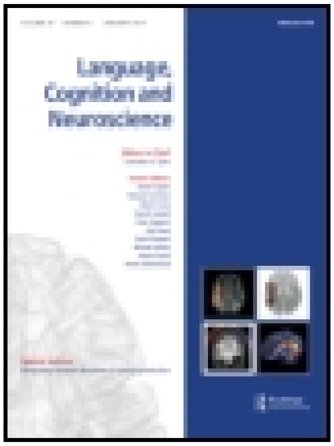

Language, Cognition and Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp21

Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition

Laura M. Morett[a] & Li-Yun Chang[b]

[a] Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

[b] Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Published online: 20 May 2014.

Language, Cognition and Neuroscience, 2014
http://dx.doi.org/10.1080/23273798.2014.923105

Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition

Laura M. Morett[a]* and Li-Yun Chang[b]

[a]*Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA;* [b]*Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA*

(*Received 11 November 2013; accepted 29 April 2014*)

Lexical tones – pitches differentiating between word meanings in tonal languages – are particularly difficult for atonal language speakers to learn. To test the hypotheses of embodied cognition and spoken word recognition, we examined whether – and how – gesture could facilitate English speakers' discrimination between Mandarin words differing in lexical tone. Words were learned with gestures conveying tone pitch contours (pitch gestures), gestures conveying word meanings (semantic gestures) or no gestures. The results demonstrated that pitch gestures enhanced English speakers' discrimination between the meanings of Mandarin words differing in tone, whereas semantic gestures hindered their identification of tones in learned words. These findings indicate that the visuospatial features of pitch gestures strengthen the relationship between English speakers' representations of Mandarin lexical tones and word meanings, supporting the predictions of spoken word recognition and embodied cognition.

Keywords: second language acquisition; word learning; gesture; lexical tone; embodied cognition; spoken word recognition

One of the most difficult aspects of learning a second language (L2) is the acquisition of sounds absent from the native language. An example of such a sound is lexical tone, which consists of a pitch superimposed upon phonemes that comprise words (Xu, 1994). Although it has been claimed that lexical tones are particularly difficult for atonal language speakers to acquire (Kiriloff, 1969; Shen, 1989), research has shown that English speakers can be trained to successfully differentiate between Mandarin tones (Wang, Spence, Jongman, & Sereno, 1999; Wong & Perrachione, 2007). There is evidence that visual illustrations of lexical tones facilitate their acquisition (Bluhme & Burr, 1971; Liu et al., 2011), and that meaning-related iconic hand gesture enhances the learning of L2 word-meaning associations (Kelly, McDevitt, & Esch, 2009; Tellier, 2008). To date, however, no research has examined whether gesture can facilitate English speakers' acquisition of Mandarin words differing in lexical tone. The current study focuses on this question, elucidating the contributions of different types of gestures to English-speaking L2 learners' representations of Mandarin words.

Mandarin has four principal lexical tones, which are described by their pitch contours and the following numbering scheme: high flat (1), rising (2), low dipping (3) and falling (4) (Chao, 1965). These pitch contours make Mandarin lexical tones especially conducive to visual depiction. Indeed, English speakers can learn to identify Mandarin lexical tones effectively when their pitch contours are illustrated using static visual depictions in combination with orthographic representations of words containing them (Liu et al., 2011). This finding is consistent with dual coding theory (Paivio, 1990), which posits that learning is reinforced when information is encoded simultaneously through the visual and verbal modalities. Dual coding theory maintains that, because the resulting memory traces are multimodal, they are richer and more robust than unimodal traces resulting from either the visual or verbal modality alone. By this logic, the addition of motor information to visual and verbal information should create memory traces that are even more multimodal, increasing their robustness to an even greater extent. Similarly, the theory of embodied cognition posits that information maintains its perceptual qualities, such that every sensory modality contributes to representations (Barsalou, 2008). Thus, the theory of embodied cognition predicts that illustration of pitch contours through hand gesture should enrich English speakers' representations of Mandarin lexical tone, resulting in improved identification and discrimination.

Research examining whether gesture can convey acoustic information has produced mixed results. In the realm of speech perception, rhythmic gestures do not reliably influence native Dutch speakers' interpretation of syllabic stress (Jesse & Mitterer, 2011), nor do they enhance English speakers' discrimination of Japanese vowel length (Hirata & Kelly, 2010). However, English speakers intuitively associate metaphorical gestures signifying height

*Corresponding author. Email: morett@pitt.edu

with musical pitch (Cassidy, 1993; Connell, Cai, & Holler, 2013; Forsythe & Kelly, 1989). These gestures are based on a spatial conceptual metaphor in which the upward direction represents high-frequency pitch and the downward direction represents low-frequency pitch (Casasanto, Phillips, & Boroditsky, 2003). This spatial conceptual metaphor of pitch is present in prelinguistic infants (Dolscheid, Hunnius, Casasanto, & Majid, 2012), indicating that it is language-independent. In contrast, mappings between rhythmic gestures, prosody and vowel length cannot be based on a language-independent temporal conceptual metaphor because prosody and vowel length only exist within a linguistic context; thus, these associations must be learned. These findings suggest that metaphorical gestures illustrating the pitch contours of Mandarin lexical tones should tap into a pre-existing, language-independent spatial conceptual metaphor of pitch, allowing them to facilitate the acquisition of these tones in speakers of atonal languages, such as English.

Previous research examining the role of gesture in L2 acquisition has focused mostly on the impact of iconic gesture on word-meaning association. Iconic gestures depict concrete entities or actions through their form and/or motion (McNeill, 2005); as such, they are ideal for illustrating L2 word meanings. In general, this work has shown that meaning-related iconic gestures enhance the learning of novel L2 words. L2 words accompanied by these gestures are learned more effectively than L2 words presented with meaning-related static images (Tellier, 2008), as well as L2 words presented only as speech (Kelly et al., 2009). Moreover, when L2 words are accompanied by meaning-incongruent iconic gestures, they are learned even less effectively than L2 words presented without gestures (Kelly et al., 2009). However, little research has addressed the impact of meaning-related iconic gestures on the acquisition of similar-sounding L2 words (but see Kelly & Lee, 2012).

To date, no research has directly examined how different type(s) of *information* conveyed via gesture can enhance specific aspects of L2 word learning. As demonstrated by the research discussed above, information conveyed via gesture can be either phonological – pertaining to sound – or semantic – pertaining to meaning. For English-speaking Mandarin learners, the theory of embodied cognition (Barsalou, 2008) predicts that gestures illustrating pitch contours should facilitate lexical tone identification by tapping into the spatial metaphor of pitch. Moreover, this theory predicts that meaning-related iconic gestures should facilitate word-meaning association by activating conceptual representations of referents concurrently with phonological representations of words, resulting in robust multimodal representations of words and their meanings. On the other hand, theories of spoken word recognition (Levelt, 1989; Weber & Broersma, 2013) predict that pitch gestures, rather than semantic gestures, should facilitate word-meaning association. This prediction arises from models and evidence indicating that phonological representations of words, which include lexical tones in Mandarin, are activated prior to conceptual representations of referents (Poss, Hung, & Will, 2008; Van Donselaar, Koster, & Cutler, 2005; Yip, Leung, & Chen, 1998). Thus, these theories predict that pitch gestures should enrich English speakers' phonological representations of Mandarin words differing in tone, facilitating phonological discrimination between them and association with their meanings.

The current study represents the first attempt to directly compare the influence of gestures conveying semantic and phonological information on L2 word learning. In particular, this study compared the impact of metaphorical gestures depicting lexical tone pitch contours (pitch gesture), iconic gestures depicting word referents (semantic gesture) and learning without gestures (no gesture) on English speakers' Mandarin lexical tone and word meaning identification. We predicted that, relative to semantic gestures and no gestures, pitch gestures would enhance both tone differentiation and word-meaning association. The word-meaning association task used in this study tested English speakers' recognition of the meanings of Mandarin words differing only in tone. Given the high phonological demands of this task, we believed that phonological representations of these words would be activated prior to semantic representations, and that tone gestures would facilitate this process, as predicted by theories of spoken word recognition (Levelt, 1989; Weber & Broersma, 2013). Because we did not believe that semantic representations of these words would be activated concurrently with phonological representations, as postulated by the theory of embodied cognition (Barsalou, 2008), we predicted that semantic gestures would not similarly enhance word-meaning association.

Methods

Participants

Fifty-seven undergraduate and graduate students (age: M = 20.33 years, SD = 3.34; 15 males) at a large public university in the Midwestern US participated. All participants were English monolinguals with no knowledge of Mandarin and reported having normal or corrected-to-normal vision and normal hearing.

Materials

Twenty Mandarin words were selected from a larger database (Chen, Chang, Chou, Sung, & Chang, 2011) for use in this research (see Table 1). These words were divided into two sets: twelve words presented in the pre-test and the learning phase (to-be-learned words) and eight words presented only in the post-test (non-learned words).

Table 1. Mandarin words and English translations used in experiment, with lexical tones.

		Pinyin	English	Tone		Pinyin	English	Tone
Learned	揮	huī	to wave	1	回	huí	to return	2
	包	bāo	to pack	1	飽	bǎo	full	3
	抽	chōu	to pump	1	臭	chòu	to stink	4
	降	xiáng	to surrender	2	想	xiǎng	to think	3
	調	tiáo	to shift	2	跳	tiào	to jump	4
	躲	duǒ	to hide	3	剁	duò	to chop	4
Novel	哆	duō	to shiver	1	奪	duó	to snatch	2
	悔	huǐ	to regret	3	賄	huì	to bribe	4
	靴	xuē	boots	1	削	xuè	to trim	4
	別	bié	to separate	2	癟	biě	shrivelled	3

To-be-learned words consisted of six minimal pairs – pairs of words differing only in tone – representing each of the possible combinations of lexical tones. The meanings of these words were conducive to illustration via iconic gesture. Non-learned words consisted of four minimal pairs of words that differed phonologically from to-be-learned words. All Mandarin words used in this experiment were monosyllabic and consisted of phonemes present in English.

Videos shown during the learning phase of the experiment were created by recording a fluent Mandarin–English bilingual repeating each to-be-learned Mandarin word and its English translation twice. Each time the bilingual repeated a Mandarin word, she simultaneously performed one of three actions: a gesture depicting the pitch contour of the word's lexical tone (pitch gesture), a gesture depicting the word's meaning (semantic gesture) or no gesture (see Figure 1).[1]

Audio recordings of another fluent Mandarin–English bilingual saying each Mandarin word presented in the experiment were used in the pre-test and post-test. These recordings featured a speaker of a different sex 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.

Procedure

The experiment consisted of three phases: pre-test, learning and post-test. Participants first completed the pre-test, in which they heard audio recordings of the to-be-learned Mandarin words from the learning phase. Upon hearing each word, participants guessed its lexical tone by pressing one of four buttons. The experimenter explained that each of the buttons was labelled with a symbol depicting the pitch contour of a Mandarin tone. To emphasise differences in tone, words from each minimal pair were presented in consecutive trials.

Next, participants completed the learning phase. First, they were provided with a short tutorial in which each of the four Mandarin lexical tones was described verbally and presented both visually, via a static image depicting its pitch contour (see Figure 2), and aurally, via pronunciation of a vowel present in both English and Mandarin (short *a*). Next, participants learned the to-be-learned Mandarin words. In each learning trial, participants viewed a video clip corresponding to their assigned learning condition (pitch gesture, semantic gesture, no gesture; see Materials section and Figure 1) for a word, and then repeated the word and its English translation aloud while re-enacting the action shown in the video. As in the pre-test, words from each minimal pair were presented in consecutive trials. All learning trials were presented within

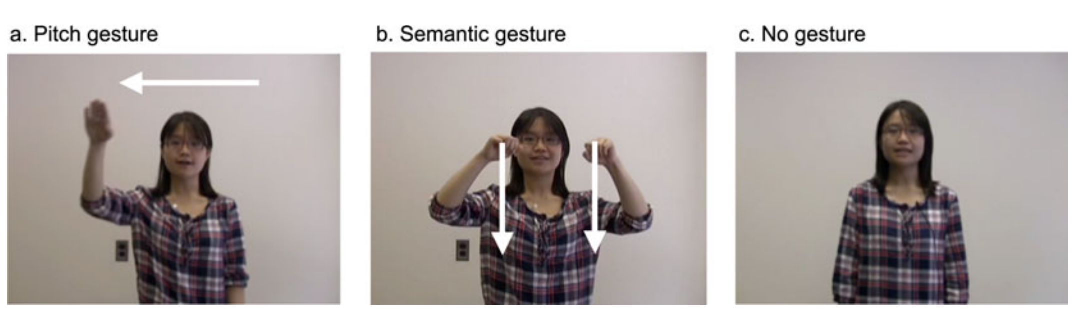

Figure 1. Screenshots of video clips from each of the three learning conditions, with arrows representing hand motion performed concurrently with pronunciation of Mandarin word. Accompanying speech (repeated twice by bilingual): "抽 (chōu) means to pump".

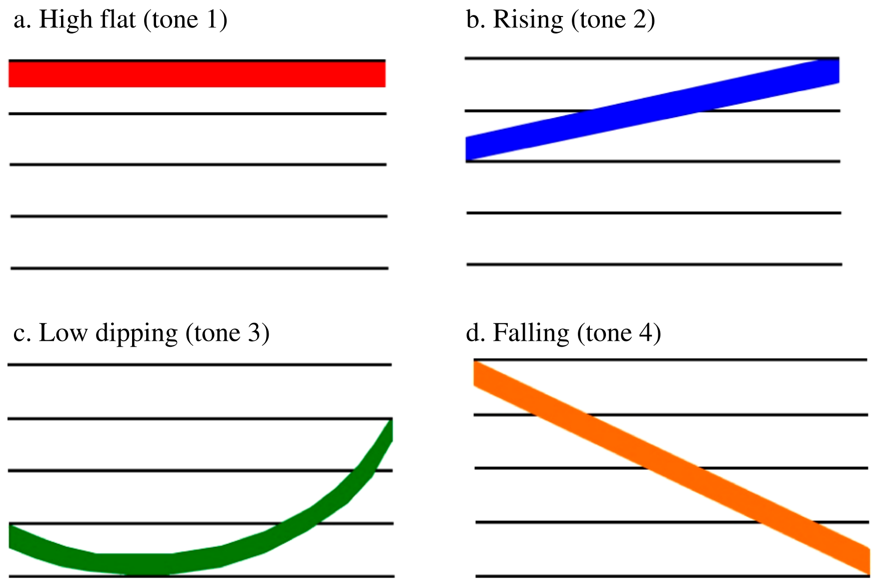

Figure 2. Static pitch contour illustrations used in beginning of learning phase. Response buttons were labelled using similar images without reference bars.

each of three blocks. After the learning phase, participants were given a five-minute break.

Participants then completed the post-test, which was divided into two counterbalanced blocks: tone identification and word-meaning association. The learned section of the tone identification block was structured identically to the pre-test. In the novel section, participants heard each of the non-learned words and pressed the button corresponding to its tone. In the word-meaning association block, participants heard each learned Mandarin word while two English words, one of which was the meaning of the presented word and one of which was the meaning of the other word in that minimal pair, were displayed on either side of the screen. Participants identified the meaning of each word by pressing the button corresponding to its English translation. In both blocks, Mandarin words from each minimal pair were presented in consecutive trials. No feedback was provided during any of the experimental tasks.

Results

There were no group differences in lexical tone identification on the pre-test, F_1 and $F_2 < 1$. Furthermore, tone identification accuracy on the pre-test did not exceed chance (.25) for participants or words assigned to any of the learning conditions: pitch gesture, $t_1 < 1$; $t_2(11) = -4.30$, $p = .001$;[2] semantic gesture, $t_1 < 1$; $t_2(11) = -1.46$, $p = .17$; no gesture, $t_1 < 1$; $t_2(11) = -1.79$, $p = .10$.

For the learned word tone identification task, there was a significant interaction of test and learning condition, $F_1(2,56) = 4.18$, $p = .02$; $\eta_p^2 = .13$; $F_2(2,22) = 5.93$, $p = .009$; $\eta_p^2 = .35$; see Figure 3. Tone identification accuracy increased significantly from pre-test to post-test under the pitch gesture condition, $F_1(1,17) = 18.76$, $p < .001$; $\eta_p^2 = .53$; $F_2(1,11) = 15.62$, $p = .002$; $\eta_p^2 = .59$, and the no gesture condition, $F_1(1,19) = 6.74$, $p = .02$; $\eta_p^2 = .26$; $F_2(1,11) = 5.30$, $p = .04$; $\eta_p^2 = .33$, but not the semantic gesture condition, F_1 and $F_2 < 1$. Performance on the post-test exceeded chance (.25) under the pitch gesture condition, $t_1(17) = 3.88$, $p = .001$; $t_2(11) = 3.38$, $p = .006$, as well as the no gesture condition, $t_1(19) = 2.93$, $p = .009$; $t_2(11) = 1.97$, $p = .07$, but not the semantic gesture condition, $t_1(19) = 1.45$, $p = .17$; $t_2 < 1$.

For the novel word tone identification task, per cent accuracy on the pre-test for the 12 to-be-learned words

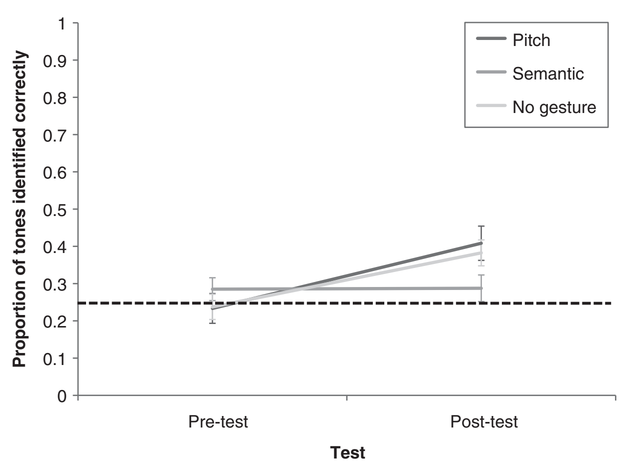

Figure 3. Lexical tone identification accuracy in to-be-learned words at pre-test and post-test for participants assigned to the pitch gesture, semantic gesture and no gesture conditions. (Dotted line represents chance; error bars represent standard error.)

was compared to per cent accuracy on the post-test for the 8 non-learned words, given that both sets of words were novel to participants at the time of testing. Across learning conditions, lexical tone identification accuracy increased significantly from pre-test to post-test, $F(1,56) = 12.15$, $p = .001$; $\eta_p^2 = .18$; see Figure 4. However, the test by condition interaction failed to reach significance, $F < 1$. Performance on the post-test exceeded chance (.25) under all three learning conditions: pitch gesture, $t(17) = 2.20$, $p = .04$; semantic gesture, $t(19) = 3.11$, $p = .001$ and no gesture, $t(20) = 2.95$, $p = .004$.

For the word-meaning association task, accuracy varied as a function of learning condition, $F_1(2,57) = 5.75$, $p = .005$, $\eta_p^2 = .17$; $F_2(2,16) = 14.15$, $p < .001$, $\eta_p^2 = .64$; see Figure 5. Word-meaning association accuracy under the pitch gesture condition exceeded accuracy under the semantic gesture ($p_1 = .02$; $p_2 = .002$) and no gesture ($p_1 = .01$; $p_2 = .02$) conditions. No differences were found between the semantic and no gesture conditions ($p_1 > .05$; $p_2 > .05$). Performance was significantly above chance (.50) under the pitch gesture condition, $t_1(17) = 3.22$, $p = .005$; $t_2(11) = 3.85$, $p = .003$. However, performance failed to exceed chance under the no gesture condition, $t_1 < 1$; $t_2(11) = -1.61$, $p = .14$, and the semantic gesture condition, $t_1 < 1$; $t_2(11) = -5.35$, $p < .001$.[3]

Discussion

The present study examined whether metaphorical gestures conveying pitch or iconic gestures conveying meaning could enhance English speakers' spoken word recognition and lexical tone discrimination in Mandarin. The results demonstrated that pitch gestures enhanced differentiation

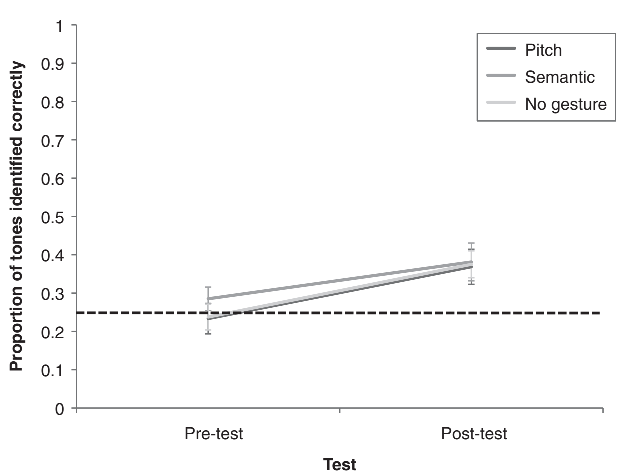

Figure 4. Lexical tone identification accuracy in novel words at pre-test and post-test for participants assigned to the pitch gesture, semantic gesture and no gesture conditions. Twelve to-be-learned words tested in pre-test; eight non-learned words tested in post-test. (Dotted line represents chance; error bars represent standard error.)

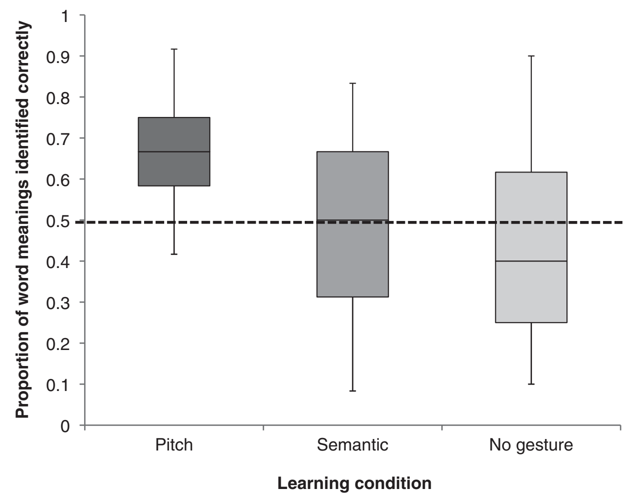

Figure 5. Performance on the word-meaning association task for participants assigned to the pitch gesture, semantic gesture and no gesture conditions. (Dotted line represents chance.)

between the meanings of Mandarin words varying only in tone, whereas semantic gestures hindered tone identification in learned Mandarin words. These findings indicate that information conveyed via gesture can affect the acquisition of L2 lexical phonology and semantics positively or negatively, and that its impact depends on the level of processing accessed at retrieval. Notably, this is the first study to demonstrate that gestures illustrating the acoustic attributes of non-native phonemes can enhance word learning.

The results of the current study reveal important information about the cognitive mechanisms underlying phonological and semantic discrimination between similar-sounding L2 words. Unlike stress and vowel length, Mandarin lexical tones are governed by a language-independent conceptual metaphor, according to which high pitch is associated with the upward direction and low pitch is associated with the downward direction (Casasanto et al., 2003). The finding that pitch gestures result in improved discrimination between the meanings of words differing in tone provides evidence that English speakers can map visuospatial information conveyed via metaphorical gestures consistent with this vertical representation of pitch onto their representations of lexical tone. These phonological representations of words are then associated with semantic representations of referents, improving discrimination between word meanings on the basis of lexical tone. This finding is consistent with theories of spoken word recognition (Levelt, 1989; Weber & Broersma, 2013) because it indicates that phonological lexical representations are accessed prior to semantic lexical representations. However, pitch gestures resulted in performance on the learned word tone identification task equivalent to performance resulting from the no gesture condition. Additionally, pitch gestures resulted in performance on the novel word tone identification task equivalent to performance resulting from both the semantic gesture and no gesture conditions. These findings are inconsistent with the theories of L2 spoken word recognition (Weber & Broersma, 2013) and embodied cognition (Barsalou, 2008), both of which maintain that visuospatial illustrations of pitch contours conveyed via gesture should enhance lexical tone identification. Pitch gesture's failure to enhance lexical tone identification relative to no gesture is consistent with previous research showing that gestures depicting vowel length do not enhance its perception relative to no gesture (Hirata & Kelly, 2010). Considered in conjunction with pitch gesture's facilitation of word-meaning association, this finding indicates that gestures depicting phonological contrasts may be most effective when the relationship between these contrasts and word meanings is clear.

In the learned word tone identification task, the semantic gesture condition hindered performance relative to the pitch gesture and no gesture conditions. This result is partially consistent with a similar finding from Kelly and Lee (2012). However, in that study, semantic gesture detracted from participants' ability to distinguish between the meanings of words differing in geminate consonants, rather than their ability to identify geminate consonants. The finding that semantic gesture impaired performance in the learned word tone identification task suggests that semantic gestures may have resulted in vivid representations of learned word meanings, which may have interfered with the phonological processing of learned words at retrieval. No such impairment in performance under the semantic gesture condition was found in the novel word tone identification task because these representations did not pertain to novel words; thus, they could not interfere with tone identification in novel words.

In the word-meaning association task, although semantic gesture failed to enhance performance, it did not impair performance relative to the no gesture condition, as in the learned word tone identification task. This pattern of results may be due to semantic gesture strengthening associations between words and their meanings subsequent to lexical tone processing. This explanation is partially consistent with the theory of embodied cognition (Barsalou, 2008), which posits that representative iconic gestures should enrich semantic lexical representations, enhancing word meaning recall. However, the similar word-meaning association performance resulting from the semantic gesture and no gesture conditions is inconsistent with prior research indicating that semantic gesture hinders L2 learners' discrimination between the meanings of words differing in consonant length (Kelly & Lee, 2012). It is worth noting that word-meaning association under the no gesture condition in the current study was less accurate than it was under the no gesture condition in this previous study. This discrepancy suggests that English speakers may be better able to tune into the acoustic differences between geminate consonants in the absence of semantic gesture, whereas semantic gesture does not interfere with semantic discrimination based on lexical tone.

The results of the present study demonstrate that, relative to semantic gestures and no gestures, pitch gestures facilitate English speakers' association of Mandarin words differing in lexical tone with their meanings, but fail to enhance their lexical tone identification. Moreover, they demonstrate that semantic gestures hinder English speakers' identification of lexical tones in learned Mandarin words. As such, these findings indicate that metaphorical gestures depicting acoustic attributes of unfamiliar phonological contrasts can enhance their perception, but only when the mappings between them and the visuospatial attributes of gestures are transparent, as well as when the correspondence between them and word meanings is clear.

Funding

This research was supported by a Federico and Rena Perlino Research Award and a Chancellor's Fellowship from the University of California, Santa Cruz to L.M.M. and a Fulbright Foreign Scholarship from the Taiwanese Ministry of Foreign Affairs to L.-Y.C. Portions of the results were presented at the twenty-first Second Language Research Forum (2012, October). We thank Kai Hwang for helping with the creation of audio test stimuli. Additionally, we thank Brian C. Becker and Adeetee Bhide, as well as our anonymous reviewers, for helpful feedback on this manuscript.

Supplemental data

Supplemental data for this article can be accessed here.

Notes

1. To assess whether the pitch contours (i.e. fundamental frequency, *f*0) of Mandarin words presented in learning videos differed between conditions, the first repetition of each word was divided into quartiles of duration, at which *f*0 was sampled. Analyses of *f*0 values failed to reveal any significant effect of condition either alone or in combination with tone or quartile (all Fs < 1).
2. This difference is negative, indicating that lexical tone identification accuracy for words assigned to the pitch gesture condition in the pre-test was significantly *below* chance.
3. This difference is negative, indicating that word-meaning association accuracy for words assigned to the semantic gesture condition was significantly *below* chance.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bluhme, H., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics*, *22*, 51–57.

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

Cassidy, J. W. (1993). Effects of various sightsinging strategies on nonmusic majors' pitch accuracy. *Journal of Research in Music Education*, *41*, 293–302. doi:10.2307/3345505

Chao, Y. R. (1965). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, H. C., Chang, L. Y., Chou, Y. S., Sung, Y. T., & Chang, K. E. (2011). Construction of a Chinese orthographic database for Chinese character instruction. *Bulletin of Educational Psychology*, *43*, 269–290.

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, *81*, 124–130. doi:10.1016/j.bandc.2012.09.005

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2012). The sound of thickness: Prelinguistic infants' associations of space and pitch. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 306–311). Austin, TX: Cognitive Science Society.

Forsythe, J. L., & Kelly, M. M. (1989). Effects of visual-spatial added cues on fourth graders' melodic discrimination. *Journal of Research in Music Education*, *37*, 272–277. doi:10.2307/3344661

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*, 298–310. doi:10.1044/1092-4388(2009/08-0243)

Jesse, A., & Mitterer, H. (2011). Pointing gestures do not influence the perception of lexical stress. In P. Cosi & R. De Mori (Eds.), *Proceedings of Interspeech 2011* (pp. 2445–2448). Baixas: International Speech Communication Association.

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, *27*, 793–807. doi:10.1080/01690965.2011.581125

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334. doi:10.1080/01690960802365567

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*, 63–67. doi:10.1159/000259274

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, *61*, 1119–1141. doi:10.1111/j.1467-9922.2011.00673.x

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

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

Poss, N., Hung, T. H., & Will, U. (2008). The effects of tonal information on lexical activation in Mandarin. In M. K. M. Chan & H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics* (pp. 205–211). Columbus: Ohio State University.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers' Association*, *24*, 27–47.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*, 219–235. doi:10.1075/gest.8.2.06tel

Van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*, 251–273. doi:10.1080/02724980343000927

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658. doi:10.1121/1.428217

Weber, A., & Broersma, M. (2013). Spoken word recognition in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5368–5375). Bognor Regis: Wiley-Blackwell.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585. doi:10.1017/S0142716407070312

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*, 2240–2253. doi:10.1121/1.408684

Yip, M. C. W., Leung, P. Y., & Chen, H. C. (1998). Phonological similarity effects in Cantonese spoken-word processing. In R. H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of ICSLP '98* (pp. 2139–2142). Sydney: Australian Speech Science and Technology Association.