Treball de Fi de Màster de Recerca

Effects of producing pitch gestures on the production of Chinese tones

Nom i Cognoms: Zhao Zeng

Màster: Lingüística Teòrica i Aplicada

Edició: 2022-2023

Directores: Dra. Pilar Prieto, Dra. Florence Baills

Any de Defensa: 2023

Col.lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 1
ABSTRACT 2
1. INTRODUCTION 3
1.1. Gestures and the acquisition of second language speech 3
1.2. Pitch gestures and the acquisition of second language speech 4
1.3. Goals and hypotheses 10
2. METHODS 12
2.1. Participants 12
2.2. Materials 12
2.2.1. Audiovisual Materials for the Tone Familiarization Phase 13
2.2.2. Audiovisual materials for the pre-test and post-test 14
2.2.3. Audiovisual materials for the tone imitation task: training phase 15
2.3. Experimental procedure 17
2.4. Data coding 18
2.4.1. Tone imitation accuracy scores: perceptual assessment of tone imitation by native speakers 18
2.4.2. Gesture imitation scores: perceptual ratings 20
2.5. Data analysis 21
3. RESULTS 22
3.1. RQ1: Effects of training on tone imitation accuracy 22
3.2. RQ2: Tone imitation accuracy during training 24
4. DISCUSSION AND CONCLUSION 26
REFERENCES 31
APPENDIX A: Descriptive statistics of tone imitation accuracy for each group in pre-test and post-test 35
APPENDIX B: Descriptive statistics of tone imitation accuracy for each group in block 1 and block 2 36
APPENDIX C: Sample of the tone imitation rating survey 37

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisors, Pilar Prieto and Florence Baills, for their professional knowledge, patience, unwavering support, and dedication of their time. Their valuable guidance has played a vital role in my research work. I am also profoundly thankful to Xiaotong Xi, who has taught me so much. Her generous help and valuable guidance in statistical analysis have been an important factor in my success in completing this study. Furthermore, I would like to extend my appreciation to the selfless help of Celia Gorba, Yao Ting and Peng Li, who assisted me a lot in the preparation of the experiment. I would also like to express my gratitude to those who participated in the rating, Yutong Na, Yunke He, Yingfei Lu, and Qianlan Zhang. Lastly, I want to convey my sincerest thanks to all my friends who have encouraged me throughout the writing of my thesis. Their companionship has been an essential driving force in my journey toward the completion of the present MA thesis.

ABSTRACT

Previous studies have shown the beneficial role of observing pitch gestures, which refer to hand movements mimicking pitch heights and contours in speech, in improving the perception of L2 tones by non-tonal language speakers. However, little is known about the potential benefits of training speakers to perform pitch gestures on tone pronunciation both during training - simultaneously imitating tone and gesture - and after training. The present study has two complementary aims. First, to investigate whether a brief 3-minute training session with performing pitch gestures while imitating the four Mandarin tones can improve non-tonal language speakers' ability to imitate the target tones after training. Second, to investigate whether the simultaneous production of the target tones during training is more accurate when accompanied by pitch gestures. In a between-subjects design, 50 Catalan-Spanish bilinguals were randomly divided into two groups: (a) the No-Gesture group (NG), in which the participants imitated tone words without gestures, and (b) the Gesture group (G), in which they imitated the same tone words while simultaneously producing pitch gestures. All participants underwent a tone imitation task before and after training. The results showed that a brief 3-minute training involving pitch gestures was not more effective than training without pitch gestures in promoting a more accurate production of the Mandarin tones after training. However, upon closer examination of the target production of the tones during training, further analyses revealed that producing pitch gestures did improve imitation accuracy of Tone 3, which is considered the most challenging tone to learn.

Keywords: *Pitch gesture, gesture production, Mandarin tones, tone production, phonetic imitation task*

1. INTRODUCTION

1.1. Gestures and the acquisition of second language speech

The paradigm of embodied cognition suggests that human cognition is rooted in sensory-motor processes and our physical body (Barsalou, 2008, Ionescu & Vasc, 2014). In human interaction, co-speech gestures (e.g., hand and body movements, as well as facial expressions) occur spontaneously and naturally during speech production. These co-speech gestures activate the sensory-motor system, forming a dynamic interaction between body and language (Macedonia & Von Kriegstein, 2012). In recent years, a growing number of studies have demonstrated that co-speech gestures are an essential component of our language abilities (e.g., Bernardis & Gentilucci, 2006; Church, Alibali, & Kelly, 2017; Cravotta, Busà, & Prieto, 2019).

In the field of second language pronunciation instruction, gestures may be used as a pedagogical tool (e.g., Hudson, 2011; Smotrova, 2017). A growing body of empirical research evidence supports the benefits of embodied training involving gestures representing segmental and suprasegmental features for boosting L2 speech production (e.g., Baills et al., 2022; Gluhareva & Prieto, 2017; Llanes-Coromina et al., 2018; Li et al., 2022; van Maastricht et al., 2019; Xi et al., 2020; Yuan et al., 2018; Zhang et al, 2020). For example, in a study conducted by van Maastricht, Hoetjes and Drie (2019), a total of 62 Dutch participants were trained in Spanish lexical stress under one of the following three multimodal conditions: audio-visual without gestures, audio-visual with beat gestures (i.e., up and down rhythmic movements aligned with the stressed syllable), or audio-visual with metaphoric gestures (i.e., gestures representing the lengthening of the stressed syllable). After training, the participants' production of Spanish lexical stress was assessed through a sentence-reading task in which participants were asked to read 28 four-word sentences that contained cognates with different stress patterns in Dutch and Spanish. The findings revealed that incorporating gestures during training was more advantageous compared to providing solely audio-visual information without gestures. Additionally, metaphoric gestures, which represented the duration feature of lexical stressed syllables in Spanish, were more beneficial than beat gestures. In two complementary studies, Zhang, Baills and Prieto (2020) and Baills and Prieto (2021) examined the potential advantages of hand clapping as a tool for acquiring pronunciation patterns of newly learned French words among Chinese adolescents and Catalan-dominant children. In Zhang, Baills and Prieto (2020), fifty Chinese adolescents aged 13 to 15 were assigned to either the clapping condition, where they repeated French words while synchronously clapping out their rhythmic structure, or the non-clapping condition, where they only repeated the words without clapping. Participants' oral production of French target words was assessed before and after training through two different measures. First, native French speakers perceptually judged the accentedness of the target words. Second, an acoustic analysis of the word-final syllable duration patterns was performed. The results revealed a nearly significant improvement in accentedness ratings for the clapping group from pre-test to post-test, while a significant improvement was observed in acoustic durational measures. In Baills & Prieto (2021), a similar training was conducted on 28 Catalan-speaking children aged 7 to 8, where one group received French word training with clapping and the other group did not. The pre- and post-test results were similar to those of the Chinese adolescents, with the clapping group of Catalan-speaking children showing significant improvement in pronunciation. The findings of these studies suggest that body movements can serve as a helpful tool to enhance foreign language pronunciation.

1.2. Pitch gestures and the acquisition of second language speech

In the present study, we will explore the beneficial effect of using pitch gestures, a type of metaphorical visuospatial gesture that depicts intonational or lexical tone pitch contours through up and down hand movements, on the learning of novel L2 tone production. A metaphoric relationship has been shown between space and pitch (Casasanto et al., 2003), where upward hand gesture movements represent a high-frequency pitch and downward gesture movements represent low-frequency pitch. Connell et al. (2013) showed that observing the musical pitch movements encoded in pitch gestures can directly affect pitch perception. The authors conducted an experiment in which a total of 32 subjects were asked to determine whether a target note produced with or without an up or down hand gesture by a singer in a video was the same as the preceding note. The results showed that while upward hand movements triggered the perception of higher pitch, downward hand movements led to the perception of lower pitch. The authors argued that the results of the study show a strong connection between pitch perception and pitch gesture movements, implying that pitch and space share a common auditory-spatial representation within our perceptual speech system. Achieving accurate tone pronunciation is an important milestone in learning L2 Chinese, and mastering tone perception and production poses a major challenge for L2 Chinese learners, especially for speakers of non-tonal languages (e.g., Kiriloff, 1969; Wang, Perfetti, & Liu, 2003b; Guo & Tao, 2008). Some studies on L2 Mandarin learners who are native English speakers have found that the difficulty of mastering the four tones of Mandarin varies. Tsai (2011) reported that students considered Tone 3 to be the most challenging, especially as it is easily confused with Tone 2. In an experimental work by Hao (2012), it was found that native English speakers and native Cantonese speakers had lower accuracy on Tone 3 and Tone 2 than on Tone 1 and Tone 4, in both the Mandarin tone identification task and the tone imitation task. Guo & Tao (2008) investigated the tone production of 16 American students over one academic year and found that Tone 3 had the lowest accuracy, followed by Tone 2, while Tone 1 and Tone 4 were produced with the greatest accuracy. In the study of He and Wayland (2010), the nine American students with 3-month Chinese learning experience and the nine students with one-year Chinese learning experience performed the worst on the production of Tone 3, while they performed the best on Tone 1, followed by Tone 4 or Tone 2. Overall, previous studies tend to highlight that Tone 3 poses a special learning challenge for non-native Mandarin speakers.

To facilitate the acquisition of Chinese tones by L2 learners, some studies have explored the role of visualization tools (Chun, Jiang & Natalia, 2013) and emphasized the importance of practice (Li, & DeKeyser, 2017), while others have investigated the role of pitch gestures for learning Chinese tones. The majority of studies assess the effects of observing and producing pitch gestures during tone training, with a specific focus on enhancing learners' tonal perception skills. These investigations aim to elucidate how pitch gestures help learners better discriminate and identify Mandarin lexical tones (e.g., Eng et al., 2013; Morett & Chang, 2015; Zhen et al., 2019). In the study conducted by Eng, Hannah, Leong and Wang (2013), 12 Canadian native English speakers were trained to perceive 80 Mandarin monosyllabic tone words under three conditions: audio-visual (AV), audio-gesture (AG), and audio-visual-gesture (AVG). Before and after the six training sessions over a 2-week period, the participants were asked to complete a tone identification task on the same 80 tone words to assess the effect of the training on tone perception. The results showed that the AVG group made greater progress in the tone identification task compared to the AV group, demonstrating that observing gestures that track pitch trajectories can enhance non-native speakers' perception of tonal contrast. Morett & Chang (2015) conducted another training study to explore the learning effects of pitch gestures on the perception of L2 tones. They divided 57 English speakers into three groups and provided training for learning the meaning of 12 minimal pairs in Chinese. The participants were required to repeat the Chinese words aloud and imitate gestures if they were present in the training video. One group imitated pitch gestures representing tone contours, another group imitated gestures conveying word meanings, and the third group received no gesture instruction. Following the training, participants were tested on Mandarin tone identification and word-meaning association tasks. The results showed that although pitch gesture training had no greater benefit than no gesture in the Mandarin tone identification task, it showed an advantage in the word-meaning association tasks, suggesting that pitch gestures can help learners to associate words with different tones and their corresponding meanings, which is beneficial for the acquisition of Mandarin words. Baills et al. (2019) replicated these findings and found that not only watching pitch gestures but also producing them themselves while pronouncing tone words helped participants improve their perception of Mandarin tones and their recall of tone words. In addition, Zhen, Heald, Goldin-Meadow, & Tian (2019) explored the effects of several different types of directional gestures on L2 tone recognition. One type involved gestures in the vertical direction to simulate pitch up and down movements, which can be categorized as congruent pitch gestures (in the same direction as the pitch) and incongruent pitch gestures (different from the pitch trajectory). The other type is rotated pitch gestures, in which horizontally oriented hands were moved away from or closer to the body to simulate pitch up and down movements. The trajectory of rotated pitch gestures is the same as the pitch trajectory, but cannot be visually aligned with it. A total of 108 native English speakers were randomly assigned to one of six training conditions: auditory only, performing congruent pitch gestures, observing congruent pitch gestures, performing rotated pitch gestures, observing rotated pitch gestures, and performing incongruent pitch gestures. During training, each group received the same auditory stimuli with the four tones of Chinese in their respective conditions. And same tone recognition task was administered in the pre-test and post-test phases conducted before and after the training. The results showed that training by observing or performing congruent pitch gestures (in the vertical plane) significantly enhanced pitch category learning compared to auditory training alone. Moreover, when the gesture was rotated (in the horizontal plane), performing them still led to a better tone recognition compared to watching them or listening-only training.

To our knowledge, no previous studies have explored the potential beneficial role of training with pitch gestures on the production of Chinese tones. The only three studies that are relevant in assessing the role of pitch gestures on the gains of suprasegmental pitch production at the sentence level are the studies by Yuan et al. (2019), Baills et al. (2022), and Li et al. (2022). In a pre-post test experimental design, Yuan and colleagues (2019) studied the role of pitch gestures for learning difficult Spanish intonation patterns. A total of 64 Chinese speakers were divided into two groups, with one group receiving intonation training that incorporated pitch gestures representing the nuclear Spanish intonation contours, while the other group received the same training in the absence of pitch gestures. Before and after the training, the participants' productions of the target words and intonation patterns were recorded. The results showed that observing pitch gestures significantly improved the intonational production outcomes of the learners. Baills et al. (2022) examined the effects of embodied prosodic training involving hand gestures that represent speech rhythm and intonation on the oral-reading pronunciation of French by Catalan learners. The results showed that embodied logatome training with gestures helps improve accentedness and suprasegmental accuracy. Similarly, Li et al. (2022) found that embodied prosodic training was effective in improving accentedness and the production accuracy of non-native vowels by Catalan learners of French.

All in all, the benefits of pitch gesture training in the acquisition of Spanish intonation patterns and French suprasegmental features may also be applicable to the learning of Mandarin tone production. Therefore, in the present study, our first goal will be to investigate the effects of producing pitch gestures on the learning of Mandarin tone production by non-tonal language speakers.

Additionally, to further understand the impact of the producing pitch gestures on the simultaneous production of target L2 tones, the second goal of this paper is to investigate whether the simultaneous production of pitch gestures during the training process leads to a more accurate tone imitation. Zheng, Hirata and Kelly (2019) examined the effects of imitating metaphoric actions, such as head nods and pitch gestures, on the simultaneous production of Mandarin tones. A total of 12 L1 Mandarin speakers and 24 L2 Mandarin speakers who were monolingual native English speakers were asked to imitate 15 minimal pairs of Mandarin tone words under three conditions: (a) speech alone, (b) speech with head nods, and (c) speech with hand gestures. Participants were asked to watch three different videos, in which the instructor produced the four Mandarin tones with hand gestures or with head nods or without any action. While watching the videos, the participants were asked to repeat what they had seen and heard after the instructor. The fundamental frequency (F0) of all participants' tone productions was analyzed and the tone production accuracy of the L2 speakers was rated by seven native Mandarin judges. Even though the perceptual ratings showed that producing hand gestures and head nods had no significant effect on tone production by L1 speakers and L2 speakers, a follow-up F0 acoustic analysis suggested that producing a pitch gesture while repeating a word with Tone 4 may have a beneficial effect on the production of tone 4 by L2 speakers.

Even though Zheng et al.'s findings suggest that the simultaneous production of pitch gestures could potentially benefit Mandarin tone production, more experimental evidence is needed to substantiate this claim. It is worth noting that a recent study by Li, Baills, Xi, and Prieto (in press) revealed the impact of gesture imitation accuracy on the simultaneous production of target consonants. Using a five-minute embodied phonetic training with fist-to-open hand gestures, Li et al.'s study explored the relationship between the accuracy in imitating gestures and the accuracy in the simultaneous imitation of non-native aspirated plosive sounds. A total of 29 Catalan speakers were instructed to produce 12 Chinese words involving aspirated plosives /p^h, t^h, k^h/ while performing fist-to-open-palm gesture. The accuracy of the participants' gesture imitation was assessed by the shape of the gesture and the temporal alignment of the gesture with the aspirated sounds. The results showed that more accurate gesture shapes and finer gesture-speech alignment predicted more accurate speech imitation. Based on the results of these previous studies, further investigations are needed to assess the potential impact of pitch gesture production on enhancing the simultaneous production accuracy of Mandarin tones, which is the second goal of the present MA thesis.

1.3. Goals and hypotheses

The goal of the present study is twofold. The first goal is to investigate whether a brief 3-minute tone training session with observing and at the same time performing pitch gestures along with the Mandarin Chinese tones can enhance the ability to imitate these tones after training. This will be achieved by comparing the pre- and post-test performance of the group trained with gestures to that of the group without gestures. Considering the benefits of the use of pitch gestures on L2 perceptual learning of lexical tones and production learning of intonational patterns, it is hypothesized that pitch gestures can also help L2 production of lexical tones. The second goal is to explore whether performing pitch gestures while simultaneously imitating target Chinese tones leads to more accurate simultaneous production of tone. This will be achieved by comparing the online tone imitation between the gesture and no-gesture training groups. Considering the close link between pitch gestures and pitch perception, as well as the influence of the embodiment on speech production, it is hypothesized that performing these pitch gestures will enable speakers to produce the tones more accurately.

In order to answer these two research questions, a total of 50 Catalan-Spanish bilinguals participated in a training study with a between-subjects and pre-post-test design. They were randomly assigned to two groups: the No-Gesture group (NG), which were trained to imitate the target tones without gestures, and the Gesture group (G), which was trained to perform pitch gestures while imitating the tones. To evaluate the training effects, all participants completed a pre-test and post-test tone imitation task before and after the training. Following up on previous research, we hypothesize that (a) a brief gesture training session will be beneficial in improving the ability to imitate tones, and (b) performing simultaneous pitch gestures during the imitation of target tone will positively impact the accuracy of tone production. Moreover, considering that Tone 3 exhibits the most complex pitch changes and is commonly identified as the most challenging to master, we expect that pitch gestures will help L2 learners in the production of this particular tone.

2. METHODS

2.1. Participants

A total of 50 Catalan-Spanish bilinguals recruited at Pompeu Fabra University participated in the experiment (45 females, M_{age} = 22.91 years, *SD* = 6.85). They all had no previous knowledge of Mandarin Chinese or any other tonal languages. None of the participants reported having any speech or hearing impairment and their vision or corrected vision was normal. Subjects were randomly assigned to either the Gesture (G, *n* = 25, 23 females, M_{age} = 21.81, *SD* = 4.42) group or the No Gesture (NG, *n* = 25, 22 females, M_{age} = 23.95, *SD* = 8.54) group. All participants volunteered to take part in this experiment and signed a written consent form allowing the researchers to analyze the data collected from the experiment.

2.2. Materials

The materials of this study were prepared and inserted into two experimental videos, one for the G group and the other for the NG group, both containing a tone familiarization phase, a pre-test phase, a tone imitation task (i.e., the training phase), and a post-test.

The audiovisual materials for the tone familiarization phase and tone imitation task were recorded by a bilingual Catalan-Chinese female speaker and a native Chinese male speaker. These recordings were initially carried out for the creation of the stimulus in Baills et al. (2019) and a selection of appropriate items was made for the present experiment. The stimulus materials for the pre-test and post-test were recorded by two other native Chinese speakers (1 female). To ensure participants' full understanding of the task, all text instructions in the videos were provided in Catalan and accompanied by a voiceover recorded by a native Catalan female speaker. All instructors were recorded at the experimental language research laboratory of Universitat Pompeu Fabra using a PDM660 Marantz professional portable digital video recorder. The videos were later edited in a professional video editing software to show the speaker's face and upper body against a white background.

2.2.1. Audiovisual Materials for the Tone Familiarization Phase

The stimuli used in this study are the four tones of Mandarin Chinese, which is a tonal language characterized by the use of a phonologically distinctive pitch pattern at the word level, using lexical tonal contrasts to distinguish the meaning of target words (Jongman et al., 2006; Xu, 1994). For example, the Chinese word *ni* means 'girl', 'mud', 'you', and 'be bored with' with Tone 1, Tone 2, Tone 3, and Tone 4, respectively. The four tones in Mandarin are phonetically distinguished in terms of pitch height and pitch contour, which can be seen by the change in fundamental frequency. Tone 1 has a high-level pitch with a flat contour while Tone 2 rises from mid-low to the highest pitch. Tone 3 is characterized by dropping to the lowest point in pitch first and then rising to a mid-high pitch. Tone 4 falls from the highest level to the lowest pitch. Furthermore, there are differences in the duration of the four tones in the isolated monosyllables. Some studies have shown that Tone 3 has the longest duration, while Tone 4 has the shortest duration. The durations of Tone 1 and Tone 2 fall between those of Tone 3 and Tone 4 (Yang, Zhang, Li, & Xu, 2017). Among the four tones, Tone 3 is widely recognized as one of the most challenging to produce (Guo & Tao, 2008; Tsai, 2011; Zheng et al., 2018).

In the familiarization phase, the goal and procedure of the tonal imitation task were first explained through written text in Catalan. The NG group was informed that they had to imitate speech, while the G group was informed that the task involved both speech and gesture imitation. The purpose of the task was clearly stated, which was to assess the participants' ability to imitate melodic features in speech. Following this, a female instructor provided a detailed explanation of the four Mandarin Chinese tones in Catalan, accompanied by visual diagrams illustrating the pitch contours and tone marks of each tone (Figure 1). Examples were given for each of the four tones: the syllable *ta* produced with Tone 1, *ba* produced with Tone 2, *da* produced with Tone 3, and *ba* produced with Tone 4. The duration of the familiarization phase for both conditions was 1 minute and 40 seconds.

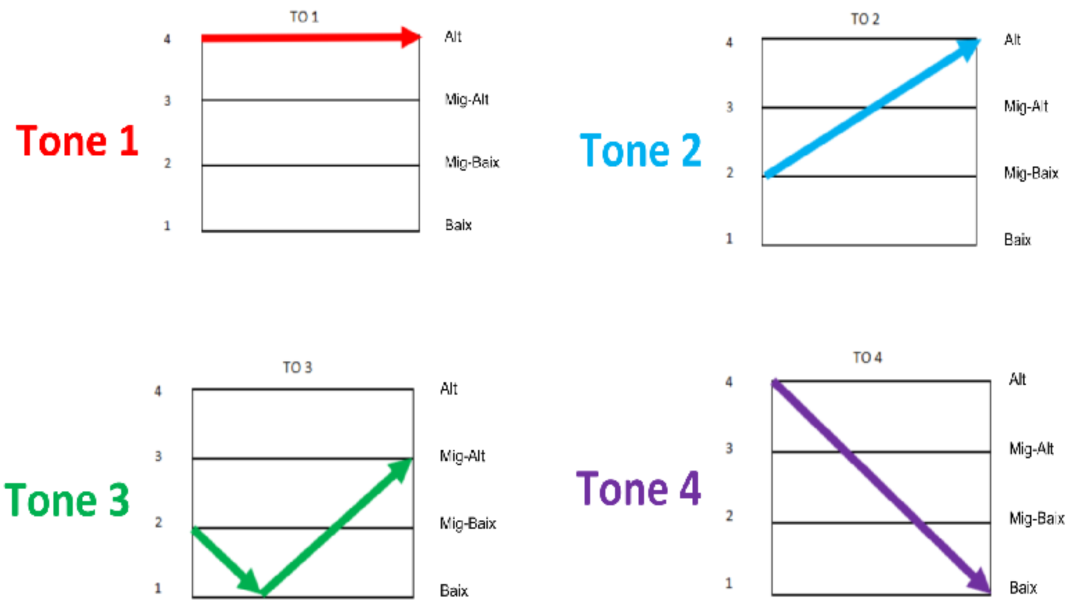


FIGURE 1. Diagrams of the four tones in Mandarin Chinese.

2.2.2. Audiovisual materials for the pre-test and post-test

The stimulus materials used in the pre-test and post-test consisted of the syllable *ni* and *la* recorded audiovisually in the four different tones each, resulting in a total of eight different words. The recordings were performed by a female native Mandarin speaker for *la4*, *la3*, *ni1*, and *ni2* and by a male speaker for *la1*, *la3*, *ni3*, and *ni4*. In the pre-test, the tone words were presented one by one in the following order: *la4*, *la1*, *la3*, *la2*, *ni1*, *ni3*, *ni2*, *ni4*. At the post-test, the order of presentation of the tone words was changed: *ni1*, *ni4*, *ni2*, *ni3*, *la4*, *la2*, *la3*, *la1*. After each tone word was produced by the speaker, a 2.5-second black screen appeared with Catalan text prompting participants to repeat the word (see Figure 2). Both the pre-test and post-test lasted 1 minute and 20 seconds and were identical in the video for the G group and the video for the NG group.

FIGURE 2. Screenshots illustrate the presentation of the target words *la4* and *la1* in the pre-test

2.2.3. Audiovisual materials for the tone imitation task: training phase

The training phase consisted of a tone imitation task where participants were audiovisually exposed to the target stimuli with or without pitch gestures. Twelve monosyllabic words (three words for each tone) were selected as stimulus material (see Table 1) taking into account compatibility with the phonotactics of Catalan., i.e. consonant sounds that are not part of the Catalan inventory were avoided. The tones were presented in two blocks. In block 1, both for the G and NG conditions, the 12 different words were presented one by one either by a male or a female speaker, without any accompanying gesture. Following each item, a 2.5-second black screen appeared, displaying instructions for the participant to repeat the word. In block 2, the same twelve words as in block 1 were presented under different conditions:

TABLE 1. Stimuli for the tone imitation task

Tone 1		Tone 2		Tone 3		Tone 4	
Pinyin	English	Pinyin	English	Pinyin	English	Pinyin	English
fā	send	má	linen	lǔ	capture	mà	insult
dī	low	lí	pear	mǐ	rice	lì	chestnut
dū	capital	ná	take	gǔ	drum	lù	deer

a. *Gesture condition*: First, a female instructor introduced the pitch gestures corresponding to the four tones of Mandarin Chinese. She provided specific examples using the words *bo1*, *fu2*, *bi3*, and *na4*. The pitch gesture for Tone 1 consisted of a movement of the right hand from left to right at the same height as the head. For Tone 2, the hand moved from the left elbow joint across the body to the right shoulder. For Tone 3, the hand first moved down from the left elbow joint to the abdomen, and then up to the right shoulder. As for Tone 4, the hand moved from the left shoulder to the right side of the abdomen. To facilitate participants in imitating gestures with their right hand, the video was digitally flipped to appear as if the right hand was being used (see Figure 3). Subsequently, during the training phase, the female and male speakers presented the same 12 tone items one by one, as outlined in block 1, albeit in a different order. Each item was accompanied by a corresponding pitch gesture. After the presentation of each item, a 2.5-second black screen appeared, instructing the participants to imitate the tones together with the hand gesture. In addition, the 12 items were repeated twice within block 2.

b. *No-Gesture condition*: Block 2 was identical to block 1, with the only difference being the item presentation order. The item order matched that of block 2 in the Gesture condition. In addition, the 12 items were trained twice.

All in all, in each condition, the 12 items were trained three times: the first repetition (block 1) involved speech-only practice, which was common for both the G and NG groups. Subsequently, the second and third repetitions (block 2) differed depending on the group: the NG group continued with the speech-only practice, while the G group involved both speech and gesture. The entire training phase lasted 2 minutes and 50 seconds in the G condition and 2 minutes and 30 seconds in the NG condition.

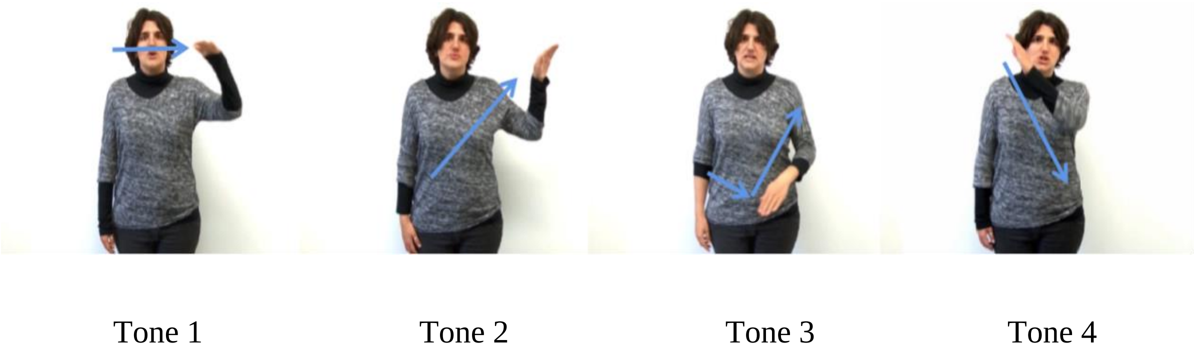


FIGURE 3. Screenshots of the female instructor illustrating the four Mandarin tones using target pitch gestures (image adapted from Baills et al., 2019). From left to right, the syllable *bo* with Tone 1, *fu* with Tone 2, *bi* with Tone 3, and *na* with Tone 4.

2.3. Experimental procedure

In this study, a between-subjects design was employed, with participants being randomly assigned to either the G group or the NG group. Participants watched the corresponding experimental video individually in a quiet room at Universitat Pompeu Fabra. Participants were seated in front of the laptop where the video was played, at a distance of approximately 60 cm from the screen. To ensure optimal audio quality, participants wore headphones, allowing them to hear the sound from the video clearly and comfortably. Participants' oral production during the pre-test, training phase, and post-test was recorded by a portable recorder (Zoom H4n Pro). The total duration of the experiment in both conditions lasted around 7 minutes. A summary of the experimental procedure can be found in Figure 4.

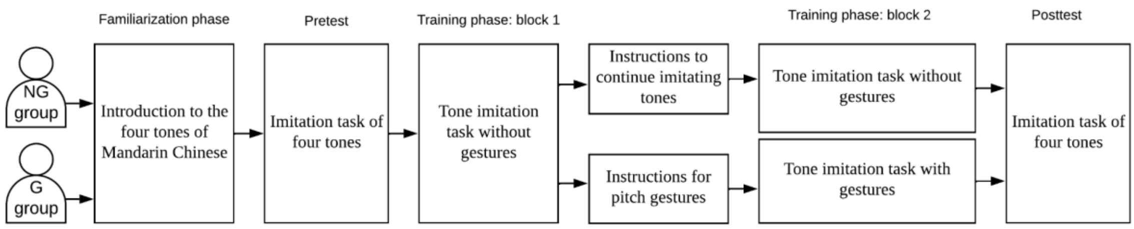


FIGURE 4. Experimental procedure.

First, participants quietly watched the tone familiarization, in order to acquire a basic understanding of the four tones in Mandarin Chinese. Then, in the pre-test, participants repeated after the instructors the four tone words with the syllable *la* and the four tone words with the syllable *ni*. Next, all participants watched the block1 of the training phase and imitated the 12 tone words without gestures. Subsequently, in block 2, participants in the NG group continued to imitate the same 12 tone words without gestures, while the participants in the G group watched an introduction to the pitch gestures representing the four tones in Mandarin Chinese and were then instructed to perform the corresponding hand gestures while producing the 12 tone words. The training phase was immediately followed by the post-test, which was similar to the pre-test.

2.4. Data coding

2.4.1. Tone imitation accuracy scores: perceptual assessment of tone imitation by native speakers

The recordings of the 50 participants (25 of NG group, 25 of G group) were organized as audio files. Each participant's recording was then edited into 5 audio clips corresponding to the pre-test, first repetition, second repetition, third repetition, and post-test. As a result, a total of 250 recordings (50 participants × 5 parts) were obtained.

To assess participants' tone imitation accuracy, perceptual evaluations were conducted using a 5-point score Likert scale. Five native Mandarin speakers aged between 22 to 24 (M_{age} = 23.6; SD = 0.8) were recruited to evaluate the participants' tone imitation accuracy. Each one received a monetary compensation of 10 euros per hour in the rating work. Before starting the rating, they received a 45-minute training session to ensure accurate and consistent tone training. During the training, they were introduced to three primary rating criteria: pitch contour, pitch height, and tone duration. Regarding pitch contour, Tone 1 should be produced with a flat pitch, Tone 2 should exhibit a rising pattern, Tone 3 should demonstrate a falling and then rising trajectory, while Tone 4 should simply fall. For pitch height, Tone 1 maintains a high pitch throughout, while Tone 2 starts from a mid-low pitch and rises to a high pitch. Tone 3 goes down from mid-low to low pitch and then rises to a mid-high pitch, while Tone 4 starts from a high pitch and falls to a low pitch. In terms of tone duration, Tone 3 is the longest and Tone 4 is the shortest. The durations of Tone 1 and Tone 2 are intermediate, falling between the durations of Tone 3 and Tone 4. The raters needed to score the global accuracy of each tone production based on these three criteria using a 5-point Likert scale. A score of 5 indicated a perfectly accurate imitation of the tone, while a score of 1 stood for a completely inaccurate imitation of the tone. After being instructed about the specific definitions of pitch contour, pitch height, and tone duration, they were shown some concrete examples for each tone. Figure 5 shows an example to illustrate the pitch contours of Tone 3, where a score of 5 indicates a drop from a high point at the beginning, to a low point in the middle, and then rises to a high point. Another example depicted a score of 3 for Tone 3, indicating a slight fall at the beginning and then a slow rise thereafter. The example with the lowest score demonstrated a pitch contour that rose and then fell, which was the exact opposite of the target pitch contour for Tone 3.

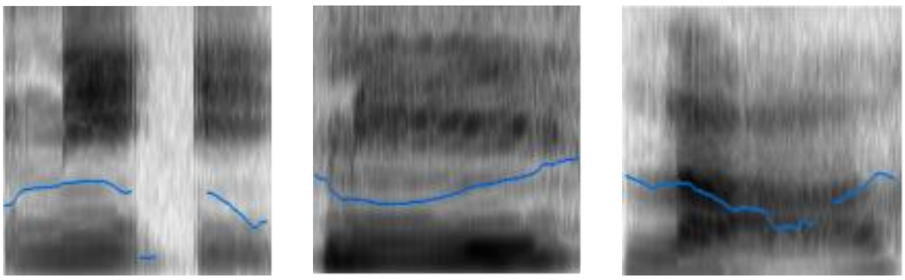


FIGURE 5. From left to right, spectrograms of the target pitch contours (F0 curves in blue) of Tone 3, produced by participants who received scores of 1 (inaccurate imitation), 3 (average accuracy), and 5 (accurate imitation) respectively.

In the final phase of the training, the 5 raters familiarized themselves with the online survey rating procedures and practiced tone imitation accuracy rating on a set of five audio samples to ensure consensus in their scoring. Once the rater training session was completed, the raters received the links to five online surveys created using the Alchemer Online Survey Platform (see Appendix B for a sample of the tone imitation rating survey). These five surveys corresponded to pre-test, repetition 1, repetition 2, repetition 3, and post-test and each contained 50 audio files. For the pre-test and post-test survey, each audio file contained eight tone productions. However, upon inspection, we found that five participants failed to produce the first tone word in the pre-test because they did not respond in time. Consequently, there were actually only 395 tone productions in the pre-test and 400 tone productions in the post-test. As for the other three surveys, each audio file contained 12 tone productions, for a total of 1800 tone productions (50 participants ×12 tone productions × 3 repetitions). All raters were advised to complete the five rating surveys in batches within a day with long pauses between them to avoid fatigue. In total, the rating task took each rater approximately four hours.

A total of 795 scores were obtained from the pre-test and post-test and 1800 scores were obtained from the training phase. To assess the inter-rater reliability among the five raters, an Intraclass Correlation Coefficient (ICC) test was used. The *irr* package version 0.84.1 (Gamer et al., 2019) was employed to calculate ICC estimates along with the 95% confidence intervals. The ICC analysis revealed an excellent level of interrater reliability for the tone imitation scores (ICC = 0.84, [0.83, 0.85]).

2.4.2. Gesture imitation scores: perceptual ratings

In order to guarantee that participants were producing gestures during the Gesture training condition and that their gestures were accurate enough, the author of the study observed their hand movements carefully and rated the accuracy of each gesture. A score of 1 means a completely inaccurate gesture, a score of 2 suggested the gesture was not very accurate and a score of 3 represented a successful execution of an accurate gesture. If a participant did not perform a gesture, no score was given. Considering that this was an online assessment and that the score was given quickly in a short interval of time, the rater gave a comprehensive score based on factors such as gesture direction, height, trajectory, and speech-temporal alignment. In the end, a total of 600 gesture scores were obtained (25 participants ×12 items × 2 repetitions). Crucially, the participants in the G group produced a gesture 100% of the times, and only 15% of their gestures were not totally accurate.

2.5. Data analysis

The statistical analysis was conducted in R (version 4.2.3) (R Core Team, 2022), using the *lme4* package (Bates, Mächler, Bolker & Walker, 2014). A single Linear Mixed-Effect Model (LME) was run with tone imitation accuracy score as the dependent variable. The fixed effects included the variables Session (4 levels: pre-test, block1, block2, post-test), Condition (2 levels: NG and G), Tone (4 levels: Tone 1, Tone 2, Tone 3, and Tone 4), as well as their interactions. As for the random structures, the model with the most complex random structures was submitted to *buildmer ()* from the *buildmer* package, version 2.9 (Voeten & Voeten, 2021) to determine the best random structure. As a result, the random structure that best fitted the data included random intercepts for particpant and item. The significance values for the fixed effects were calculated using Type II Wald chi-square tests by using the *Anova()* from the *car* package, version 3.1-2 (Fox & Weisberg, 2011). Additionally, pairwise comparisons were conducted using the *emmeans* function with Bonferroni adjustment to examine significant differences between conditions, sessions, and tones.

3. RESULTS

The results of the LME analysis (see Table 2) revealed significant main effects of Session and Tone, significant two–way interactions of Condition × Tone and Session × Tone, and a significant three-way interaction of Condition × Session × Tone. The significant three-way interaction ($\chi 2(9) = 23.09$, $p = .006$) supports the existence of conditional effects on the tone imitation accuracy across every session for each tone. We conducted post hoc pairwise comparisons on the three-way interaction so as to address Research Question 1 and Research Question 2 in the following.

Table 2. Results from the linear mixed model of tone imitation accuracy scores

	Chisq	Df	Pr(>Chisq)
Condition	0.24	1	0.623
Session	49.00	3	**< .001**
Tone	26.06	3	**< .001**
Condition × Session	0.30	3	0.960
Condition × Tone	9.16	3	**0.027**
Session × Tone	57.33	3	**< .001**
Condition × Session × Tone	23.09	9	**0.006**

3.1. RQ1: Effects of training on tone imitation accuracy

For the effect of training on tone imitation accuracy, we report the post-hoc comparisons of the results of tone imitation score across Test (pre-test and post-test), Condition (G and NG groups), and Tone.

For Tone 1, both groups improved significantly from pre-test to post-test (NG group, estimate = 0.572, z = -8.225, p < .001; G group, estimate = -0.388, z = -5.579, p < .001). The score of the NG group was significantly lower than that of the G group at pre-test (estimate = -0.200, z = 2.614, p = 0.009), while there was no significant difference between the two groups at post-test. For Tone 2, both groups improved significantly from pre-test to post-test (NG group, estimate = -0.284, z = -4.084, $p < .001$; G group, estimate = -0.388, z = -0.200, $p = .024$), but there were no significant differences between the groups in either the pre-test or the post-test. As for Tone 3, the mean score of the G group was significantly lower than that of the NG group at pre-test (estimate = -0.196, z = -2.555, $p = 0.011$), while there was no significant difference between the two groups at post-test. Moreover, the NG group showed a significant decrease (estimate = -0.280, z = 4.026, $p < 0.001$) from pre-test to post-test, while there were no significant differences between tests for the G group. Regarding Tone 4, no significant differences were found either between pre-test and post-test for the same group or between the groups for the same test. A visual comparison of the two groups' mean accuracy scores of tone imitation on each tone at pre-test and post-test can be seen in Figure 6 and more detailed descriptive statistics can be found in Appendix A. All in all, the results revealed that the group of participants who trained tone imitation with pitch gestures did not outperform the NG group after training.

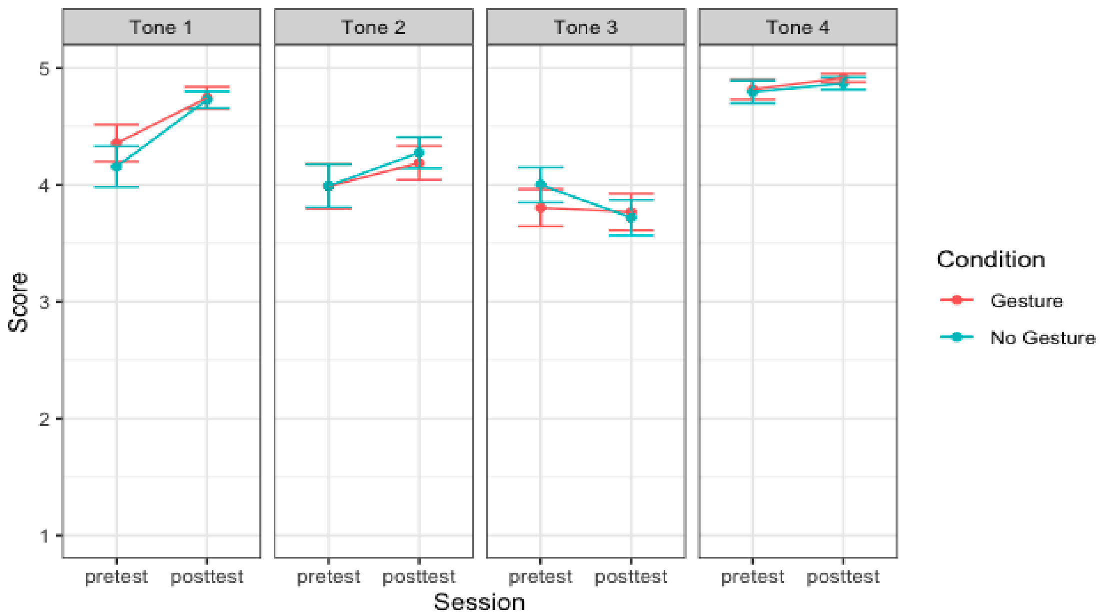


FIGURE 6. Mean accuracy scores obtained for each tone across Condition (NG and G) and Test (pre-test and post-test).

3.2. RQ2: Tone imitation accuracy during training

For the effect of producing pitch gestures on the accuracy of simultaneous tone imitation, we report the post-hoc comparisons of the results of tone imitation scores across Test (between block 1 and block 2), Condition (G and NG groups), and Tone.

Figure 7 illustrates the mean accuracy score obtained of each tone for the No-gesture group and Gesture group during the 2 phases of training, namely block 1 and block 2 (also can be seen in Appendix B). Post hoc contrasts revealed that there were no significant differences observed between the two phases of training across the G and NG groups for the production of Tone 1 and Tone 4. Regarding the Tone 2 imitation accuracy, the G group was found to have a significant drop from block 1 to block 2 (estimate = -0.191 , z = 3.884, $p < .001$), while the NG group did not (estimate = -0.104 , z = 2.108, $p = 0.210$). However, the imitation score for Tone 2 did not differ between the two groups at block 2 (estimate = -0.068, z = -1.325; $p = 0.185$). Crucially, for the Tone 3 imitation, a significant improvement of both groups was found, with the improvement of the G group (estimate = 0.317 , z = -6.453, $p < .001$) beinglarger than the NG group (estimate = 0.162 , z = -3.308, $p = 0.006$). In the first training phase without gestures, there was no significant difference in Tone 3 production between the two groups (estimate = -0.042, z = -0.650, $p = 0.516$). However, in the second training phase, the G group exhibited significantly a higher accuracy score of Tone 3 compared to the NG group (estimate = 0.112, z = 2.183, $p = 0.029$). In summary, for Catalan-Spanish speakers, performing pitch gestures while imitating Mandarin tones had no effect on their accuracy in producing Tone 1, Tone 2, and Tone 4. However, performing pitch gestures proved to be beneficial in enhancing the imitation accuracy of Tone 3.

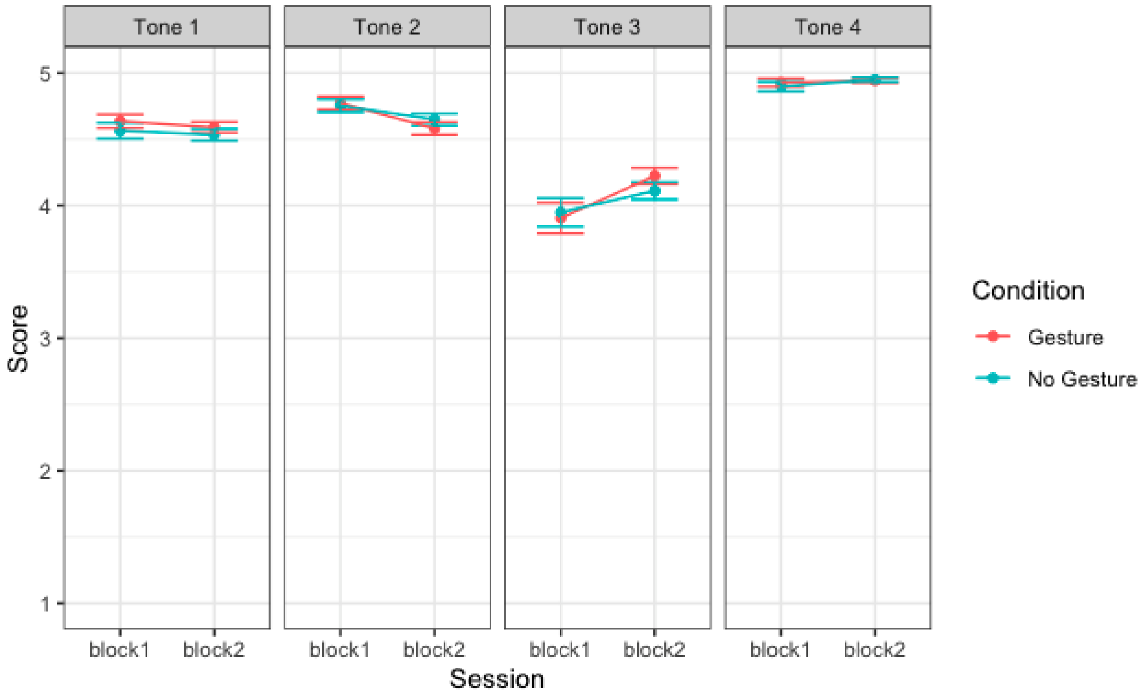


FIGURE 7. Mean accuracy scores obtained for each tone imitation across Condition (NG and G) and Session (block 1: tones are trained without pitch gestures in either group; block 2: tones are trained with pitch gestures in the G group and without pitch gestures in the NG group).

4. DISCUSSION AND CONCLUSION

The present study investigated the effect of producing pitch gestures on the imitation of non-native Mandarin tones by non-tonal language speakers. A total of 50 Catalan-Spanish speakers participated in a short training session involving the imitation of the four Mandarin tones. Participants were divided into two groups and trained under different conditions: one group received training with both speech and gesture, while the other group received training with speech only. Additionally, the participants completed a tone imitation task both before and after the training session. On the one hand, we aimed to determine whether a 3-minute pitch gesture training could enhance participants' ability to imitate Mandarin tones. On the other hand, we investigated whether performing pitch gestures while imitating the tones contributes to better simultaneous tone production. First, contrary to our expectation, the brief training with pitch gestures was not more effective than training without pitch gestures to improve non-tonal language speakers' tone production. Our results show an improvement between pre- and post-test for Tone 1 and Tone 2, and with no significant difference between the G and NG groups. Second, the results showed that performing pitch gestures during tone production significantly helped the production of the most difficult tone (Tone 3) compared to imitating the tone without gesture.

Regarding the effects of training, the results of the pre-test and post-test showed that after a short 3-minute training session, participants in both the G and NG groups improved their imitation of Tone 1 and Tone 2. However, there was no significant difference between the groups, showing that pitch gestures did not have any enhancing learning effect. Thus, improvement on Tone 1 and Tone 2 may be largely attributed to the effects of the practice. For Tone 4, there was no significant improvement from pre-test to post-test in either group, due to the fact that all participants performed well enough in the pre-test (NG group, Mean= 4.80; G group, Mean= 4.82). Interestingly, for the most difficult Tone 3, the NG group performed worse in the post-test than in the pre-test while the G group's performance remained the same, not improving but not regressing either. One possible reason for this is that the tone words selected for the training were different from those used in the pre- and post-test, which means the effects of the training may not generalize well to novel syllables, especially for those with complex pitch changes like Tone3. Moreover, the fact that the NG group worsen while the G group did not suggest that pitch gesture training may have played some positive role in preventing participants from getting worse on the production of new syllables with Tone 3.

All in all, a longer training session or more training sessions may be required to observe the beneficial effect of pitch gestures on the production of four tones. In the study by Yuan and colleagues (2019), the duration of the training session was also very short (around three minutes), however, the Chinese learners of Spanish had already undergone a 2-month Spanish phonetics course prior to participating in the experiment. In Baills et al. (2022) and Li et al. (2022), the Catalan learners of French took part in three 30-minute sessions to practice basic sentence prosody. In the case of the present study, participants who had never been exposed to Mandarin Chinese before had to learn four different tones, which may be more difficult and require more time and/or repetitions. In addition, a delayed post-test may be crucial to reveal the effects of gestures on learning, as demonstrated in Li et al. (2020).

Secondly, the study investigated the effects of imitating pitch gestures on simultaneous tone production. Our results partially replicate and expand the results of previous studies conducted with English native speakers (e.g., Guo et al., 2008; Zheng et al., 2018). First, Catalan-Spanish speakers demonstrated ease in producing Tone 4, while Tone 3 proved to be the most challenging for them to master. In the first phase of the training, the NG group achieved an average tone imitation accuracy scores of 4.57, 4.75, 3.95, and 4.9 for Tone 1, Tone 2, Tone 3, and Tone 4, respectively. Similarly, the first phase of the training from the G group obtained average scores of 4.35, 3.99, 3.8, and 4.82 for the corresponding tones. Clearly, for Tone 3, which exhibits the poorest performance and thus has the greatest room for improvement, the benefits of using pitch gestures were most noticeable, while for those tones that were easier to master, the influence of pitch gestures may be negligible.

Furthermore, it is worth noting that our findings regarding the beneficial effect of performing pitch gestures on the production of Tone 3 contrasts with the study of Zheng et al. (2018). In their study, the benefits of using pitch gestures failed to be supported by the perceptual assessments, and an acoustic analysis revealed a benefit of imitating pitch gestures while producing Tone 4 production by native English speakers. It is important to consider the strictness of the evaluation methodology employed in Zheng et al., where native speakers were instructed to judge tone productions as either correct or incorrect. In this way, as long as one judge disagreed, the item would be judged incorrect. This evaluation procedure may have been overly stringent, as individual biases of judges may have an undue influence on the final results and might have failed to fully account for the diversity and subtle variations in tone production. By contrast, our study aimed to explore the extent to which participant's tone productions approached the target pitch pattern, using a 5-point Likert scale. This approach allowed for a more nuanced evaluation of tone imitation accuracy. By refining the evaluation criteria, five native speakers were invited to assess participants' tone productions based on pitch contour, pitch height, and tone duration. In this case, the benefits of producing pitch gestures on the production of Tone 3 could be pointed out through the assessment of the native speakers' ears.

Why is it that Tone 3 in our study was the one obtaining significant benefits from the use of pitch gestures? Research on tone production has shown that L2 learners with similar pitch patterns in their native language are better able to produce target tones (Yang, 2019). Although Catalan and Spanish are not tonal languages, they frequently employ pitch patterns like rising (LH) or falling (HL) in their intonational phonology (Prieto, 2014; Yuan et al., 2019). Presumably, the presence of these pitch accentual contours allows them to more easily reproduce Tone 1, Tone 2 and Tone 4, as these tones closely match pitch patterns in their native languages. However, the pitch contour of Tone 3, which involves a complex pitch change from high to low and then to high (HLH) within a single syllable, does not exist in Catalan and Spanish. This circumstance may explain why Tone 3 is quite difficult for L2 learners. By performing pitch gestures that indicate pitch changes, learners can connect the sensory and motor systems. This allows them to more intuitively feel the trajectory of Tone 3 pitch changes and combine this perception with speech output. In the first phase of the training without pitch gestures, participants of both groups tended to produce the HLH pitch pattern of Tone 3 with a slight drop at the beginning followed by a gradual rise (see Figure 6 above), which is more like the pitch change of Tone 2. By contrast, during the second phase of the training, the use of pitch gestures helped participants of the G group to produce a significant drop in pitch during the first half of Tone 3, followed by a significant rise in the second half. However, the tone imitation accuracy in the NG group (the group not using pitch gestures) did not exhibit improvements. This may explain why the simultaneous production of pitch gestures promotes the accuracy of Tone 3 imitation. Another complementary explanation is that performing the pitch gesture helped participants to reach a good temporal pitch location of the low point of Tone 3, i.e. the turning point. The temporal location of the L turning point in the HLH Tone 3 contour is key to the success of Tone 3 production (Shen & Lin, 1991). The earlier the turning point starts and the less low it is, the more the target production sounds like Tone 2. Performing pitch gestures with a distinct turning point together with Tone 3 allowed participants to align the turning point of their productions with that of the pitch gesture they were performing, thereby achieving a more accurate pronunciation of Tone 3.

Some limitations of the present study should be mentioned. First, the 3-minute pitch gesture training was found to be too short to be effective, indicating that a longer pitch imitation training might be needed to observe more clear effects on production after training. Second, since research has shown that a more accurate gesture imitation leads to a more accurate L2 speech imitation (according to the research of Li et al. (in press), future research should take this factor into account. In the present study, we assessed participants' gesture performance through online ratings while they were performing the task. The results indicated that while all participants produced all gestures in an accurate way, still 15% of the gestures were not performed at the highest precision. These inaccurate gestures may not only have failed to enhance the tone imitation accuracy, but could also have had a negative impact due to inaccurate simulation of pitch changes. In future research, participants' gesture performance needs to be more strictly controlled. In this way, we would be able to explore the influence of the accuracy of imitating pitch gestures on the accuracy of simultaneous tone production. Third, the present study only examined the effect of imitating pitch gestures on isolated tones at the word level. However, the production of tones at the sentence level poses a greater challenge in learning Mandarin tones. Importantly, confusion of tones in near-natural conversations is a common error among L2 learners (Guo & Tao, 2008). Therefore, an interesting avenue of future research would be to explore the effects of performing pitch gestures on sentence-level tone production.

In conclusion, the present MA thesis provides empirical support for the beneficial role of pitch gestures in non-native tone production. While the results indicate that a brief 3-minute pitch gesture training was insufficient to enhance the non-tonal language speakers' ability to imitate Chinese tones after this training, a significant effect was observed when speakers were engaged in performing pitch gestures during tone production. Crucially, performing pitch gestures while imitating tones specifically enhanced the production of the most difficult tone, Tone 3. These findings support the close relationship between body and speech and highlight the potential of incorporating pitch gestures as a valuable tool for improving non-native tone production.

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.

Baills, F., Alazard-Guiu, C., & Prieto, P. (2022). Embodied Prosodic Training Helps Improve Accentedness and Suprasegmental Accuracy. *Applied Linguistics*, *43*(4), 776-804.

Baills, F., Suarez-Gonzalez, N., Gonzalez-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*(1), 33-58.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*(2), 178-190.

Casasanto, D., Phillips, W., & Boroditsky, L. (2003). Do we think about music in terms of space? Metaphoric representation of musical pitch. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25(25)*, 1323.

Chun, D. M., Jiang, Y., & Natalia, A. Ì. (2013). Visualization of tone for learning Mandarin Chinese. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Aug. 2012. (pp. 77-89). Iowa State University.

Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, *1*(1), 86-114.

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, *81*(1), 124-130.

Cravotta, A., Busà, M. G., & Prieto, P. (2019). Effects of encouraging the use of gestures on speech. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3204-3219.

Eng, K., Hannah, B., Leong, L., & Wang, Y. (2013, June). Can co-speech hand gestures facilitate learning of non-native tones? In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060225). Acoustical Society of America.

Fox, J., & Weisberg, S. (2011). An {R} companion to applied regression (2nd ed.). Thousand Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609-631.

Guo, L., & Tao, L. (2008, April). Tone production in Mandarin Chinese by American students: A case study. In M.K.M. Chan and H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)* (Vol. 1, pp. 123-138). The Ohio State University.

Han, Y., Goudbeek, M., Mos, M., & Swerts, M. (2019). Mandarin tone identification by tone-naïve musicians and non-musicians in auditory-visual and auditory-only conditions. *Frontiers in Communication*, *4*, 70.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of phonetics*, *40*(2), 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*(4), 3225-3225.

Hongwei, D. I. N. G., Hoffmann, R., & Jokisch, O. (2011). An investigation of tone perception and production in German learners of Mandarin. *Archives of Acoustics*, *36*(3), 509-518.

Hirata, & 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(2), 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*(6), 2090-2101.

Hudson, N. (2011). *Teacher gesture in a post-secondary English as a second language classroom: A sociocultural approach* (Doctoral dissertation, University of Nevada, Las Vegas). UNLV Theses, Dissertations, Professional Papers, and Capstones.

Ionescu, T., & Vasc, D. (2014). Embodied cognition: challenges for psychology and education. *Procedia-Social and Behavioral Sciences*, *128*, 275-280.

Jongman, A., Wang, Y., Moore, C., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In *The Handbook of East Asian Psycholinguistics* (pp. 209–217). Cambridge University Press.

Kelly, S. D. (2017). Exploring the boundaries of gesture-speech integration during language comprehension. *Why gesture*, 243-265.

Kelly, S., Bailey, A., & Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts. Collabra. *Psychology, 3*(1), 7.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*(2-4), 63-67.

Lee, Y. C., Wang, T., & Liberman, M. (2016). Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in psychology*, *7*, 1058.

Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., & Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology*, *172*, 168-188.

Li, M., Chen, X., Zhu, J., & Chen, F. (2022). Audiovisual Mandarin Lexical Tone Perception in Quiet and Noisy Contexts: The Influence of Visual Cues and Speech Rate. *Journal of Speech, Language, and Hearing Research*, *65*(11), 4385-4403.

Li, P., Baills, F., Baqué, L., & Prieto, P. (2022). The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research*, First Online. https://doi.org/10.1177/02676583221124075

Li. P, Baills, F., Xi, X., & Prieto, P. (in press). Hand gesture production accuracy predicts simultaneous L2 sound production accuracy. In A. Brown, & S. W. Eskildsen (Eds.). *Multimodality and Multilingual Interaction Across Epistemologies.* Routledge.

Morett, L. M., & Chang, L. Y. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*(3), 347-353.

Prieto, P. (2014). The intonational phonology of Catalan. *Prosodic Typology*, 2, 43-80.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34(2), 145-156.

Smotrova, T. (2017). Making pronunciation visible: Gesture in teaching pronunciation. *Tesol Quarterly*, *51*(1), 59-89.

Tsai, R. (2011). Teaching and learning the tones of Mandarin Chinese. *Scottish Languages Review*, *24*, 43-50.

Maastricht, L.V., Hoetjes, M., & Drie, E.V. (2019). Do gestures during training facilitate L2 lexical stress acquisition by Dutch learners of Spanish? Proceedings of *the 15th*

International Conference on Auditory-Visual Speech Processing (pp. 6-10). ISCA Archive.

Macedonia, M., & Von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics, 6*(3-4), 393–416.

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.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, *95*(4), 2240-2253.

Yang, C. (2019). The effect of L1 tonal status on the acquisition of L2 Mandarin tones. *International Journal of Applied Linguistics*, 29(1), 3-16.

Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the duration of Mandarin tones. In Proceedings of the Interspeech conference (pp. 1407-1411). ISCA Archive.

Yuan, C., Gonzalez-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, *41*(1), 5-32.

Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, *24*(5), 666-689.

Zheng, A., Hirata, Y., & Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 Mandarin tone production. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2179-2195.

Zhen, A., Van Hedger, S., Heald, S., Goldin-Meadow, S., & Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition*, *187*, 178-187.

Zheng, A., Hirata, Y., & Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 Mandarin tone production. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2179-2195

APPENDIX A: Descriptive statistics of tone imitation accuracy for each group in pre-test and post-test

Tone 1

Condition	Session	Mean	SD	SE	95%CI
Gesture	Pre-test	4.36	1.27	0.08	[4.10,4.61]
	Post-test	4.74	0.76	0.05	[4.59,4.90]
No Gesture	Pre-test	4.16	1.40	0.09	[3.87,4.44]
	Post-test	4.73	0.58	0.04	[4.61,4.84]

Tone 2

Condition	Session	Mean	SD	SE	95%CI
Gesture	Pre-test	3.99	1.53	0.10	[3.68,4.30]
	Post-test	4.19	1.16	0.07	[3.95,4.42]
No Gesture	Pre-test	3.99	1.48	0.09	[3.70,4.29]
	Post-test	4.28	1.06	0.07	[4.06,4.49]

Tone 3

Condition	Session	Mean	SD	SE	95%CI
Gesture	Pre-test	3.80	1.28	0.08	[3.55,4.06]
	Post-test	3.77	1.26	0.08	[3.52,4.02]
No Gesture	Pre-test	4.00	1.21	0.08	[3.76,4.24]
	Post-test	3.72	1.23	0.08	[3.47,3.97]

Tone 4

Condition	Session	Mean	SD	SE	95%CI
Gesture	Pre-test	4.82	0.66	0.04	[4.68,4.95]
	Post-test	4.92	0.29	0.02	[4.86,4.97]
No Gesture	Pre-test	4.80	0.78	0.05	[4.64,4.95]
	Post-test	4.87	0.42	0.03	[4.78,4.95]

APPENDIX B: Descriptive statistics of tone imitation accuracy for each group in block 1and block 2

Tone 1

Condition	Session	Mean	SD	SE	95%CI
Gesture	Block 1	4.64	0.51	0.03	[4.55,4.72]
	Block 2	4.59	0.56	0.02	[4.53,4.66]
No Gesture	Block 1	4.57	0.59	0.03	[4.47,4.66]
	Block 2	4.54	0.63	0.02	[4.46,4.61]

Tone 2

Condition	Session	Mean	SD	SE	95%CI
Gesture	Block 1	4.77	0.48	0.02	[4.69,4.85]
	Block 2	4.58	0.66	0.02	[4.51,4.66]
No Gesture	Block 1	4.75	0.48	0.02	[4.67,4.83]
	Block 2	4.65	0.62	0.02	[4.58,4.72]

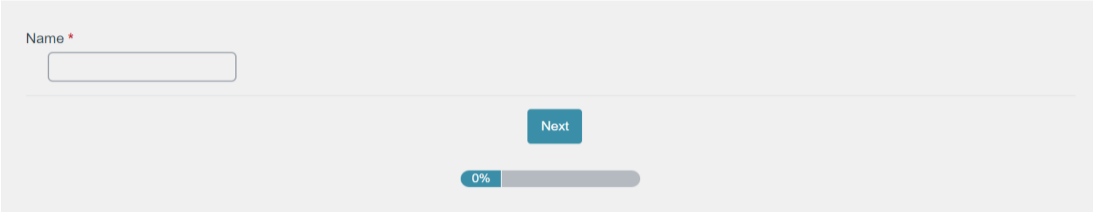
Tone 3

Condition	Session	Mean	SD	SE	95%CI
Gesture	Block 1	3.91	1.14	0.06	[3.72,4.09]
	Block 2	4.22	0.85	0.03	[4.13,4.32]
No Gesture	Block 1	3.95	1.06	0.05	[3.78,4.12]
	Block 2	4.11	0.91	0.03	[4.01,4.22]

Tone 4

Condition	Session	Mean	SD	SE	95%CI
Gesture	Block 1	4.93	0.31	0.02	[4.88,4.98]
	Block 2	4.94	0.25	0.01	[4.92,4.97]
No Gesture	Block 1	4.90	0.36	0.02	[4.84,4.96]
	Block 2	4.95	0.23	0.01	[4.92,4.98]

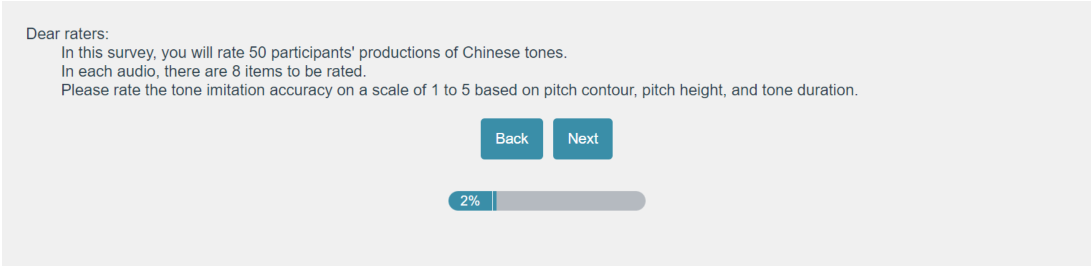
APPENDIX C: Sample of the tone imitation rating survey



Name *

Next

0%

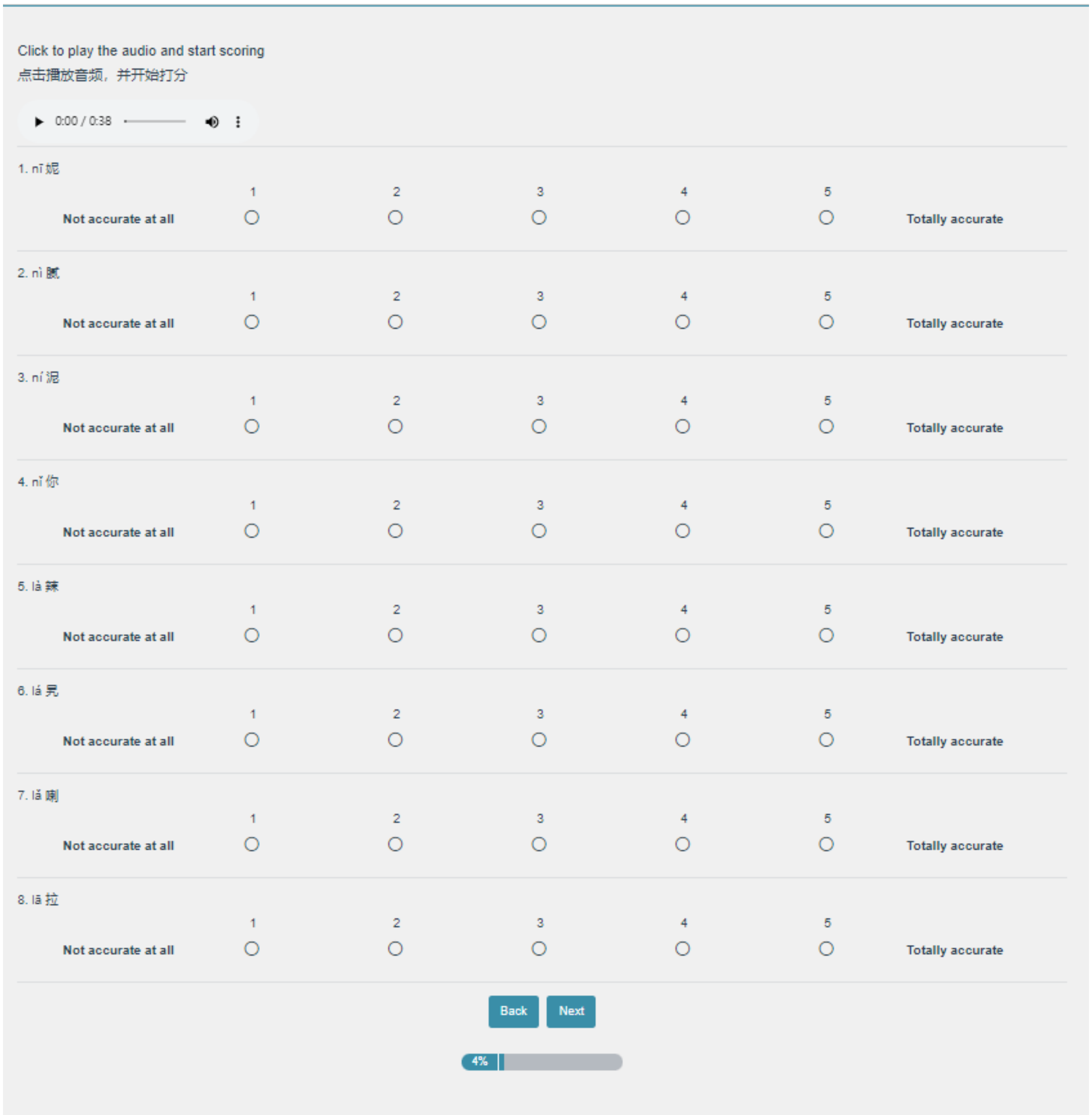


Dear raters:

In this survey, you will rate 50 participants' productions of Chinese tones.
In each audio, there are 8 items to be rated.
Please rate the tone imitation accuracy on a scale of 1 to 5 based on pitch contour, pitch height, and tone duration.

Back Next

2%



Click to play the audio and start scoring
点击播放音频，并开始打分

0:00 / 0:38

1. nī 妮

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

2. nì 腻

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

3. ní 泥

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

4. nǐ 你

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

5. là 辣

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

6. lá 旯

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

7. lǎ 喇

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

8. lā 拉

	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Totally accurate

Back Next

4%