Studies in Second Language Acquisition **41** (2019), 33–58
doi:10.1017/S0272263118000074

Research Article

OBSERVING AND PRODUCING PITCH GESTURES FACILITATES THE LEARNING OF MANDARIN CHINESE TONES AND WORDS

Florence Baills*

Universitat Pompeu Fabra

Nerea Suárez-González

Universitat Pompeu Fabra

Santiago González-Fuente

Universitat Pompeu Fabra

Pilar Prieto

Institució Catalana de Recerca i Estudis Avançats and Universitat Pompeu Fabra

Abstract

This study investigates the perception and production of a specific type of metaphoric gesture that mimics melody in speech, also called *pitch gesture*, in the learning of L2 suprasegmental features. In a between-subjects design, a total of 106 participants with no previous knowledge of Chinese were asked to observe (Experiment 1) and produce (Experiment 2) pitch gestures during a short

Florence Baills and Santiago González-Fuente are predoctoral researchers at the Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona. Pilar Prieto is an ICREA researcher and a professor at the Department of Translation and Language Sciences, Universitat Pompeu Fabra.

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 first author has a predoctoral research grant awarded by the Department of Translation and Language Sciences, Universitat Pompeu Fabra. The third author also acknowledges a FPU 2012-05893 grant awarded by the Spanish Ministry of Science and Innovation.

We are deeply grateful to all the people who patiently participated in the creation of the material for this experiment: Rita Zaragoza Jové, Chenjie Yuan, Feifei Li, and Yuan Yuan. Many thanks also to Joan Carles Mora for his comments and suggestions on the first draft of this article and to Joan-Borràs Comes for his crucial help all along.

*Correspondence concerning this article should be addressed to Florence Baills, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. E-mail: florence.baills@upf.edu

multimodal training session on Chinese tones and words. In both experiments they were tested on (a) tone identification and (b) word learning. Results showed the positive effect of a training session with pitch gesture observation compared to a training session without it (Experiment 1) and the benefits of producing gestures compared to only observing them and repeating the words aloud (Experiment 2). A comparison of the results of the two experiments revealed that there was no significant difference between the simple observation of pitch gestures and the production of speech accompanied by pitch gestures in facilitating lexical tone identification and word learning. Thus, both perception and production tasks with pitch gestures can be regarded as beneficial learning strategies for the initial stages of tones acquisition in the Chinese as a Second Language classroom.

LITERATURE OVERVIEW

Tonal languages like Mandarin Chinese, as opposed to intonational languages like English or Catalan, use pitch variations at the word level—that is, lexical tone contrasts—to distinguish meanings between otherwise segmentally identical words (Xu, 1994). For speakers of nontonal languages, acquiring these lexical tones has been shown to be particularly difficult (e.g., Kiriloff, 1969; Wang, Perfetti, & Liu, 2003b). Despite this intrinsic difficulty, there is evidence that speakers of both tonal and nontonal languages can be trained with success in both the perception and production of L2 tonal systems (e.g., Francis, Ciocca, Ma, & Fenn, 2008; Hao, 2012; Li & DeKeyser, 2017, among many others). Laboratory research has shown that learners of nontonal languages can be successfully trained to discriminate Mandarin tones by using short auditory tone training procedures consisting of paired combinations of tones both in perception (e.g., Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2003a; Wong & Perrachione, 2007) and in production (Wang et al., 2003a). Very recently, Li and DeKeyser (2017) showed the importance of specificity of practice in the learning of tones, in the sense that training in perception or production led to progress in only that skill area, not both. They found that after a three-day training session, participants who learned 16 Mandarin tone words in the perception condition obtained better results in perception posttasks, while participants trained in the production condition obtained better results in production posttasks.

In general, a challenge for educational research is to assess the procedures that can reinforce the teaching of a different prosodic system, such as the use of visualizers, gestures, or supporting transcription systems. In this respect, Liu et al. (2011) showed that having the support of visual illustrations depicting the acoustic shape of lexical tones (together with pinyin spelling of the spoken syllables) can help facilitate their acquisition. Research in gestures and second language acquisition has described the positive effects of observing iconic gestures on vocabulary learning (e.g., Kelly, McDevitt, & Esch, 2009, among others) as well as the positive effects of beat gestures on both L2 pronunciation learning and vocabulary acquisition (e.g., Gluhareva & Prieto, 2017; Kushch, Igualada, & Prieto, 2018, among others). However, little is known about the supportive use of gestures in learning pitch modulations in a second language, as well as potential differences between the benefits of perception and production practices. This study examines the role of *pitch gestures*, a specific type of metaphoric gesture that mimics melody in speech, in the learning of L2 tonal features, and focuses on the potential benefits of observing versus producing these gestures in the context of pronunciation learning.

MULTIMODAL CUES AND LEXICAL TONE PERCEPTION

Research in second language acquisition has shown that access to audiovisual information enhances nonnative speech perception in general (see Hardison, 2003, for a review). A series of studies have reported that when it comes to learning novel speech sounds, language learners benefit from training that includes both speech and mouth movements compared to just speech alone (e.g., Hardison, 2003; Hirata & Kelly, 2010; Wang, Behne, & Jiang, 2008). With respect to the learning of novel tonal categories, research has shown that having access to visual information about facial articulators has beneficial effects on tone perception for both tonal-language speakers in their native language (e.g., Burnham, Ciocca, & Stokes, 2001; Reid et al., 2015) and nontonal language speakers (e.g., Chen & Massaro, 2008; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Reid et al., 2015; Smith & Burnham, 2012). For example, Reid et al. (2015) tested the role of visual information on the perception of Thai tones by native speakers of typologically diverse languages, namely three tonal languages (Thai, Cantonese, and Mandarin), a pitch-accented language (Swedish), and a nontonal language (English). The results of a tone discrimination test in audio only (AO), audiovisual (AV), and visual only (VO) conditions showed a significant increase in tone perception when auditory and visual (AV) information was displayed together. Similarly, eyebrow movements (Munhall et al., 2004) and the visible movements of the head, neck, and mouth have been found to play a beneficial role in the perception of lexical tones (Chen & Massaro, 2008).

GESTURES AND L2 WORD LEARNING

It is becoming increasingly clear that cospeech gestures (i.e., the hand, face, and body movements that we produce while we speak) are an integral aspect of our language faculty and form an integrated system with speech at both the phonological (i.e., temporal) and semantic-pragmatic levels (e.g., Bernardis & Gentilucci, 2006; Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992, 2005). Concerning cospeech hand gestures in particular, there is ample evidence of the cognitive benefits of their use in educational contexts (e.g., Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Cook, & Mitchell, 2009). A growing body of experimental research in second language acquisition has shown that cospeech gestures can be used as an effective tool to help students improve their language skills (Gullberg, 2006, 2014; see Gullberg, deBot, & Volterra, 2008, for a review on gestures in L1 and L2 acquisition).

According to McNeill (1992), cospeech gestures comprise a broad category that includes iconic gestures, metaphoric gestures, deictic gestures, and beat gestures. Whereas iconic gestures use space to mimic concrete objects or actions (e.g., using one’s hand to form a spherical shape to represent a ball), metaphorical gestures use space to represent something abstract (e.g., fingers forming a heart shape to represent the idea of affection). Experimental and classroom research in the last few decades has stressed the benefits of observing (and producing) both iconic and metaphoric gestures for word recall in a first language and word learning in a second language. Kelly et al. (2009) reported that observing congruent iconic gestures was especially useful for learning novel words in comparison to observing the same content presented only in speech, or in speech associated with incongruent iconic gestures. In a study involving 20 French children (average age 5.5) learning English, Tellier (2008) asked them to learn eight common words (*house*, *swim*, *cry*, *snake*, *book*, *rabbit*, *scissors*, and *finger*). Four of the items were associated with a picture while the other four items were illustrated by a gesture that the children saw in a video and then enacted themselves. The results showed that the enacted items were memorized better than items enriched visually by means of pictures. In a recent study, Macedonia and Klimesch (2014) looked at the use of iconic and metaphoric gestures in the language classroom in a within-subject longitudinal study lasting 14 months. They trained university students to learn 36 words (nine nouns, nine adjectives, nine verbs, and nine prepositions) in an artificial language corpus. For 18 items, participants only listened to the word and read it. For the other 18 items, participants were additionally instructed to perform the gestures proposed by the experimenter. Memory performance was assessed through cued native-to-foreign translation tests at five time points. The results showed that enacting iconic gestures significantly enhanced vocabulary learning in the long run. Goldin-Meadow, Nusbaum, Kelly, and Wagner (2001) suggested that "gesturing may prime a speaker's access to a temporarily inaccessible lexical item and thus facilitate the processing of speech" (p. 521)—an idea consistent with the Lexical Retrieval Hypothesis proposed by Krauss, Chen, Gottesmen, and McNeill (2000; see also Krauss, Chen, & Chawla, 1996, for a review).

However, gestures need not be semantically related to words to boost word learning and recall. Studies investigating beat gestures (rhythmic hand gestures that are associated with prosodic prominence) have demonstrated that watching these gestures also favors information recall in adults (Kushch & Prieto, 2016; So, Sim Chen-Hui, & Low Wei-Shan, 2012) and children (Austin & Sweller, 2014; Igualada, Esteve-Gibert, & Prieto, 2017), as well as second language novel word memorization (Kushch et al., 2018).

PRODUCING VS. PERCEIVING GESTURES

Under the approach of embodied cognition, cognitive processes are conditioned by perceptual and motor modalities (Borghi & Caruana, 2015). In other words, any knowledge relies on the reactivation of external states (perception) and internal states (proprioception, emotion, and introspection) as well as bodily actions (simulation of the sensorimotor experience with the object or event to which they refer). Much research on this domain, especially in neuroscience, has shown brain activation of motor and perception networks when participants were engaged in different tasks involving abilities such as memory, knowledge, language, or thought (see Barsalou, 2008, for a review). By highlighting the importance of appropriate sensory and motor interactions during learning for the efficient development of human cognition, embodied cognition has crucial implication for education (see Kiefer & Trumpp, 2012;

Wellsby & Pexman, 2014, for reviews). We believe that hand gestures can be investigated from this perspective.

In general terms, the production of hand gestures by learners has been found to be more effective than merely observing them for a variety of memory and cognitive tasks (Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; for a review of the effects of enactment and gestures on memory recall, see Madan & Singhal, 2012). Goldin-Meadow et al. (2009) investigated how children extract meaning from their own hand movements and showed that children who were required to produce correct gestures during a math lesson learned more than children that produced partially correct gestures, who in turn learned more than children that did not produce any gestures at all. Furthermore, recent neurophysiological evidence seems to show that self-performing a gesture when learning verbal information leads to the formation of sensorimotor networks that represent and store the words in either native (Masumoto et al., 2006) or foreign languages (Macedonia, Müller, & Friederici, 2011). However, mere observation of an action without production also seems to lead to the formation of motor memories in the primary motor cortex (Stefan et al., 2005), which is considered a likely physiological step in motor learning. Stefan et al. (2005) contend that the possible engagement of the same neural mechanisms involved in both observation and imitation might explain the results of behavioral experiments on embodied learning. For example, Cohen (1981) tested participants on their ability to recall actions following training under three conditions: They either performed the actions, observed the experimenter performing the same actions, or simply heard and read the instructions for these actions. He found that participants remembered actions better when they were performed either by themselves or by the instructor than when the actions were simply described verbally. Notwithstanding, Engelkamp, Zimmer, Mohr, and Sellen (1994) showed that self-performed tasks led to superior memory performance in recognition tasks for longer lists of items (24–48 items) but not for shorter lists (12 items).

GESTURES AND L2 PRONUNCIATION LEARNING

Little is known about the potential benefits of using cospeech gestures in the domain of L2 pronunciation learning, and specifically the potential differences between the effectiveness of observing versus producing gestures in the L2 classroom. A handful of studies have focused on the potential benefits of observing cospeech gestures for pronunciation learning, with contradictory results. For example, Hirata and Kelly (2010) carried out an experiment in which English learners were exposed to videos of Japanese speakers who were producing a type of rhythmic metaphoric gesture to illustrate the Japanese short and long vowel phoneme contrasts, namely using a vertical chopping movement or a long horizontal sweeping movement, respectively. Their results showed that observing lip movements during training significantly helped learners to perceive difficult phonemic contrasts while the observation of hand movements did not add any benefit. The authors thus speculated that the mere observation of hand movement gestures might not have any impact on the learning of durational segmental contrasts. Hirata, Kelly, Huang, and Manansala (2014) explored specifically whether similar types of metaphoric gestures can play a role in the auditory learning of Japanese length contrasts. For this purpose, they carried out an experiment in which English speakers were trained to learn Japanese bisyllabic words by either observing or producing gestures that coincided with either a short syllable (one quick hand flick), a long syllable (a long horizontal sweeping movement), or a mora (two quick hand flicks). Basing themselves on a previous study (Hirata & Kelly, 2010), they hypothesized that producing beat gestures rather than merely observing them would enhance auditory learning of both syllables and moras. Although training in all four conditions yielded improved posttest discrimination scores, producing gestures seemed to convey no particular advantage relative to merely observing gestures in the overall amount of improvement, regardless of whether the gesture accompanied a syllable or a mora. All in all, the results reported by this line of work have shown that hand gestures do not make a difference when learning phonological contrasts like length contrasts in Japanese (but lips do).

By contrast, positive results of hand gestures have been documented for learning suprasegmental functions, for example, highlighting prosodic prominence of words within a sentence. A recent study with a pretest/posttest design by Gluhareva and Prieto (2017) found positive effects of observing beat gestures placed on prosodically prominent segments on pronunciation results in general. Catalan learners of English were shown rhythmic beat gestures (simple up-and-down or back-and-forth motions of the hands) that highlighted the relevant prosodic prominence positions in speech during pronunciation training. The instructor replicated naturally occurring cospeech gestures as much as possible, placing beat gestures on words that carried the most semantic and prosodic weight. After training, the participants who had observed the training with beat gestures significantly improved their accentedness ratings on a set of difficult items.

PITCH GESTURES

In this study we will focus on the effects of observing versus producing another type of cospeech gesture sometimes used by second language instructors, namely pitch gestures, on the learning of lexical tones in a second language. Pitch gestures (a term coined by Morett & Chang, 2015) are a type of metaphoric visuospatial gesture in which upward and downward hand movements mimic melodic high-frequency and low-frequency pitch movements. How can pitch gestures, frequently used in CSL (Chinese as a Second Language) classrooms, promote the learning of lexical tones? Experimental evidence has shown that pitch and space have a shared audiospatial representation in our perceptual system. The metaphoric representation of pitch was first investigated by Casasanto, Phillips, and Boroditsky (2003) in a nonlinguistic psychophysical paradigm. Native subjects viewed lines "growing" vertically or horizontally on a computer screen while listening to varying pitches. For stimuli of the same frequency, lines that grew higher were estimated to be higher in pitch. Along these lines, Connell, Cai, and Holler (2013) asked participants to judge whether a target note produced by a singer in a video was the same as or different from a preceding note. Some of the notes were presented with the corresponding downward or upward pitch gestures, while others were accompanied by contradictory spatial information, for example, a high pitch with a falling gesture. The results showed that pitch discrimination was significantly biased by the spatial movements produced in gesture, such that downward gestures induced perceptions that were lower in pitch than they really were, and upward gestures induced perceptions of higher pitch. More recently, Dolscheid, Willems, Hagoort, and Casasanto (2014) explored the link between pitch and space in the brain by means of an fMRI experiment in which participants were asked to judge whether stimuli were of the same height or shape in three different blocks: visual, tactile, and auditory. The authors measured the amount of activity in various parts of subjects' brain regions as they completed the tasks and found significant brain activity in the primary visual cortex, suggesting an overlap between pitch height and visuospatial height processing in this modality-specific (visual) brain area.

We therefore surmise that the strong cognitive links between the perception of pitch and visuospatial gestures can have an important application in the learning of melody in a second language.

PITCH GESTURES AND THE LEARNING OF TONAL WORDS AND INTONATION PATTERNS

Relatively little experimental work has been conducted thus far on the potential beneficial effects of pitch gestures on the learning of L2 tones and words in a tonal language. CSL teachers report that pitch gestures are commonly used in the classroom and that there may be variability in the gesture space used to allow more or less ample pitch movements, and in the articulators used to perform the pitch gesture, which can vary from the whole arm to a simple head movement. However, in all these gestures the spatial metaphor to describe pitch certainly remains the same.

Two longitudinal studies by Jia and Wang (2013a, 2013b) showed a positive effect of teachers' pitch gestures on the perception and production of tones by elementary-level learners of Mandarin. In a longitudinal study, Chen (2013) showed that 40 learners perceiving and producing "tonal gestures" (as he labeled them) seemed to have significantly superior communicative skills and performed significantly better in tonal production with a higher frequency of accurate responses, regardless of their tonal or nontonal background. Moreover, the learners displayed a wider pitch range when producing Mandarin words together with gesture. Nonetheless, Chen's study was a classroom training study with no experimental control of (a) the materials used in the training session, (b) the perception and production activities during training, and (c) the participants' language background.

To our knowledge, four recent experimental studies have been carried out on the potential benefits of pitch gestures on the learning of L2 tones and/or intonation, with positive results. Three of these studies dealt with the effects of observing pitch gestures. Hannah, Wang, Jongman, and Sereno (2016) looked at how pitch gestures affect nonnative Mandarin tone perception by testing 25 English speakers who listened to two monosyllabic words with the four tones under four conditions: audio-facial/congruent, audio-facial/incongruent, audio-facial-gestural/congruent, and audio-facial-gestural/incongruent. After each pair of words, participants had to immediately indicate whether they had heard a level, "mid-dipping," "rising," or "falling" tone. The authors found that participants in the audio-facial-gestural/congruent condition obtained significantly better scores in tone identification than participants in the audio-facial/congruent condition. In the second of these studies, Kelly, Bailey, and Hirata (2017) explored the effect of two types of metaphoric gestures on the perception of length and intonation features of Japanese phonemic contrasts by 57 English-speaking participants that had no previous knowledge of Japanese. They found that when visuospatial gesture depicting intonation were congruent with the auditory stimuli, accuracy was significantly higher than the control no gesture condition. Moreover, when the gesture was incongruent, accuracy was significantly lower than the control condition. The third study, by Yuan, González-Fuente, Baills, and Prieto (2018), tested whether pitch gesture observation would help the learning of difficult Spanish intonation pattern by 64 Chinese basic-level learners. Half of the participants received intonation training without gestures while the other half received the same training with pitch gestures representing nuclear intonation contours. Results showed that observing pitch gestures during the learning phase improved learner's production outcomes significantly more than training without gestures. By contrast, rather than focusing on observing, the fourth experimental study (Morett & Chang, 2015) tested the potential benefits of producing pitch gestures on the learning of L2 tones. In a between-subjects experimental design, 57 English speakers were divided into three groups and then trained to learn the meaning of 12 minimal pairs in Chinese. They had to repeat aloud the 12 Chinese words and imitate the gestures they saw performed by an instructor in a video in three conditions. One group of subjects saw and mimicked pitch gestures depicting the lexical tone pitch contours while hearing the Mandarin tones; the second group saw and mimicked gestures conveying word meanings (semantic gestures); and the third group were taught without gestures. Then participants were tested on a Mandarin lexical tone identification task and a word-meaning association task. The results showed that, in comparison with semantic gestures and no gestures, producing pitch gestures facilitated the learning of Mandarin words differing in lexical tone, but failed to enhance their lexical tone identification. These findings suggested that the visuospatial features of pitch gestures strengthen the relationship between English speakers' representations of Mandarin lexical tones and word meanings. However, the null results found in the lexical tone identification task challenge the belief that the production of pitch gestures can enhance lexical tone acquisition. Furthermore, because all participants in the gesture groups had to both observe and produce pitch gestures or semantic gestures (depending on the group) one cannot disentangle the potential effects of observing versus producing gestures. Thus, an open question that was not addressed by any of these four studies is whether it is observing or producing pitch gestures that has the stronger impact on L2 phonological acquisition.

GOALS AND HYPOTHESES

The present study represents the first attempt to compare the effects of observing versus producing pitch gestures on the initial learning of tones and lexical items in Mandarin Chinese. First, we aim to enrich the debate on embodied cognition by exploring the respective roles of observing and producing gestures. Second, on a more practical level, we would like to determine the most advantageous pedagogical approach for the teaching of lexical tones to beginning learners of Mandarin Chinese. The study comprises two complementary between-subjects experiments. While Experiment 1 investigates the effects of observing pitch gestures on learning tones and words in Mandarin Chinese, Experiment 2 investigates the effects of producing such gestures. In both experiments, subjects without any previous knowledge of a tonal language were randomly assigned to the Gesture (experimental) condition or the Non-Gesture (control) condition. Both experiments included two parts, an audiovisual perceptual tone training session with minimal pair combinations of the four Mandarin Chinese tones, and an audiovisual vocabulary training session focused on monosyllabic Mandarin Chinese words differing only in lexical tone. While after the tone-learning session, participants were asked to complete a lexical tones identification task, after the vocabulary training session they were asked to complete a word-meaning recall task and a word-meaning association task. First, based on previous findings, we predicted that observing pitch gestures would produce greater benefits for tone and word learning than not observing them, and second, given the literature on enactment and embodied learning, we predicted that the benefits of producing pitch gestures would be greater than the benefits of just observing them.

EXPERIMENT 1

The main goal of Experiment 1 was to assess the effect of pitch gesture observation on the learning of Chinese tones and words. The experiment consisted of a between-subjects training procedure with newly learned Chinese tones and words.

PARTICIPANTS

A total of 49 undergraduate and graduate students (age: M = 19.86, SD = 1.44; 15 males, 34 females) were recruited at the Communication Campus at the Universitat Pompeu Fabra in Barcelona, Spain. All participants were native speakers of Catalan and considered Catalan to be their dominant language relative to Spanish (mean percentage of Catalan in total daily language use = 72%, SD = .664). All were right-handed and reported no previous knowledge of Mandarin Chinese or any other tonal language. All had normal or corrected-to-normal vision and normal hearing. Participants were assigned to either the control No Gesture (NG) group or the experimental Gesture Observe (GO) group. In the NG condition, the instructors in the training video remained still and the participant remained still and silent while viewing the video. In the GO condition, the instructors in the training video performed gestures while teaching the tones and the participant remained still and silent while viewing the video. The groups were comparable in terms of the number of participants (24 in the NG group, 25 in the GO group), age (M = 19.88 in the NG group, M = 19.68 in the GO group), gender distribution (71% female, 29% male in the NG group and 68% female, 32% male in in the GO group), the amount of Catalan spoken in daily use (M = 72.8% in the NG group, M = 71.2% in the GO group), and results on a memory span test (M = 5,88 words in both groups). Participants were informed that the experiment consisted of an introductory tutorial on Mandarin Chinese tones and words and that they would learn how to pronounce the tones and some vocabulary. They were therefore unaware of the real purpose of the study. They signed a written consent form and received 10 euros for their participation.

MATERIALS

The experiment consisted of three consecutive phases, first a tone familiarization phase containing introductory information on Mandarin tones, then two consecutive training sessions, one focusing on tones and the other on vocabulary items, and finally the corresponding test tasks. As will be explained in the following subsections, audiovisual stimuli were prepared for use in the two training sessions and auditory items were prerecorded for the tone identification and word-meaning recall and word-meaning association tasks.

Audiovisual Materials for the Tone Familiarization Phase

All the audiovisual materials for the three phases of the experiment were recorded by a male native speaker of Chinese and a female bilingual Catalan-Chinese speaker. The video recordings were carried out at the experimental language research laboratory of the Universitat Pompeu Fabra's Department of Translation and Language Sciences using a PDM660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone. The two instructors were recorded against a white background and the video clips for all the recordings showed the speaker's face and the upper half of their body so that participants could see all hand and face movements.

With narration in Catalan, the familiarization video first illustrated the four Mandarin tones both verbally and visually with the help of the 4-scale diagram shown in Figure 1 (adapted from Zhu, 2012). Mandarin Chinese distinguishes between four main lexical tones which are numbered according to their pitch contours: high flat-level (tone 1), rising (tone 2), low falling and rising (tone 3), and high-falling (tone 4) (Chao, 1968). For example, the syllable <ma> can have four different meanings according to the tone used: <ma>1 means *mother*, <ma>2 means *hemp*, <ma>3 means *horse*, and <ma>4 means *scold*. Two different videos were produced for the habituation phase, one for the GO condition, the other for the NG condition. Both lasted around 8 minutes. The monosyllabic words presented in the familiarization phase were all different from the words in the subsequent training phase, and they were accompanied in the video by subtitles showing their orthographic transcription (generally in pinyin) and tones.

One instructor was a native Mandarin Chinese speaker and the other was an experienced CSL teacher for Catalan speakers. When performing the pitch gestures used in both the familiarization and training videos of the GO condition, the instructors used their right hand to gesture from left to right. They were also asked to produce the target words naturally while keeping their body and articulators like eyebrows, head, and neck totally still. Later the videos were digitally flipped to allow participants to observe the gestures from their left to their right. Figure 2 shows four stills from the videos illustrating the four target Mandarin tones (tones 1, 2, 3, 4) in the GO condition. Importantly, the two

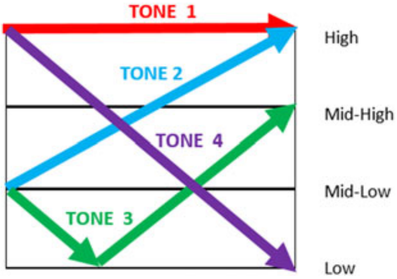

FIGURE 1. Diagram representing the four lexical tones in Mandarin Chinese.

instructors were trained to use clear visuospatial hand gestures, making sure that the hand movements accurately mimicked the pitch variations and the natural duration corresponding to each lexical tone. To do this, we relied on the visual pitch line obtained in Praat (Boersma & Weenink, 2017) for each word in the stimulus recordings. For spatial consistency across renditions, the imaginary space for the hand movements was divided into four areas: the high tonal range corresponded to the face level, the mid-tonal range to the shoulder level, the mid-high frequency range to the chest level, and finally the mid-low frequency range to the area of the hips. The duration of the tones, which can be a clue to determining what tone is being used, was left to the instincts of the instructors.

To guarantee that the speech characteristics in the NG and GO conditions would not differ, recordings of the same item in the two conditions were performed consecutively. Following González-Fuente, Escandell-Vidal, and Prieto (2015), mean pitch and duration cues were calculated for each speech file. Mean F0 was extracted from Praat for each item and computed in a Generalized Linear Mixed Model (GLMM) test using IBM SPSS Statistics 23 (IBM Corporation, 2015) to determine whether there were significant differences in speech duration between the NG and GO conditions. PITCH was set as the fixed factor and SUBJECT and TONE were set as random factors. Results reveal no significant differences of mean pitch between the two conditions. MSD (mean syllable duration, in ms) was calculated by dividing the total duration of the target sentence by the number of syllables. A GLMM test was run with TIME set as the fixed factor and SUBJECT and TONE set as random factors. We found a significant difference of duration [F(1, 66) = 5,134, p = .027], with speech in the GO condition lasting significantly longer than speech in the NG condition (M = 63 ms). Nevertheless, on the assumption that this difference was a consequence of the extra time required to produce the gesture in the GO condition, we decided not to modify the recorded stimuli in any way to keep the stimuli as natural as possible.

Audiovisual Materials for the Tone Training Session

A total of 36 monosyllabic Mandarin words (18 minimal pairs differing only in tone) were chosen as stimuli for both tone training phases (see Table 1). Words were selected

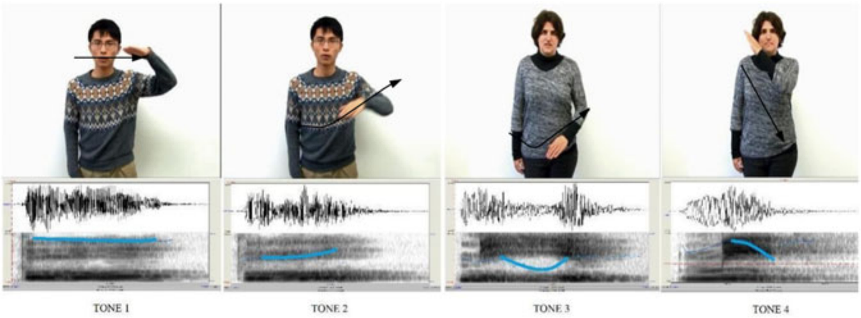

FIGURE 2. Screenshots illustrating the four target Mandarin tones in the Gesture Observe, with the corresponding sound waves and pitch tracks. The two left panels show the target syllable “puo” produced with tones 1 and 2 by the male speaker and the two rightmost panels show the target syllable “mi” produced with tones 3 and 4 by the female speaker.

so that all minimal pair words shared the same phonological shape (except for tone) and the same grammatical category. There were a total of 5 pairs of verbs, 10 pairs of nouns, and 3 pairs of adjectives. All words conformed to the phonotactic restrictions of Catalan (Prieto, 2004) to avoid additional difficulty. The words were presented in orthographic form following the pinyin orthographic conventions, except when this would cause difficulty for Catalan speakers (the forms in brackets in Table 1 are the forms that participants were shown).

The 36 stimuli were recorded and presented in pairs to heighten contrast perception (Kelly, Hirata, Manansala, & Huang, 2014). In total, each of the four lexical tones was repeated nine times. The video recordings for the GO condition were produced with pitch gestures and the video recordings for the NG condition were produced without. After recording, the videos were edited using Adobe Premiere Pro CC 2015 software to produce six videos in which the six tonal contrasts (each composed of three pairs of stimuli) were put into sequences in different orders to avoid primacy and recency effects (i.e., each video started and ended with different pairs of tonal stimuli).

Audiovisual Materials for the Vocabulary Training Session

A total of 12 targets were selected from the list of words in Table 1 (the minimal pairs selected appear in bold), which consisted of six minimal pairs of words differing only in their lexical tones. In each pair, Catalan translations of the two words were matched for mean log frequency per million words using NIM, an online corpus search tool that is

TABLE 1. Pairs of stimuli for the tone training and vocabulary training sessions (18 pairs; 36 words)

Tonal Contrast	Pinyin	English	Tonal Contrast	Pinyin	English
1–2	**bō**	**wave**	2–4	**má**	**linen**
	bó [puo]	**uncle**		**mà**	**insult**
	chī	eat		ná	take
	chí [txi]	pool		nà	sodium
	fā	send		lí	pear
	fá	raft		lì	chestnut
2–3	fú	fortune	1–4	tī	stairs
	fǔ	axe		tì [thi]	shave
	bí	nose		pō	slope
	bǐ [pi]	pen		pò [phuo]	spirit
	tá	**battery**		**gē**	**song**
	tǎ [tha]	**tower**		**gè [ke]**	**piece**
1–3	tū	bald	3–4	**mǐ**	**rice**
	tǔ [thu]	soil		**mì**	**honey**
	dī	**taxi**		lǔ	prisoner
	dǐ [ti]	**background**		lù	deer
	chū	first		gǔ	drum
	chǔ [txu]	storage		gù [ku]	hire

Note: When the orthographic form of the syllable presented to the participants differed from the pinyin orthography, the orthographic form is specified here within brackets. In bold, the words selected for the vocabulary training.

useful for establishing word frequencies in Spanish, Catalan, or English (Guasch, Boada, Ferré, & Sánchez-Casas, 2013). The target minimal pairs were video-recorded in consecutive pairs following the same procedure described for the materials used for tone training. After the recordings, the pairs of stimuli were edited using Adobe Premiere Pro CC 2015 software. Items were repeated in randomized order within three blocks. In total, participants ended up seeing and hearing each vocabulary item (Catalan meaning + Chinese word) a total of three times. Six different videos containing the trials in different orders were created and distributed among the participants to avoid any primacy or recency effects.

Auditory Materials for the Test Tasks (Tone Identification, Word-Meaning Recall, and Word-Meaning Association Tasks)

For the tonal identification task, eight items (four pretrained: "mì," "fŭ," "txí," "dī"; four new: "té," "nù," "lā," "txě") were chosen as real syllables or pseudosyllables respecting Catalan phonotactic rules. Auditory materials were recorded by three native speakers of Mandarin Chinese, two of them male and one female, who were not the instructors. The files were then uploaded on an online survey builder (https://www.surveygizmo.com) that automatically randomized the order of items.

For the vocabulary tests (word-meaning recall and word-meaning association tasks) the 12 items from the training session were used. The recordings featured a speaker of a different sex than in the training session to ensure that posttest performance reflected learners' ability to identify Mandarin lexical tones across word tokens rather than their recall of the specific token produced during the learning phase.

PROCEDURE

Participants were tested individually in a quiet room. They were randomly assigned to one of the two between-subjects groups, 24 in the NG condition and 25 in the GO condition. Participants were asked to sit in front of a laptop computer equipped with earphones and mark their answer to the tone identification task on a sheet of paper next to the computer. First, a word memory span test (Bunting, Cowan, & Saults, 2006) adapted to the Catalan language was carried out to control for short-term working memory capacity. After completing the memory span task, participants in both conditions were instructed to remain silent and listen carefully to the audiovisual stimulus recordings as they played back on the computer. Participants in the experimental (GO) group were additionally asked to pay attention to the gestures conveying the melodic movements. No feedback was provided at any point during the experimental tasks.

As mentioned previously, the experiment consisted of three phases (see Figure 3). In the familiarization phrase, participants were presented with a video consisting of a short introduction to the Chinese tones (8 min). After this, participants went on to view the tone training video (5 min), which was followed by a tone identification task (10–12 min). Finally, participants were shown the vocabulary training video (6 min), which was followed by two tasks, namely a word-meaning recall task and a word-meaning association task (15–20 min).

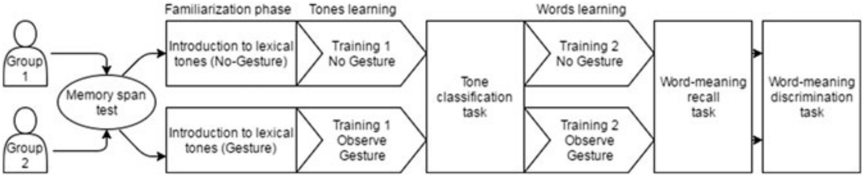

FIGURE 3. Experimental procedure for Experiment 1.

Tone Training and Tone Identification Task

After familiarization, participants were trained to discriminate between pairs of Mandarin Chinese lexical tones. The tone training video contained a total of 18 units which each consisted of pairs of target tones (see Table 1). Within each unit, participants were exposed to the following sequence (see Figure 4): (a) the target pairs of tones to be discriminated; (b) the orthographic form of the pairs of Mandarin words together with their tone marks; and (c) the pair of video clips of these words as produced by the instructor.

Immediately after viewing the training video, participants were asked to complete a tone identification task by listening to eight audio-only items. They were instructed to listen to the syllable and then write down what they had heard together with the correct tone mark. They could only listen to the syllable once. When they finished writing their answer, they had to go to the next screen to listen to the next syllable. The answers were afterward coded as 0 if the tone mark was incorrect or 1 if it was correct, regardless of the orthographic form of the word written by the participants.

Vocabulary Training and Word-Meaning Recall Tasks

In the vocabulary training session, participants were asked to learn 12 words. The vocabulary training video contained a total of six units each containing minimal pair words, which were presented in three consecutive blocks with the stimuli in different orders. In total, they listened to the same stimuli three times. Within each unit, and for each of the word pairs, they were exposed to the following temporal sequence (see Figure 5): (a) the orthographic form of the Catalan word corresponding to the target Mandarin word to be learned; and (b) the video clip with the target word as produced by the instructor.

After they had viewed the training video, participants carried out a word-meaning recall task in which they were instructed to listen to the 12 target Mandarin Chinese words and translate each one of them into Catalan. They could only listen to each word once before writing down their answer and then going on to the next screen to listen to the next word. Subsequently, they carried out a word-meaning association task involving the same 12 Mandarin words. Here they were shown the Catalan translations of the two words of a minimal pair but only heard one of the two and were asked to select the correct translation.

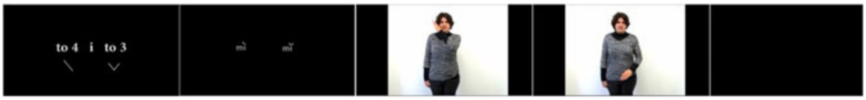

FIGURE 4. Example of a trial sequence of the tone training video in the Gesture Observe condition involving tones 4 and 3 over the syllable “mi”.

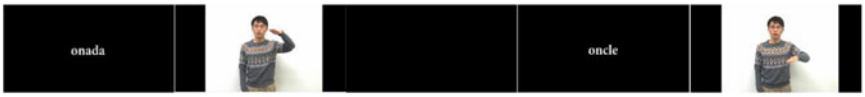

FIGURE 5. Example of a unit sequence during the vocabulary training session in the Gesture Observe condition with the minimal pair of Mandarin Chinese words bō “Cat. onada - Eng. wave’” and bó “Cat. oncle—Eng. uncle”.

STATISTICAL ANALYSES

A total of 392 experimental responses were obtained (49 participants × 8 tone identification questions) for the tone identification task and a total of 588 responses were obtained (49 participants × 12 words) for each of the word-learning tasks. Statistical analysis of the results of the three tone and vocabulary tasks (e.g., the tone identification task, the word-meaning recall task, and the word-meaning association task) was carried out using IBM SPSS Statistics v. 24 (IBM Corporation, 2016) by means of three GLMM. Results of the memory span task revealed that all participants behaved within a normal range in short-term working memory capacity (M = 5.88 items remembered, SD = .712), and thus all of them were included in the analysis.

In each of the three models, ACCURACY of response was set as the dependent variable (two levels: Correct vs. Incorrect), which was modeled with a Binomial distribution and a Logit link. CONDITION (two levels: NG vs. GO) was set as a fixed factor. One random effects block was specified, in which we controlled for subject intercept, with the type of tone defined as a random slope (covariance type: variance components).

RESULTS

The GO group scored higher than the NG group in the three tasks (see Table 2) and for all four tones. Results of the three GLMM models revealed a significant main effect of CONDITION in the tone identification task, F(1, 390) = 3.890 (β = .657, *SE* = .333, *p* = .049, *Exp*(β) = 1.929), in the word-meaning recall task, F(1, 586) = 4.789 (β = .683, *SE* = .312, *p* = .029, *Exp*(β) = 1.980), and in the word-meaning association task, F(1, 586) = 10.365 (β = 1.043, *SE* = .324, *p* = .001, *Exp*(β) = 2.834), meaning that the GO experimental group significantly outperformed the NG control group in all three tasks. Calculating odd ratios (*Exp*(β), reported previously) is a reliable method to analyze effect sizes with logistic regressions. Odd ratios represent the odds that an outcome will occur given a particular exposure, compared to the odds of

TABLE 2. Means and standard deviations of accuracy (based on accuracy means per participant) for the three tasks in Experiment 1

	No Gesture		Gesture Observe	
	M	SD	M	SD
Tone identification	.70	.19	.80	.23
Word-meaning recall	.49	.22	.64	.25
Word-meaning association	.74	.19	.89	.12

the outcome occurring in the absence of that exposure (Szumilas, 2010). Odd ratios superior to 1 are associated with higher odds of outcome. In the three tasks, the GO condition received a much higher probability of obtaining more accurate values than the NG condition (specifically, compared to the NG control condition, the odds of obtaining correct answers is 1.929 higher in the GO condition in the tone identification task, 1.980 higher in the word-meaning recall task, and 2.834 higher in the word-meaning association task).

In sum, results of the tone identification task show that observing pitch gestures significantly improved tonal perceptual learning in participants without any prior knowledge of Mandarin Chinese. Similarly, results from the two word-learning tasks demonstrate that a short vocabulary training session in which they observe pitch gestures may enhance L2 students' vocabulary learning in a tonal language like Chinese, at least at an initial stage of learning, and thus confirm the role of merely observing this specific type of gesture regarding the learning of words with tones.

EXPERIMENT 2

The main goal of Experiment 2 was to assess the effect of pitch gesture production on the learning of Chinese tones and words. The experiment consisted of a between-subjects training procedure with newly learned Chinese tones and words.

PARTICIPANTS

Fifty-six undergraduate students (age M = 19.93 years, SD = 1.414; 9 males, 47 females) were recruited at the Universitat Pompeu Fabra in Barcelona, Spain. None of them had been subjects in Experiment 1. All were native speakers of Catalan and considered Catalan to be their dominant language relative to Spanish (mean percentage of Catalan in total daily language use = 68.4%, SD = .794). All of them reported no previous knowledge of Mandarin Chinese or any other tonal language.

In the control group (No Gesture Produce condition, henceforth NGP), the instructors in the training video performed gestures while teaching the tone words and the participant was instructed to repeat the tone words after the instructor but not to perform any hand movement. In the experimental group (Gesture Produce condition, henceforth GP), the instructors in the training video performed gestures while teaching the tone words and the participant was instructed to repeat the tone words after the instructor and at the same time mimic the gesture performed by the instructors. The rationale for adding a control group where participants had to produce speech was that it required some form of active learning, which would be more accurately comparable to a condition where participants have to produce gestures. The groups were comparable in terms of the number of participants (28 in the NGP group, 28 in the GP group), age (M = 19.71 in the NGP group, M = 20.14 in the GP group), gender distribution (81% female, 19% male in the NGP group and 86% female, 19% male in the GP group), the amount of Catalan spoken in daily use (M = 67.8% in the NGP group, M = 68.6% in the GP group), and results on the memory span test (M = 5.54 words in the NGP group, M = 5.66 words in the GP group). They went through the same preliminary steps as in Experiment 1.

TABLE 3. Means and standard deviations of accuracy (based on accuracy means per participant) for the three tasks of Experiment 2

	No Gesture Produce		Gesture Produce	
	M	SD	M	SD
Tone identification	.59	.21	.72	.25
Word-meaning recall	.40	.21	.57	.22
Word-meaning association	.74	.17	.82	.14

MATERIALS

In Experiment 2, observing pitch gestures only was compared with observing and producing those gestures. Therefore, the video stimuli were the same for both conditions and identical to those used in the GO condition of Experiment 1 except that the instructions were different. Here, in both control group and experimental group, participants were instructed to repeat the Mandarin words they heard spoken by the instructor on the video. However, those in the GP condition were additionally instructed to mimic the pitch gestures illustrated by the instructor with their own right hand as they heard and repeated them (as in Kelly et al., 2014). To allow them enough time to repeat the Mandarin word and produce the gesture, a 5-second black screen followed the modeling of each tone by the instructors in the video. In the two conditions, the training video was the same, but participants were asked to respond differently.

PROCEDURE

As in Experiment 1, Experiment 2 consisted of three phases. In the initial familiarization phase, the experimenter initially informed the participants about the general procedure of the training session, after which they were presented with a short video introducing the Chinese tones (8 min). Here, they were also familiarized with the pitch gestures by repeating two monosyllabic items for each tone, for a total of eight familiarization items. In the NGP condition, they were asked to repeat the word and pay attention to the gesture, while in the GP condition, they were asked to repeat the word and mimic the pitch gesture. There was no feedback on the pronunciation of the tones; however, at this stage, the experimenter could offer some feedback on the production of gesture if needed. Next, they viewed a tone training video (8 min), which was followed by a tone identification task (10–12 min). They then watched a vocabulary training video (9 min), which was followed by two tasks, a word-meaning recall task and a word-meaning association task (15–20 min).

In the NGP condition, for each minimal pair they were first presented with the two Chinese syllables in written form, and then heard the instructor produce the target syllable with both tones and the corresponding gestures. When the screen subsequently went black they had to repeat the syllable aloud only.

Participants in the GP condition watched the same video as in the NGP condition and repeated the target syllables; additionally, however, they were asked to copy and perform the pitch gesture.

Accuracy of speech during the training was not measured and no feedback was provided during the training. However, the experimenter was present in the room and could thus make sure that the participants were performing the gestures/speech appropriately depending on the condition.

STATISTICAL ANALYSES

A total of 416 responses were obtained (26 participants × 2 conditions × 8 tone identification questions) for the tone identification task and a total of 624 responses were obtained (26 participants × 2 conditions × 12 words) for each of the word-learning tasks. Statistical analysis of those results (tone identification task, word-meaning recall task, and word-meaning association task) was carried out using IBM SPSS Statistics v. 24 (IBM Corporation, 2016) by means of three GLMMs. Results of the memory span tasks revealed that all subjects behaved within a normal range in short-term working memory capacity (M = 5.88 items remembered, SD = .712), and thus the experimental data from all of them were included in the analysis.

In each of the three models, ACCURACY of response was set as the dependent variable (two levels: Correct vs. Incorrect), which was modeled with a Binomial distribution and a Logit link. CONDITION (two levels: NGP vs. GP) was set as a fixed factor. One random effects block was specified, in which we controlled for subject intercept, with the type of tone defined as a random slope (covariance type: variance components).

RESULTS

The GP group scored higher than the NGP group in the three tasks (see Table 3).

The results of the three GLMM models revealed a significant main effect of CONDITION in the three models, namely in the tone identification task, F(1, 446) = 4.550 (β = .769, *SE* = .331, *p* = .033, *Exp(*β*)* = 2.158), in the word-meaning recall task, F(1, 670) = 7.360 (β = .827, *SE* = .305, *p* = .007, *Exp(*β*)* = 2.287), and in the word-meaning association task, F(1, 670) = 4.237 (β = .535, *SE* = .260, *p* = .040, *Exp(*β*)* = 1.707), indicating that the GP experimental group outperformed the NGP control group in the learning of both tones and words. In the three tasks, the GP condition received a much higher probability of obtaining more accurate values than the NGG condition (specifically, compared to the NGG control condition, the odds of obtaining correct answers is 2.158 higher in the GP condition in the tone identification task, 2.287 higher in the word-meaning recall task, and 1.707 higher in the word-meaning association task).

All in all, the results revealed that the group of participants who produced the pitch gestures performed significantly better in all three tasks, namely the tone-learning task and both word-learning tasks. Note that our results partially contrast with those obtained by Morett and Chang (2015), who did not find that producing pitch gestures significantly helped lexical tone identification compared to other types of gesture. However, results from the vocabulary tasks support Morett and Chang's (2015) results on the role of pitch gestures in vocabulary learning.

Comparing the effects of pitch gesture observation and pitch gesture production to further compare perception and production of gestures, we statistically compared the effects of passively observing pitch gestures with the effects of a more "enacted" training condition, that is, observing pitch gestures and additionally mimicking them while repeating the tonal words. Because the training procedures and tone perception and vocabulary tests were the same in every other respect across both experiments, we set out to perform a direct comparison between the GO condition from Experiment 1 and the GP condition from Experiment 2.

As before, we ran three GLMMs, one for each dependent variable, that is, the proportion of correct responses in the tone identification task, the word-meaning recall task, and the word-meaning association task. In each of the three models, ACCURACY of response was set as the dependent variable (two levels: Correct vs. Incorrect), which was modeled with a Binomial distribution and a Logit link. CONDITION (two levels: GO vs. GP) was set as a fixed factor. One random effects block was specified, in which we controlled for subject intercept, with the type of tone defined as a random slope (covariance type: covariance components). Results of the GLMM did not reveal any significant main effect of CONDITION in any of the tasks.

Given these results, it is necessary to explore why the benefit of producing pitch gestures seen in Experiment 2 is no longer visible when data from the two experiments are compared. The main difference between the experimental conditions of Experiment 1 (GO) and the control condition of Experiment 2 (NGP) being the production of speech, we compared the scores in these conditions and found that the NGP group had significantly lower scores than the GO group in the tone identification task, F(1, 422) = 14.724 (β = - 1.236, *SE* = .322, *p* = .000, *Exp(*β*)* = 0.290), in the word-meaning recall task, F(1, 634) = 10.604 (β = -1.132, *SE* = .348, *p* = .001, *Exp(*β*)* = 0.322), and in the word-meaning association task, F(1, 634) =12.198 (β = -1.035, *SE* = .296, *p* = .001, *Exp(*β*)* = 0.355). Therefore, it seems that repeating the tonal words while watching the gesture during the training had a negative outcome on scores in all the tasks.

DISCUSSION AND CONCLUSIONS

The present study has added more evidence in favor of the use of pitch gestures to learn tones in a second language and, crucially, has assessed the potential differences between gesture perception and production in facilitating tone and word learning. The study comprised two experiments that examined whether the learning of Mandarin lexical tones and words would be enhanced by: (a) a short training session where participants merely observe pitch gestures (Experiment 1) or (b) a short training session where participants mimic pitch gestures (Experiment 2). The results demonstrated that both the observation and the production of pitch gestures showed a beneficial effect in subsequent tone-learning and word-learning test tasks in comparison with the control nongesture condition. Specifically, while the results of Experiment 1 demonstrated that a short training session involving observing pitch gestures enhanced the acquisition of Mandarin Chinese tones and words more than a comparable short training session without gestures, the results of Experiment 2 showed that a short training session in which subjects produced pitch gestures while repeating the words enhanced the acquisition of Mandarin Chinese tones and words more than just observing the gestures and repeating the words.

The results of our study add more evidence in favor of the benefits of pitch gestures for learning L2 tones and intonation (Hannah et al., 2016; Kelly et al., 2017; Morett & Chang 2015; Yuan et al., 2018). Specifically, our results partially replicate and extend the findings by Morett and Chang (2015). Their experimental results showed that while the production of pitch gestures by participants facilitated the learning of words differing in lexical tones in Mandarin Chinese, they failed to enhance lexical tone identification performance. By contrast, our results showed an amplified effect of pitch gestures in that not only producing but also just observing pitch gestures triggers an enhancement of both tone identification and word-learning scores. These experimental results support the findings from Chen's (2013) longitudinal classroom study, where students who saw and used gestures were more accurate in answering their instructors' tonal queries than students taught with the traditional 5-scale tone chart (Chao, 1968), and the findings seen in Jia and Wang (2013a, 2013b).

In more general terms, these results add more evidence about the importance of using different types of supporting gestures for L2 instructional practices. As we have seen before, semantically related iconic gestures have also been found to enhance novel word acquisition (Kelly et al., 2009; Macedonia et al., 2011; Tellier, 2008; Thompson, 1995). However, pitch gestures do not convey semantic information per se. So why is it that they produce these beneficial effects?

We believe that the metaphorical visuospatial properties of pitch gestures are visually encoding one of the essential phonological features of words in a tonal language, namely their lexical tone. It is presumably the enrichment of these phonological properties through visual means that provides a positive supporting channel for the acquisition of novel words in tonal languages. Moreover, the benefits of pitch gestures for tone identification provide further evidence for theories claiming that pitch perception is fundamentally audio-spatial in nature (e.g., Cassidy, 1993; Connell et al., 2013; Dolscheid et al., 2014) as well as supporting the spatial conceptual metaphor of pitch (Casasanto et al., 2003).

In contrast with the positive results obtained in various studies on the role of pitch gestures on the acquisition of second language tones or intonation (Hannah et al., 2016; Kelly et al., 2017; Morett & Chang, 2015; Yuan et al., 2018), there is to date no clear view on how other types of metaphoric (and beat) gestures affect phonological learning. In contrast with the positive effects of pitch gestures for learning L2 tones and intonation, the results of studies targeting the effectiveness of what are called "length gestures" to learn duration contrasts in a second language are not so clear. In various studies, Kelly, Hirata, and colleagues (Hirata & Kelly, 2010; Hirata et al., 2014; Kelly & Lee, 2012; Kelly et al., 2017) have explored the role of two types of gestures that metaphorically map the duration of a vowel sound in Japanese duration contrasts without thus far detecting any positive effects. For example, Hirata and Kelly (2010) investigated the role of cospeech gesture perception in the auditory learning of Japanese vowel length contrasts. In the study, participants were exposed to videos of Japanese speakers producing Japanese short and long vowels with and without hand gestures that were associated with vowel length. A short vertical chopping movement was used to mark short vowels and a long horizontal sweeping movement was used to mark long vowels. The results of the experiment showed that there was no noticeable benefit for participants when they learned vowel length by viewing videos showing length gestures as opposed to viewing videos that did not show such gestures. More recently, Kelly et al. (2017) suggested that it may be possible to safely narrow down the effective use of perhaps the utility of visuospatial gestures in pronunciation learning is limited to the use of pitch gestures for the learning of intonation patterns (but not thus excluding the use of various types of metaphoric gestures for the study of duration). There might be a set of several possible reasons that can explain for the discrepancy between the results of the previously mentioned studies. First, as Kelly et al. (2017) noted, pitch gestures tend to have a stronger effect on learning L2 pitch differences than length/duration gestures on learning durational differences. Indeed, Kelly et al. (2017) explored the potential differences in the effect of length and pitch gestures on learning length and pitch phonological contrasts, respectively. In this study, English-speaking adult participants were exposed to videos with a trainer producing Japanese length contrasts and sentence-final intonation distinctions accompanied by congruent metaphoric gestures, incongruent gestures, or no gestures. The results showed that for intonation contrasts, congruent metaphoric gestures (i.e., pitch gestures) had a positive effect, as identification was more accurate in comparison to other conditions. For the length contrast identification, however, similar results were not obtained, and no clear and consistent pattern emerged. In fact, the use of congruent metaphoric gestures seemed to make length contrast identification more difficult.

We would like to suggest that the type of metaphorical gestures used by Kelly, Hirata, and colleagues (Hirata & Kelly, 2010; Hirata et al., 2014; Kelly & Lee, 2012; Kelly et al., 2017) may have had an influence too. Specifically, the mora gestures used in the studies of Kelly, Hirata, and colleagues (e.g., the short vertical chopping movements) might have come across as "nonintuitive" to English speakers and thus did facilitate (or even hindered) their learning of durational information in the second language (see also the comments on the lack of effectiveness of length gestures in Kelly et al., 2017). The fact that other studies like Gluhareva and Prieto (2017) have found that observing other types of rhythmic gestures (e.g., beat gestures) has a positive effect on general pronunciation results leads us to suspect that perhaps the pitch gestures must seem natural to have positive results.

Another goal of the present study was to compare the effects of observing versus producing pitch gestures on learning Chinese tones and words. Results from a variety of studies have suggested that the production of gestures by the learners is more effective than observing them alone in various learning contexts (Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; Macedonia et al., 2011; Masumoto et al., 2006; Saltz & Donnenwerth-Nolan, 1981). Regarding pitch gestures specifically, Morett and Chang (2015) did explore their effect, but all the participants in their study had to perform pitch gestures, and thus the study could not disentangle the potential effects of observing versus producing gestures. In our data, a comparison of results from the GO group in Experiment 1 and the GP group from Experiment 2 revealed that training with mere observation and training with production of both speech and gesture had equally beneficial effects in both tone- and word-learning tasks. One explanation for this effect can be the specificity of practice effect explored by Li and De Keyser (2017). Their study provides strong evidence that tone-word perception and production skills each depend on the practice used to develop them. In our experiments, the tasks used to evaluate participants' acquisition of tones in Mandarin after training exclusively targeted perception, which may explain why the results obtained from the GO group were as good as those obtained from the GP group, and why the results from the NGP group were so low.

Another explanation could be related to the effects of using gesture on the speaker's cognitive load. Whereas some studies have suggested that gestures help reduce the cognitive load or processing cost by conveying the same message through an additional modality (Goldin-Meadow, 2011; Wagner, Nusbaum, & Goldin-Meadow, 2004) and thus function as a compensatory and facilitating device in the acquisition of a second language (Gullberg, 1998; McCafferty, 2002), other studies have found that when learning higher aspects of a L2 such as semantics, syntax, or phonetics, observing (Kelly & Lee, 2012) and producing gestures (Kelly et al., 2014; Post, Van Gog, Paas, & Zwaan, 2013) only helps when cognitive demands are not too high, otherwise becoming counterproductive and/or distracting.

In our study, participants in the NGP group might have experienced such a cognitive overload. It seems reasonable to think that for participants with no previous knowledge of Chinese, having to learn new words while having to repeat them and at the same time not mimic the target pitch gestures might be a demanding task. This may be borne out by the fact that the mean accuracy for the GO group (Table 2, Experiment 1) was much higher than the NGP group (Table 3, Experiment 2). In other words, repeating the words while seeing the words produced with pitch gesture was altogether the less effective strategy to learn both tones and words. These results may be interpreted as the consequence of a disconnection between the perceptive modality (seeing the gesture) and the productive modality (repeating speech). Because gesture and speech are highly integrated and interdependent, it is possible that this disconnection produced cognitive overload.

In general, the evidence reported in this article adds to the growing body of evidence in favor of using gestures in vocabulary and pronunciation learning, thus reinforcing the embodied cognition paradigm. This paradigm theorizes that the human perceptual and motor system play an important role in cognition and underlines the importance of body movements and multimodal supporting channels in cognition and in favoring memory traces (see Barsalou, 2008; Barsalou, Simmons, Barbey, & Wilson, 2003; Paivio, 1990). According to the dual coding theory (Paivio, 1990), learning is reinforced when the visual modality is added to the verbal modality. Dual coding theory supports the idea that multimodal memory traces are richer and stronger than unimodal traces that result from either the visual or verbal modality alone. Empirical evidence that mere observation of an action, like in our GO group, leads to the formation of motor memories in the primary motor cortex supports the predictions made by these theories (Stefan et al., 2005), in the sense that the addition of visual information to verbal information should create stronger memory traces.

LIMITATIONS AND FUTURE DIRECTIONS OF RESEARCH

Several limitations of this study can be identified. First, in Experiment 1, the slight increase in duration found in the auditory signal of the training items corresponding to the GO condition (mean of + 63 ms) may play a role to some extent in the positive results favoring the tones' acquisition in the gesture observation condition. Therefore, further research could try to assess the mechanisms behind the effects of gesture, that is, whether the gesture alone could obtain an effect, or whether it is both the auditory and the visual properties (e.g., the auditory signal that is naturally modified by the production of the gesture) that are responsible for the effect.

Though our results confirm that pitch gestures can be useful for learning Chinese tones at a basic level (our participants were completely new to Mandarin Chinese), our study cannot tell whether pitch gestures will have such strong effects with more proficient learners. It would be very interesting to test the effectiveness of pitch gestures using more complex phrasal contexts such as two-syllable words and with participants that have some prior knowledge of Mandarin Chinese.

Another limitation of the study lies in the lack of a productive task in the posttests. Indeed, it would have been helpful to verify the specificity of practice effect suggested by Li and DeKeyser (2017) by exploring whether participants in the GP condition showed any advantage in productive tasks. Finally, it would be interesting to assess more precisely the respective roles of perceiving versus producing pitch gestures and determine how to use this information best to achieve particular pedagogical goals.

These limitations notwithstanding, our study shows that, at least for initial levels of L2 tone learning, observing or producing pitch gestures can be equally effective to help students perceive the tones of the target language and learn new tonal words. From a pedagogical perspective, our findings support the use of teaching and learning methods that implement more active audio-visual and embodied cognition strategies in the second language classroom. On this basis, for example, teachers of CSL could use pitch gestures while teaching the tones for the first time or, when teaching a new word, asking learners to pay attention to the gesture while listening to the word, therefore enhancing discrimination abilities and memorization. Once learners have gained some knowledge of Chinese tones and tonal words and have observed the teacher performing pitch gestures, the teacher could ask them to repeat the words accompanied with the pitch gesture to practice oral skills. Though more applied research is clearly needed, these results constitute an incentive to start implementing more effective multimodal approaches in the CSL classroom.

REFERENCES

Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*, 92–103.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W., Simmons, W. K., Barbey, A., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.

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

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer software]. Version 6.0.33. Retrieved from http://www.praat.org/.

Borghi, A. M., & Caruana, F. (2015). Embodiment theory. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., pp. 420–426). Amsterdam, The Netherlands: Elsevier.

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*, 1691–1700.

Burnham, D., Ciocca, V., & Stokes, S. (2001). Auditory-visual perception of lexical tone. In P. Dalsgaard, B. Lindberg, H. Benner, & Z.-H. Tan (Eds.), *Proceedings from Eurospeech 2001: 7th European Conference on Speech Communication and Technology* (pp. 395–398). Aalborg, Denmark: Center for Personkommunikation.

Casasanto, D., Phillips, W., & Boroditsky, L. (2003). Do we think about music in terms of space? Metaphoric representation of musical pitch. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science*, *10*, 1323.

Cassidy, J. W. (1993). Effects of various sightsinging strategies on non-music majors' pitch accuracy. *Journal of Research in Music Education*, *41*, 293–302.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, C. M. (2013). Gestures as tone markers in multilingual communication. In I. Kecskes (Ed.), *Research in Chinese as a second language* (pp. 143–168). Boston, MA, and Berlin, Germany: De Gruyter Mouton.

Chen, T. H., & Massaro, D. W. (2008). Seeing pitch: Visual information for lexical tones of Mandarin-Chinese. *The Journal of the Acoustical Society of America*, *123*, 2356–2366.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*, 267–281.

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

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*, 1047–1058.

Dolscheid, S., Willems, R. M., Hagoort, P., & Casasanto, D. (2014). The relation of space and musical pitch in the brain. *The 36th Annual Meeting of the Cognitive Science Society*, *3*, 421–426.

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory and Cognition*, *22*, 34–39.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*, 268–294.

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

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

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 595–607.

Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 1–11.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science: A Journal of the American Psychological Society*, *20*, 267–272.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science: A Journal of the American Psychological Society*, *12*, 516–522.

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

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, *45*, 765–771.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish*. Lund, Sweden: Lund University Press.

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

Gullberg, M. (2014). Gestures and second language acquisition. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body, language, communication: An international handbook on multimodality in human interaction* (*Vol. 2*, pp. 1868–1875). Berlin, Germany, and Boston, MA: Mouton De Gruyter.

Gullberg, M., deBot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture*, *8*, 149–179.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and nontonal language speakers. *Journal of Phonetics*, *40*, 269–279.

Hannah, B., Wang, Y., Jongman, A., & Sereno, J. A. (2016). Cross-modal association between auditory and visual-spatial information in Mandarin tone perception. *The Journal of the Acoustical Society of America*, *140*, 3225.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*, 495–522.

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

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

IBM Corporation. (2015). IBM SPSS statistics for Windows, version 23.0 [Computer software]. Armonk, NY: IBM Corporation.

IBM Corporation. (2016). IBM SPSS statistics for Windows, version 24.0 [Computer software]. Armonk, NY: IBM Corporation.

Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. *Journal of Experimental Child Psychology*, *156*, 99–112.

Jia, L., & Wang, J. (2013a). On the effects of visual processing of tone production by English-speaking learners of Chinese. *TCSOL Studies*, *52*, 63–104.

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

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, *27*, 793–807.

Kelly, S. D., Bailey, A., & Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of nonnative phonemic contrasts. *Collabra: Psychology*, *3*, 7.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334.

Kelly, S. D., Hirata, Y., Manansala, M., & Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Frontiers in Psychology*, *5*, 1–11.

Kendon, A. (2004). *Gesture: Visible action as utterance*. New York, NY: New York University Press.

Kiefer, M., & Trumpp, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education*, *1*, 15–20.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*, 63–67.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology*, *28*, 389–450.

Krauss, R. M., Chen, Y., Gottesman, R. F., & McNeill, D. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). New York: Cambridge University Press.

Kushch, O., & Prieto, P. (2016). The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of speech prosody 8* (pp. 922–925). Boston: ICSA.

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favor second language novel word learning. *Language, Cognition and Neuroscience*. Available at https://doi.org/10.1080/23273798.2018.1435894.

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*, 593–620.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, *61*, 1119–1141.

Macedonia, M., & Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education*, *8*, 74–88.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*, 982–998.

Madan, C. R., & Singhal, A. (2012). Using actions to enhance memory: Effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology*, *3*, 2010–2013.

Masumoto, K., Yamaguchi, M., Sutani, K., Tsunetoa, S., Fujitaa, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain Research*, *1101*, 102–109.

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *Modern Language Journal*, *86*, 192–203.

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

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

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

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15*, 133–137.

Paivio, A. (1990). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Post, L. S., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior*, *29*, 1450–1455.

Prieto, P. (2004). *Fonètica i fonologia. Els sons del català*. Barcelona, Spain: EdiUOC.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., et al. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception and Psychophysics*, *77*, 571–91.

Saltz, E., & Donnenwerth-Nolan, S. (1981). Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior*, *20*, 322–332.

Smith, D., & Burnham, D. (2012). Facilitation of Mandarin tone perception by visual speech in clear and degraded audio: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, *131*, 1480–1489.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*, 665–681.

Stefan, K., Cohen, L. G., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., et al. (2005). Formation of a motor memory by action observation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*, 9339–9346.

Szumilas, M. (2010). Explaining odd ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*, 227–229.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*, 219–235.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging*, *10*, 215–228.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is hand waving spatial? *Journal of Memory and Language*, *50*, 395–407.

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, *124*, 1716–1726.

Wang, Y., Jongman, A., & Sereno, J. A. (2003a). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*, 1033–1043.

Wang, M., Perfetti, C. A., & Liu, Y. (2003b). Alphabetic readers quickly acquire orthographic structure in learning to read Chinese. *Scientific Studies of Reading*, *7*, 183–208.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*, 3649–3658.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 1–10.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, *95*, 2240–2253.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2018, in press). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 1–28. Available at https://doi.org/10.1017/S0272263117000316.