Producing Pitch Gestures enhances the acquisition of Mandarin Chinese Tones

Florence Baills

Màster: Lingüística Teórica y Aplicada
Edició: 2015-2016
Directors: Dra Pilar Prieto Vives
Any de defensa: 2016

Col·lecció: Treballs de fi de màster
Programa oficial de postgrau
Departament de Traducció i Ciències del Llenguatge
Acknowledgments

My most profound thanks and my gratitude go to Dr. Pilar Prieto for opening the doors of experimental research for me, for sharing her knowledge and wisdom with me and for providing guidance and support in this project.
I am very grateful to Santiago González for his crucial help on so many aspects of this investigation, for teaching me so much and for his moral support throughout the project.
I am happy to have experienced, with Nerea Suárez, the adventure of thinking, recording and data collecting.
Thank you to all the members of the GrEP group for their help, their nice spirits and for simply being there!
Finally, I am deeply grateful to all the people who patiently participated in the creation of the material for this experiment and without whom nothing would have been possible: Rita Zaragoza Jové and Chenjie Yuan were our instructors in the stimuli videos. Feifei Lian and Yuanyuan Sun were our voices for the auditory items in the tests.

And to Pedro, Núria and Guillem, you are my world.
Abstract

Research in second language acquisition has described the positive effects of observing iconic gestures on vocabulary acquisition (Kelly, McDevitt, & Esch, 2009; Macedonia, Müller, & Friederici, 2011; Tellier, 2008) and recent studies have also proven that observing speech with beat gestures can improve L2 learners’ production of suprasegmental features in a foreign language (Gluhareva & Prieto, in press), as well as in word recall (Kushch, Igualada, & Prieto, submitted). However little is known about the role of a specific type of metaphoric gesture that mimicks melody in speech (e.g. the so-called pitch gestures) in the learning of L2 intonational variations across languages. Recently, Morett & Chang (2014) demonstrated that the production of visuospatial features of pitch gestures by English learners of Mandarin enhanced discrimination between the meanings of Mandarin words differing only in tone. However, it is not clear whether the specific gains obtained in this study can be mostly attributed to the production or to the perception of these gestures by second language learners. The goal of this investigation is to explore the gain of observing vs. producing pitch gestures during a short multimodal training of Mandarin Chinese tones on tonal discrimination abilities and on learning newly presented Mandarin words that only differ in tone. Fifty-seven Catalan-dominant native speakers participated in a short training on Mandarin Chinese tones where the instructor produced pitch gestures. The results show that producing gestures favors tone discrimination and word recall more than merely repeating the word and viewing the gesture. Accordingly, the production of pitch gestures by beginning learners of Mandarin Chinese can serve as a facilitator for learning tones and vocabulary.

Keywords: Second language acquisition, gesture, pitch, embodied learning, tones, Mandarin Chinese
1 Introduction

Research in second language acquisition agrees that gesturing helps learners express themselves, to convey meaning, and to compensate for speech difficulties; conversely, native interlocutor’s movements also play a potential role for comprehension and learning (see Gullberg, 2006 for a review). In the foreign language classroom, teachers typically use gestures as a tool to improve the language acquisition process (Taleghani-Nikazm, 2008).

The Verbotonal method, originally developed by Gurbina (1939) for the acquisition of oral language by deaf people and extended to second language acquisition under the name audio-visual-global structural (AVGS) method (Gurbina, 1984), emphasizes the importance of rhythm, intonation, and gestures as optimal factors in the acquisition of a second language. In theory, a student must learn pronunciation, syntax, and gestures altogether thanks to simultaneous auditory and visual stimuli. In practice, however, adepts of this method did not develop the use of gestures and concentrated on the phonetic and prosodic aspects. In the 1960s, Asher (1966) carried out studies exploring the effects of the “total physical response” (TPR) learning strategy using the Japanese and Russian languages with adults and children. Learners had to perform actions while they were hearing the corresponding order, starting with single action verbs to gradually increase to more complex imperative sentences. The author suggested that acting out during retention tests dramatically facilitated the acquisition of listening skills. However, this line of research didn’t become influential.

Research in recent decades has stressed the benefits of observing iconic and metaphoric gestures\(^1\) for word recall during first language and second language acquisition. A seminal study on the impact of observing iconic and metaphoric gestures on memory for verbal information in a foreign language was conducted by Quinn-Allen (1995). She taught English-speaking students 10 French expressions by accompanying the expressions with illustrative, semantically related gestures typical of French culture. For example, the gesture paired with

---

\(^1\) In this work, we use McNeill’s (1992) classification system of hand co-speech gestures as iconics, metaphorics, deictics, and beats).
the sentence “Do you want something to drink?” was performed by pointing the thumb toward the open mouth. Results revealed that learning gestures simultaneously with French expressions lead to greater immediate recall and a smaller decay in recall after 2 months than learning these expressions without the gestures. Kelly et al. (2009) trained 28 young adults on recalling 12 Japanese verbs conveying common everyday meanings. The words were presented according to four modes: speech, speech + congruent gesture, speech + incongruent gesture, and repeated speech. The results showed that participants memorized the largest number of words in the speech + congruent gesture mode, followed by the repeated speech mode, and the least number of words was memorized when they were accompanied by an incongruent gesture.

Similarly, studies investigating beat gestures (e.g., hand rhythmic gestures that are associated with prosodic prominence), have demonstrated that they can favor language comprehension and recall in adults (So, Sim Chen-Hui, & Low Wei-Shan, 2012) and also second language novel word acquisition (Kusch et al., submitted). Another recent study by Gluhareva & Prieto (in press) showed that training with a beat gesture observation by Catalan learners of English significantly improved the participants’ accentedness ratings on the set of difficult (more discourse-demanding) items, thus favoring better pronunciation².

However, very little experimental work has been carried out to assess to what extent producing co-speech gestures can help learners in the acquisition of a new language. In a study with 20 French children (average age 5.5) learning English, Tellier (2008) presented 8 common words (house, swim, cry, snake, book, rabbit, scissors, and finger). Four items were associated with a picture and four items were illustrated by a gesture that the children saw in a video and they thereafter performed. Enacted items were better memorized than items enriched visually by the pictures. In a recent study, Macedonia and Kimesch (2014) looked at the use of gesture in the language classroom in a within-subject longitudinal study lasting 14 month. They trained university students to learn 36 words (9 nouns, 9 adjectives, 9 verbs and 9 prepositions) in an artificial language corpus. For 18 items, participants only listened to the word and read it. For another 18 items, participants were additionally instructed to perform

² However, beat gestures haven’t yet proven to systematically improve perception in a foreign language. In Hirata and Kelly (2010), hand gestures represented the rhythm of short and long vowels, a vertical chopping movement and a long horizontal sweeping movement, respectively. Contrary to their predictions, participants in the speech-gesture condition did not learn to perceive the short/long vowel contrasts any better than those in the speech alone condition.
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 symbolic gestures significantly enhance vocabulary learning in quantity and over time. Verbal items belonging to different word categories benefited from gesture use during the learning phase. Furthermore, the authors claim that under Klimesch's connectivity model (CM) of information processing (Klimesch, 1987), a word is better integrated into long-term memory if it is deep, that is, if it is comprised of many interconnected components.

Little is known about the potential learning gains of self-producing actions and gestures as opposed to just observing them in others. First, in various instructional settings, producing gestures has been found more effective than observing them alone (S. D. Goldin-Meadow, Cook, & Mitchell, 2009; S. Goldin-Meadow, 2014). Goldin-Meadow and colleagues (S. D. Goldin-Meadow et al., 2009) investigated how children extract meaning from their hand movements and showed that children 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 gesture. Second, recent results from neuroscientific experiments seem 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 native (Masumoto, 2006) and foreign languages (Macedonia, Müller, & Friederici, 2011). However, mere observation of an action 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. The authors contend that the possible engagement of the same neural mechanisms involved in both observation and imitation could explain some results in the behavioral experiments on embodied learning where memory performance was found equivalent under the “self-performed task” (SPT) condition and the “experimenter-performed task” condition (Cohen, 1981). Notwithstanding, Engelkamp and colleagues (Engelkamp, Zimmer, Mohr, & Sellen, 1994) showed that SPTs lead to superior memory performance in recognition tasks for longer lists of items; between 24 and 48 but not for short lists (12 items).
This study is concerned with a different type of gesture, namely pitch gestures, or gestures that reproduce the melodic F0 curve in speech (see Morett & Chang, 2014) and their effects in learning lexical tones. Lexical tones are known to be particularly difficult to acquire for speakers of non-tonal languages (M. Wang, Perfetti, & Liu, 2003). One of the challenge is that the spoken syllable in a tonal language consists of a segment overlapped by a tone resulting in a complex perceptual input. In the case of Mandarin Chinese, differentiating between the second and the third tone is particularly challenging (Kiriloff, 1969). There is evidence that addressing this difficult feature directly with an auditory training consisting of paired combinations of tones increases identification accuracy significantly between pre- and posttests (Wang, Jongman, & Sereno, 2003). Moreover, visual illustrations of the pitch contours depicting the acoustic shape of the lexical tones together with pinyin spelling of the spoken syllables has also demonstrated to facilitate their acquisition (Liu et al., 2011).

Even though gestures are widely used by teachers of Mandarin Chinese to teach lexical tones, little experimental work has been conducted on the positive effect of these gestures on the perception and production of tones. A set of longitudinal studies by Lin and Jianqin (Lin & Jianqin, 2013a, 2013b) showed a positive effect of teachers’ gestures on the perception and production of tones by elementary level learners. An extensive longitudinal study (Chen, 2013) focused on the positive role of gestures as tone markers in the tonal achievements of 40 learners of Chinese as a second language in the context of a communicative classroom both during drills and face-to-face interactions. Results showed better tonal production and a higher frequency of accurate responses by learners who saw and produced gestures as tone markers than by learners following the traditional numeric notational system of Mandarin tones. To our knowledge, the only experimental study on the potential benefits of pitch gestures on the learning of L2 tones was done by Morett & Chang (2014). In a between-

---

3 An interesting experiment by Connell et al. (Connell, Cai, & Holler, 2013) demonstrated that pitch and space have a shared representation, usually described as “high” and “low”, and that the mental representation of pitch is audio-spatial in nature. These findings led us to hypothesize that gestures reproducing pitch contours might help learners of tonal languages (and also of intonational languages) to learn the intonational properties of a second language faster.

4 Mandarin tonal pitch was found to be highly correlated with the shape of the F0 contour in early linguistic studies (Chao, 1930)
subjects experiment, 57 participants were asked to repeat aloud the 12 Chinese words presented in minimal pairs and their translation in English and also to imitate the gestures performed by the instructor in the video in three conditions (semantic or iconic gestures, pitch gestures, and no gestures). Participants in the pitch gesture condition had better results at a word-meaning association task than participants in the semantic gesture and no gesture conditions. Interestingly, the authors found that pitch gestures failed to enhance lexical tone identification compared to the other conditions. Thus, their results show that producing pitch gestures facilitates English speakers’ association of Mandarin Chinese words differing in lexical tone with their meanings. Interestingly, Chen and Massaro (2008) also reported an improvement in the production of lexical tones after a short training where learners’ awareness of the visible movements of the neck, head, and mouth of the speaker is raised.

However, the abovementioned studies do not tease out whether the tone learning improvement is due to observing or producing the target pitch gestures. The main goal of this investigation is to explore in more detail the gains of producing pitch gestures as opposed to just observing them. We hypothesize that learners producing the pitch gestures will obtain more benefits from the task than repeating them out loud after the instructor. We carried out a between-subjects experiment with two parts. In the first part, we trained and tested Catalan-dominant speakers’ perception of Mandarin Chinese lexical tones by presenting participants with 18 minimal lexical tone pairs in two conditions, e.g. Pitch Gesture Observe vs. Pitch Gesture Produce. In the second part of the experiment, we trained and tested the same participants on word acquisition of 6 pairs of words, which only differed in their lexical tone composition. Moreover, since musical experience has been correlated with tonal perception abilities (e.g., Wong & Perrachione, 2007), the study controlled for participants’ musical knowledge as well as pitch and melodic abilities.

2 Methods

In this study we address the question of whether beginner second language learners of a tone language (e.g. Mandarin Chinese), can enhance their acquisition of lexical tones after a short training with pitch gestures where they must either perform the pitch gestures themselves (Produce Pitch Gesture / Produce Speech Condition) or simply watch and repeat the speech
without producing the pitch gestures (Observe Pitch Gesture / Produce Speech Condition). A between-subjects short training experiment was conducted with these two main conditions.

2.1. Participants

Fifty-seven undergraduate students (age M= 19, 93 years, SD = 1, 414, 9 males) were recruited at Universitat Pompeu Fabra in Barcelona, Spain. All were Catalan-Spanish bilinguals (Catalan-dominant), 59.6 % reported a daily usage of Catalan superior to 75%, 24.6% reported a daily usage of 50-75% and 14% reported a daily usage of 25-50%. All of them were right-handed (and thus they produce the gesture with their dominant hand) and manifested no previous knowledge of Mandarin Chinese or any other tonal language. All participants submitted written consent to take part in the study and filled out a questionnaire. As musical expertise can be directly related with pitch perception abilities (e.g. Sadakata & Sekiyama, 2011), participants were asked about their musical knowledge, as well as foreign language expertise and motivation to learn Chinese. All participants were paid 10 euros for their participation in the study, which lasted approximately one hour.

2.2. Materials

The experiment consisted of three consecutive phases, e.g., a first habituation phase containing an introduction to Mandarin tones followed by two consecutive trainings, namely the pitch training and the vocabulary training (see Figure 3 with the procedure).

All the target audiovisual materials for the three phases of the experiment (as well as for the specific tasks) were recorded by a native speaker of Chinese and by a Catalan-Chinese bilingual speaker. The video recordings were performed at the Sala de Recerca Experimental en Llenguatge of the Department of Translation and Language Sciences (Universitat Pompeu Fabra) with a PDM660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone. The two instructors were recorded against a white wall and from the upper half of the body.

5 Laboratory for Experimental Research on Language
Mandarin Chinese distinguishes four main lexical tones described according to their pitch contours along a numbering scheme: high flat-level (1), rising (2), low falling and rising (3) and high-falling (4) (Chao, 1968). For example, the syllable /ma/ can have four different meanings according to the tone used when spoken out: /ma/1 means mother, /ma/2 means hemp, /ma/3 means horse and /ma/4 means scold. The habituation video first illustrated the four Mandarin tones verbally in Catalan and visually, with the help of the following 4-scale diagram (see Figure 1, Zhu, 2012).

![4-scale diagram representing the 4 lexical tones in Mandarin Chinese](image)

Two different videos were produced for the habituation phase, one to prepare the participants to repeat the target tones without doing any gesture (in the Gesture Observe Condition) and the other to prepare the participants to repeat the target tones with pitch gestures (in the Gesture Produce Condition). For the production of the pitch gestures, the instructors used their right hand to naturally gesture from left to right and the videos were digitally flipped to allow participants to observe the gestures from left to right or to mirror the gestures with their own right hand. Importantly, the two instructors were trained to use clear visuospatial hand gestures making sure that the hand movements corresponded to the pitch variations and the natural duration corresponding to each lexical tone (see Figure 2). For spatial consistency across renditions, the imaginary space for the hand gestures was divided in four areas: the high tonal range corresponded to the face level, the mid tonal range to the shoulder level, the

6 There is a neutral, mid-flat tone (often labeled as 5), that we don’t take into account in this study since it doesn’t involve a difference in meaning.
mid-high frequency range corresponded to the chest level, and finally the mid-low frequency range corresponded to the hips. The duration of the tones, which can be a clue to determining what tone is used, was left natural.

Figure 2: Video examples (with corresponding sound waves and pitch tracks) for the syllable “puo” pronounced with tones 1 and 2 by a male speaker (left panels) and of the syllable “mi” pronounced with tones 3 and 4 by a female speaker (right panels).

2.2.2. Audiovisual stimuli for the pitch training

Table 1 shows the stimuli for the pitch training, which consisted of 18 minimal pair words contrasting only in lexical tones. All of the target monosyllabic Mandarin Chinese words syllables consisted of phonemes that constituted acceptable combinations for Catalan-Spanish bilingual speakers. The words were typically presented in orthographic form following the pinyin orthographic conventions, except when this would cause difficulty for Catalan speakers (see the actual transcription used in brackets). The stimuli were presented in pairs, following the list of stimuli shown in Table 1 to improve contrast perception (Kelly, Hirata, Manansala, & Huang, 2014). In total, each of the four lexical tones was repeated 9 times. There were 6 groups with differing orders of presentation in order to avoid primacy and recency effects (see Appendix 1).
Table 1: Stimuli for the pitch training (36 words; 18 pairs). The orthographic form of the syllable presented to the participants, when different from the pinyin orthography, is specified within brackets.

Table 2: Stimuli for the vocabulary training (36 words; 18 pairs)

7 http://psico.fcep.urv.es/utilitats/nim/index_cat.php
2.2.4. Auditory stimuli for the tonal classification task and the word meaning recall and word meaning association tasks

For the tonal classification task, 8 items (4 trained, 4 untrained) were chosen as real syllables or pseudo-syllables respecting Chinese phonotactic rules. Auditory materials were recorded by three native speakers of Mandarin Chinese and uploaded on the online survey builder Survey Gizmo.

<table>
<thead>
<tr>
<th>TRAINED ITEMS</th>
<th>UNTRAINED ITEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE SPEAKER</td>
<td></td>
</tr>
<tr>
<td>mi T4</td>
<td>te T2</td>
</tr>
<tr>
<td>fu T3</td>
<td>nu T4</td>
</tr>
<tr>
<td>FEMALE SPEAKER</td>
<td></td>
</tr>
<tr>
<td>txi T2</td>
<td>la T1</td>
</tr>
<tr>
<td>pi T1</td>
<td>txe T3</td>
</tr>
</tbody>
</table>

Table 3: Items for the tonal classification task

For the vocabulary tasks, recordings featured a speaker of a different sex than in the training to ensure that post-test 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.

2.3. Procedure

The participants were tested individually in the Sala experimental (see section 2.2). The experimenter was always present in the room in which the experiment took place to ensure that participants attended to the stimuli and did not omit any part of the training, i.e., repetition or gesture imitation. No feedback was provided during any of the experimental tasks. Prior to realizing the tasks, participants filled out a questionnaire to certify that Catalan was their first language, to specify their daily use of Catalan language, to know the number of foreign languages they spoke, their musical knowledge and number of years of practice.

Figure 3 summarizes the experimental procedure. After signing a written consent and before starting the experiment proper, a span memory test (Bunting, Cowan, & Saults, 2006) in

---

8 https://www.surveygizmo.com/
Catalan was carried out. Participants were also informed that the experiment consisted in a beginner tutorial of Mandarin Chinese and that they will learn how to pronounce words and some vocabulary.

Participants were randomly assigned to one of the two between-subject groups. First participants were presented a video consisting of a short introduction to the Chinese tones, which lasted 8 minutes. After this, a first pitch training was given (5 minutes), which was followed by a tone classification task. Then, participants were shown a second vocabulary training (6 minutes), which was followed by two tasks, namely a word meaning recall task and a word meaning discrimination task. Finally, all participants were asked to participate in an online evaluation of musical skills for melody and pitch perception (Law & Zentner, 2012).⁹

![Figure 3: Schema of the experimental procedure](image)

2.3.1 Pitch training

Participants were trained to discriminate between pairs of Mandarin Chinese lexical tones (see Table 3). Participants in the two conditions viewed the same training video that only differed in the way in which they were reacting to the audiovisual materials.

In the Gesture Observe/Speech Produce condition, participants watched the video and repeated the syllable orally (with no pitch gesture). Before starting the training, the following instructions were given orally by the experimenter to complete the written instructions in the video: “Now you are going to practice the tones in pairs. You will have first to watch and

---

⁹ Profile of Music Perception Skills: modular proms (18 questions each), melody and pitch, available at:
https://wwwu.ibk.ac.at/psychologie/forschung/tests_and_diagnostics/proms/about-the-proms/
listen to the instructor produce the two syllables, then, when the screen goes black, it will be your turn to repeat the syllable aloud without moving”.

Participants in the Gesture Produce/Speech Produce condition watched the video, and repeated the syllable and re-enacted the pitch gesture. Beforehand, they were instructed the following: “Now you are going to practice the tones in pairs. You will have first to watch and listen to the instructor produce the two syllables, then, when the screen goes black, it will be you turn to mimick the gesture while repeating the syllable aloud”. Therefore, the difference between the Gesture Observe and the Gesture Produce conditions doesn’t lie in the stimuli but in the instructions given to the participants.

Figure 4 illustrates the learning sequence of the target syllable "mi" produced with tones 4 and 3. Three successive target syllables were presented for each tone contract (see table 1)

1) Gesture Observe (GO): Participants were first presented with the two Chinese syllables in written form, heard the instructor pronounce the syllable with both tones and perform the two corresponding gestures. Then when the screen went black they had to repeat the syllable aloud only.

2) Gesture Produce (GP): Participants were first presented with the two Chinese syllables in written form, heard the instructor pronounce the syllable with both tones and perform the two corresponding gestures. Then when the screen went black they repeated the syllable aloud and imitated the gesture performed by the instructor.

Figure 4: Example of the pitch training with tones 4 and 3 over the syllable “mi”.

Tone classification task

Immediately after the training, participants were asked to complete a tone classification task with 8 audio-only items (see table 2). 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 program randomized the items automatically so that the presentation of items was different for each participant.

2.3.2 Vocabulary training

In the vocabulary training, participants learned 12 words in the same conditions as in the first experiment. Words for each minimal pair were presented in consecutive trials and all learning trials were presented three times in different orders. For each word item to be learnt, participants first saw the written word in Catalan and then viewed a video clip with the instructor producing the pitch gestures. In the Gesture Observe condition, they just repeated the word in Chinese aloud without re-enacting the pitch gesture. In the Gesture Produce condition, they repeated the word in Chinese aloud while re-enacting the pitch gesture.

![Figure 5: Example of a word learning trial for the vocabulary training with the minimal pair of Mandarin Chinese words bō onada 'wave' and bō oncle 'uncle'](image)

**Word-learning tasks**

In a word meaning recall task, we instructed participants to listen to the 12 target Mandarin Chinese words and to translate each one of them into Catalan. They could only listen to each word once, write down their answer and then go to the next screen to listen to the next word.

After that, in a word meaning association task, participants heard each Mandarin word presented in the vocabulary training while two Catalan words - the correct translation into Catalan and the translation of the other word in that minimal pair (same syllable, different tone) - were presented as options. Participants identified the meaning of each word by circling the correct word between the two.
3. Results

Three Generalized Linear Mixed Models (GLMM) were performed using IBM SPSS Statistics 23.0 (IBM Corporation 2016), one for each dependent variable. The dependent variables were the results of the three tasks, namely tone classification, word meaning recall and word meaning association task. We used binary coding (0 = incorrect or missing, 1 = correct) to quantify the participants ‘answers. In all three models, the fixed factor was the experimental Condition (two levels: Gesture Observe and Gesture Produce). Random factors were age, gender and number of foreign languages spoken.

The result of the first GLMM model on the proportion of tones classified correctly revealed a main effect of Condition (F(1, 455) = 7.275, p < 0.07). The mean proportion of correctly classified tones was of 59.1% (SD = 0.04112) in the Gesture Observe Condition and of 71.2% in the Gesture Produce Condition.

![Figure 6: Mean accuracy in the tone classification task for the two conditions (Gesture Produce and Gesture Observe).](image)

The GLMM model for the word meaning recall performance revealed a main effect of condition (F(1,682) = 21,057, p < 0.001). As we can see in Figure 7, the mean performance was of 39.9% (SD = 0.05030) in the Gesture Observe Condition and of 57.6% (SD = 0.4177) in the Gesture Produce Condition.
Finally, the GLMM result for the word meaning association task revealed a main effect of Condition (F(1,682) = 6.747, p < 0.01). Mean results reached 73.8% (SD = 0.03338) in the Gesture Observe Condition and 81.5% (SD = 0.02809) in the Gesture Produce Condition.
For each task, we also checked for possible effects of memory aptitude, musical knowledge, experience and abilities by adding these as fixed factor and as fixed 2-factors effects together with condition in a GLMM model.

In the tones classification task, we didn’t find any significant effect of memory aptitude ($F(2, 451) = 0.629, p = 0.534$). For musical knowledge, results showed an effect of condition for musicians but not for non-musicians ($F(1,453) = 6.127, p = 0.014$), that the years of practicing music were also significant for accuracy results for this task ($F(4, 439) = 3.902, p = 0.04$). There were 43% (SD = 0.27) of musicians in the Speech Produce Condition and 52% in the Gesture Produce Condition.

In the word meaning recall task, we didn’t find any significant effect of memory aptitude ($F(2, 678) = 0.111, p = 0.895$). For musical knowledge, there was no significant difference between musician and non-musician ($F(1,680) = 0.986, p = 0.321$). The years of practicing music were not significant for accuracy results in this task ($F(2, 439) = 3.902, p = 0.04$).

In the word meaning association task, we didn’t find any significant effect of memory aptitude ($F(2, 678) = 0.917, p = 0.400$). For musical knowledge, there was no significant difference between musician and non-musician ($F(1,680) = 0.986, p = 0.321$). The years of practicing music were not significant for accuracy results in this task ($F(12, 666) = 1.382, p = 0.169$).

4. Discussion and Conclusions

The present study examined whether a short training with pitch gestures could enhance the learning of Mandarin lexical tones and words by Catalan speakers in three tasks, namely (1) a tone classification task, (2) a word-meaning recall task, and (3) a word meaning association task. The results of a short between-subject training study revealed that the group of participants who repeated the pitch gestures obtained significantly higher results in the three tasks as compared to participants who did not repeat the pitch gestures.
First, the finding that producing pitch gestures resulted in improved tonal classification scores\textsuperscript{10} provides evidence that the participants who repeated the actions themselves were able to more accurately map visuospatial information conveyed via metaphorical gestures, namely the representation of pitch on a vertical axis where high-frequency pitch is represented upward and low-frequency pitch is represented downward, on their representation of lexical tone. These results contrast with Morett & Chang (Morett & Chang, 2014) findings that producing pitch gestures did not significantly enhance lexical tone identification. This difference may be due to the role of musical knowledge and musical experience, which were significant for accuracy in tone classification and may also explain the different results between the studies. Nevertheless, in the two studies it is clear that producing pitch gestures significantly helped participants to better remember the words differing only in lexical tone composition.

The results of the current study support the embodied learning paradigm. Research on embodied learning has demonstrated that physically produced actions enhance learning and memory. Cohen (1981) set up the subject-performed task (SPT) paradigm by comparing subject-performed tasks (SPTs) recall to words recall, using tasks which the subjects performed (SPTs) as to-be-recalled list items. SPT recall obeyed one law of word recall in showing a strong positive recency effect in immediate free recall, which appeared to be due to these events being in a temporary state of high accessibility. However SPT recall did not show the primacy effect associated with word recall, and subjects did not report using active memorization strategies for SPT lists as they did for word lists. These results indicate that SPTs should be regarded as a different class of memory event than words, and that memory models dealing with SPT recall should de-emphasize the importance of encoding, stressing instead retention and retrieval operations. Saltz and Donnenwerth-Nolan (1981) showed that both motoric enactment and visual imagery significantly facilitated retention compared to a verbal-only control group but also that the effects of motoric enactment on sentence recall were disrupted by a motoric competition task, while the effects of visual imagery on recall were not. They also demonstrated that a verbal competition task disrupted the verbal-only technique for sentence memory but did not affect the facilitative effects of motoric enactment. Similarly, a motoric competition task again disrupted the facilitative effects of motoric

\textsuperscript{10} E.g. to what extent participants were able to remember the four tones and recognize them auditorily.
enactment of sentences, but had little or no effect on sentences learned by means of verbal-only instructions. The authors suggest that motoric enactment is effective in sentence recall because it leads to the storage of some type of motoric trace or image. These studies show that enactment may stimulate visual imagery or verbal mediation. We suggest that pitch gestures may reinforce neural representations of tones by illustrating metaphorically the notion of pitch.

In general, the findings of this study offer support of the benefits of multimodal instruction on lexical tones and vocabulary in Mandarin Chinese. Previous research has shown that multimodal input may be useful when learning novel speech sounds in L2 (Hardison, 2003; Hirata & Kelly, 2010; Y. Wang, Behne, & Jiang, 2008). Craik & Tulving (Craik & Tulving, 1975) defined the principle of depth of processing according to which mental representations of linguistic information would be more durable when the information is elaborated at the time of presentation. Greater recall is a measurable indication of depth of processing and of more durable mental representations.

This study may provide additional support for multimodal teaching in the L2 classroom, especially in tonal languages, where the acquisition of suprasegmental features is linked to lexical knowledge. A short training with pitch gesture production seems to allow for the reinforcement of the phonological representation of words and facilitate their association with their semantic content, improving recall and discrimination of words differing in lexical tones. In practice, teachers of Mandarin Chinese at a beginner level could proceed first to a pitch training to teach the students the four tones and how they integrate with syllables. Afterwards, new words would be presented in connection to the related pitch gestures and students would be encouraged to repeat the gesture themselves.

It would be interesting to measure the effects of pitch gestures on memory in the long term. Macedonia and Klimesch (Macedonia & Klimesch, 2014) found that items enacted with semantic gestures experienced slower decay than items learned under audio-visual condition. Indeed, there is even more interest in implementing a learning tool that ensures that the words acquired in a second language are not only learned faster but also retained longer.

Another factor that could be investigated in future experiments might be a control for different learning styles. If gestures help learning altogether, it may be the case that producing gestures will help learners with a certain type of learning style, whereas observing them will be more profitable to another (Chen, 2015).
In sum, during early stages of acquisition, perceptual training accompanied by pitch gestures is beneficial in learning lexical tones. Subsequently, this knowledge is beneficial for learning words that are in competition - only distinguishable by tones together with pitch gestures. Within the wider framework of grounded cognition, a tight connection has been established between gestures and language (Barsalou, 2008). This study comes as an enrichment of the theory of embodied language processing, stating that moving one’s body while learning impacts how we process and comprehend information.
REFERENCES


Appendix 1: Order of presentation of stimuli in the pitch training

We have controlled for:
* Primacy and recency: all the combinations in first and last position are different.
* All the combinations are different inside each color group.
* There is never the same combination in the same position for each group.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>T1/T2</th>
<th>T4/T3</th>
<th>T4/T1</th>
<th>T1/T3</th>
<th>T1/T3</th>
<th>T4/T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2</td>
<td>T1/T3</td>
<td>T4/T2</td>
<td>T3/T2</td>
<td>T1/T2</td>
<td>T1/T2</td>
<td>T4/T3</td>
</tr>
<tr>
<td>Group 3</td>
<td>T4/T1</td>
<td>T2/T3</td>
<td>T3/T4</td>
<td>T2/T4</td>
<td>T2/T4</td>
<td>T3/T1</td>
</tr>
<tr>
<td>Group 4</td>
<td>T3/T2</td>
<td>T1/T4</td>
<td>T3/T1</td>
<td>T3/T4</td>
<td>T3/T4</td>
<td>T2/T1</td>
</tr>
<tr>
<td>Group 5</td>
<td>T2/T4</td>
<td>T3/T1</td>
<td>T2/T1</td>
<td>T4/T1</td>
<td>T4/T1</td>
<td>T2/T3</td>
</tr>
<tr>
<td>Group 6</td>
<td>T3/T4</td>
<td>T2/T1</td>
<td>T4/T2</td>
<td>T3/T2</td>
<td>T3/T2</td>
<td>T1/T4</td>
</tr>
</tbody>
</table>

Appendix 2: Frequency of Catalan vocabulary for the vocabulary training

<table>
<thead>
<tr>
<th>Word</th>
<th>Relative Frequency</th>
<th>Log</th>
<th>Absolute Frequency</th>
<th>Nb of letters</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>onada</td>
<td>17,209</td>
<td>1.26</td>
<td>882</td>
<td>5</td>
<td>noun adj</td>
</tr>
<tr>
<td>oncle</td>
<td>58,845</td>
<td>1.777</td>
<td>3016</td>
<td>5</td>
<td>noun</td>
</tr>
<tr>
<td>taxi</td>
<td>6,087</td>
<td>0.85</td>
<td>312</td>
<td>4</td>
<td>noun</td>
</tr>
<tr>
<td>cançó</td>
<td>63,391</td>
<td>1.809</td>
<td>3249</td>
<td>5</td>
<td>noun</td>
</tr>
<tr>
<td>fons</td>
<td>252,119</td>
<td>2.403</td>
<td>12922</td>
<td>4</td>
<td>noun adj adv</td>
</tr>
<tr>
<td>peça</td>
<td>48,348</td>
<td>1.693</td>
<td>2478</td>
<td>4</td>
<td>noun adj</td>
</tr>
<tr>
<td>pila</td>
<td>21,267</td>
<td>1.348</td>
<td>1090</td>
<td>4</td>
<td>noun</td>
</tr>
<tr>
<td>torre</td>
<td>43,977</td>
<td>1.653</td>
<td>2254</td>
<td>5</td>
<td>noun vrb</td>
</tr>
<tr>
<td>lli</td>
<td>8,507</td>
<td>0.978</td>
<td>436</td>
<td>3</td>
<td>noun adv</td>
</tr>
<tr>
<td>insult</td>
<td>5,307</td>
<td>0.8</td>
<td>272</td>
<td>6</td>
<td>noun</td>
</tr>
<tr>
<td>arròs</td>
<td>24,818</td>
<td>1.412</td>
<td>1272</td>
<td>5</td>
<td>noun</td>
</tr>
<tr>
<td>mel</td>
<td>30,203</td>
<td>1.494</td>
<td>1548</td>
<td>3</td>
<td>noun adj</td>
</tr>
</tbody>
</table>
Appendix 3: Order of presentation of stimuli within each bloc for the vocabulary training

| 1st bloc | puo puo mi mi tha tha ke ke ma ma ti ti T1 T2 T3 T4 T3 T2 T1 T4 T2 T4 T4 T1 T3 |
| 2nd bloc | tha tha ke ke ti ti ma ma puo puo mi mi T2 T3 T4 T1 T3 T1 T4 T2 T2 T4 T1 T4 T3 |
| 3rd bloc | ma ma ti ti mi mi puo puo ke ke tha tha T2 T4 T3 T1 T3 T4 T1 T2 T1 T4 T3 T2 |

Appendix 4: Order of presentation of stimuli for the vocabulary training

<table>
<thead>
<tr>
<th>GROUP 1</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP 2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>GROUP 3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>GROUP 4</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>GROUP 5</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GROUP 6</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>