as lexical tone activates additional semantic areas only in Mandarin speakers, whereas f0 perceived as intonation engages bilateral brain regions common to tonal and non-tonal language speakers, regardless of the language-specific realization of intonation (Chien et al., 2020). These findings may elucidate why certain previous studies (e.g., Liang & Heuven, 2007; Gussenhoven & Chen, 2000; Tsukada et al., 2018) did not observe significant f0 processing differences in intonation between listeners of tonal and non-tonal languages.

The ongoing discussion on f0 perception not only reveals complexities in perception but also challenges the scope of the cue-weighting transfer hypothesis. Prior research (e.g., Francis & Nusbaum, 2002; Holt & Lotto, 2006) has shown that the weight of a cue used to distinguish phonetic categories in one's L1 transfers to their L2. This transfer is suggested to occur across different types of linguistic categories (e.g., Kim & Tremblay, 2021: from Gyeongsang Korean lexical pitch accents to English lexical stress; Kim & Tremblay, 2022: from South Korean segmental contrast to English lexical stress). However, applying this hypothesis to the transfer of f0 cues from L1 Mandarin to L2 intonation reveals theoretical constraints. The distinct sensitivities of Mandarin listeners to f0 cues in lexical tone versus intonation (Chien et al., 2020) complicate predictions regarding which prosodic category' f0 weight in L1 is transferred to their L2 intonation. Moreover, findings that Mandarin listeners prioritized vowel quality over f0 cues significant in their L1 when perceiving English lexical stress (Chrabaszcz, Winn, Lin, & Idsardi, 2014; Zhang & Francis, 2010) also question the scope of cue-weighting transfer. These findings underscore the need for more in-depth exploration into the transfer of suprasegmental cues across different L1-L2 categories, especially in contexts involving tonal and non-tonal language systems.

1.3. The dynamic nature of speech perception

The weight of input acoustic dimensions that serve for object recognition is not fixed. Listeners can flexibly shift their reliance on acoustic cues depending on their prior history of experience and the demands of the listening environment (Holt et al., 2018; H. Zhang et al., 2022). For instance, when the primary cue of the stimulus was absent or weakened, listeners could increase the weight of secondary cues to improve the accuracy of identifying phonetic categories. (e.g., Feng et al., 2019; Holt & Lotto, 2006; Peng et al., 2012; H. Zhang et al., 2022). The way multiple cue weights are dynamically adjusted or reweighted is frequently described using phonetic trading relations—a phenomenon in which one acoustic cue's shift of values can be compensated by another cue's opposing changes, such that the original percept is preserved (see Repp, 1982, for more details). This compensatory behavior based on cue trade-off regularities has been observed in the perception of various auditory objects, including lexical tones (Liu & Samuel, 2004; H. Zhang et al., 2022) and intonation (Feng et al., 2019; Peng et al., 2012).

The extent of auditory compensation by listeners through cue weight adjustment is related to their sensitivity to the acoustic dimensions that define the internal structure of phonetic categories (Hodgson & Miller, 1996). Studies have shown that individuals with superior F1 (first formant frequency) discrimination skills tend to show greater compensatory responses to formant perturbations in English vowels (Nault & Munhall, 2020; Villacorta et al., 2007). Notably, such compensatory ability varies across linguistic backgrounds. For example, English L1 listeners demonstrated greater susceptibility to f0-duration trade-offs in question-statement identification compared to Mandarin learners of English (Feng et al., 2019). This observation might imply, in line with the established link between perceptual compensation and auditory sensitivity (e.g., Hodgson & Miller, 1996; Nault & Munhall, 2020), that English L1 listeners were more sensitive to these cues. However, Feng et al. (2019) reported contradictory results, challenging this assumption. These inconsistencies highlight the need for further cross-linguistic research to better understand how listeners of tonal and non-tonal languages differentially utilize cue weight adjustments in response to dynamic acoustic changes in intonation.

1.4. Acoustic conflict between prosodic categories

Mandarin intonation, as outlined in Section 1.1, is susceptible to the lexical tone identity. Research has shown that Mandarin yes/no questions with a final rising tone (Tone2) are harder to recognize than those ending with other tone types (Liu et al., 2022; Yuan, 2011; Yuan & Shih, 2004). This difficulty in perception has been primarily attributed to the overlap of f0 contours in Tone2 and question intonation, leading to potential conflicts in f0′s functional use (Liu, 2018; Liu et al., 2022; Wu & Ortega-Llebaria, 2017). Similar f0 conflicts in perceiving lexical tone and intonation are observed in various Sinitic tonal languages, including Cantonese (Kung et al., 2014) and Tianjin Mandarin (Zhang, 2018), with variations in their specific mechanisms (see Chen, 2022 for a comprehensive review).

However, such acoustic conflicts are not exclusive to these specific categories and languages. Given that numerous phonetic objects can be conveyed through similar acoustic resources, inevitable competition arises for the coding spaces (Xu, 2004). In languages like Spanish, where stress and intonation both rely on f0 and duration cues, acoustic competition is inherent (Ortega-Llebaria & Prieto, 2011). The overlap in encoding these distinct prosodic events may impede the processing of acoustic cues necessary for intonation, especially when stress and intonation are present in the same phonetic unit. Given the perceptual difficulty identified in Mandarin questions ending with a Tone2 (Liu et al., 2022; Yuan, 2006; 2011; Yuan & Shih, 2004), we wonder if Spanish questions would also be more difficult to discern when ending with an oxytone word, where the functional load of f0 and duration in the final stressed syllable is not unique. To explore this, we analyzed the perceptual differences of intonation between one-word sentences with and without stress in the final syllable.

1.5. The present study

Building on earlier discussions, this study aims to investigate several key aspects of intonation cue weighting among native Spanish listeners and Mandarin L2 learners of Spanish. To achieve this, we have designed two perceptual tests, in which the stimuli were generated by gradually manipulating the final f0 contour from falling to rising directions, with concurrent adjustments in either duration (Test 1) or intensity (Test 2). Our acoustic manipulations are strategically centered on the sentence-final position. This choice is informed by prior research (Face, 2005, 2007), which identifies the final intonational contour as the primary indicator for perceiving Spanish statements and questions. To exclude the potential influence of the intonational prenucleus, we have used sentences with a single stressed word to create the stimuli. Overall, by conducting two perceptual identification tasks, our study sets out to address the following research questions (RQs), with the hypotheses for each question detailed thereafter:

RQ1 and Hypothesis 1: How do variations in f0, duration, and intensity cues affect intonation perception among Spanish L1 listeners and Mandarin L2 learners of Spanish, and are there notable differences in perceptual weighting between the two language groups? This question focuses on listeners' cue-weighting strategies in intonation perception, emphasizing the role of secondary cues which have received less attention in prior research. Based on the literature reviewed in Sections 1.1 and 1.2, we hypothesize that Mandarin L2 learners will show a sensitivity to f0 linear changes comparable to Spanish L1 listeners. Additionally, we posit that Spanish L1 listeners can use duration effectively in recognizing intonation, but their reliance on intensity cues is expected to be limited. Conversely, for Mandarin L2 learners, both duration and intensity cues are predicted to have minimal influence on their intonation perception, based on prior findings for Mandrin learners of English (Feng et al., 2019).

RQ2 and Hypothesis 2: Do Spanish L1 listeners and Mandarin L2 learners adjust their cue weights to compensate for the fluctuating variations in intonation, and are there any differences between their compensatory strategies? This question explores the perceptual trade-offs between L1 and L2 listeners, exploring how their compensatory behavior correlates with their sensitivity to the covarying acoustic cues. Drawing on the discussions in Section 1.3, we hypothesize that L1 and L2 listeners will exhibit varying degrees of auditory compensation based on their respective sensitivities to the acoustic cues under continuous changes.

RQ3 and Hypothesis 3: How does the stress pattern of the one-word sentence influence listeners' use of acoustic cues to distinguish Spanish intonation contrasts? This question examines the interplay between stress and intonation processing within a single-word context. The hypothesis tested is that paroxytone words, with the stress on the penultimate syllable, are more likely categorized as questions compared to oxytone words under the same acoustic condition. This assumption is based on the premise that paroxytone, being the most frequent and unmarked stress pattern in Spanish, typically demands less cognitive processing during perception (Defior & Serrano, 2017; Roca, 2019). Conversely, in oxytone words, where both stress and intonation are encoded through identical acoustic signals in the final syllable, the processing of these overlapping cues for lexical stress might occupy the acoustic space allocated for identifying question intonation, hereby reducing listeners' efficacy in processing the intonational cues.

2. Methods

2.1. Participants

The participants were recruited by word of mouth or through advertisements posted on social media accounts of the Phonetics Laboratory of the Universitat de Barcelona or elsewhere. All participants gave informed consent electronically before inclusion in the test, and the study design was approved by the local ethics committee. Generally, but not always, participants were compensated monetarily for their time. A total of forty-eight Spanish-native (hereafter, SN) listeners and ninety-five Mandarin-native (hereafter, MN) speakers of Spanish participated in the experiment. All SN listeners were born in Spain and were living in Spain at the time of testing. In contrast, MN listeners were born in mainland China and confirmed that Peninsular Spanish was the language variety to which they had been predominantly exposed during the learning process. To provide a detailed description of the language profile of the L2 speakers, information on the MN listeners' immersion experience in the target language environment is given in Table 1. However, in this study, we did not strictly control for the MN listeners' language immersion status and length of immersion in the native language environment, as these two variables have been reported to be poor predictors of L2 phonological and phonetic accuracy, especially when the amount of L2 use and interaction with native speakers by learners was unclear (Nagle, 2013; Shively, 2008; Trofimovich & Baker, 2006).

We performed data cleaning before statistical modeling. Data on two MN listeners who started studying Spanish before the age of 16 were removed to control for the effect of age of acquisition. Data on two SN listeners aged over 60 and on five MN listeners with an A1 or A2 proficiency in Spanish were also excluded. All MN listeners were reported to have an intermediate (B1), advanced (B2), or superior (C1) level of proficiency in Spanish. The Spanish proficiency of most MN listeners (roughly 60%) was assessed using the level information of the last Spanish certificate DELE (Diploma of Spanish as a Foreign Language) that they held. As for the rest of the MN listeners who did not have such a diploma, they were asked to self-evaluate their Spanish proficiency based on the Common European Framework of Reference for Languages (CEFR). The CEFR divides language proficiency into six levels, ranging from A1 to C2, and provides detailed descriptions of the knowledge and skills required at each level. Overall, in Test 1, there were 39 SN listeners and 78 CN listeners, whose ages ranged from 18 to 59 years (N = 117, 89 women, 28 men, M_{age} = 28.17, SD = 8.34). Test 2 had 33 SN listeners and 77 CN listeners, whose ages ranged from 19 to 58 years (N = 110, 84 women, 26 men, M_{age} = 28.03, SD = 8.49). No participants reported any history of hearing or communication disorders.

2.2. Stimulus synthesis

The intonation of Spanish yes/no questions and broad focus statements differs in the intonational prenucleus and nucleus (Face, 2011). Since the focus of this study is the utterance-final position, we have only used sentences with a single stressed word as source materials such that the prenucleus effect could be excluded. Each sentence comprised one trisyllabic word: *Sevilla*[2] (penultimate syllable stress—paroxytone) and *Alcalá*[3] (final syllable stress—oxytone). The original recordings of the two items were obtained using a discourse completion task (Félix-Brasdefer, 2010). A female native speaker of Peninsular Spanish (age at the time of the recording: 31) was required to produce the two one-word sentences in a broad focus statement context. The speaker was asked to speak at a normal rate. The recordings were done in a quiet room using a Rode Smartlav + microphone connected to a Scarlett Solo interface. Speech files were digitized at a sampling rate of 44.1 kHz with a quantization precision of 16 bits. The final syllable of the one-word sentences was chosen for acoustic manipulation. The target stimuli were created by manipulating f0, duration, and intensity. Prior to pitch manipulation, the non-final segments of the two one-word sentences were normalized at a 70 dB sound pressure level and used as a reference for the parametric variation of intensity in Test 2.

2.2.1. F0 manipulation

The f0 contour of the word-final syllable was replaced by a multi-step f0 continuum using the "to manipulation" function in Praat (Boersma & Weenink, 2020). The f0 contour of the last syllable was stylized into two pitch points (defined as A1 and

[2] The first test word, *Sevilla*, is the name of the largest city in the Spanish autonomous community of Andalusia and is pronounced [seˈβiʎa] in Peninsular Spanish.

[3] The second test word, *Alcalá*, is the name of a city in the Spanish autonomous community of Madrid and is pronounced [alkaˈla] in Peninsular Spanish.

Table 1
Descriptive statistics of MN listeners' language immersion status and living time in Spain.

	Living in Spain or not		Living time in Spain				
	Yes	No	< 3 months	3 ∼ 8 months	6 months ∼ 1 year	1 year ∼ 2 years	> 2 years
Test 1	51	27	4	2	14	17	41
Test 2	46	31	4	2	17	20	34

A2), and the values between them were defined by interpolation. The start point, A1, was fixed at the beginning of the final syllable's vowel, keeping the pitch height similar to the original utterance. Thus, the value of A1 for *Alcalá* (see Fig. 1) and *Sevilla* (see Fig. 2) was set at 196 Hz and 200 Hz, respectively. The endpoint, A2, was anchored at the last regular glottal pulse observed in the spectrogram and had an f0 value nearly identical to A1 (difference less than 5 Hz). The f0 continuum of A2 was manipulated upward nine times and downward once,[4] with a 20 Hz step size (greater than the slightest pitch variation normal-hearing adults can perceive). Thus, 11 f0 steps spanning over 200 Hz were generated for the final syllable of each test word. The 22 f0 stimuli (2 stress patterns*11 f0 steps) with different offset frequencies were further used as the basis for manipulating the duration and intensity.

In selecting the scale for f0 manipulation, we have opted for linear Hertz because, compared to other logarithmic scales (such as semitones or ERB-rate), a (close-to) linear scale has proven to be a fairer and more reliable choice when addressing perceived equivalence under conditions of constant baseline for f0 height manipulation (Jeon & Heinrich, 2022). This approach is particularly relevant to our study, where the f0 contour of the adjacent, non-manipulated contour remains unchanged. Additionally, the linear Hertz provides a finer representation of the mapping relationship between f0 variations and perceived intonation categories. This precision stems from its ability to facilitate direct comparisons of different f0 values without necessitating any additional scaling or transformation. Moreover, the widespread use of linear Hertz in intonation perception research makes it a practical choice, enabling more straightforward comparisons with prior studies that have investigated cross-linguistic differences in f0 perception using this scale.

2.2.2. Duration manipulation

In Test 1, the 22 f0 stimuli previously detailed in Section 2.2.1 were used to add a duration manipulation. Three duration levels were established for this study: short, medium (original), and long. As shown in Fig. 1 and Fig. 2, the medium duration was defined using the original final syllable vowel duration for each word. To generate the long-duration stimuli, we first extracted a sequence of glottal cycles with an interval of 50 ms from the center of the final syllable's original vowel nucleus in each word. This 50-ms segment was then appended to the end of the 5th glottal cycle of the original vowel, thereby lengthening the duration of both words.[5] To avoid clicks or spectral discontinuities when cutting or pasting audio, segment boundaries were precisely placed at the zero crossings of the acoustic signals. The choice of + 50 ms was made based on previous acoustic evidence, which found the final vowel in Spanish yes/no questions to be roughly 40–70 milliseconds longer than that in statements, varying with the stress position in the final word (Romera Barrios et al., 2007). The short-duration stimuli were similarly constructed by extracting a sequence of regular glottal cycles with an interval of 40 ms from the corresponding position in the original vowel nucleus. This reduced duration was determined based on the shortest production of the statement by the same Spanish speaker. Schematic representations of the duration manipulation are included in Appendix A, and Table 2 lists the precise vowel lengths for the two test words after the duration manipulation. Thus, in Test 1, each duration level was paired with 11 different pitch contours, culminating in a total of 66 stimuli (2 stress patterns*3 duration levels*11 f0 steps).

2.2.3. Intensity manipulation

In Test 2, the same set of 22 f0 contours was used to manipulate intensity. The intensity of the final syllable in each word was altered using the "constant amplification" function in the Cool Edit Pro 2.1 software (Syntrillium Software Corporation, 2003). We established three intensity levels by applying changes of −7 dB, 0 dB, and + 7 dB to the final syllable's vowel. A schematic representation of this intensity manipulation is included in Appendix A . The adjusted values were based on the normalized intensity level of the sentence's non-final part (70 dB, as outlined in Section 2.1.1) as a baseline. Hence, the low, medium (original), and high intensities were set at 63 dB, 70 dB, and 77 dB, respectively. Therefore, Test 2 comprised 66 auditory stimuli, categorized into 2 stress patterns, 3 intensity levels, and 11 f0 steps.

2.3. Procedure

The data for the perceptual experiment were gathered through an online survey, developed using the web-based tool

[4] Acoustic studies of Peninsular Spanish indicate that the average f0 change on the last syllable of broad focus statements is usually between −20 Hz and +20 Hz from the preceding adjacent syllable (Romera Barrios et al., 2007; Salamanca et al., 2005). Based on these findings, we have implemented a single-step downward adjustment of 20 Hz in the f0 contour of the word-final syllable. This approach aims to align our synthesized stimuli more closely with the intonation patterns commonly observed in natural Spanish speech.

[5] We applied a uniform duration treatment to both stress patterns for several reasons. The first and most important is that this approach allows us to explore whether listeners differ in their use of acoustic cues for question perception depending on stress patterns. Second, the lack of empirical evidence in Spanish as to whether oxytone words require a larger duration increase compared to paroxytone words for question perception renders the application of customized duration adjustments for each stress pattern impractical. In our study, despite the uniform duration treatment, paroxytone words consistently had longer final vowels than oxytone words across all conditions (see Table 2). Such design aligns with previous acoustic observations in Spanish, which showed that the duration of the vowel in stressed and accented syllables following an intonation phrase boundary was longer than that in unaccented stressed syllables, both in declarative and interrogative sentences (e.g., Ortega-Llebaria, 2006; Ortega-Llebaria et al., 2013; Ortega-Llebaria & Prieto, 2011; Romera Barrios et al., 2007).

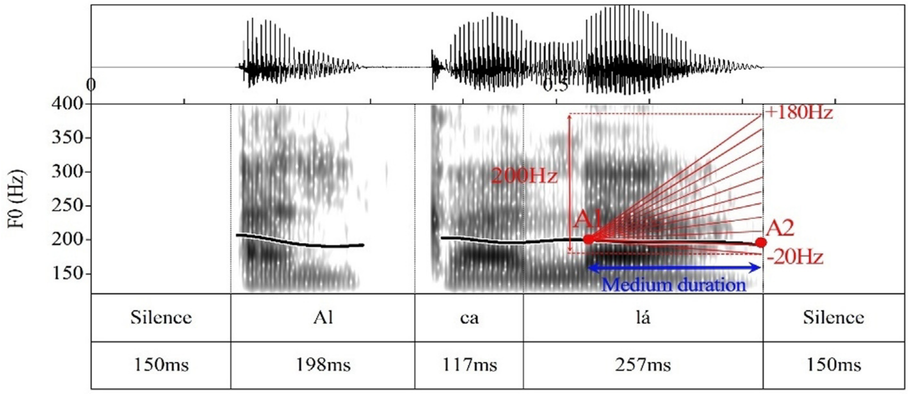

Fig. 1. Schematic representation of the f0 manipulation in the oxytone word *Alcalá*. The start point A1 was 196 Hz. The endpoint A2 was manipulated from 176 Hz to 376 Hz, with a 20 Hz step size. The original duration of the final vowel (184 ms) was set as the medium duration for the stimulus *Alcalá*.

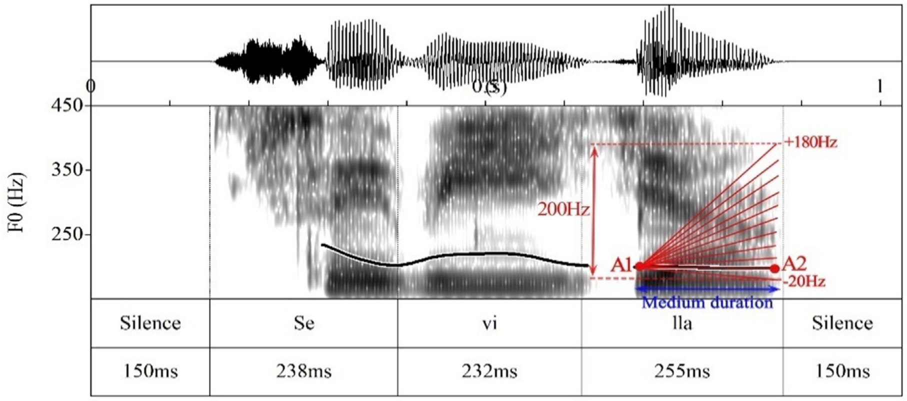

Fig. 2. Schematic representation of the f0 manipulation in the paroxytone word *Sevilla*. The start point A1 was 198 Hz. The endpoint A2 was manipulated from 178 Hz to 378 Hz, with a 20 Hz step size. The original duration of the final vowel (171 ms) was set as the medium duration for the stimulus *Sevilla*.

Table 2
Detailed values of the final vowel length for the two test words across the three duration levels.

Final vowel length	Alcal<u>á</u>	Sevill<u>a</u>
Short	144 ms	131 ms
Medium (original)	184 ms	171 ms
Long	234 ms	221 ms

Alchemer. The survey was divided into three sections: the first section collected socio-demographic information, while the second and third sections presented the 66 auditory stimuli for Test 1 and Test 2, respectively. Administered in the participant's preferred language (Mandarin or Spanish), the survey randomized the order of stimuli in each test, and presented them on-screen without punctuation marks. Participants had the option to partake in either one or both tests, based on their interest. Those opting for a single test were randomly assigned to one of the two auditory tests. They were instructed to use earphones in a quiet room and to listen to each stimulus once, with the provision to replay if technical issues occurred. Prior to the formal test, participants performed a practice trial to acquaint themselves with the procedure. Diverging from the binary choices (statement vs. question) typical in categorical perception tasks, we adopted a five-point Likert response scale for a more nuanced assessment of how multiple acoustic signals map onto intonation categories. This approach can better capture the rich internal structure in the representation of phonetic categories (Holt et al., 2018).

During the test, participants were presented with five response options for each stimulus: "statement," "more statement than question," "either statement or question," "more question than statement," and "question." They were asked to select the option that most closely matched their perception of the stimulus. Fig. 3 illustrates the distribution of perceptual responses for both tests, revealing a bimodal pattern with the highest frequencies at the two ends of the scale.

2.4. Statistical analysis

The perceptual responses were fitted using a logistic sigmoid model, as shown in Eq. (1). This approach was chosen considering the probabilistic distribution of the results and the close-to S-shaped curve observed in the question-statement identification function. The five response options in the perceptual tests were assigned values of 0, 0.25, 0.5, 0.75, and 1, which represented the probability from 0 to 1 that a stimulus was perceived as a yes/no question. Then, the identification results from each participant at each duration and intensity level were modeled as a function of the f0 contour, using the following equation:

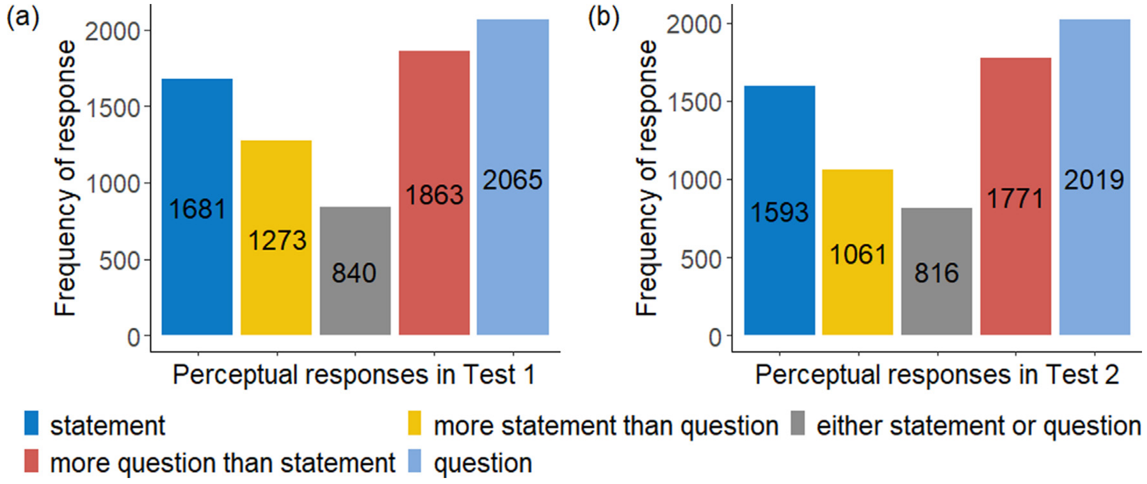

Fig. 3. Frequency distribution of participants' perceptual responses in Tests 1 and 2.

$$p = \frac{1}{1 + e^{(-(x - x_0)/b)}} \tag{1}$$

In this model, p represents the probability of perceiving a one-word sentence as a yes/no question. The variable x denotes the degree of f0 change applied to the final f0 contour, which varied from -20 Hz to 180 Hz in increments of 20 Hz. The parameter x_0 indicates the point at which the question-statement boundary was at 50% identification. Higher x_0 values suggest a higher f0 contour needed for question recognition and vice versa. The parameter b is a steepness indicator which is inversely related to the slope of the identification curve. The smaller the b, the steeper the identification curve and the more sensitive listeners to f0 linear changes in intonation. Model fitting was executed in MATLAB (The Mathworks, 2022), utilizing the Levenberg-Marquardt algorithm (Moré, 1978), a hybrid technique widely employed for extracting model parameters, especially in nonlinear least-squares curve-fitting problems (Gavin, 2022). Through an iterative process, the algorithm sought to determine the optimal values for the fitting parameters x_0 and b in relation to the independent variable x, thereby minimizing the objective function F(x_0, b). This function represents the sum of squared deviations between N pairs of actual data points (p) and their corresponding model predictions, as specified in Eq. (2). The code used for this data modelling process is provided in Appendix A .

$$[x_0, b]^* = \underset{x_0,\ b}{\operatorname{argmin}} F(x_0, b) = \underset{x_0, b}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| p - \frac{1}{1 + e^{-(x - x_0)/b}} \right\|^2 \tag{2}$$

To enhance the reliability of our curve fitting results, we excluded any instances where the Root Mean Square Error (RMSE) exceeded 0.03, as this indicated a less accurate model prediction. Additionally, we built four linear mixed models (LMMs, see Table 3) using the *lme4* package for R (R Core Team, 2021) to further examine the hypotheses of the study. Mandarin learners with varying proficiency levels in Spanish (B1, B2, and C1) were grouped together for analysis. This decision stemmed from the results of a backward elimination process, detailed in Appendix A , which revealed statistically no significant differences among the three proficiency levels of Mandarin learners in the outcomes of the perceptual tests (all p values > 0.1). Thus, the current study did not categorize Mandarin participants into subgroups based on their L2 proficiency. Table 3 shows that the dependent variables of the four LMMs were the question-statement identification boundary x_0 or the steepness indicator b.[6] The fixed effects included acoustic condition (Duration: short < medium < long; Intensity: 63 dB < 70 dB < 77 dB), language group (MN vs. SN), stress pattern (oxytone vs. paroxytone), and all their interactions. To account for individual variability among participants, each subject was included as a random effect. For analyzing the main effects, Type III ANOVA was employed, and the *emmeans* package (Lenth et al., 2019) facilitated multiple comparisons of interaction effects.

Table 3
Linear mixed models built for the curve fitting data in Tests 1 and 2.

Test 1	Model 1 < lmer(x_0 ~ Duration*Language group*Stress pattern + (1|Subject), data = Test1, REML = F)
	Model 2 < lmer(b ~ Duration*Language group*Stress pattern + (1|Subject), data = Test1, REML =F)
Test 2	Model 3 < lmer(x_0 ~ Intensity*Language group*Stress pattern + (1|Subject), data = Test2, REML = F)
	Model 4 < lmer(b ~ Intensity*Language group*Stress pattern + (1|Subject), data = Test2, REML = F)

Specifically, to test Hypothesis 1, we first examined the weighting of duration and intensity by analyzing their interaction with language group in all models, especially in models 1 and 3. Significant changes in x_0 and b, ranging from lower to higher levels of duration and intensity within a particular group, would indicate a strong reliance on these cues for intonation recognition, while minimal variations would suggest a

[6] x_0 and b are independent parameters that contribute uniquely to the model's behavior. As shown in the subsequent Section 3.1, altering the value of x_0 results in a horizontal shift of the sigmoid function along the x-axis, thereby modifying its location but not its slope or overall shape. Conversely, varying the value of b affects the steepness of the sigmoid curve's slope, without impacting its x-axis positioning. This independence is further supported by our internal consistency analysis. The analysis, revealing a Cronbach's alpha of less than 0.5 in both Tests 1 and 2, suggesting that x_0 and b do not have a consistent relationship and they may represent different underlying construct. Given their independent nature and the absence of a consistent correlation (detailed in Appendix A), we justify the inclusion of x_0 and b as separate dependent variables in our LMMs.

lower perceptual weight of the secondary cues. Then, we analyzed the outputs of models 2 and 4. If the variable b shows no significant differences between language groups under diverse conditions, this would support a similar level of sensitivity to intonational f0 cues among SN and MN listeners. In testing Hypothesis 2, we compared the compensatory behaviors of SN and MN listeners, focusing on their extent of changes in x_0 and b in response to secondary cue variations. The hypothesis would be confirmed if the group exhibiting higher sensitivity to these covarying cues also shows greater adjustments in x_0 or b. Finally, Hypothesis 3 was evaluated by analyzing the impact of stress pattern on x_0 and b under different conditions. A significantly lower x_0 value when perceiving paroxytone words would confirm our predictions that words with penultimate stress require lower f0 contours for question recognition, compared to those with final stress. Additionally, a notably lower b value when perceiving paroxytone words would suggest greater sensitivity among listeners to f0 linear transitions in paroxytones as opposed to oxytones. Further details for the data analysis can be found in Section 3.

3. Results

3.1. Results of Test 1

The results of Model 1 showed a statistically significant three-way interaction between duration, language group and stress pattern [$\chi^2(2) = 6.20$, $p < .05$], and a significant two-way interaction of duration with language group [$\chi^2(2) = 8.55$, $p < .05$] and stress pattern [$\chi^2(2) = 14.62$, $p < .001$]. In contrast, there was no significant interaction between stress pattern and language group [$\chi^2(1) = 0.21$, $p > .1$].

The post-hoc test of the two-way interaction between duration and language group (see Table 4) indicated that SN listeners had a significantly higher question-statement boundary (x_0) than MN listeners over the three duration levels, implying that SN listeners needed higher f0 contours to identify one-word sentences as questions than MN listeners. Moreover, as revealed by the three coefficients of the contrast in Table 4, differences in the value of x_0 between MN and SN listeners gradually decreased as the duration of the word-final syllable increased to the highest level (see Fig. 4). This is mainly due to that SN listeners strongly lowered the f0 threshold for question identification when listening to stimuli with a long duration compared to those with a short or medium duration (see Table 5). A similar effect of duration on x_0 was observed in the MN group. For example, Table 5 shows that MN listeners exhibited a significantly lower x_0 for stimuli with a long duration compared to those with a short duration. However, the magnitude of change in x_0 value was less pronounced for the MN group than for the SN group when comparing the longest to the shortest durations, as indicated by the contrast coefficients in Table 5. This suggests that MN listeners made fewer auditory f0 compensations in response to decreased duration and were less sensitive to duration cues than SN listeners.

On the other side, the post-hoc analysis of the two-way interaction between stress pattern and duration revealed that paroxytone words were perceived with significantly lower question-statement boundary (x_0) than oxytone words across the three duration levels (all $ps < 0.001$). The three-way interaction analysis indicated that the simple interaction between stress pattern and duration was different across language groups. Specifically, Fig. 5 shows that the question-statement boundary (x_0) for SN listeners was significantly lower in paroxytone words at the three duration levels compared to oxytone words (all $ps < 0.001$). In contrast, the effect of stress pattern was significant for MN listeners only at the short duration level [$t(579) = 4.03$, $p < .001$].

In Model 2 fitted for the steepness indicator b, we observed significant simple main effects of language group [$\chi^2(2) = 5.30$, $p < .05$] and stress pattern [$\chi^2(1) = 6.38$, $p < .05$], as well as a significant two-way interaction between language group and stress pattern [$\chi^2(1) = 12.03$, $p < .001$]. Our further analysis of the effect of language group across each level of duration and stress pattern (Table 6) reveals that compared to MN listeners, SN listeners exhibited significantly smaller b values, which correspond to steeper question-statement identification functions and higher f0 sensitivities, specifically when perceiving oxytone words with a short duration. For other duration and stress conditions, the results showed no significant variance in b between the SN and MN groups. Additionally, pairwise comparisons of the stress patterns across each category of language group indicated that MN listeners had significantly stepper identification functions (i.e., smaller b) when perceiving paroxytones compared to oxytones [$t(581) = 6.45$, $p < .0001$], whereas SN listeners had similar steepness of identification curves in the two stressed words [$t(579) = -1.25$, $p > .1$].

3.2. Results of Test 2

The analysis of Model 3 with the question-statement identification boundary (x_0) as the dependent variable revealed a statistically significant main effect of language group [$\chi^2(1) = 31.46$, $p < .001$]. However, the effects of intensity, stress pattern and other interactions did not reach statistical significance. We performed pairwise comparisons between language groups across each level of intensity and stress pattern. Table 7 illustrates that SN listeners consistently exhibited a significantly higher question-statement boundary (x_0) than MN listeners in each assessed condition, aligning with the findings from Test 1.

Furthermore, our examination of multiple comparisons of intensities across language groups and stress patterns, despite the main effect of intensity being non-significant, is depicted in Fig. 6. It reveals a statistically marginal trend for both MN [$t(532) = 0.81$, $p > .05$] and SN [$t(532) = 1.84$, $p > .05$] listeners to assign lower question-statement boundaries (x_0) to the perception of paroxytone words when the intensity of word-final syllables was altered from 63 dB to 77 dB. Specifically, the contrast coefficients between 63 dB and 77 dB for each language group indicated a relatively stronger intensity effect on the perception of paroxytone words for SN listeners ($\beta = 6.57$, $SE = 3.58$) compared to their MN counterparts ($\beta = 1.91$, $SE = 2.36$). Additionally, our interactional analysis concerning stress pattern by language group showed that both MN and SN listeners tended to exhibit higher

Table 4
Pairwise comparisons of the effect of language group on the question-statement identification boundary (x_0) across each level of duration (*t*-stat. ****p* <.001; ** *p* <.01; * *p* <.05; '.' *p* <.1).

Duration	Contrast	Estimate	*SE*	*df*	*t*	*p*
Short	MN – SN	−27.3	4.23	179	−6.44	< 0.001***
Medium	MN – SN	−25.7	4.23	179	−6.08	< 0.001***
Long	MN – SN	−12.1	4.23	179	−2.85	< 0.001**

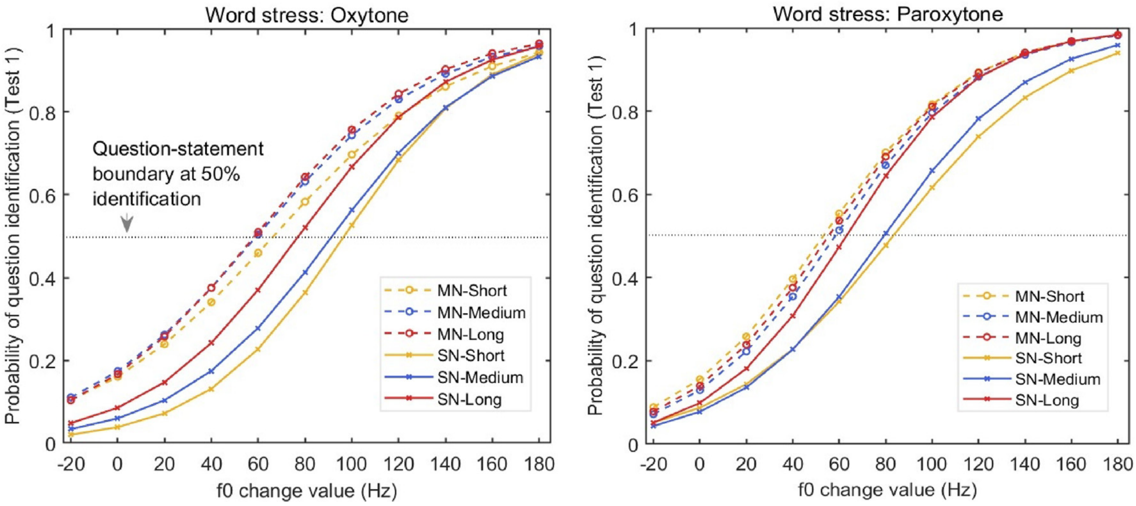

Fig. 4. Average curve fitting results for listeners' question-statement identification as a function of the f0 change in the word-final syllable across the three duration levels.

Table 5
Pairwise comparisons of the effect of duration on the question-statement identification boundary (x_0) at each level of language group (*t*-stat. ****p* <.001; ** *p* <.01; * *p* <.05; '.' *p* <.1).

Language Group	Contrast	Estimate	*SE*	*df*	*t*	*p*
MN	Short – Medium	1.65	1.86	580	0.89	0.65
	Medium – Long	3.29	1.87	580	1.76	0.18
	Short – Long	4.94	1.86	580	2.64	< 0.05*
SN	Short – Medium	3.19	2.60	579	1.23	0.44
	Medium – Long	16.96	2.60	579	6.52	< 0.001***
	Short – Long	20.15	2.59	579	7.78	< 0.001***

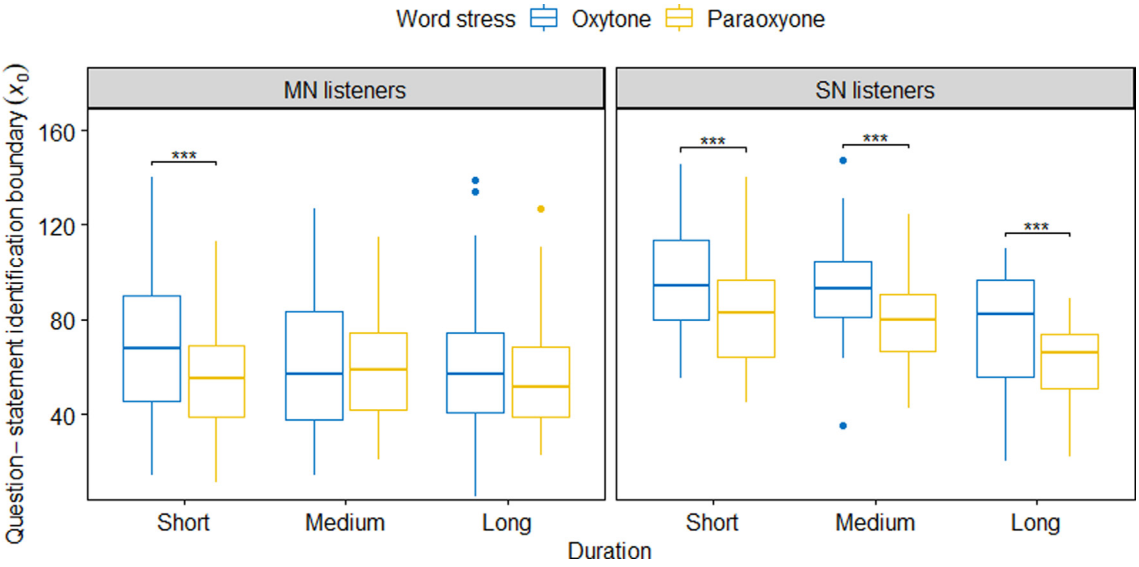

Fig. 5. Effect displays for the three-way interaction between duration, language group and stress pattern on the question-statement identification boundary (x_0).

Table 6
Pairwise comparisons of the effect of language group on the steepness indicator (b) at each level of duration and stress pattern (t-stat. ***p <.001; ** p <.01; * p <.05; '.' p <.1).

Stress pattern	Duration	Contrast	Estimate	*SE*	*df*	*t*	*p*
Oxytone	Short	MN – SN	9.74	4.27	296	2.28	< 0.05*
	Medium	MN – SN	2.67	4.29	301	0.62	0.53
	Long	MN – SN	4.15	4.27	297	0.97	0.33
Paroxytone	Short	MN – SN	−4.51	4.27	296	1.23	0.29
	Medium	MN – SN	−3.68	4.25	293	6.52	0.39
	Long	MN – SN	−1.88	4.26	295	7.78	0.69

Table 7
Pairwise comparisons of the effect of language group on the question-statement boundary (x_0) at each level of intensity and stress pattern (t-stat. ***p <.001; ** p <.01; * p <.05; '.' p <.1).

Intensity	Stress pattern	Contrast	Estimate	*SE*	*df*	*t*	*p*
63 dB	Oxytone	MN – SN	−24.5	4.41	294	−5.56	< 0.001***
	paroxytone	MN – SN	−23.1	4.39	290	−4.65	< 0.001***
70 dB	Oxytone	MN – SN	−22.3	4.36	283	−5.13	< 0.001***
	paroxytone	MN – SN	−20.4	4.39	290	−4.65	< 0.001***
77 dB	Oxytone	MN – SN	−30.1	4.34	281	−6.93	< 0.001***
	paroxytone	MN – SN	−18.5	4.33	278	−4.27	< 0.001***

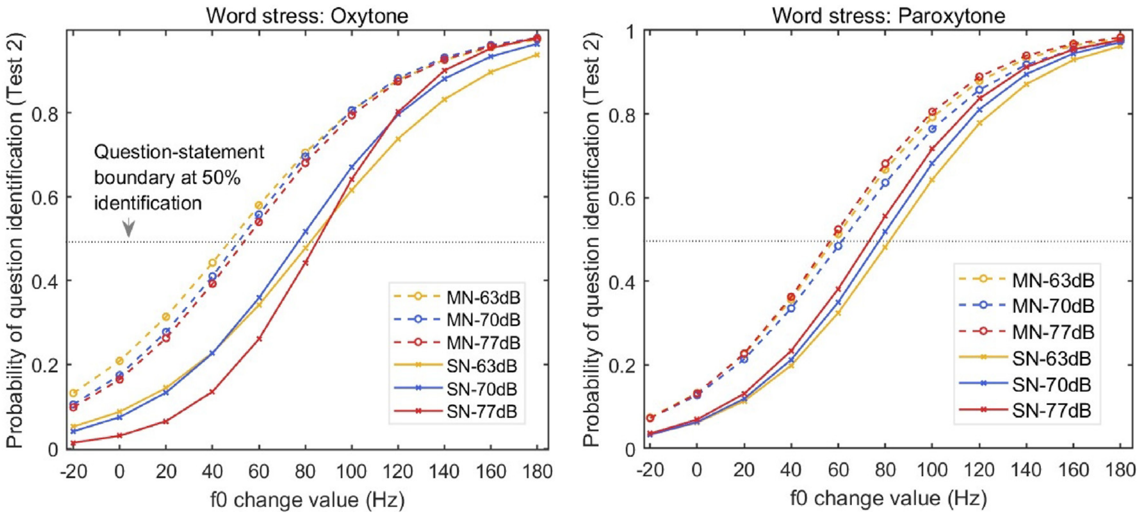

Fig. 6. Average curve fitting results for listeners' question-statement identification as a function of the f0 change in the word-final syllable at the three intensity levels.

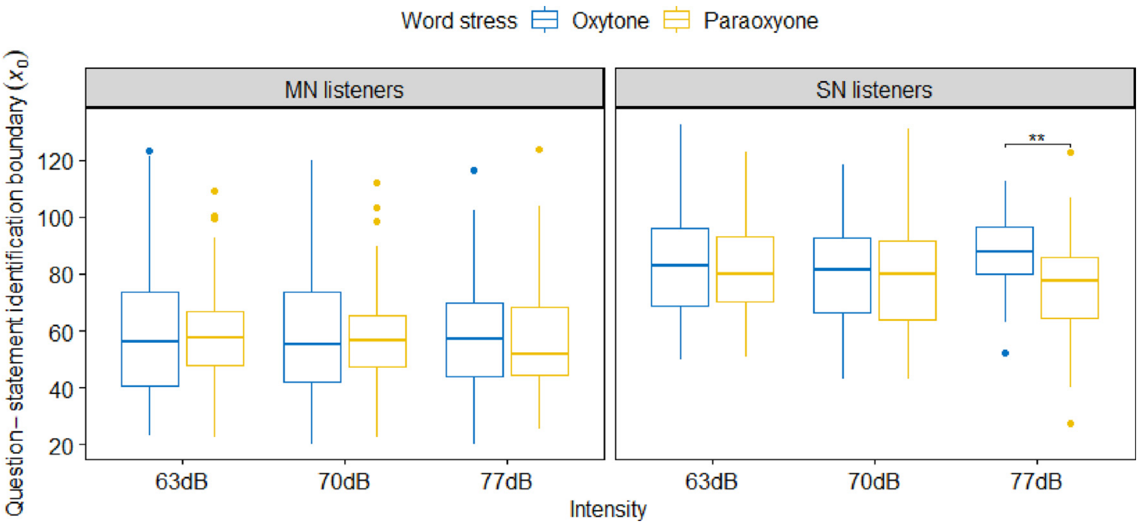

Fig. 7. Effects display for the three-way interaction between intensity, language group and stress pattern on the question-statement identification boundary (x_0).

Table 8
Pairwise comparisons of the effect of language group on the steepness indicator (*b*) at each level of intensity and stress pattern (*t*-stat. ****p* <.001; ** *p* <.01; * *p* <.05; '.' *p* <.1).

Stress pattern	Intensity	Contrast	Estimate	*SE*	*df*	*t*	*p*
Oxytone	63 dB	MN – SN	1.92	3.65	290	0.53	0.60
	70 dB	MN – SN	2.42	3.60	280	0.67	0.50
	77 dB	MN – SN	5.85	3.60	278	1.63	0.11
Paroxytone	63 dB	MN – SN	0.76	3.64	287	0.21	0.83
	70 dB	MN – SN	5.65	3.63	286	1.56	0.12
	77 dB	MN – SN	2.54	3.58	275	0.71	0.48

question-statement boundaries (x_0) in recognizing oxytone words compared to paroxytone words. Notably, this perceptual difference between stress patterns was significant only for SN listeners at an intensity of 77 dB [*t*(530) = 3.24, *p* <.01], as shown in Fig. 7.

On the other side, our analysis of Model 4 with the steepness indicator (*b*) as the dependent variable revealed a marginally significant main effect of stress pattern [$\chi^2(1) = 3.61$, *p* =.057]. While the overall main effect of language group on the steepness indicator (*b*) was not statistically significant, a closer examination of the contrasts between SN and MN listeners demonstrates a consistent yet non-significant trend in intonation perception. As outlined in Table 8, positive coefficients across all intensities and stress patterns were observed, suggesting a tendency for SN listeners to display steeper identification curves for question-statement perception. This pattern may indicate a potential for enhanced sensitivity in SN listeners to intonational f0 cues relative to their MN counterparts. Nevertheless, this observed trend did not reach the threshold of statistical significance, and thus, any interpretations drawn should be considered as indicative rather than conclusive.

We also performed pairwise comparisons between intensity conditions across each level of the other variables, aiming to assess the intensity weighting in intonation perception. SN listeners were observed to consistently exhibit a steeper slope for the identification function (i.e., smaller *b*) when the intensity of word-final syllables increased from 63 dB to 77 dB. However, this intensity effect was marginally significant only for perceiving oxytone words [*t*(531) = 2.06, *p* =.09], as depicted in Fig. 6. In contrast, for MN listeners, the impact of intensity on the steepness indicator (*b*) was irregular and not statistically significant (all *ps* > 0.1). Furthermore, when comparing stress patterns at each level of intensity and within each language group, MN listeners were found to show a marginally steeper identification curve for paroxytone words as opposed to oxytone words in the 63 dB condition [*t*(533) = 1.88, *p* =.06]. Conversely, for SN listeners, the identification curve's steepness did not differ significantly between the two stress patterns across all intensity levels (all *ps* > 0.1).

4. Discussion

The aim of this study was to examine the dynamic multi-cue weighting in the perception of statements and yes/no questions in context situations involving tonal and non-tonal languages. The overall findings provide valuable insights into the cross-linguistic differences between L1 and Mandarin L2 speakers of Spanish in their utilization of fine-grained acoustic details across multiple dimensions to perceive intonation categories.

4.1. The role of f0 cue in cross-linguistic intonation perception

Our study aligns with existing research (Chandrasekaran et al., 2007b; Feng et al., 2019; Peng et al., 2012; Shang et al., 2022) in establishing a robust positive correlation between f0 and question perception within specific pitch intervals, as illustrated in Fig. 4 and Fig. 7. Both SN and MN groups displayed an increased likelihood of identifying questions as the f0 level elevated at sentence-final positions. Specifically, regarding RQ1, our analysis of the curve fitting results revealed no significant differences in the slope of the identification curves between SN and MN groups across most conditions, supporting our initial hypothesis that MN listeners, despite of their long-term experience with a tonal language, exhibit sensitivity to f0 linear modulations in Spanish intonation comparable to that of SN listeners from a non-tonal language background. Thus, the superior f0 sensitivity typically shown by Mandarin listeners to contrastive tone patterns with particular curvatures and direction changes (e.g., Deroche et al., 2019; Hallé et al., 2004; Ortega-Llebaria et al., 2017; Xu et al., 2006) did not extend to their processing of f0 linear transitions in L2 intonation. Similarly, Bildelman et al. (2013) noted that Cantonese listeners' enhanced ability to discriminate large f0 incongruences did not transfer to their perception of subtle f0 deviations smaller than the f0 differences between Cantonese-level tones. These findings challenge the traditional claim of tonal language benefit in f0 perception and diverge from existing cue-weightings studies that suggested an auditory transfer of L1 cue weights to L2 across prosodic categories (e.g., Kim & Tremblay, 2021, 2022). The disparity in auditory skills necessary for processing different pitch events is thought to underlie why MN listeners did not consistently show enhanced sensitivity in f0 perception. (Bidelman et al., 2011; Deroche et al., 2019; Krishnan et al., 2010). Neurobehavioral studies support this view, showing distinct neural processing pathways for tone and intonation. While f0 processing as lexical tone activates semantic areas specific to tonal language listeners (Chien et al., 2020; Friederici, 2011; Kreitewolf et al., 2014), f0 processing as intonation engages bilateral cortical areas in both tonal and non-tonal language listeners, regardless of their L1 intonation realizations (Chien et al., 2020). These insights support our findings on the perception of f0 cues, prompting future research to carefully assess the scope and conditions under which cue-weighting transfer occurs, particularly across different prosodic categories and contrast types.

Furthermore, it is important to note that SN listeners, albeit not reaching statistical significance, tended to show steeper identification functions (thus, higher f0 sensitivities) than MN listeners. This trend was consistently observed in the perception of different stimuli in Test 2 as well as in the perception of short-duration oxytone words in Test 1. Although the applicability of this trend to a larger population remains uncertain, it has led us to consider why, despite sharing similar neural networks for intonational f0 processing (Chien et al., 2020), MN listeners displayed relatively lower f0 sensitivity compared to SN listeners. In our study, we have proposed several factors might contribute to this phenomenon. The first reason for the diminished f0 sensitivity observed in MN listeners stems from their exposure to a non-native language context where they possessed less experience in processing specific f0 contours along the lines of language-specific and well-defined intonation categories, unlike SN listeners. Additionally, previous studies indicated that tonal language listeners' inclination to initially decode f0 information related to word meanings could reduce their sensitivity to f0 variations in sentence intonation (Gussenhoven & Chen, 2000; Liang & Heuven, 2007). This leads us to speculate that MN listeners' primary focus on processing L2 stress, crucial for word recognition in Spanish, may hinder their ability to effectively discern f0 cues vital for L2 intonation categories. Moreover, the divergent reliance of tonal and non-tonal language listeners on different f0 dimensions for expressing and identifying intonation categories could also contribute to varying f0 performance. In Spanish, intonation types are distinguished by the shape of the f0 contour, whereas in Mandarin, intonation is mainly conveyed by pitch range and register, especially in sentences without final particles (Wu & Ortega-Llebaria, 2017; Yuan, 2006, 2011). The multidimensional manipulation of the synthesized f0 contours in our study, encompassing f0 slope, f0 range, and f0 direction, limits our ability to pinpoint the specific dimension where the differences in f0 sensitivity primarily manifest. Therefore, further research incorporating controls for characteristic pitch parameters, such as shape, height, slope, and range, is necessary to delineate the cue-weighting distinctions in specific f0 dimensions between MN and SN listeners.

Beyond findings on f0 sensitivity, our research revealed an unexpected contrast in the way MN and SN listeners map f0 contours onto Spanish intonation categories. Compared to MN listeners, SN listeners showed a significantly higher f0 threshold for the question-statement identification boundary (x_0), consistent across all durations, intensities, and stress patterns. This implies that SN listeners required a higher f0 contour to categorize an utterance as a yes/no question than MN listeners. While Feng et al. (2019) encountered similar results, they did not provide an explanation for the disparity in f0 levels used by English L1 and Mandarin L2 speakers to perceive yes/no questions. Our study suggests this perceptual discrepancy could be influenced by the categorization norms for f0 contours in Peninsular Spanish, which presents a nuanced three-way contrast in intonational phrase (IP) final boundary tones—namely, low (L%), high (H%), and two mid-level tones (!H% for uncertainty statements and L!H% for statements of obviousness) (Estebas-Vilaplana & Prieto, 2010). Therefore, the unique mapping between f0 levels and intonation specific to SN listeners may have prompted them to preferentially employ the highest boundary tone (H%) to distinctly identify questions, as opposed to statements—whether they are broad focus statements encoded with L%, or epistemically biased statements encoded with!H% or L!H%. Conversely, in Mandarin, intonation does not rely on surface f0 movements, which may lead MN listeners to generalize any IP-final boundary tone that is not strictly low as high, thereby categorizing it as a question. Put simply, MN listeners, perhaps owing to their limited experience in differentiating the subtle f0 variations for L2 intonation categories, tended to relate both mid- and high-level f0 contours with question intonation, thereby displaying a lower question-statement identification boundary (x_0) than SN listeners.

4.2. The role of secondary cues in cross-linguistic intonation perception

Contrary to our partial assumptions for RQ1, the results demonstrated that both language groups, including Mandarin L2 learners of Spanish, significantly depended on duration cues for distinguishing statements from yes/no questions. Long-duration patterns significantly decreased the final f0 contour levels required for question identification, thereby increasing the probability of question responses compared to stimuli with short durations. Notably, this duration effect was more pronounced in the SN group, indicating a relatively higher sensitivity for SN listeners to duration variations compared to MN listeners. Similar cross-linguistic differences in duration weighting are observed in Feng et al. (2019), which showed that Mandarin learners of English were less sensitive to duration cues in intonation compared to English L1 listeners. Furthermore, unlike f0 and duration cues that reliably contributed to intonation perception in both language groups, intensity emerged as a non-significant cue in MN listeners' auditory assessment of L2 intonation. Conversely, SN listeners showed a marginally significant response to intensity variations based on the final stress pattern of the sentence. Particularly, SN listeners demonstrated steeper identification functions at 77 dB compared to 63 dB, but this trend was significant only for the perception of oxytone words. They also made greater perceptual adjustments to the final f0 contour levels in response to intensity changes compared to MN listeners, although this effect did not reach statistical significance. These observations align with the phonetic cue weighting of Spanish stress (Ortega-Llebaria et al., 2007; Ortega-Llebaria & Prieto, 2011), suggesting that intensity weighting is limited in SN listeners' intonation perception and influenced by the sentence's internal prosodic structure. As for MN listeners, the finding that intensity is an unnecessary cue in perceiving intonation seems to provide evidence for cue redundancy in speech perception (e.g., Carter, 2011; Jiao & Xu, 2019 for cue redundancy perception of stop consonants and intonation, respectively).

One possible explanation for the perceptual differences in secondary cues between SN and MN listeners is related to their ability to compensate for f0 deviations between the linear f0 contours synthesized in this study and the actual intonation patterns of Spanish statements and questions produced in natural speech, which are not entirely linear at the utterance end. We posit that SN listeners, due to their extensive exposure to the f0 distributions in Spanish intonation, are more adept at increasing the salience of other concurrent secondary cues (e.g., duration or intensity) to compensate for the loss of f0 information in resynthesized stimuli. Beyond compensatory strategies, listeners' prior history of experience about how informative these input acoustic dimensions are in their L1 categories can shape their perceptual weighting in a new language environment (Holt et al., 2018). Therefore, the reduced reliance MN listeners place on duration and intensity cues, compared to SN listeners, may stem from the overwhelming weight of f0 in Mandarin prosody (Lin, 1988; Ortega-Llebaria et al., 2017; Wang & Xu, 2011). This cue-weighting practice in their L1 may lead MN listeners to overlook other less informative acoustic cues in L2 intonation perception (Feng et al., 2019). Overall, these findings corroborate previous studies demonstrating that listeners from different languages utilize multiple acoustic cues differently when recognizing the best exemplar of a phonetic category (e.g., Feng et al., 2019; Holt & Lotto, 2006; Holt et al., 2018; Kuang & Cui, 2018; Peng et al., 2012).

4.3. The cue weight adjustment as a compensatory strategy for intonation perception

Regarding RQ2, it is generally found that the weight of acoustic dimensions in the auditory processing of intonation was not constant across different contexts. This observation aligns with previous studies (e.g., Gussenhoven & Zhou, 2013; Holt et al., 2001; Holt & Lotto, 2006; Holt et al., 2018; Kuang & Cui, 2018; Peng et al., 2012), which showed that individuals could adaptively shift their reliance on multiple cues based on a general compensation mechanism in both production and perception. Such compensatory behavior in speech is often explained using phonetic trading relations, whereby changes in one acoustic dimension can be compensated for by making opposite changes to other important dimensions of the target speech category so that the phonetic quality could be preserved (Repp, 1982). In support of this notion of perceptual compensation and our second hypothesis, we observed that MN and SN listeners significantly up-weighted the f0 cues by increasing the question-statement identification boundary (x_0) when the duration of word-final syllables was decreased to the lowest level. SN listeners, in addition, were able to flexibly adjust their cue weights to compensate for the continuous acoustic fluctuations between f0 and intensity cues. In contrast, MN listeners did not demonstrate significant compensatory adjustments in response to dynamic changes between f0 and intensity. This finding provides evidence for our prediction that SN and MN listeners perceptually compensated for acoustic variations in intonation to varying degrees.

Furthermore, when comparing the extent of cue weight adjustments between the two language groups, we found that SN listeners were more adept at compensating for acoustic–phonetic variations in Spanish intonation than MN listeners. This superior compensatory behavior in SN listeners may stem from their relatively enhanced sensitivity to modulations in the three acoustic dimensions correlated with target intonation categories, making them more responsive to the perceptual trading relations between f0, duration, and intensity cues. Empirical research supports a positive correlation between perceptual sensitivity and compensatory ability (e.g., Hodgson & Miller, 1996; Nault & Munhall, 2020; Villacorta et al., 2007), suggesting that the degree to which listeners can make auditory compensations depends on their sensitivity to the input acoustic dimensions, which should be consistently correlated with the perception of the phonetic object. These insights, along with our study' findings, may exemplify how perceptual compensatory strategies can be effectively utilized as tools for assessing listeners' auditory sensitivity to multiple acoustic cues within specific linguistic categories.

4.4. The role of stress in cross-linguistic intonation perception

In accordance with our hypothesis for RQ3, the results indicated that the perception of Spanish intonational cues was influenced by the final stress position of the sentence. Listeners, especially those in the SN group, perceived paroxytone words as questions using significantly lower f0 contours compared to oxytone words. That is, paroxytone words were more likely to be identified as yes/no questions than oxytone words at the same f0 level in sentence-final positions. This perceptual distinction can be partly attributed to the principle of least effort (Zipf, 2016), which holds that humans naturally gravitate towards the course of action that's the least cognitively demanding even though our minds are capable of processing complex patterns. Paroxytone, being the most common and unmarked stress pattern in Spanish (and certainly the most frequent and familiar stress for Mandarin learners), typically demands less effort in both production and perception (Defior & Serrano, 2017; Roca, 2019). Consequently, it is not surprising that SN and MN listeners were more inclined to categorize paroxytone words as questions rather than oxytone words when presented under identical pitch conditions.

Another factor contributing to the differences observed between both stress patterns relates to their required duration cues in intonation processing. As depicted in Fig. 1 and Fig. 2, the final vowel in the oxytone word "Alcalá" had a longer duration (184 ms) than that in the paroxytone word "Sevilla" (171 ms) in the statement context. This observation aligns with prior studies that noted longer durations of accented stressed syllables at the IP boundary compared to unaccented stressed counterparts (e.g., Navarro-Tomás, 1974; Ortega-Llebaria, 2006; Ortega-Llebaria et al., 2013). This discrepancy in duration further increased in our yes/no questions, with the final vowel length reaching 190 ms in the oxytone word and decreasing to 154 ms in the paroxytone word (see Appendix A). While direct empirical examination of this pattern is currently lacking, our data suggest that oxytone words require a more significant duration increase to encode both stress and question intonation in the sentence-final syllable effectively. A possible effect of this is that the same amount of duration changes in "Alcalá" is perceived as less salient compared to those in "Sevilla". Thus, to ensure the perceived salience of the oxytone word "Alcalá", our listeners may have increased the f0 levels of the sentence-final syllable based on the perceptual trade-offs between acoustic cues.

On the other side, when comparing the slopes of the identification curve across stress patterns, it is observed that MN listeners exhibited less sensitivity to f0 modulations in words with final stress compared to those with penultimate stress, especially in Test 1. This discrepancy, again, can be interpreted in the context of the principle of least effort in human behavior (Zipf, 2016). Specifically, we propose that the preference of MN listeners for processing word-level meanings in their L1 (Gussenhoven & Chen, 2000; Liang & Heuven, 2007) might transfer to their perception of L2 Spanish intonation, leading them to prioritize or put extra efforts on decoding f0 cues for L2 lexical stress. This negative transfer from L1 was thought to be a relevant factor for MN listeners' reduced efficiency in processing the intonational f0 cues in oxytone words. In contrast, the processing of f0 cues in paroxytone words seems to be less complicated because, in these cases, the f0 conflict at the sentence-final syllable is minimal, and the primary function of the final f0 contours is to signal intonation contrasts. However, despite positing that stress interference could be an important factor in the perceptual differences observed between the two stressed words, the exact mechanisms of how stress and intonation are encoded simultaneously through changes in the same acoustic dimensions, and how stress processing might influence the cue-weighting functions in intonation perception, are yet to be fully understood.

4.5. Limitations and future directions

This study leaves several important questions open for further investigation, stemming from both its limitations and the insights it provides. A methodological limitation constraint is the psychophysically unbalanced perceptual impact of our manipulation in the three acoustic cues: f0, duration, and intensity. The magnitude of changes applied to these dimensions did not have equivalent perceptual salience, which may restrict our ability to draw definitive conclusions about the relative weights of the three intonational cues. However, it is important to note that this limitation does not invalidate our findings regarding the perceptual differences between SN and MN listeners, as both groups were consistently exposed to the same amount of acoustic variations in intonation. This ensures that any cross-linguistic differences in perceptual cue sensitivity can still be accurately compared and analyzed. To address this limitation, future studies should adopt perceptually comparable measures when manipulating acoustic properties across multiple dimensions. This can be achieved, for instance, by establishing perceptually equivalent units based on the just noticeable difference (JND) for each acoustic parameter of intonation (see Koffi, 2019, 2020, 2021; Long, 2014, for a detailed understanding of f0, duration, and intensity JNDs).

Furthermore, our results indicating differential reliance on f0 cues for intonation processing of various stressed words are interpreted as auditory compensation for the less prominent duration changes in oxytone words compared to paroxytone words. This interpretation presupposes that in Spanish, oxytone words exhibit more pronounced duration increases in the final vowel than paroxytone words, particularly in encoding yes/no questions. While our study confirms this duration discrepancy in the two test words (in both statement and question forms) uttered by a native speaker, it remains unclear if this feature is a widespread characteristic of Spanish intonation. Consequently, future studies are warranted to investigate the acoustic properties of Spanish intonation further. Such research would enhance our understanding of the perceptual interplay between stress and intonation at the IP-final boundary of Spanish questions.

5. Conclusion

This study reveals important cross-linguistic commonalities and distinctions in the perceptual processing of acoustic details across multiple dimensions relevant for categorizing Spanish intonation. It is found that both f0 and duration cues were significant in intonation perception among MN and SN listeners, while intensity emerged as a redundant cue, exerting limited influence on SN listeners' auditory judgements. The perceptual weighting of duration and intensity differed between L1 and L2 listeners from distinct language backgrounds. Specifically, SN listeners demonstrated a heightened sensitivity to variations in secondary cues compared to MN listeners. Contrary to the cue-weighting transfer hypothesis, our findings do not support the notion of MN listeners with a tonal language background transferring their f0 perception advantage from lexical tones to the processing of f0 linear transitions in L2 intonation. Instead, tonal and non-tonal language listeners showed similar perceptual sensitivities to f0 cues processed as sentence intonation. Additionally, the variability in perceptual weighting across different acoustic conditions suggests that listeners are capable of flexibly adjusting their reliance on various cues to accommodate modulations in speech. The disparate performances of listeners in perceiving oxytone and paroxytone words also indicate that the cue-weighting functions in intonation are influenced not only by prior linguistic experience and acoustic environment but also by word-level suprasegmental constituents.

Funding

This study was supported by the Beijing Institute of Technology Research Fund Program for Young Scholars (No. 3240012222308) and the Ministry of Education University-Industry Collaborative Education Program (No. 230824085307241).

CRediT authorship contribution statement

Peizhu Shang: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Paolo Roseano:** Supervision, Writing – review & editing. **Wendy Elvira-García:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Peizhu Shang reports financial support was provided by the Beijing Institute of Technology and the Ministry of Eudcation University-Industry Collaborative Education Program.

Acknowledgments

We extend our thanks to the participants for their involvement in this study and to the editor and anonymous reviewers for their invaluable insights and feedback.

Appendix A

Supplementary materials for this article are available online at https://osf.io/ex8k7/?view_only=6c9fcf0687e5408da57f7c877ad3f774.

References

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425–434. https://doi.org/10.1162/jocn.2009.21362.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program] (Version 5.3.82). Retrieved from http://www.praat.org.

Bosch, M. M. V., & Fernández-Soriano, O. M. (2013). Variation at the interfaces in Ibero-Romance. Catalan and Spanish prosody and word order. *Catalan journal of linguistics, 12*, 253–282 http://hdl.handle.net/10486/673068.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS One, 8*(4). https://doi.org/10.1371/journal.pone.0060676 e60676.

Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg (Ed.), *Universals of Human Language Volume 2: Phonology* (pp. 471–524). Stanford University Press.

Brown, E. L., & Rivas, J. (2011). Subject-verb word order in Spanish interrogatives: A quantitative analysis of Puerto Rican Spanish. *Spanish in Context, 8*(1), 23–49. https://doi.org/10.1075/sic.8.1.02bro.

Braun, B., Galts, T., & Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research, 30*(3), 323–350. https://doi.org/10.1177/0267658313510926.

Carter, N. R. (2011). *The effect of acoustic cue redundancy on the perception of stop consonants by older and younger adults*. University of British Columbia (Doctoral dissertation).

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007a). Mismatch negativity to pitch contours is influenced by language experience. *Brain research, 1128*(1), 148–156. https://doi.org/10.1016/j.brainres.2006.10.064.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport, 18*(18), 1963–1967. https://doi.org/10.1097/wnr.0b013e3282f213c5.

Chen, Y. (2022). Mind the subtle f0 modifications: The interaction of tone and intonation in Sinitic varieties. *Stellenbosch Papers in Linguistics Plus, 62*(2), 113–136. https://doi.org/10.5842/62-2-904.

Chien, P. J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping, 41*(7), 1842–1858. https://doi.org/10.1002/hbm.24916.

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of speech, language, and hearing research, 57*(4), 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279.

Defior, S., & Serrano, F. (2017). Learning to Read Spanish. In L. Verhoeven & C. Perfetti (Eds.), *Learning to Read across Languages and Writing Systems* (pp. 243–269). Cambridge University Press.

Deroche, M. L. D., Lu, H. P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S. C., Limb, C. J., Lin, Y. S., & Chatterjee, M. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports, 9*(1), 1–12. https://doi.org/10.1038/s41598-018-36393-1.

Doherty, C. P., West, W. C., Dilley, L. C., Shattuck-Hufnagel, S., & Caplan, D. (2004). Question/statement judgments: An fMRI study of intonation processing. *Human Brain Mapping, 23*(2), 85–98. https://doi.org/10.1002/hbm.20042.

Escandell-Vidal, V. (2002). Echo-syntax and metarepresentations. *Lingua, 112*(11), 871–900. https://doi.org/10.1016/S0024-3841(02)00051-7.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 17–48). München: Lincom Europa.

Face, T. L. (2005). F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana, 3*(2), 49–65 https://dialnet.unirioja.es/servlet/articulo?codigo=1342296.

Face, T. L. (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de Fonética Experimental, 16*, 186–225.

Face, T. L. (2011). *Perception of Castilian Spanish Intonation: Implications for Intonational Phonology*. München: Lincom Europa.

Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. In A. M. Flor & E. U. Juan (Eds.), *Speech Act Performance: Theoretical, Empirical and Methodological Issues* (pp. 69–82). Amsterdam: John Benjamins.

Feng, J., Tao, S., Wu, X., Alsbury, K., & Liu, C. (2019). The effects of amplitude and duration on the perception of English statements vs questions for native English and Chinese listeners. *Journal of the Acoustical Society of America, 145*(5), EL449–EL455. https://doi.org/10.1121/1.5109046.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human perception and performance, 28*(2), 349–366. https://doi.org/10.1037//0096-1523.28.2.349.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews, 91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011.

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 22*, 765–768. https://doi.org/10.1121/1.1908022.

Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics, 11*(2), 149–175. https://doi.org/10.1016/S0095-4470(19)30813-7.

Gandour, J. T. (2009). Neural substrates underlying the perception of linguistic prosody. In C. Gussenhoven & T. Riad (Eds.), *Experimental Studies in Word and Sentence Prosody* (pp. 3–26). Berlin: De Gruyter Mouton.

Gavin, H. P. (2022). *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. Duke University: Department of Civil and Environmental Engineering.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In *6th International Conference on Spoken Language Processing* (pp. 91–94). Xi'an, China.

Gussenhoven, C., & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. In 1*4th Annual Conference of the International Speech Communication Association* (pp. 1365-1369). Lyon, France.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0.

Haverkate, H. (2006). Aspectos pragmalingüísticos de la interrogación en español con atención especial a las secuencias de preguntas. *Cultura, Lenguaje y Representación, 3*(3), 27–40 https://www.e-revistes.uji.es/index.php/clr/article/view/1317.

Heeren, W., & Heuven, V. V. (2009). Perception and production of boundary tones in whispered Dutch. In *10th Annual Conference of the International Speech Communication Association* (pp. 2411–2414). Brighton, UK. https://doi.org/10.21437/INTERSPEECH.2009-302.

Hodgson, P., & Miller, J. L. (1996). Internal structure of phonetic categories: Evidence for within-category trading relations. *Journal of the Acoustical Society of America, 100*(1), 565–576 https://psycnet.apa.org/doi/10.1121/1.415867.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119*(5), 3059–3071. https://doi.org/10.1121/1.2188377.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America, 109*(2), 764–774. https://doi.org/10.1121/1.1339825.

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research, 366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014.

Jeon, H.-S., & Heinrich, A. (2022). Perceptual asymmetry between pitch peaks and valleys. *Speech Communication, 140*, 109–127. https://doi.org/10.1016/j.specom.2022.04.001.

Jiao, L., & Xu, Y. (2019). Whispered Mandarin has no production-enhanced cues for tone and intonation. *Lingua, 218*, 24–37. https://doi.org/10.1016/j.lingua.2018.01.004.

Kim, H., & Tremblay, A. (2020). Testing the Cue-Weighting Transfer Hypothesis with Korean listeners' perception of English lexical stress. *Journal of the Acoustical Society of America, 148*(4), 2812. https://doi.org/10.1121/1.5147839.

Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. Journal of. *Phonetics, 87*, 101059. https://doi.org/10.1016/j.wocn.2021.101059.

Kim, H., & Tremblay, A. (2022). Intonational cues to segmental contrasts in the native language facilitate the processing of intonational cues to lexical stress in the second language. *Frontiers in Communication, 7*, 845430. https://doi.org/10.3389/fcomm.2022.845430.

Koffi, E. (2019). A comprehensive review of f0 and its various correlations. *Linguistic Portfolios, 8*(1), 2 https://repository.stcloudstate.edu/stcloud_ling/vol8/iss1/2.

Koffi, E. (2020). A comprehensive review of intensity and its linguistic applications. *Linguistic Portfolios, 9*(1), 2 https://repository.stcloudstate.edu/stcloud_ling/vol9/iss1/2.

Koffi, E. (2021). A comprehensive review of the acoustic correlate of duration and its linguistic implications. *Linguistic Portfolios*, *10*(1), 2. https://repository.stcloudstate.edu/stcloud_ling/ vol10/iss1/2.

Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage, 102*, 332–344. https://doi.org/10.1016/j.neuroimage.2014.07.038.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics, 23*(1), 81–95. https://doi.org/10.1016/j.jneuroling.2009.09.001.

Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia, 53*, 293–309. https://doi.org/10.1016/j.neuropsychologia.2013.11.020.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics, 71*, 194–214. https://doi.org/10.1016/j.wocn.2018.09.002.

Lenth, R., Singmann, H., & Love, J. (2019). Emmeans: Estimated marginal means, aka least-squares means. *R package version, 1*(3), 4 https://cran.r-project.org/web/packages/emmeans.

Liang, J., & Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.). *Experimental Studies in Word and Sentence Prosody* (Volume 2, pp. 27–62). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110207576.1.27.

Lin, M. (1988). Putonghua shengdiao de shengxue texing he zhijue zhengzhao [The acoustic characteristics and perceptual cues of tones in Standard Chinese]. *Zhongguo Yuwen [Chinese Linguistics], 204*(3), 182–193.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29*(1), 3–11. https://doi.org/10.1177/002383098602900102.

Liu, M. (2018). *Tone and intonation processing: From ambiguous acoustic signal to linguistic representation (Doctoral dissertation)*. Leiden University.

Liu, M., Chen, Y., & Schiller, N. O. (2022). Context Matters for Tone and Intonation Processing in Mandarin. *Language and Speech, 65*(1), 52–72. https://doi.org/10.1177/0023830920986174.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech, 47*(2), 109–138. https://doi.org/10.1177/00238309040470020101.

Long, M. (2014). Human perception and reaction to sound. In *Architectural acoustics* (pp. 81–127). New York: Academic press. https://doi.org/10.1016/B978-0-12-398258-2.00003-9.

Ma, J. K. Y., Ciocca, V., & Whitehill, T. L. (2008). Acoustic cues for the perception of intonation in Cantonese. In *9th Annual Conference of the International Speech Communication Association* (pp. 520–523). https://www.isca-speech.org/archive_v0/archive_papers/ interspeech_2008/i08_0520.pdf.

Ma, J.-K.-Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *Journal of the Acoustical Society of America, 129*(2), 1012–1023. https://doi.org/10.1121/1.3531840.

Moré, J. J. (1978). The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical analysis* (pp. 105–116). Berlin, Heidelberg: Springer.

Morrow, K., & Liu, C. (2013). Intonation perception in English: Effects of stimulus amplitude and listeners′ language background. *Journal of Acoustical Society of America, 133*(5). https://doi.org/10.1121/1.4800156.

Nagle, C. (2013). A reexamination of ultimate attainment in L2 phonology: Length of immersion, motivation, and phonological short-term memory. In *Selected Proceedings of the Second Language Research Forum* (pp. 148–161).

Nault, D. R., & Munhall, K. G. (2020). Individual variability in auditory feedback processing: Responses to real-time formant perturbations capacity and their relation to perceptual acuity. *Journal of the Acoustical Society of America, 148*(6), 3709–3721. https://doi.org/10.1121/10.0002923.

Navarro-Tomás, T. (1974). *Manual de entonación española* (Vol. 175) Madrid: Ediciones Guadarrama.

Niebuhr, O. (2007). *Categorical perception in intonation: A matter of signal dynamics? In 8th Annual Conference of the International Speech Communication Association* (pp. 642–645). Antwerp: Belgium.

Ortega-Llebaria, M. (2006). Phonetic cues to stress and accent in Spanish. In *Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 104–118). Bloomington, USA.

Ortega-Llebaria, M., Prieto, P., & Vanrell, M. D. M. (2007). In *Perceptual evidence for direct acoustic correlates of stress in Spanish* (pp. 1121–1124). Germany: Saarbrücken.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and speech, 54*(1), 73–97. https://doi.org/10.1177/0023830910388014.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics, 41*(3–4), 186–197. https://doi.org/10.1016/j.wocn.2013.01.006.

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive "Rose?" vs. *"Rose". Bilingualism, 20*(2), 367–383. https://doi.org/10.1017/S1366728915000723.

Ortega-Llebaria, M., Olson, D. J., & Tuninetti, A. (2019). Explaining cross-language asymmetries in prosodic processing: The cue-driven window length hypothesis. *Language and Speech, 62*(4), 701–736. https://doi.org/10.1177/0023830918808823.

Peng, S. C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification, 16*(2), 67–82. https://doi.org/10.1177/1084713812451159.

Qin, Z., Chien, Y. F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics, 38*(3), 541–570. https://doi.org/10.1017/S0142716416000321.

R Core Team (2021). *R: A language and environment for statistical computing [Computer program]*. Vienna, Austria: Foundation for Statistical Computing.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*(1), 81–110 https://psycnet.apa.org/doi/10.1037/0033-2909.92.1.81.

Roca, I. (2019). Spanish Word Stress: An updated multidimensional account. In R. Goedemans, J. Heinz, & H. van der Hulst (Eds.), *The Study of Word Stress and Accent: Theories, Methods and Data* (pp. 256–292). Cambridge University Press.

Romera Barrios, L., Fernández-Planas, A. M., Salcioli Guidi, V., Carrera Sabaté, J., & Montes de Oca, D. R. (2007). Una muestra del español de Barcelona en el marco AMPER. *Estudios de Fonética Experimental, 16*(16), 147–184.

Salamanca, F. J. Z., de la Red, M. C., & Matías, M. del M. M. (2005). Variación geoprosódica en dos entonaciones de Castilla y León: análisis de frases declarativas e interrogativas sin expansión. *Estudios de Fonética Experimental*, 126–139.

Shang, P., Elvira-García, W., & Roseano, P. (2021). La modalidad interrogativa en español y en chino: un enfoque funcionalista. *Círculo de Lingüística Aplicada a la Comunicación*, 2021, *88*, 235-254. doi: 10.5209/clac.78313.

Shang, P., Elvira-García, W., & Li, X. (2022). Cue weighting differences in perception of Spanish sentence types between native listeners of Chinese and Spanish. In *proceedings of the 11th International Conference on Speech Prosody* (pp. 644-648). Lisbon, Portugal.

Shively, R. L. (2008). L2 Acquisition of [β], [δ], and [γ] in Spanish: Impact of Experience, Linguistic Environment and Learner Variables. *Southwest Journal of Linguistics, 27* (2), 79–114.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53* (2), 273–293. https://doi.org/10.1177/0023830909357156.

Syntrillium Software Corporation. (2003). Cool Edit Pro. Version 2.1. Phoenix.

Tillman, G., Benders, T., Brown, S. D., & van Ravenzwaaij, D. (2017). An evidence accumulation model of acoustic cue weighting in vowel perception. *Journal of Phonetics, 61*, 1–12. https://doi.org/10.1016/j.wocn.2016.12.001.

The Mathworks (2022). MATLAB and statistics toolbox release [Computer program]. Natick: The MathWorks Inc. Available from https://es.mathworks.com/products/matlab.html.

Tremblay, A., Broersma, M., & Coughlin, C. E. (2018). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition, 21*(3), 640–652. https://doi.org/10.1017/S136672891700030X.

Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021). Dutch listeners' perception of English lexical stress: A cue-weighting approach. *Journal of the Acoustical Society of America, 149*(6), 3703–3714. https://doi.org/10.1121/10.0005086.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1–30. https://doi.org/10.1017/S0272263106060013.

Tsukada, K. (2018). The perception of Mandarin lexical tones by native speakers of Burmese. *Language and Speech, 62*(4), 1–16. https://doi.org/10.1177/0023830918806550.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America, 122*(4), 2306–2319. https://doi.org/10.1121/1.2773966.

Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics, 39*(4), 595–611. https://doi.org/10.1016/j.wocn.2011.03.006.

Wang, M., Zi, G., Xiong, W., & Lin, M. (2013). Pitch declination of Chinese spontaneous speech. *Journal of Chinese Information Processing, 27*(6), 128–133.

Wiener, S. (2017). Changes in early l2 cue-weighting of non-native speech: Evidence from learners of Mandarin Chinese lexical tone. In 1*8th Annual Conference of the International Speech Communication Association* (pp. 1765-1769). Stockholm, Sweden.

Wu, Z., & Ortega-Llebaria, M. (2017). Pitch shape modulates the time course of tone vs pitch-accent identification in Mandarin Chinese. *Journal of the Acoustical Society of America, 141*(3), 2263–2276. https://doi.org/10.1121/1.4979052.

Xu, Y. (2004). Transmitting Tone and Intonation Simultaneously-The Parallel Encoding and Target Approximation (PENTA) Model. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages* (pp. 215–220). Beijing, China.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America, 120*(2), 1063–1074. https://doi.org/10.1121/1.2213572.

Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling*. Cornell University (Doctoral dissertation).

Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In *Proceedings of the 5th International Symposium on Chinese Spoken Language Processing* (pp. 19-30). Singapore. https://doi.org/10.1007/11939993_7.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of the Acoustical Society of America, 130*(6), 4063–4069. https://doi.org/10.1121/1.3651818.

Yuan, J., & Shih, C. (2004). Confusability of Chinese intonation. In *Proceedings of the 2nd International Conference on Speech Prosody* (pp. 131–134). Nara, Japan.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161.

Zhang, C. (2018). *Tones and Tunes in Tianjin Mandarin*. University of Oxford (Doctoral dissertation).

Zhang, G., Shao, J., Zhang, C., & Wang, L. (2022). The perception of lexical tone and intonation in whispered speech by Mandarin-speaking congenital amusics. *Journal of Speech, Language, and Hearing Research, 65*(4), 1331–1348. https://doi.org/10.1044/2021_jslhr-21-00345.

Zhang, H., Wiener, S., & Holt, L. L. (2022). Adjustment of cue weighting in speech by speakers and listeners: Evidence from amplitude and duration modifications of Mandarin Chinese tone. *Journal of the Acoustical Society of America, 151*(2), 992–1005. https://doi.org/10.1121/10.0009378.

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38*(2), 260–271. https://doi.org/10.1016/j.wocn.2009.11.002.

Zipf, G. K. (2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books.

Zubizarreta, M. L. (1999). Word order in Spanish and the nature of nominative case. In K. Johnson & I. Roberts (Eds.), *Beyond Principles and Parameters* (pp. 223–250). Dordrecht: Springer.