Observing Pitch Gestures Favors the Learning of Spanish Intonation by Mandarin Speakers

Chenjie Yuan, *Universitat Pompeu Fabra*
Santiago González-Fuente, *Universitat Pompeu Fabra*
Florence Baills, *Universitat Pompeu Fabra*
Pilar Prieto, *Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Pompeu Fabra*

Abstract

Recent studies on the learning of L2 prosody have suggested that pitch gestures can enhance the learning of the L2 lexical tones. Yet it remains unclear whether the use of these gestures can aid the learning of L2 intonation, especially by tonal-language speakers. Sixty-four Mandarin speakers with basic-level Spanish were asked to learn three Spanish intonation patterns, all involving a low tone on the nuclear accent. In a pre-post test experimental design, half of the participants received intonation training without the use of pitch gestures (the control group) while the other half received the same training but with pitch gestures representing nuclear intonation contours (the experimental group). Musical (melody, pitch) abilities were also measured. The results revealed that (a) the experimental group significantly improved intonational production outcomes, and (b) even though participants with stronger musical abilities performed better, those with weaker musical abilities benefited more from observing pitch gestures.

Introduction [1]

It is widely accepted that learning the pronunciation of a second language is especially difficult for L2 learners in their adulthood, even for those who have achieved nativelike performance in other linguistic aspects, such as vocabulary and morphosyntax (e.g., Bongaerts, Van Summeren, Planken, & Schils, 1997; Mennen & de Leeuw, 2014; Morales, 2008). In part, this is because successful L2 pronunciation involves the ability to reproduce not only individual phonemes but also the unique prosodic patterns of the target language. Importantly, recent research on the acquisition of L2 pronunciation has shown that segmental errors tend to be less predictive of the listeners' comprehension of L2 speech than prosodic errors (e.g., Anderson-Hsieh, Johnson, & Koehleret, 1992; Munro & Derwing, 1995; Rasier & Hiligsmann, 2007; Trofimovich & Baker, 2006).

Over the past decades, studies within the embodied cognition paradigm have shown that cospeech gestures are tightly integrated with speech (e.g., Bernardis & Gentilucci, 2006; Goldin-Meadow, 2003; Kendon, 1980, 2004; Levinson & Holler, 2014; McNeill, 1992). Gullberg

*Correspondence concerning this article should be addressed to Chenjie Yuan, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. Email: chenjie.yuan@upf.edu

[1] This research has been funded by two research grants awarded by the Ministry of Science and Innovation (FFI2015-66533-P) and the Generalitat de Catalunya (2014 SGR-925), both to the Prosodic Studies Group. The second author also acknowledges a FPU 2012-05893 grant awarded by the Spanish Ministry of Science and Innovation. Our gratitude goes to Ms. Huan Zhang, Mr. Wei Cao, and Ms. Nan Huang, who provided us with multimodal experimental classrooms and helped us to collect the oral data at Xi'an International Studies University in Xi'an, China. We are indebted to Joan Carles Mora, who read a first version of the manuscript and offered us many insightful comments and suggestions. We would also like to thank Joan Borràs-Comes, who helped us with the statistics. Special thanks also go to two anonymous reviewers and the editors, Susan Gass and Bill VanPatten, for invaluable comments and feedback.
The experiment in this article earned an Open Materials badge for transparent practices. The materials are available at https://www.iris-database.org/iris/app/home/detail?id=york:932719

(2006) suggested an important connection between gestures and L2 acquisition, underlining that gestures may provide learners with additional cues to aid comprehension and overall acquisition. However, little is known about the potential effects of the use of cospeech gestures in the domain of L2 pronunciation, and the few studies that have addressed this question have led to contradictory results. First, a recent study by Gluhareva and Prieto (2017) investigated the effects of observing beat gestures (a type of gesture characterized as a rhythmic up-and-down movement of the hand/finger that is also associated with prosodic prominence in speech; see McNeill, 1992) on the acquisition of nativelike speech patterns in English. They found that training involving the observation of beat gestures significantly improved the participants' accentedness ratings. Similarly, Hirata, Kelly, Huang, and Manansala (2014) found that participants trained with beat gestures (in this case, mimicking the syllabic-rhythm patterns in speech) did show an auditory improvement in identifying and reproducing L2 vowel length contrasts in Japanese, but this positive effect was rather limited. By contrast, the results reported earlier by Hirata and Kelly (2010) had revealed that participants trained by means of gestures did not perceive the vowel length contrasts any better than those in the no-gesture condition. Overall, nonetheless, the results obtained by Gluhareva and Prieto (2017) and Hirata et al. (2014) seem to suggest that beat gestures are useful for improving L2 pronunciation abilities. In both experiments, beats could be considered visuospatial metaphors that visually reflect the auditory rhythm of a second language and thus make it easier for L2 learners to grasp this abstract prosodic property. As postulated by McCafferty (2006), the beneficial use of hand gestures for prosodic L2 learning is because they act as a type of metaphorical representation that can help to establish a physicalized sense of the prosodic features

of the L2. The idea of viewing gestures as a type of spatial metaphor dates to Lakoff (1993), who considered spatial metaphors a special case of image metaphor in which the knowledge from the visuospatial domain is mapped into the perceptual domain. A more concrete proposal on the relationship between hand gestures and the acquisition of L2 prosody is found in McCafferty (2006), who proposed that the mental plane is built upon activity in the material plane and that the action representation of acoustic features by hand gestures help to consolidate a conceptual foundation of the prosodic modalities in the mental plane.

The beneficial results of beat gestures on rhythmic learning lead us to question whether the learning of other prosodic features, such as lexical tones and intonational patterns, might also benefit from the use of matched representative hand gestures, in other words, pitch gestures. According to McNeill (1992) and Morett and Chang (2015), pitch gestures are a type of visuospatial gesture in which upward and downward hand movements mimic high-frequency pitch and low-frequency pitch movements, respectively. Recent experimental evidence has also shown that pitch and space have a shared representation such that the mental representation of pitch is audiovisual in nature. For example, Connell, Cai, and Holler (2013) asked participants to judge whether a target vocal note accompanied by a hand gesture in a video clip was the same as or different from a preceding note. The video showed either an upward or a downward hand gesture, independently of the direction in the change in pitch between the two notes. The results showed that pitch discrimination was significantly biased by the direction of the gesture, such that downward gestures tended to induce a perception that a fall in pitch had occurred and upward gestures tended to induce the opposite effect. More recently, in an fMRI-based experiment, Dolscheid, Willems, Hagoort, and Casasanto (2014)

investigated different types of spatial representations that may be linked to musical pitch. Participants were asked to judge stimuli that varied in spatial height in both the visual and tactile modalities, as well as auditory stimuli that varied in pitch height. The pitch activations were examined with respect to whether they were presented in the modality-specific region or the multimodal region of the brain. The results showed that the judgments of musical pitch would activate unimodal visual areas, confirming the hypothesis that the perception of musical pitch would involve spatial representations. All these findings suggest that the strong cognitive links between the perception of pitch and corresponding visuospatial gestures can have an important application in the learning of prosody and melody in a second language.

The role of pitch gestures in the acquisition of L2 prosody has been empirically investigated in a handful of studies on learning Mandarin lexical tones. In a classroom-based longitudinal experiment, Jia and Wang (2013a) tested 31 native English-speaking CSL (Chinese as a Second Language) learners'perception of the four Mandarin lexical tones before and after a three-week training session. Sixteen of them were taught the target items with speech and cospeech pitch gestures (i.e., the gesture condition) while the other 15 were taught the same items but with speech only (i.e., the no-gesture condition). The results showed that training with gestures had a superior effect on tone perception than training without gestures. In a similar experiment, but this time focused on production, Jia and Wang (2013b) tested 28 English-speaking CSL learners' performance in producing Mandarin lexical tones after a three-week training session. Fifteen of them were trained with speech and cospeech pitch gestures whereas the remaining 13 participants were taught the same items with speech only. Here too, the results showed the effectiveness of using cospeech gestures to facilitate the production

of Mandarin lexical tones by elementary-level learners. Similarly, Morett and Chang (2015) showed that the use of pitch gestures can help the memorization of target Mandarin words with lexical tones. In their experiment, 57 English-speaking CSL students were asked to identify the correct tones as well as the corresponding meaning of the target words both before and after a training phase. Three blocks of Mandarin words were taught respectively with (a) pitch gestures, (b) semantic/iconic gestures, and (c) no gestures in the training session. The results showed that observing pitch gestures representing tones while learning helped English-speaking CSL learners to distinguish between target Mandarin words differing exclusively in lexical tones; conversely, iconic gestures were proved to negatively influence learners' performance in word identification.

Relatedly, Baills, Suárez, González-Fuente, and Prieto (ms.) ran two between-subjects experiments to test the effects of observing and producing pitch gestures on the learning of Mandarin lexical tones and words by speakers of Catalan. In Experiment 1, the control group of 24 participants were taught words by speech only, that is, without being either shown pitch gestures to represent tones or taught to reproduce them. In the same training phase, another group of 25 was taught the words by speech and simultaneously observing pitch gestures. In Experiment 2, the same items were learned either by observing the gestures and repeating the words (28 participants, control group) or by repeating the words and additionally mimicking the gestures performed by the instructor (28 participants). The results revealed that participants in the two experimental groups (observing gestures or mimicking gestures) performed significantly better at (a) perceiving Mandarin tones and (b) discriminating between the meanings of Mandarin words differing only in tone than the two corresponding

control groups (speech only or observing gestures but not mimicking them). The authors finally compared the results from both observing gesture and producing gesture conditions and found no significant differences between the two experimental conditions. These results reported by Baills et al. (ms.), on the one hand, are consistent with Jia and Wang (2013a) and Morett and Chang (2015) in showing that pitch gestures do play a positive role in the learning of L2 lexical tones; on the other hand, they seem to suggest that the benefit of either observing or producing gestures for the learning of tonal contrasts in Chinese is equivalent.

While the previously mentioned studies have demonstrated the positive role of pitch gestures in the learning of lexical tones, little is known about whether using pitch gestures could help learners to acquire the intonational properties of a second language faster. Given that intonational patterns vary as a function of the same physical property of sound as lexical tones, namely pitch, then if pitch gestures can help speakers of nontonal languages learn a tonal language, they might also be useful for speakers of tonal languages to learn the intonation of a nontonal language. Many studies over the past decades have shown that speakers of intonational languages face difficulties in learning intonational patterns of an L2 because they tend to transfer the intonational patterns of their L1 to the L2, not only in perception (e.g., Flege, 1991; He, van Heuven, & Gussenhoven, 2011, 2012) but also in production (e.g., Braun & Tagliapietra, 2010; Cruz-Ferreira, 1989; Ortega-Llebaria & Colantoni, 2014; Ortega-Llebaria, Nemogá, & Presson, 2015; see Mennen, 2015, for a review). Moreover, speakers of tonal language like Mandarin have been reported to have more systematic difficulty in learning L2 intonation than speakers of intonational languages. Several studies have revealed that Mandarin L1

speakers encounter systematic problems in not only perceiving the target L2 intonation patterns (e.g., Cortés Moreno, 1997, 2001, 2004; Ji, 2010) but also reproducing them (Cortés Moreno, 1997, 2004). Cortés Moreno (1997, 2004) investigated the perception and production patterns of L2 Spanish intonation by Mandarin-speaking Taiwanese students. In fact, a comparison of the results of these two studies reveals that Taiwanese students had considerably more difficulty producing than perceiving the L2 intonation. Furthermore, the two studies suggest that the different intonational patterns of statements, questions, and exclamations implied different degrees of difficulty for Mandarin-speaking students, whether in perception or production. Along the same lines, through a well-designed experiment, Ji (2010) tested the perception and production of L2 English intonation by Mandarin-speaking Chinese students and found that production caused more systematic problems than perception. More specifically, although some advanced learners of L2 showed performance equivalent to native speakers in perceiving the L2 intonation, they were still observed to have problems in producing the L2 intonation patterns accurately. Given these findings, it is worth reviewing now the research on the specific problems that Mandarin speakers face when producing intonational languages.

One of the most salient errors in English intonation produced by Mandarin learners is the tendency to systematically produce a high tone (henceforth H*) on the target nuclear accent when it should be a low tone (henceforth L*) instead. This tendency was first reported by Ji, Wang, and Li (2009) and later explored in Xu (2009), Ji (2010), Hong (2012), and Barto (2015). Such a phenomenon has also been reported for Mandarin speakers learning other intonational languages like Spanish, as reported by Liu (2003). An example of this is illustrated in Figure 1, which compares the waveform, spectrogram, and pitch track produced

by a Spanish native speaker uttering the non-biased information-seeking yes-no question in (1) (left panel) with those produced by a Mandarin-speaking basic-level learner of Spanish (right panel). As can be seen, while the Spaniard produces the L* H% nuclear configuration (Figure 1, left panel) that is typical of Spanish native speakers (Hualde & Prieto, 2015), the basic-level Mandarin-speaking learner of Spanish systematically produces a H* on the nuclear accent (Figure 1, right panel).

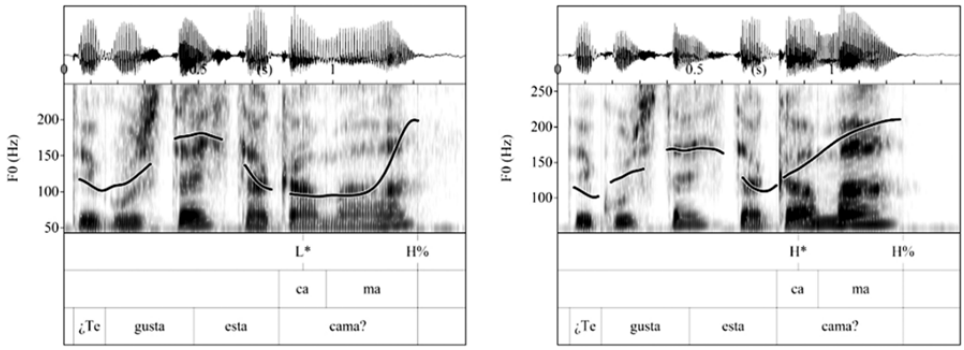

FIGURE 1. Waveform, spectrogram, and pitch track of the target interrogative sentence ¿Te gusta esta cama? "Do you like this bed?" as produced by a native Spanish speaker (left panel) and by a Mandarin-speaking beginner learner of Spanish (right panel).

(1) ¿Te gusta esta cama?
you-DAT like-3-SG this-DEM bed-FEM
Do you like this bed?

The production of the H* by the Mandarin speaker instead of the L* tone is most probably the result of transfer from the L1 (Mandarin) to the L2 (Spanish). Unlike intonational language speakers, Mandarin speakers rely less on intonation because in their own language the lexical

tone constitutes the minimal prosodic unit. Due to the lack of a steady L* tone in their lexical tone inventory, Mandarin speakers tend to systematically treat lexical stress in intonational languages as a H* tone, which results in the use of H* in all nuclear configurations (see Barto, 2015; Cortés Moreno, 2004; Ji, 2010).

In the light of all these findings, the main goal of the present investigation is to test whether observing pitch gestures can help Mandarin speakers to learn L2 Spanish intonation and specifically to reverse their tendency to produce a high tone in nuclear accents instead of a low tone. To our knowledge, there is no empirical research thus far that explores the relationship between manual pitch gestures and the acquisition of L2 intonation patterns. However, there have been previous studies on the potential utility of using electronic visualizers [2] in the teaching of L2 intonation, research that is relevant to our study because pitch gestures and visualizers both depict pitch movements visually. In general, using visualizers has been proved to enhance the learning of L2 intonation. For example, both Taniguchi and Abberton (1999) and Shimizu and Taniguchi (2005) tested Japanese-speaking EFL (English as a Foreign Language) learners' perception and production of English intonation both before and after a longitudinal training session with and without visualizers (a Laryngograph Processor in the former study and Speech Filing System in the latter). The results showed that students trained with visualizers performed significantly better in the posttest in both perception and production tasks. By contrast, when Ostrom (1997) tested Thai-speaking EFL learners' intonation production before and

[2] Visualizers are a type of software or electronic device that can generate animated imagery of changes in the loudness and fundamental frequency (i.e., pitch) of an acoustic signal in real time. Although such devices were originally employed in music training, in the last decade of the twentieth century linguists and language instructors started to explore their potential use in the teaching of prosody.

after longitudinal training with and without the use of a visualizer (VisiPitch), the results showed no significant difference between the control group in which target items had been taught with traditional methods, using speech and pitch gestures, and the experimental group, which had received interactive visual feedback provided by the visualizer. But note that Ostrom (1997) used pitch gestures in the control group, which might have had a positive effect on their learning of the L2 intonation. We may thus infer from Ostrom's (1997) results that both pitch gestures and visual feedback (provided by visualizers) favor the learning of L2 intonation. In the present study, following up on Jia and Wang (2013a, 2013b), Morett and Chang (2015), and Baills et al. (ms.), we hypothesized that having Mandarin learners of Spanish observe cospeech gestures will benefit their production of target intonation patterns involving a low tone and improve their overall learning of L2 intonation. To test this hypothesis, a between-subjects training experiment with a pre-post design was run with 64 Chinese basic-level learners of Spanish, who were randomly assigned to a no-gesture control group and a gesture experimental group.

Additionally, we hypothesized that the musical ability of participants would positively correlate with their ability to acquire L2 intonation regardless of whether they were exposed to gestures while learning. Neuroscience research studies have shown that language (and especially its prosody) shows interesting parallels with music (see, among many others, Brown, Martinez, & Parsons, 2006; Heffner & Slevc, 2015; Ludke, Ferreira, & Overy, 2014; McMullen & Saffran, 2004; Slevc, 2012; Slevc & Miyake, 2006). Patel (2011, 2014) summarized the empirical findings in former research and proposed the so-called OPERA (Overlap-Precision-Emotion-Repetition-Attention) hypothesis. According to this hypothesis, though there is an anatomical overlap in

the brain networks that are used to process acoustic features in both music and speech, music places higher demands on these shared networks than does speech in terms of the precision of processing, and thus it is expected that individuals with higher musical expertise will be more precise in acoustic and phonological perception. Several studies have proven that musical expertise tends to facilitate language learning in general (see Chobert & Besson, 2013, for a review) and favor the perception of prosodic patterns (e.g., among many others, Marques, Moreno, Castro, & Besson, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007). Much research has been done to explore the effects of musical ability on the perception of lexical tones by intonational language speakers (e.g., Alexander, Wong, & Bradlow, 2005; Bidelman, Gandour, & Krishnan, 2011; Chandrasekaran, Krishnan, & Gandour, 2009; Lee & Huang, 2008; Wong et al., 2007). Recently, for example, Zhao and Kuhl (2015) investigated whether prior music training influences the perceptual learning of lexical tones. In their experiment, three groups of participants (20 monolingual English-speaking musicians, 20 monolingual English-speaking nonmusicians, and 20 native Mandarin-speaking nonmusicians) were instructed to first complete an AX discrimination task in which they had to judge whether a sound X was identical to a sound A or not. This was followed by an AXB identification task in which they were asked to judge whether a sound X was more similar to A or B. The results showed that English-speaking musicians performed significantly better than nonmusicians in both identifying Mandarin lexical tones and distinguishing between them. Marques et al. (2007) tested how well adult French-speaking musicians (i.e., who had at least 14 years of prior musical training) and nonmusicians (i.e., who had no training in music) perceived pitch variation in sentence-final words in Portuguese. Participants were

presented with both congruous (i.e., spoken at normal pitch height) and incongruous (i.e., spoken respectively at weakly increased [35%] or sharply increased [120%] pitch heights) pronunciations of sentence-final words. Results revealed that musicians were statistically better at perceiving pitch deviations, especially with those coming from weak prosodic incongruities. However, as pointed out by Fonseca-Mora, Jara-Jiménez, and Gómez-Domínguez (2015), much remains to be investigated about the connection between musical ability and L2 acquisition. Specifically, as far as we know, no empirical research has been conducted to investigate the potential relationship between musical ability and the use of pitch gestures in the process of learning L2 prosody. Drawing upon the positive findings of Zhao and Kuhl (2015) and Marques et al. (2007), among others, our hypothesis would be that students with higher musical abilities have an advantage when it comes to learning L2 intonation. In our case, Mandarin learners with a high musical ability level are predicted to perform better at mimicking intonation patterns correctly, and we hypothesize that this ability will benefit their learning of Spanish intonation. As for the relation between musical ability and the use of cospeech gestures, because it has not been explored before, no specific hypothesis will be put forth at this point.

Thus, to sum up, the aim of the present study is to investigate (a) whether observing pitch gestures can benefit the learning of target Spanish intonation patterns by Mandarin learners and improve their performance in intonation production; and (b) whether the learners' musical ability will positively correlate with their learning of L2 intonation and interact with the learning procedure (i.e., with or without gestures). To test these hypotheses, we first tested all participants for their musical abilities and then carried out a between-subjects study in which we analyzed the participants' production of target intonation

patterns both before and after a short training session. While half of the participants were presented with the audiovisual training stimuli that did not show pitch gestures (no-gesture condition), the other half were presented with the same audiovisual training materials but this time accompanied by pitch gestures (gesture condition).

METHOD

PARTICIPANTS

Eighty-nine students at Xi'an International Studies University in Xi'an, China, volunteered to participate in our experiment. Inquiries about their language background revealed that 64 of these students (M_{age} = 19.797; SD = 1.299; age range 18–23) used Mandarin habitually in their daily lives ($M_{language\ use\ per\ day}$ = 89.266%; SD = 8.105) and were therefore selected to be participants in our study. The mother tongues of the remaining 25 proved to be other Chinese languages, and these students were excluded from the study to control for L1. The final group of 64 participants were all beginning (A1 level) ELE (*Español como Lengua Extranjera*, "Spanish as a Foreign Language") learners who had only attended the first 8 to 10 weeks (M_{time} = 9.125 weeks; SD = 0.740; time range 8–10 weeks) of a first-year Spanish phonetics course offered by their university [3] and had no significant contact with native Spanish speakers. They signed a written consent giving permission to process their recorded data.

[3] In accordance with the *Programa de enseñanza para cursos básicos de las especialidades de lengua española de escuelas superiores chinas* (Comisión orientada de la enseñanza de lenguas extranjeras en las universidades, sección de español, 1998), all undergraduates undertaking a degree in Spanish language at Chinese universities receive a course in the phonetics of Spanish which lasts between 8 and 12 weeks. This regulation helped us recruit qualified participants for our experiment.

MATERIALS

The L* pitch accent constitutes the testing target in our present research. Accordingly, we selected three non-biased nuclear pitch configurations in Castilian Spanish that were also employed in the experiments of Cortés Moreno (1997, 2001, 2004). The three patterns contain a L* on the nuclear accent, specifically, a low-fall (or L* L%) in statements, a low-rise (or L* H%) in yes-no questions, and a low-rise-fall (or L* HL%) in requests. A schematic description of the target nuclear configurations is provided in Figure 2 (see also Hualde & Prieto, 2015, for the Sp_ToBI transcription of these pitch contours).

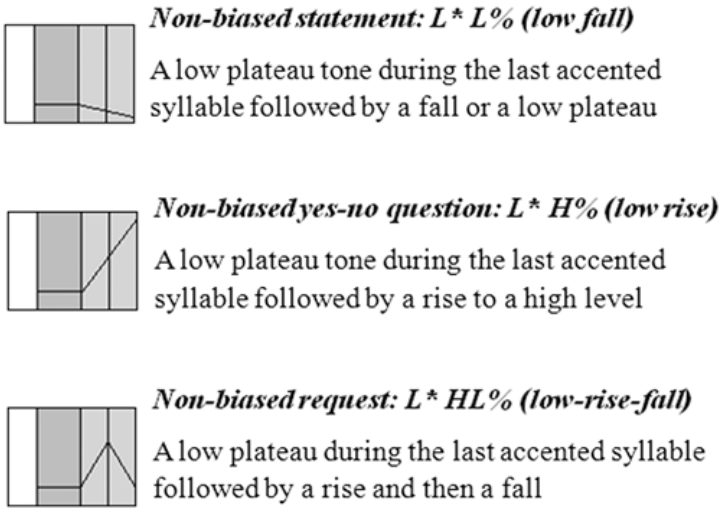

FIGURE 2. Schematic description of the three target nuclear pitch patterns containing a L* pitch accent.

Five words were chosen as target nuclear words. Because Su and Hu (2011) and Gao and Qin (2012) showed that Mandarin learners' production of L2 intonation is strongly affected by the position of lexical stress as well as the number of syllables in the nuclear word, we decided to use only disyllabic CVCV nouns with stress on the first syllable,

namely *casa* "house/home," *cama* "bed," *coche* "car," *silla* "chair," and *leche* "milk." Also, all target Spanish words (as well as target phrases) conveyed common everyday meanings and were easy to understand. They were chosen from the first ten lessons of the textbook *Español Moderno, Libro del Alumno I* (*Modern Spanish, Student's Book I*; Dong & Liu, 2014). [4]

As illustrated in Table 1, a total of 15 target sentences were prepared by combining the five target words (located in the nuclear/final position in the sentence) with three target intonation patterns (statement, yes-no question, and request).

TABLE 1. Target intonation contours in the training stimuli

[4] *Español Moderno* is a series of Spanish language textbooks (six books for students and six companion books for both students and teachers) that has been adopted for language instruction by almost all the departments of Spanish in Chinese universities. The beginning-level book *Español Moderno*, *Libro del Alumno I* (*Modern Spanish, Student's Book I*) consists of 8 lessons of basic phonetic training and 12 lessons of basic grammar and vocabulary and is typically used during the first term of the first year.

Target Words	Target Intonation Patterns	Target Sentences
Casa "home"	statement	*Ahora no estoy en casa.* "Now I am not at home."
	yes-no question	*¿Tus padres están en casa?* "Are your parents at home?"
	request	*¡Vente a mi casa!* "(Please,) come to my home!"
Cama "bed"	statement	*La puse en tu cama.* "I put it on your bed."
	yes-no question	*¿Te gusta esta cama?* "Do you like this bed?"
	request	*¡Vamos a la cama!* "(Please,) let's go to bed!"
Coche "car"	statement	*No conozco la marca de este coche.* "I don't know the make of this car."
	yes-no question	*¿Conoces la marca de este coche?* "Do you know the make of this car?"
	request	*¡Préstame tu coche!* "(Please,) lend me your car!"
Silla "chair"	statement	*Yo prefiero la silla.* "I prefer the chair."
	yes-no question	*¿Es muy cara esta silla?* "Is this chair very expensive?"
	request	*¡Cómprame la silla!* "(Please,) buy me the chair!"
Leche "milk"	statement	*Todas las mañanas desayuno un vaso de leche.* "Every morning I have a glass of milk."
	yes-no question	*¿Te has acabado la leche?* "Have you finished the milk?"
	request	*¡Prueba esta leche!* "(Please,) Try this milk!"

To create the audiovisual stimuli for the experiment, two native Castilian Spanish speakers, one female and one male, from central-northern Spain were video-recorded as they uttered the 15 target sentences in Table 1 in two conditions, the no-gesture condition and the gesture condition. This yielded a total of 60 recorded utterances (5 target words $\times$ 3 intonation patters $\times$ 2 speakers $\times$ 2 conditions). The video recordings were carried out in a classroom at Pompeu Fabra University in Barcelona with a PMD-660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone, and later edited with Adobe Premiere Pro CS6 and Audacity 2.1.2. To obtain natural pronunciations of the F0 contours, each target sentence was

elicited using a discourse completion task [5] (Billmyer & Varghese, 2000; Blum-Kulka, House, & Kasper, 1989; Félix-Brasdefer, 2010). For example, the discourse context in (2) was used to elicit the target sentence *¿Te gusta esta cama?* "Do you like this bed?"

(2) Discourse context for the target sentence *¿Te gusta esta cama?* "Do you like this bed?":
English version: Imagine that your husband/wife would like to buy a single bed for your son. One day when you are both at IKEA, you have been comparing a lot of beds and he/she seems to pay a lot of attention to one of them in particular. Nevertheless, you are not sure whether he/she really likes it or is just talking a lot about it. Please ask him/her whether he/she likes it.

As noted, each of the two speakers was videotaped producing the five target words in the context of three different intonation patterns and in two conditions. In the no-gesture condition, the speakers produced the target intonation patterns only in speech and without pitch gestures, as illustrated in the right panel of Figure 3. In the gesture condition, the speakers produced the target intonation patterns simultaneously with associated pitch gestures, as illustrated in the left panel of Figure 3.

[5] The discourse completion task is an inductive method that has been successfully applied in linguistic research for many years and recently also in prosodic research (see, e.g., Prieto & Roseano, 2010).

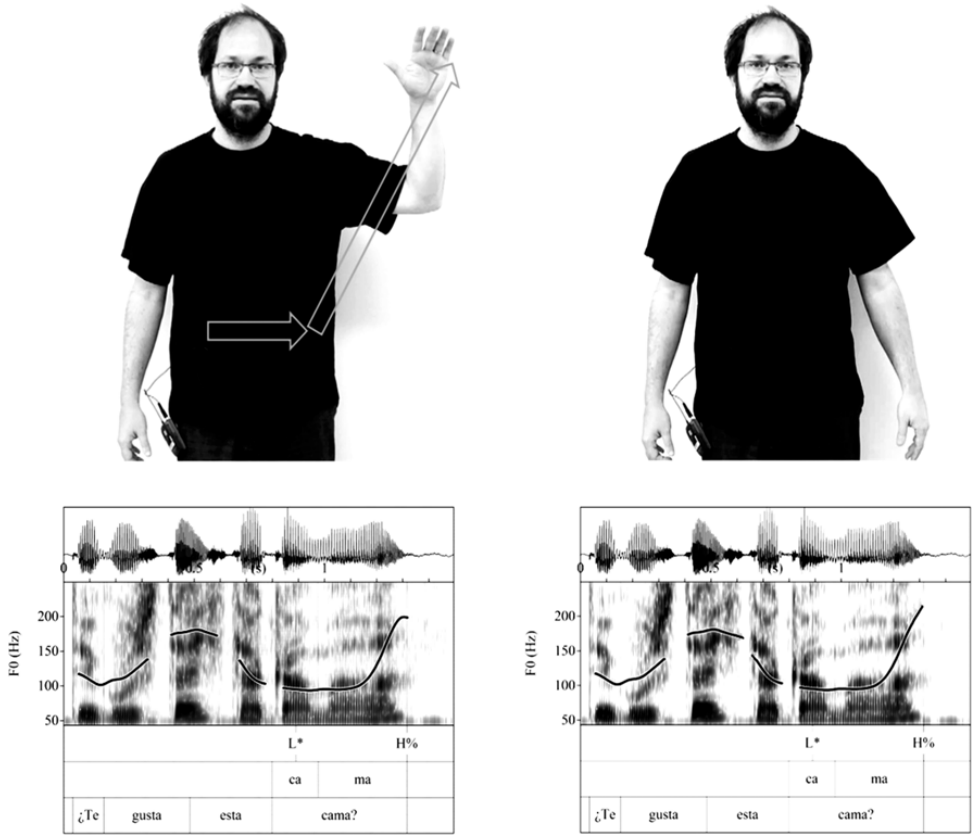

FIGURE 3. Video stills (top) from the production of the target sentence ¿Te gusta esta cama? "Do you like this bed?" in the gesture condition (left panel) and in the no-gesture condition (right panel) with corresponding waveforms and F0 contours (bottom). The arrows on the photo in the left panel represent the dynamic pitch gestures performed over the target nuclear pitch configuration L* H%.

Before the recordings, the two speakers were trained to use the cospeech gestures intended to illustrate the pitch movements of the three target nuclear configurations (statements, yes-no questions, and requests), as illustrated in Figure 4. The pitch gesture for the low L* tone (see arrow marked with L* in Figure 4) remained constant in the three configurations but was combined with different gestures representing the different boundary tones (see arrows indicating the rise

H%, rise-fall HL%, and fall L% in Figure 4).

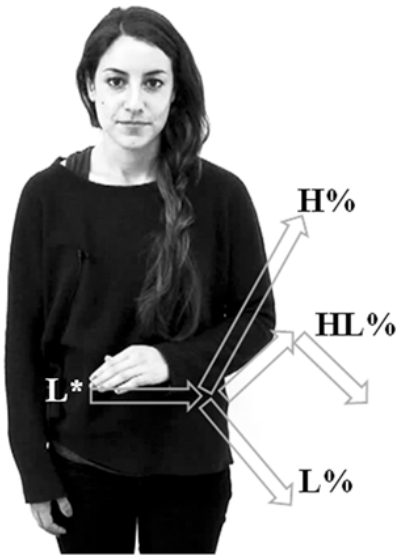

FIGURE 4. Illustration of the pitch gestures associated with the three target intonation patterns.

To ensure that spoken utterances were kept similar across the gesture and no-gesture conditions, the two speakers were recorded producing first the target sentence in the gesture condition and immediately thereafter producing the same sentence in the no-gesture condition. After the recordings, the first author auditorily checked that the pairs of sentences in the gesture and no-gesture conditions were pronounced consistently. Following González-Fuente, Escandell-Vidal, and Prieto (2015), four acoustic cues were calculated for each speech file, namely Mean F0, F0 Variability, MSD (mean syllable duration in ms., calculated by dividing the total duration of the target sentence by the number of syllables), and Mean Intensity. Four generalized linear mixed model (GLMM) tests were run using IBM SPSS Statistics 23 (IBM Corporation, 2015) for each speaker. The experimental condition (2 levels: gesture and no-gesture) was set as the fixed factor, and the

dependent variables were the four acoustic features. No significant effect of condition was reported in any of the GLMM tests, which confirmed that the training materials across conditions differed only in the presence or absence of pitch gesture, not in the acoustics of the speech.

PROCEDURE

The experiment consisted of five phases, a musical ability test, an introductory video, a pretest, the training session (which was different for the two groups), and the posttest. Before the pretest, participants were required to take a musical ability test. This was followed by a brief habituation session that consisted of an overview of the experimental procedure as well as a brief introduction to the three target intonation patterns. A diagram of the procedure is given in Figure 5. The whole experiment lasted around 45 minutes and was conducted in a single day in two multimedia classrooms at Xi'an International Studies University with two groups of 32 students.

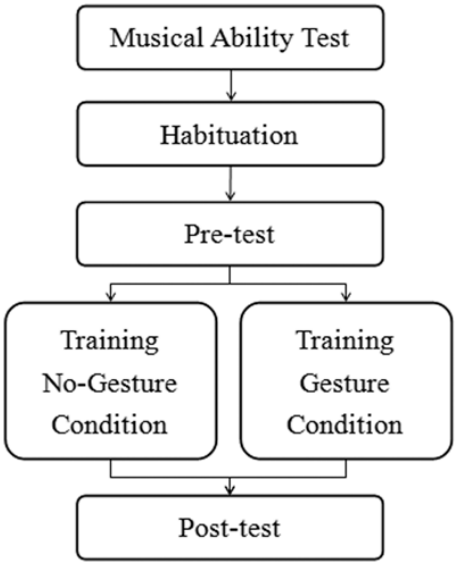

FIGURE 5. Experimental procedure.

MUSICAL ABILITY TEST

Following Law and Zentner (2012), musical abilities can be defined as the perceptual sensitivity to multiple modalities, such as timbre, tuning, melody, pitch, rhythm, and tempo. In relation to our current experiment, we were interested in assessing sensitivity to melody and pitch, which are the two acoustic properties that are mostly related to intonation in speech. Consequently, we selected two specific tasks targeting pitch and melody from the PROMS (Profile of Music Perception Skills). For more details about the PROMS and how different properties are controlled across various subtests, see the PROMS web page [5] and the article by Law and Zentner (2012) on how musical ability is measured and validated using the PROMS test.

The musical ability test applied here thus consisted of a melody subtest and a pitch subtest. In both subtests, participants were asked to listen to a series of audio pairs of different complexities through AKG K27 Mk II headphones. For each audio pair, they heard a reference sound clip twice followed by a test sound clip played once and were asked to rate how similar the test sound was to the reference sound by selecting one of five responses: *definitely the same*, *probably the same*, *I don't know*, *probably different*, *definitely different*. In the melody subtest they were asked to judge the similarity of two melodies, and in the pitch test they were asked to compare the pitches of two pure tones. [6] The

[6] A complete trial consisted of two pieces of audios, namely a reference sound clip, which is played twice, and a test sound clip, which is only played once. For instance, an easy trial from the melody subtest consists of a tonal melody (upper part) as opposed to a complex trial, which is atonal (lower part). In a melody test trial, a reference sound clip (with standard melody) is played twice, followed by a test sound clip (with comparison melody). Participants are then required to determine whether the melody of test sound clip is identical to that of the reference sound clip. Likewise, in relation to the pitch test, sinusoids (or pure tones) are used and the difficulty level is manipulated by varying the degree of the pitch difference between the standard pitch and comparison pitch. In a complete trial of pitch test, a reference sound clip (with standard pitch) is displayed twice, followed by a test sound clip (with comparison pitch).

musical ability subtests together lasted around 20 minutes. The final scores were automatically generated for each participant and later collected by the test administrator (first author of this study).

HABITUATION PHASE

Immediately after the musical ability test, each participant watched a 15-minute video recording offering an explanation of the experimental procedure and explaining the three target intonation patterns that they would have to learn. Each intonation pattern was presented in the video by means of a written text in Mandarin Chinese describing the intonation pattern and its usual context, with the intonation illustrated by means of arrows pointing up or down. The description then offered an appropriate discourse context for the use of the target intonation pattern, and a sentence in Spanish containing that intonation, which the viewers then heard being spoken by a native speaker of Spanish (see Figure 6). Participants viewed the introductory video on a Lenovo E99 laptop computer using the same AKG K27 Mk II headphones as in the musical ability test. They were asked to only watch and listen to the stimuli quietly.

Afterward, participants are asked whether the pitch (of pure tones) in the test sound clip is the same as that in the reference sound clip.

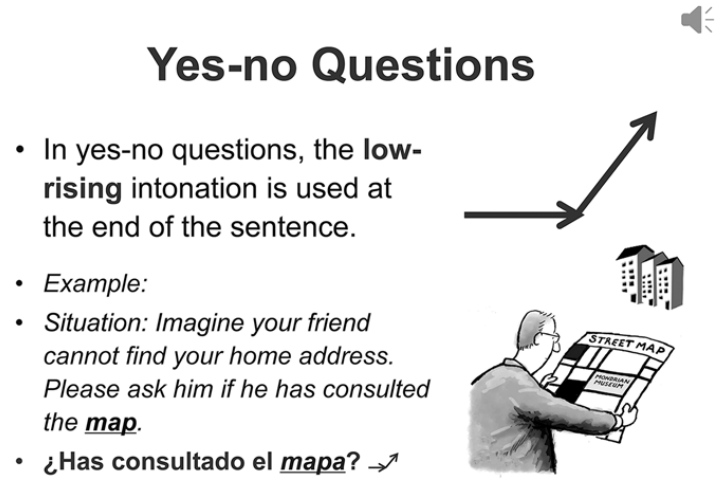

FIGURE 6. English translation of a screenshot from the introductory video illustrating the target nuclear pitch configuration L* H% of yes-no questions with the sentence ¿Has consultado el mapa? "Have you consulted the map?" Explanatory text in the actual test material was in Mandarin.

PRETEST

The pretest phase consisted of a set of 15 discourse completion trials (5 target words × 3 target intonation patterns) in the form of a PowerPoint presentation that participants interacted with individually by means of the computer keyboard and an AKG C417 PP microphone. For each pretest trial, the participant saw a slide containing the description of a discourse context written in Chinese with a drawing by way of illustration (see Figure 7). The description of the context was followed by a prompt in Chinese telling the participant what to do, such as "Ask your mother if your notebook is on the table" (as in Figure 7). Target words in Spanish were shown in the top right corner of the slide (e.g., *mesa* "table" in Figure 7). Participants were instructed to use this word as the last word in their own response. After reading the context, participants clicked on the button "Record" to record their spoken

response to the discourse prompt. When finished, they clicked the same button a second time to stop the recording and store the audio data in the computer. For each trial, a maximum of 60 seconds was given to perform the recording, which proved to be ample time, as no item was left without a response. As soon as participants had recorded their response, they could proceed to the next trial by clicking on the screen. Each participant viewed 15 pretest trials, yielding a total of 15 recorded sentences per speaker. To avoid primacy and recency effects on memory, pretest trials were presented in a randomized order by combining blocks each of which contained one instance of the three different intonation contours, that is, a statement, a yes-no question, and a request. [7] Participants were given a 5-minute break between the pretest and the training session that followed.

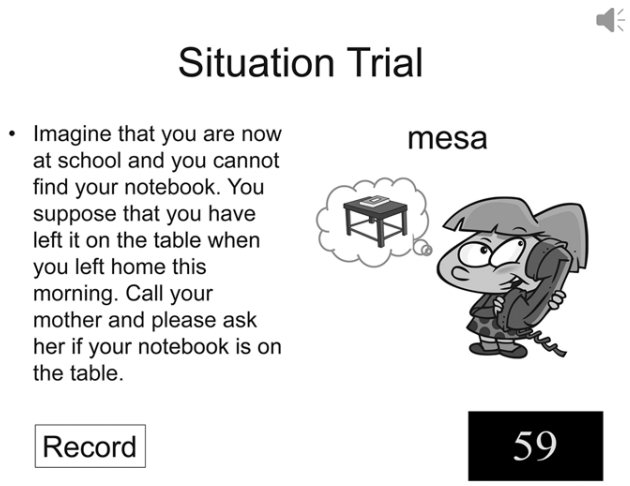

FIGURE 7. English translation of the pretest trial slide intended to

[7] Randomization followed the following criteria: (a) the same intonation pattern could not appear within the same block; (b) the same intonation pattern could not be adjacent to itself across blocks; (c) the sequence order of one block could not be identical in the subsequent block; (d) the same target words could not co-occur either within or across blocks; and (e) one target word's left neighbor could not be the same as its right neighbor both within and across blocks. The same method was also applied to randomize the order of trials in both the training session and the posttest recordings.

elicit the target nuclear pitch configuration L* H% of yes-no questions with the sentence ¿Mi cuaderno está en la mesa? “Is my notebook on the table?”

TRAINING SESSION

Prior to the training session, participants were randomly assigned to one of the two between-subjects groups, with 32 participants per group. Participants in the experimental (gesture) group watched a set of training videos intended to teach them the three Spanish intonation patterns under study by means of both speech and gesture. They saw the two Spanish native speakers each producing 15 utterances with accompanying gestures intended to illustrate the relevant intonation patterns, to which participants were reminded to pay close attention. By contrast, participants in the control (no-gesture) group watched a set of training videos of the two Spanish speakers each saying the same 15 utterances exemplifying the three intonation patterns, but without accompanying gestures. In both cases, participants thus heard each token twice, once spoken by the female Spaniard and once spoken by the male. Also, in both groups, before they heard the target sentences, participants could first read them on the screen of their computer. Figure 8 illustrates the sequence followed for each of the training trials, with a still of the written form followed by a video clip of one of the Spanish speakers producing the utterance, then a video clip of the other Spanish speaker doing the same.

FIGURE 8. Screenshots of the training sequence for the target sentence ¿Te gusta esta cama? "Do you like this bed?" in the gesture condition.

The target sentences in the training phase were played to participants in the same order as they were presented within embedded discourse contexts in the pretest phase. As in the habituation phase, in the training phase participants were asked to only watch and listen to the stimuli quietly, without either repeating what they heard (in both conditions) or mimicking the gestures they saw (in the gesture condition). The total time for the training session was 3 minutes and 17 seconds. Participants were given another 5-minute break between the training session and the posttest that followed.

POSTTEST

To test the participants' learning results, in other words, to determine how much they could apply the intonation patterns they had learned to both old and new materials, participants were asked to record the intonation patterns of the same 15 items they had been exposed to in the training and pretest phases (which we will call "related" items) plus 15 new items (which we will call "unrelated" items). The randomization, elicitation, and recording procedures were the same as those used in the pretest. Again, participants were asked to respond to the situations in the randomized discourse contexts by using the target words provided in the top right corner of each slide (see Figure 7).

ANNOTATION

Both pretest and posttest recordings by all participants ([15 pretest items + 30 posttest items] × 64 participants) were analyzed and coded

following the Spanish_ToBI labeling system (Hualde & Prieto, 2015). The transcription focused on the intonational forms of the target nuclear pitch configurations L* H%, L* HL%, and L* H%. If the participant produced a L* pitch accent, it was rated as accurate and marked "1," whereas those wrongly produced with a H* pitch accent were marked "0." Pitch accents like L+H* and H+L* were also found and were respectively rated as equivalent to H* and L*. Boundary tones were not considered.

RESULTS

MAIN GLMM ANALYSES

First, a compound musical ability score was obtained for each participant by calculating the mean scores of the melody and pitch musical ability subtests. [8] Following Tavakoli (2013), a TwoStep Cluster was applied using these scores so that the participants were automatically sorted into three different musical ability levels, namely high ($M_{scores} = 12.844$, $SD = 0.902$), mid ($M_{scores} = 9.779$, $SD = 0.797$), and low ($M_{scores} = 7.443$, $SD = 0.914$). The dataset then was submitted to a GLMM analysis with accuracy of intonation set as the dependent variable and condition (two levels: gesture vs. no-gesture), test (two levels: pretest vs. posttest), intonation pattern (three levels: statement, yes-no question, and request), musical ability level (three levels: high, mid, and low), as well as their interactions were set as fixed factors.

[8] The different subtest scores of the PROMS test are usually averaged to provide a composite musical ability test score (e.g., Law, 2012). One advantage of the PROMS test is that the individual subtest scores are also generally reported in case researchers are interested in a particular musical ability. Following Faßhauer, Frese, and Evers (2015), we took participants' mean scores on the melody and pitch subtests to be representative of their general musical ability.

Subject and task were set as random factors. The post-hoc pairwise comparisons were performed by applying sequential Bonferroni comparisons.

The results of the GLMM model showed significant main effects of condition (F (1, 2860) = 5.880, p = .015), test (F (1, 2860) = 118.854, p < .001), intonation pattern (F (2, 2860) = 47.760, p < .001), and musical ability level (F (2, 2860) = 10.994, p < .001), as well as five significant two-way interactions: condition × test (F (1, 2860) = 26.846, p < .001), intonation pattern × test (F (2, 2860) = 70.936, p < .001), intonation pattern × musical ability level (F (4, 2860) = 11.266, p < .001), test × musical ability level (F (2, 2860) = 3.246, p = .039). We found two random slopes (subject and task) and only one significant random intercept for subject (p < .001).

MAIN EFFECTS

Condition

The significant main effect of condition confirmed that participants trained with gestures performed significantly better than those in the control group ($M_{accuracy}$ (no-gesture) = 0.418 < $M_{accuracy}$ (gesture) = 0.577).

Test

The test effect indicated that all participants improved after watching the training stimuli ($M_{accuracy}$ (pretest) = 0.309 < $M_{accuracy}$ (posttest) = 0.686).

Intonation Pattern

The significant intonation pattern main effect showed that a significant difference existed between all pairs of target intonation patterns. This

result indicates that the three intonation patterns implied different degrees of difficulty for participants, which was further supported by the mean accuracy of each pattern (M_{accuracy} (yes-no question) = .690 > $M_{accuracy}$ (request) = .527 > $M_{accuracy}$ (statement) = .281).

Musical Ability Level

The main effect of musical ability level was also reported to be significant, suggesting that participants with different musical ability levels performed differently in the production test. The post-hoc comparisons showed that this effect was found between all three musical ability levels. The mean accuracy of participants in each musical ability level ($M_{accuracy}$ (high) = .739 > $M_{accuracy}$ (mid) = .410 > $M_{accuracy}$ (low) = .330) further revealed that that those who were better at music performed much better at reproducing intonation patterns in Spanish. [9]

---

[9] On the recommendation of an anonymous reviewer, statistical analyses were also done to test the effect of the independent musical ability skills, namely, melody and pitch. For both of the musical skills, a TwoStep Cluster was run so that the participants were automatically sorted into three different musical ability levels, melody high (M_{scores} = 9.125, SD = 1.597), melody mid (M_{scores} = 9.119, SD = 2.859), and melody low (M_{scores} = 8.111, SD = 1.988); and pitch high (M_{scores} = 10.531, SD = 2.021), pitch mid (M_{scores} = 9.870, SD = 2.049), and pitch low (M_{scores} = 9.833, SD = 2.302). The data were later submitted to a GLMM where melody level and pitch level, as well the main factors (condition, test, pattern) and their mutual interactions were set as fixed factors. Subject and task were again set as random factors. The results showed a main significant effect of pitch level (F (2, 2850) = 3.727, p = .024), and three significant interactions, namely condition × pitch level, pattern × pitch level, and pattern × melody level. We had two random slopes (subject and task) but only one significant intercept, for subject (p < .001). As for the condition × pitch level interaction, post-hoc comparisons showed that the significant effect of pitch level existed in both conditions (gesture condition: F = 4.180, p = .015; no-gesture condition: F = 7.451, p = .001). Post-hoc comparisons for the pattern × pitch level interaction showed that the pitch level effect was found in statement (F = 12.758, p < .001; M_{scores} (high) = 0.444 > M_{scores} (mid) = 0.202 > M_{scores} (low) = 0.187) but not in yes-no question or request. This seems to suggest that the acquisition of statements is more dependent on the ability to detect pitch accent, a finding consistent with the difference between the three target patterns, where the L* pitch accent in statements is followed by a L%, while in the other two intonation patterns, the L* pitch accent is much more clearly contrasted with a H% or a HL% boundary tone and thus much easier to detect. In relation to the pattern × melody interaction, post-hoc analysis showed that the effect of melody level was present in yes-no question (F =

TWO-WAY INTERACTIONS

Importantly, five two-way interactions were also found to be significant: condition × test (F (1, 2860) = 26.846, $p < .001$), intonation pattern × test (F (2, 2860) = 70.936, $p < .001$), intonation pattern × musical ability level (F (4, 2860) = 11.266, $p < .001$), test × musical ability level (F (2, 2860) = 3.246, $p = .039$), and condition × item type (F (1, 1888) = 55.471, $p < .001$).

Condition × Test (Significant)

As for the interaction condition × test, post-hoc comparisons showed that the effect of condition was only found in posttest (F (1, 2860) = 18.742, $p < .001$), not in pretest. Conversely, a significant test effect was found in both conditions (no-gesture condition, F (1, 2860) = 41.515, $p < .001$; gesture condition, F (1, 2860) = 169.556, $p < .001$), which suggests that in both gesture and no-gesture conditions, participants improved after watching their respective training videos. Interestingly, the mean accuracy results further revealed that participants in the gesture condition generally improved more ($M_{accuracy}$(gesture condition) = .473) than those in the no-gesture condition ($M_{accuracy}$(no-gesture condition) = .254). This further supports our hypothesis about the positive role of pitch gestures because participants showed no group difference in the pretest, but after a short training session those in the

11.050, $p < .001$) and request ($F = 8.427$, $p < .001$), but not in statement. This indicates that the acquisition of the latter two patterns is more tied with the melody ability than is the statement pattern. This is not surprising because melody is concerned with the general pitch variation of the sentence pattern, and accordingly participants with a higher melody ability level are predicted to perform better in learning complex patterns such as yes-no questions (L*H%) or requests (L*HL%) than simple patterns like statements (L*L%). We will not go into more detail here about the effect of independent musical skills, but it is clear they influence the acquisition of intonation differently and deserve to be further explored in future research.

gesture condition performed significantly better than those in the no-gesture condition. Figure 9 shows the mean proportion of correctly produced L* pitch accents in the two conditions (gesture vs. no-gesture) in the pretest and posttest items.

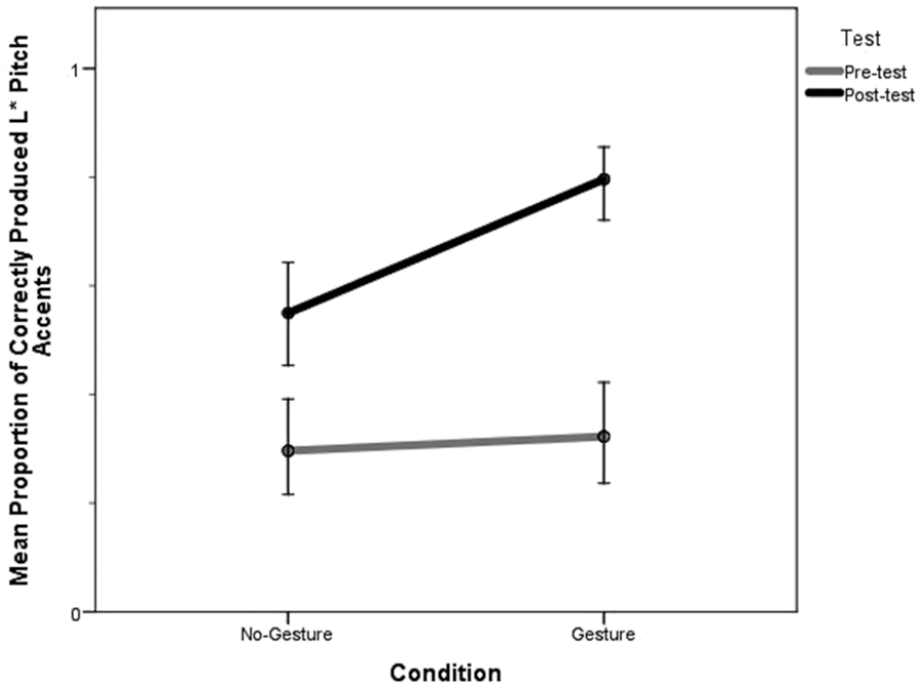

FIGURE 9. Mean proportion of correctly produced L* in both conditions (no-gesture vs. gesture) and both tests (pretest [gray line] and posttest [black line]).

Intonation Pattern × Test (Significant)

Post-hoc comparisons for the intonation pattern × test interaction revealed that the test effect was present in both the yes-no question pattern (F (1, 2860) = 21.726, $p < .001$) and the request pattern (F (1, 2860) = 575.759, $p < .001$), but not in the statement pattern. These results suggest that although participants improved after the training session, their improvement was significant for only two of the three

intonation patterns. A comparison of the mean accuracy for each intonation pattern in both tests (pretest: $M_{accuracy}$ (yes-no question) = .566 > $M_{accuracy}$ (statement) = .278 > $M_{accuracy}$ (request) = .152; posttest: $M_{accuracy}$ (request) = .874 > $M_{accuracy}$ (yes-no question) = .791 > $M_{accuracy}$ (statement) = .284) showed even clearer differences between target intonation patterns. Figure 10 shows the mean proportion of correctly produced L* pitch accents in the two tests (pretest and posttest) for the three intonation patterns (statement, yes-no question, and request).

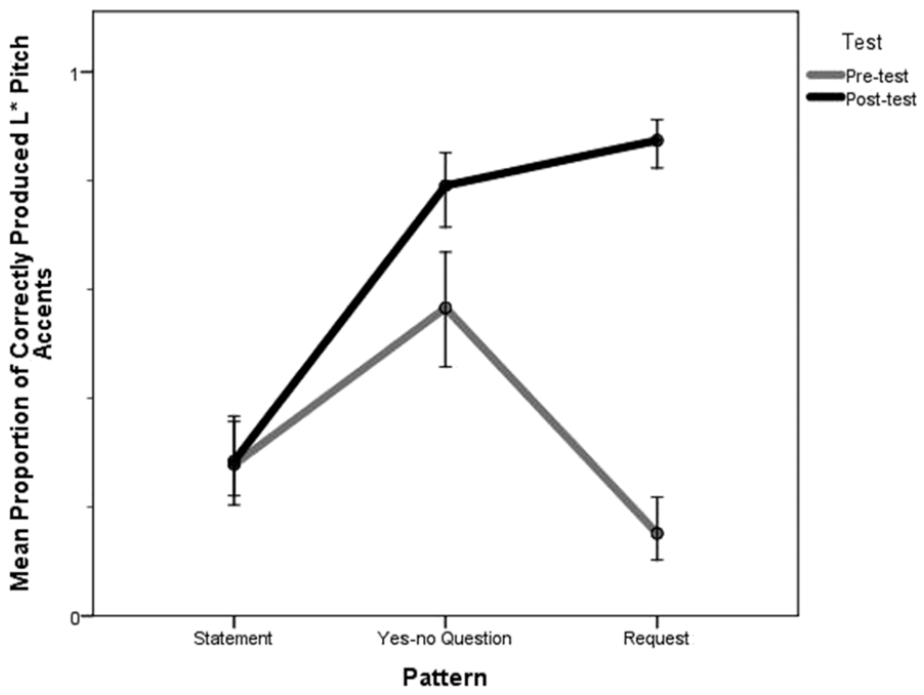

FIGURE 10. Mean proportion of correctly produced L* in the three intonation patterns (statement, yes-no question and request) and in both tests (pretest [gray line] and posttest [black line]).

Test × Musical Ability Level (Significant)

For the two-way interaction involving test and musical ability level, the

post-hoc comparisons revealed that the test effect was found in groups of each musical ability level (high level: $F\,(1,\ 2860) = 24.560$, $p < .001$; mid level: $F\ (1,\ 2860) = 134.558$, $p < .001$; low level: $F\ (1,\ 2860) = 45.708$, $p < .001$), which shows that all the participants at each musical level improved after the training session. The post-hoc analyses showed a general significant effect of musical ability level in both pretest ($F\,(2,\ 2860) = 4.924$, $p = .007$) and posttest ($F\,(2,\ 2860) = 25.729$, $p < .001$). More specifically, the pretest effect was found between high-mid, and high-low, but not between mid-low, whereas the posttest effect was found between any two musical ability levels. The mean accuracy scores further illustrated these results (pretest: $M_{\,accuracy}\,(\text{high}) = .515$, $M_{\,accuracy}\ (\text{mid}) = .237$, $M_{\,accuracy}\,(\text{low}) = .214$; posttest: $M_{\,accuracy}\,(\text{high}) = .883$, $M_{\,accuracy}\ (\text{mid}) = .609$, $M_{\,accuracy}\ (\text{low}) = .471$) and indicated that participants with a high musical ability level generally performed much better than the participants with mid and low levels in both tests, whereas mid-level participants performed almost the same as low-level students in pretest, but much better in posttest. Figure 11 shows the mean proportion of correctly produced L* accents as a function of test (pretest and posttest) and musical ability (high, mid, and low).

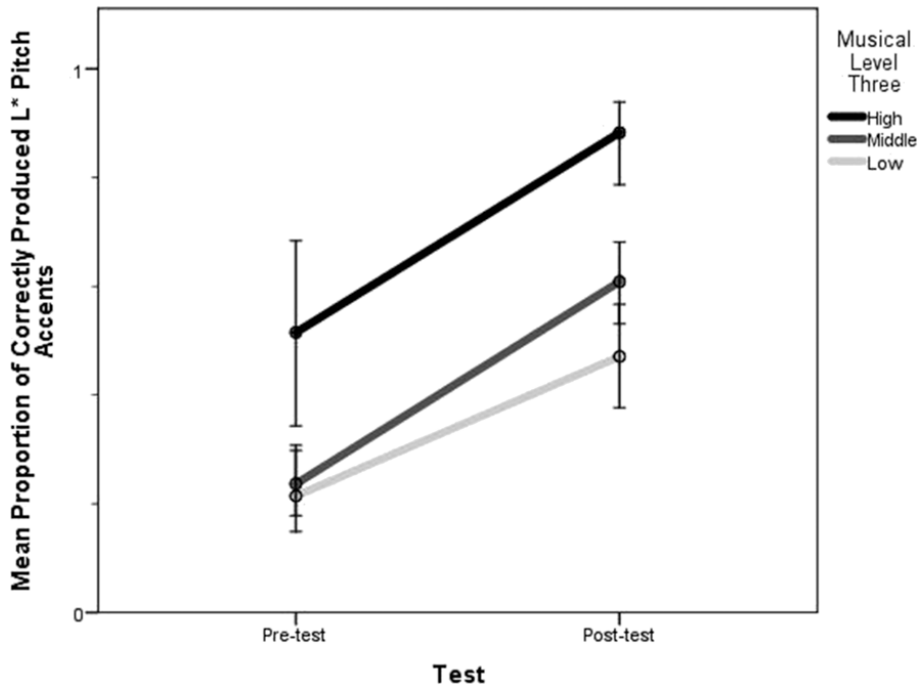

FIGURE 11. Mean proportion of correctly produced L* in both tests (pretest and posttest) and in the three musical ability levels (high [black line], mid [strong gray line], and low [soft gray line]).

Intonation Pattern × Musical Ability Level (Significant)

As for the intonation pattern × musical ability level interaction, post-hoc analyses showed that the musical ability level effect was present in both yes-no question ($F\,(1, 2860) = 21.726$, $p < .001$) and request ($F\,(1, 2860) = 575.759$, $p < .001$), but not in statement. However, an intonation pattern effect was present between all the musical abilities levels: high level, $F\,(2, 2860) = 25.365$, $p < .001$); mid level, $F\,(2, 2860) = 52.774$, $p < .001$; low level, $F\,(2, 2860) = 4.167$, $p = .016$). These results suggest that the three target patterns may imply different degrees of difficulty for participants of different musical ability levels. The mean accuracy scores for each pattern at each musical ability level (statement: $M_{\,accuracy}\,(\text{high}) = 0.399 > M_{\,accuracy}\,(\text{low}) = 0.205 > M_{\,accuracy}$

(mid) = 0.259; yes-no question: $M_{accuracy}$(high) = 0.937 > $M_{accuracy}$(mid) = 0.569 > $M_{accuracy}$(low) = 0.358; request: $M_{accuracy}$(high) = 0.696 > $M_{accuracy}$(mid) = 0.496 > $M_{accuracy}$(low) = 0.381) provided positive evidence for our hypothesis that musical ability plays a role in producing L2 intonation and further revealed that the positive effect of musical ability was present for all three intonation patterns.

Condition × Musical Ability Test (Nonsignificant)

Interestingly, though the interaction between condition and musical ability level was not significant, results of the post-hoc analyses showed a significant effect of condition on mid-level participants (mid: F (1, 2860) = 8.341, p = .004), but not on high-level or low-level participants. To confirm the post-hoc results, three separate GLMMs were run for the different musical ability levels of the participants. The results confirmed that condition had a significant effect only on mid-level participants (F (1, 2878) = 74.247, p < .001). All this suggests that while mid-level participants benefited from the observation of pitch gestures, high-level and low-level participants did not.

OTHER RELEVANT STATISTICAL ANALYSES

With the purpose of exploring whether there was a significant difference in accuracy between the 15 related items and the 15 unrelated items, which would measure learners' ability to apply learned intonation patterns to completely new items, the whole set of data was submitted to another GLMM statistical test, in which the data from the pretest were eliminated and the accuracy of intonation contours in the posttest was set as the dependent variable. Item type in the posttest (two levels: related and unrelated), as well as the three two-way interactions involving item type (item type × condition, item type × pattern, and

item type × musical ability level) were set as fixed factors. Subject and task were again set as random factors. The post-hoc pairwise comparisons were performed by applying the sequential Bonferroni for adjusting for multiple comparisons.

We obtained two random slopes (again, subject and task) and only one significant random intercept for subject ($p < .001$) in both GLMM tests. The results showed a significant interaction for only condition × item type (F (1, 1888) = 55.471, $p < .001$). Post-hoc analyses showed that the condition effect was found for both related and unrelated items in the posttest, indicating that participants in the experimental group performed significantly better at producing not only the related items (F (1, 1888) = 46.661, $p < .001$) but also the unrelated items (F (1, 1888) = 64.280, $p < .001$). The nonsignificant effect of item type suggests that the improvement in the posttest is not merely a mimicking of the target patterns taught by native speakers in the training session under both conditions. Moreover, the significant interaction between condition and item type further indicates that the use of gesture benefited the learning of not only related items (i.e., items taught in the training session) but also unrelated items in the posttest, which were totally new to them.

DISCUSSION AND CONCLUSIONS

The present study examined whether observing pitch gestures depicting the intonation pattern of accompanying speech could enhance Mandarin basic-level ELE learners' production of Spanish intonation, especially in learning the most difficult pitch accent (L*) in the target language. Overall, the results showed a significant gain in performance after a short training session, especially in those who underwent the training

with cospeech pitch gestures. This demonstrated that visuospatial gestures signaling pitch movements have the potential to aid Mandarin basic-level speakers in learning the intonational melodies of Spanish. More importantly, the improvement of participants in producing Spanish intonation after a short training session cannot be attributed merely to memorization because it was shown to be easily generalized to new items. Thus, the results of this research extend and complement the recent results reported by Jia and Wang (2013a, 2013b), Morett and Chang (2015), and Baills et al. (ms.) regarding the beneficial role of pitch gestures in teaching lexical tones in a tonal language because we have demonstrated that gestures can also enhance the learning of L2 intonation by tonal language speakers. We may safely conclude that observing pitch gestures that illustrate spoken pitch movements favors the learning of L2 tonal and intonational systems in general.

Interestingly, our results showed that this positive effect of pitch gesture was not systematically obtained for all the three types of pitch configurations because pitch gestures aided significantly in the learning of the L* H% (yes-no questions) and the L* HL% (requests) patterns, but did not overtly benefit the learning of L* L% (statements). These results seem to suggest a hierarchy of difficulty in the learning of Spanish intonation by Mandarin-speaking beginner ELE learners whereby statement is the most difficult pattern, followed by yes-no question, and then request (statement > yes-no question > request). In general, this hierarchy is consistent with the proposals put forth by Liu (2003) and Cortés Moreno (1997, 2004). Such a ranking could be related to the perceptual salience of the target intonational movements, such that the request intonation pattern (involving a complex fall-rise-fall movement) is the most salient one and thus most easily perceived by learners, whereas the statement intonation pattern (involving a nonalternating

sequence of low tones) is the least salient and thus least readily perceived. In other words, intonation patterns involving more H-L contrasts would be aurally (and visually) more perceptually salient to L2 learners and this would accelerate the rate of their acquisition. Further research is required, however, to explore whether the perceptual salience of different intonation patterns is related to the number of H-L contrasts as we predicted and how far this explanation can be extended to account for the difficulties involved in learning Spanish intonation.

Another interesting finding of this study is related to the correlation between musical abilities and the learning of Spanish intonation by Mandarin speakers. Our results clearly show that stronger musical abilities facilitate the learning process of intonation. Though participants of all musical levels improved their Spanish intonation skills after receiving intonation training, participants with higher levels of musical ability generally showed more success in producing the target intonation patterns. This is consistent with the perceptual findings reported by Marques et al. (2007) and Zhao and Kuhl (2015). As for the relation between musical ability and the presence or absence of gestures to illustrate intonational movements, however, though no significant interaction was found between these two factors, our statistical analyses suggested that the presence of gestures was useful for learners with moderate musical ability but less so for learners with a high degree or a low degree of musical ability. These analyses showed that in the pretest, there were no differences between the control and experimental groups at each level of musical ability, whereas in the posttest, such group differences emerged in the learners of moderate musical ability but not in those with high or low ability. The finding that gestures enhance intonational learning less in learners with high musical ability could be accounted for from two angles. On the one hand, having already

performed very well in the pretest (because they are especially sensitive to pitch movements), there was less room available for learners with a high musical ability to demonstrate improvement in the posttest. On the other hand, the effect could be due to a potential ceiling effect between hypersensitivity to acoustic parameters and the observation of pitch gestures. Because participants with high musical abilities are presumably more sensitive to acoustic variations in pitch (Zhao & Kuhl, 2015), a training stimulus that requires visual attention may distract them and even have a negative effect. The fact that the use of gestures did not benefit intonational learning in learners with low musical ability skills seems to suggest that regardless that pitch gestures tend to help learning intonation, auditory sensitivity is also an important skill to take into consideration because individuals need to build a connection between the two perceptual channels (i.e., visible pitch gestures and auditory pitch variation in speech). In other words, unlike learners with moderate or high musical ability skills, the main difficulty faced by learners with low musical abilities is the level of perceptual sensitivity to auditory pitch variations. With the data obtained from our experiment, we can't speculate about which of these explanations is more likely for learners with a high or a low musical ability level, but we may safely conclude that learners with a strong musical ability generally perform better in producing target intonation patterns and such an ability most likely favors their learning of prosody in general.

To sum up, the present study demonstrates that a short training session involving hand gestures representing pitch movements does help Mandarin-speaking basic-level ELE learners to learn a set of target intonation contours containing a L* pitch accent, which has been reported to be one of the most difficult intonational features for Mandarin learners of Spanish. In general, these findings confirm the

hypothesis proposed by McCafferty (2006) that actional representation in the form of pitch gestures helps learners to acquire the conceptual foundations of pitch variations in the mental plane. Additionally, it also shows that musical abilities (particularly sensitivity to pitch and melody) play a positive role in the learning of target intonation patterns, confirming claims that the processing of acoustic features in both music and speech is implemented by the same anatomical dimension in brain networks (overlap in the OPERA hypothesis, Patel, 2011).That is, due to anatomical overlap, because participants with higher musical abilities tend to process the target acoustic features, namely, melody (pitch variations) and pitch (pure tones) much faster and accurately, they are predicted to detect the corresponding acoustic features of speech more quickly and more accurately. By contrast, according to the precision concept in the OPERA hypothesis proposed by Patel (2011), speech places lower demands on the shared networks between music and speech and, consequently, participants with a strong musical ability are predicted to perform better under both conditions and across all three target intonation patterns. In fact, in our data, even the most difficult intonational pattern (i.e., statement) implied little difficulty for high-level participants. Interestingly, participants with higher musical abilities seemed to rely less on the aid provided by the visual processing.

From a pedagogical perspective, our findings hold implications for the teaching of L2 intonation because they clearly suggest a more embodied and multimodal approach in the L2 classroom. Specifically, the use of gestures in the instructional setting can constitute a helpful resource for representing prosodic modalities such as pitch and rhythm, making the learning process more playful and interesting. It would be of considerable practical interest to assess whether cospeech gestures can outperform the effectiveness of digital acoustic signal visualizing tools

like the Laryngograph Processor, Speech Filing System, and so forth in L2 teaching of prosodic features. Regardless, the technique of using gestures to imitate the rhythmic and pitch properties of speech in a second language is easy to apply, given that it does not involve the use of any additional supportive tools. Moreover, some basic musical training strategies such as ear training to differentiate basic pitch tones and melodic patterns also deserve to be promoted in L2 classrooms, given the probability that they significantly help students to perceive and acquire the target L2 prosodic patterns. It is also worth pointing out that musical education in childhood could promote children's future development not only in music ability but also in language learning.

REFERENCES

J. A. Alexander, P. C. M. Wong, & A. R. Bradlow. (2005). Lexical tone perception in musicians and non-musicians. In INTERSPEECH (Ed.), *Proceeding of the (Eurospeech) 9th European Conference on Speech Communication and Technology* (pp. 397–400). Lisbon: ISCA.

J. Anderson-Hsieh, R. Johnson, & K. Koehler. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segments, prosody, and syllable structure. *Language Learning*, **42**, 529–555.

F. Baills, N. Suárez, S. González-Fuente, & P. Prieto. (under review). Observing and producing pitch gestures enhance the acquisition of Mandarin Chinese tones and words.

K. A. Barto. (2015). *Mandarin speakers' intonation in their L2 English* (Unpublished doctoral dissertation). Tucson, AZ: University of Arizona.

P. Bernardis, & M. Gentilucci. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, **44**, 178–190.

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

K. Billmyer, & M. Varghese. (2000). Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics*, **21**, 517–552.

S. Blum-Kulka, J. House, & G. Kasper. (1989). Investigating cross-cultural pragmatics: An introductory overview. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 1–34). Norwood, NJ: Ablex.

T. Bongaerts, C. Van Summeren, B. Planken, & E. Schils. (1997). Age and ultimate attainment in the pronunciation of a foreign languages. *Studies in Second Language Acquisition*, **19**, 447–465.

B. Braun, & L. Tagliapietra. (2010). On-line interpretation of intonational meaning in L2. *Language and Cognitive Processes*, **26**, 224–235.

S. Brown, M. J. Martinez, & L. M Parsons. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, **23**, 2791–2803.

B. Chandrasekaran, A. Krishnan, & J. T. Gandour. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, **108**, 1–9.

J. Chobert, & M. Besson. (2013). Musical expertise and second language learning. *Brain Sciences*, **3**, 923–940.

Comisión orientada de la enseñanza de lenguas extranjeras en las universidades, sección de español. (1998). *Programa de Enseñanza*

para Cursos Básicos de las Especialidades de Lengua Española en Escuelas Superiores Chinas. Shanghai: Editorial de la enseñanza de lenguas extranjeras de Shanghai.

L. Connell, Z. G. Cai, & J. Holler. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, **81**, 124–130.

M. Cortés Moreno. (1997). Sobre la percepción y adquisición de la entonación española por parte de hablantes nativos de chino. *Estudios de Fonética Experimental*, **5**, 67–134.

M. Cortés Moreno. (2001). Percepción y adquisición de la entonación española en enunciados de habla espontánea: El caso de los estudiantes taiwaneses. *Estudios de Fonética Experimental*, **11**, 90–119.

M. Cortés Moreno. (2004). Análisis acústico de la producción de la entonación española por parte de sinohablantes. *Estudios de Fonética Experimental*, **13**, 80–110.

M. Cruz-Ferreira. (1989). Non-native interpretive strategies for intonational meaning: An experimental study. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 103–120). Dordrecht, The Netherlands: Foris.

S. Dolscheid, R. M. Willems, P. Hagoort, & D. Casasanto. (2014). The relation of space and musical pitch in the brain. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *36th annual meeting of the cognitive science society (CogSci 2014)* (pp. 421–426). Quebec: Cognitive Science Society.

Y. S. Dong, & J. Liu. (2014). *Español Moderno, Libro del Alumno I.* Beijing: Foreign Language Teaching and Research Press.

C. Faßhauer, A. Frese, & S. Evers. (2015). Musical ability is associated with enhanced auditory and visual cognitive processing. *BMC*

Neuroscience, **16**, 59.

J. C. Félix-Brasdefer. (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp. 41–56). Amsterdam, The Netherlands, and Philadelphia, PA: John Benjamins.

J. Flege. (1991). Orthographic evidence for the perceptual identification of vowels in Spanish and English. *Quarterly Journal of Experimental Psychology*, **43**, 701–731.

M. D. C. Fonseca-Mora, P. Jara-Jiménez, & M. Gómez-Domínguez. (2015). Musical plus phonological input for young foreign language readers. *Frontiers in Psychology*, **6**, 286.

G. Gao, & H. W Qin. (2012). The influence of Chinese experience on English phonetic acquisition at suprasegmental level. *Shandong Foreign Language Teaching Journal*, **151**, 52–57.

D. Gluhareva, & P. Prieto. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, **21**, 609–631.

S. Goldin-Meadow. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

S. González-Fuente, V. Escandell-Vidal, & P. Prieto. (2015). Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics*, **90**, 26–47.

M. Gullberg. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *IRAL—International Review of Applied Linguistics in Language Teaching*, **44**, 103–124.

X. L. He, V. J. van Heuven, & C Gussenhoven. (2011). Choosing the optimal pitch accent location in Dutch by Chinese learners and

native listeners. In M. Wrembel, M. Kul, & K. Dziubalska-Kolaczyk (Eds.), *Achievements and perspectives in SLA of speech: New sounds 2010* (pp. 125–137). Frankfurt am Main: Peter Lang Verlag.

X. L. He, V. J. van Heuven, & C. Gussenhoven. (2012). The selection of intonation contours by Chinese L2 speakers of Dutch: Orthographic closure vs. prosodic knowledge. *Second Language Research*, **28**, 283–318.

C. C. Heffner, & L. R. Slevc. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology*, **6**, 1962.

Y. Hirata, & S. D. Kelly. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, **53**, 298–310.

Y. Hirata, S. D. Kelly, J. Huang, & M. Manansala. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, **57**, 2090–2101.

W. Hong. (2012). *An experimental study of Chinese students' English intonation pattern* (Unpublished doctoral dissertation). China: Nankai University.

J. I. Hualde, & P. Prieto. (2015). *Intonational variation in Spanish: European and American varieties*. In S. Frota & P. Prieto (Eds.), *Intonation in romance* (pp. 350–391). Oxford: Oxford University Press.

IBM Corporation. (2015). *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, NY: IBM Corporation.

X. L. Ji. (2010). *Acquisition of intonation by Chinese EFL learners—An empirical study based on experimental phonetics* (Unpublished master's thesis). China: Zhejiang University.

X. L. Ji, X. Wang, & A. J. Li. (2009). Intonation patterns of yes-no

questions for Chinese EFL learners. In IEEE (Ed.), *2016 Conference of the Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*. New York: Curran Associates, Inc.

L. Jia, & J. Q. Wang. (2013a). The effects of visual processing on tone perception by native English-speaking learners of Chinese. *Chinese Teaching in the World*, **27**, 548–557.

L. Jia, & J. Q Wang. (2013b). On the effects of visual processing on tone production by English-speaking learners of Chinese. *TCSOL Studies*, **52**, 30–34.

A. Kendon. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.

A. Kendon. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

G. Lakoff. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–251). Cambridge: Cambridge University Press.

L. Law. (2012). *Assessing and understanding individual differences in music perception abilities* (Unpublished doctoral dissertation). York: University of York.

L. N. Law, & M. Zentner. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PloS One*, **7**, e52508.

C. Y. Lee, & T. H. Huang. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, **124**, 3235–3248.

S. C. Levinson, & J. Holler. (2014). The origin of human communication.

Philosophical Transactions of the Royal Society, B: Biological Sciences, **369**, 227–246.

Y. H. Liu. (2003). *La entonación del español hablado por taiwaneses* (Unpublished doctoral dissertation). Barcelona: Universitat de Barcelona.

K. M. Ludke, F. Ferreira, & K. Overy. (2014). Singing can facilitate foreign language learning. *Memory and Cognition*, **42**, 41–52.

C. Marques, S. Moreno, S. L. Castro, & M. Besson. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, **19**, 1453–1463.

S. G McCafferty. (2006). Gesture and the materialization of second language prosody. *IRAL—International Review of Applied Linguistics in Language Teaching*, **44**, 197–209.

E. McMullen, & J. R. Saffran. (2004). Music and language: A developmental comparison. *Music Perception: An Interdisciplinary Journal*, **21**, 289–311.

D. McNeill. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

I. Mennen. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact* (pp. 171–188). Berlin and Heidelberg: Springer.

I. Mennen, & E. de Leeuw. (2014). Beyond segments, prosody in SLA. *Studies in Second Language Acquisition*, **36**, 183–194.

S. P. Morales. (2008). Enseñanza de la pronunciación del español en estudiantes chinos: La importancia de las destrezas y los contenidos prosódicos. In S. P. Cesteros & S. R. Martin (Eds.), *La evaluación en el aprendizaje y la enseñanza del español como lengua*

extranjera/segunda lengua: XVIII Congreso Internacional de la Asociación para la Enseñanza del Español como lengua Extranjera (ASELE) (pp. 497–503). Alicante: Universidad de Alicante.

L. M. Morett, & L. Y. Chang. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, **30**, 347–353.

M. J. Munro, & T. M. Derwing. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, **45**, 73–97.

M. Ortega-Llebaria, & L. Colantoni. (2014). L2 English intonation, relations between form-meaning association, access to meaning, and L1 transfer. *Studies in Second Language Acquisition*, **36**, 331–353.

M. Ortega-Llebaria, M. Nemogá, & N. Presson. (2015). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese–English bilinguals perceive "rose?" vs. "rose." *Bilingualism: Language and Cognition*, **11**, 1–17.

A. B. W. Ostrom. (1997). *Acquisition of American English intonation patterns by non-native speakers: Use of real-time computer-mediated visual feedback* (Unpublished doctoral dissertation). Austin, TX: The University of Texas at Austin.

A. D. Patel. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, **2**, 142.

A. D. Patel. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, **308**, 98–108.

P. Prieto, & P. Roseano. (2010). *Transcription of intonation of the Spanish language*. München: Lincom.

L. Rasier, & P. Hiligsmann. (2007). Prosodic transfer from L1 to L2: Theoretical and methodological issues. *Cahiers de Linguistique Française*, **28**, 41–66.

M. Shimizu, & M. Taniguchi. (2005). Reaffirming the effect of interactive visual feedback on teaching English intonation to Japanese learners. Paper presented at *the Phonetics Teaching and Learning Conference 2005*. London: University of London.

L. R. Slevc. (2012). Language and music: Sound, structure, and meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, **3**, 483–492.

L. R. Slevc, & A. Miyake. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, **17**, 675–681.

Z. B. Su, & D. Hu. (2011). An empirical study on SLA learners' intonation acquisition of general questions, focusing on the syllable number of the last word. *Journal of Hubei University of Education*, **28**, 94–97.

M. Taniguchi, & E. Abberton. (1999). Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speech, Hearing and Language: Work in Progress*, **11**, 76–89.

H. Tavakoli. (2013). *A dictionary of research methodology and statistics in applied linguistics*. Tehran: Rahanama Press.

The Personality, Emotion and Music Laboratory. (2017, July 30). The Profile of Music Perception Skills (PROMS). Retrieved July 30, 2017, from http://www.zentnerlab.com/psychological-tests/the-profile-of-music-perception-skills.

P. Trofimovich, & W. Baker. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language*

Acquisition, **28**, 1–30.

P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees, & N. Kraus. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, **10**, 420–422.

H. M. Xu. (2009). *A survey study of Chinese EFL learners' acquisition of English intonation: A functional perspective* (Unpublished master's thesis). China: Jiangsu University.

T. C. Zhao, & P. K Kuhl. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, **137**, 1452–1463.