Task-Based and Individual Differences Influence the Effect of Gesture Observation on Novel L2 Speech Sound Learning

Laura M. Morett[1] (lmorett@ua.edu)
[1]Department of Educational Studies, University of Alabama, Tuscaloosa, AL 35401

Abstract

This study sought to replicate the effect of observing pitch gesture and clarify the effect of observing representational gesture on L2 lexical tone learning and to explore the influences of individual differences in lexical and non-lexical tone perception on these effects. The results revealed that observing representational gestures facilitates lexical tone discrimination, albeit to a lesser extent than observing pitch gestures, suggesting that task difficulty may influence its effect. Moreover, they revealed that individual differences in non-speech tone perception predict discrimination of lexical tones learned by observing pitch gesture and no gesture, but not representational gesture. Together, these findings suggest that task difficulty as well as individual differences in sensitivity to non-speech sounds influence the effects of observing gesture on novel L2 speech sound learning.

Keywords: gesture; speech sound learning; L2 acquisition

Introduction

Acquisition of novel speech sounds is a challenging aspect of learning a second language (L2), particularly for adults. Thus, lexical tone, a speech sound present in many world languages that consists of pitches differentiating between word meanings and grammatical inflections (Gussenhoven, 2004; Maddieson, 2013; Yip, 2002), is challenging for native speakers of atonal languages, such as English, to acquire in an L2 (Pelzl, 2019). Mandarin, the most prevalent tonal language, has four lexical tones with distinct pitch contours: Tone 1 (high-flat); Tone 2 (rising); Tone 3 (low or low-dipping); and Tone 4 (falling; Chao, 1965; Ho, 1976; Howie, 1974). Therefore, successful acquisition of Mandarin lexical tones entails recognizing their pitch contours, which is essential to differentiating between words differing minimally in lexical tone (Wong & Perrachione, 2007).

Despite the challenges associated with L2 Mandarin lexical tone acquisition for native English speakers, brief, focused auditory training can facilitate it (Wang et al., 1999, 2003; Wong & Perrachione, 2007). Moreover, the addition of visual depictions of pitch contours can further facilitate L2 Mandarin lexical tone acquisition by native English speakers (Bluhme & Burr, 1971; Godfroid et al., 2017; Liu et al., 2011), suggesting that they may result in multimodal representations of lexical tone, as hypothesized by dual coding theory (Paivio, 1990). Furthermore, observing pitch gestures, which convey pitch contours haptically as well as visually, also facilitates L2 Mandarin lexical tone acquisition by native English speakers (Baills et al., 2019; Hannah et al., 2017; Morett et al., 2022; Morett & Chang, 2015; Zhen et al., 2019). Notably, the effects of observing pitch gestures on L2 lexical tone acquisition are more robust than the effects of observing gestures conveying other phonological contrasts on their acquisition in other L2s (Hirata et al., 2014; Hoetjes & Van Maastricht, 2020; Kelly et al., 2014; Xi et al., 2020). This difference may be due to pitch gesture's basis in the vertical conceptual metaphor of pitch, according to which high pitch is associated with the upward direction and low pitch is associated with the downward direction (Casasanto & Boroditsky, 2003; Connell et al., 2013; Morett et al., 2022). Thus, observing pitch gesture may elicit implicit mental simulation of lexical tone, as hypothesized by theories of embodied cognition (Shapiro, 2019).

Further evidence that pitch gesture's effect on L2 lexical tone learning is based on the vertical conceptual metaphor of pitch comes from work demonstrating that pitch gestures congruent with the lexical tones of Mandarin words facilitate L2 lexical tone acquisition, whereas pitch gestures incongruent with their lexical tones hinder it (Morett et al., 2022). These findings parallel findings demonstrating that representational gestures congruent with the meanings of L2 words enhance memory for these words, whereas representational gestures incongruent with their meanings decrease memory for them (Garcia-Gamez & Macizo, 2019; Kelly et al., 2009). In both cases, the effect of congruency on learning is due to iconicity, which conveys meaning via visual resemblance to referents, whether they are lexical tones or word meanings (Perniss et al., 2010).

In contrast to the effects of observing pitch gestures, the effects of observing representational gestures on novel L2 speech sound acquisition are less clear. Some evidence indicates that observing representational gestures conveying the meanings of Mandarin words differing minimally in lexical tone at learning hinders subsequent lexical tone identification in these words (Morett & Chang, 2015). Other evidence indicates that observing representational gestures conveying the meanings of Japanese words differing minimally in vowel length at learning neither hinders nor facilitates subsequent vowel length identification in these words (Kelly & Lee, 2012). Moreover, some evidence indicates that observing representational gestures conveying the meanings of Mandarin words differing minimally in lexical tone facilitates association of these words with their meanings (Baills et al., 2019; Morett & Chang, 2015), whereas other evidence indicates that it hinders association of Japanese words differing minimally in vowel length with their meanings (Kelly & Lee, 2012). These mixed findings concerning the effects of representational gestures conveying word meanings on differentiation between L2 words differing in an unfamiliar speech sound contrast with overwhelming evidence that these representational gestures enhance memory for phonologically dissimilar L2 words (Allen, 1995; Garcia-Gamez & Macizo, 2019; Kelly et al., 2009; Macedonia et al., 2011; Porter, 2016; Tellier, 2008), leading some to propose that they facilitate the learning of L2 words that are phonologically dissimilar, but not phonologically similar, to one another (Kelly & Lee, 2012).

One factor providing a possible explanation for the mixed results concerning the effects of observing representational gesture on differentiation between unfamiliar L2 speech sounds in words is task difficulty. In Morett and Chang (2015), four lexical tones were learned, whereas in Kelly and Lee (2012), only two vowel lengths were learned, suggesting that observing representational gesture may not hinder differentiation between unfamiliar L2 speech sounds when it is less challenging, whereas it may hinder it when it is more challenging. If this is true, the difficulty of the task used to assess learning may also affect the impact of observing representational gesture on unfamiliar L2 speech sound differentiation. To date, all studies examining the effects of gesture observation on unfamiliar L2 speech sound differentiation have used classification tasks to assess learning, which tend to be quite challenging. A less challenging alternative that could be used for the same purpose is a same-different task, in which a target word containing a learned L2 speech sound is compared with a prime word containing either the same or a different sound. Examining the effects of observing representational and pitch gesture on unfamiliar L2 speech sound learning using a same-different task and comparing it to prior results of classification tasks would provide further insight into the extent to which task difficulty modulates these effects.

Another factor that may affect the impact of gesture on L2 lexical tone learning is individual differences in lexical and non-speech tone perception prior to learning. Such differences can be taken into account by assessing non-speech tone perception via a standardized measure administered before the main experimental task (Morett et al., 2022) and lexical tone perception via a pre-test administered prior to training (Morett et al., 2022; Morett & Chang, 2015; Zhen et al., 2019). In previous work using a pre-test, lexical tone perception has been assessed in the same way, typically using the same stimuli, as in the post-test. This may have resulted in practice effects, which may also be influenced by individual differences in tone perception or more general cognitive abilities such as working memory. In addition, pre-tests used in previous work have assessed lexical tone perception solely in the auditory modality, which may not account for individual differences in accuracy of association of visual depictions of tone contours with lexical tones.

The goals of the current study were twofold: (1) To replicate the effect of observing pitch gesture and further probe the effect of observing representational gesture on acquisition of L2 lexical tone by atonal language speakers; (2) To explore the effects of individual differences in lexical and non-speech tone perception prior to learning on these effects of gesture on L2 lexical tone acquisition. Based on the findings discussed above, we predicted that observing both pitch and representational gesture would facilitate subsequent discrimination of lexical tones and that individual differences in lexical and non-speech tone perception prior to lexical tone learning would contribute significantly to these effects.

Methods

Participants

42 adult native English speakers without knowledge of any tonal languages (age range: 18-32 years; 29 females, 13 males) participated on a volunteer basis or in return for partial course credit. All participants had normal hearing and normal or corrected-to-normal vision. Additionally, participants had no documented speech, language, or learning disorders.

Materials

Six pairs of monosyllabic Mandarin words differing minimally in lexical tone from Morett and Chang (2015) were used in this experiment (see Table 1). Each possible combination of lexical tones was represented in pairs, and words comprising each pair had meanings that could be conveyed transparently via representational gesture.

Videos for use during learning were created by recording a female native Mandarin speaker fluent in English in a headshot configuration saying each Mandarin word and its English translation twice. While saying each Mandarin word, the speaker either produced a pitch gesture conveying the pitch contour of the word's lexical tone, produced a representational gesture conveying the word's meaning, or kept her hands still (see Figure 1).

Videos for use during the pre-test were created by recording another female native Mandarin speaker producing pitch gestures. In these videos, a torso configuration was used to ensure that performance was not influenced by facial movements, and audio tracks were removed.

Audio recordings used in pre- and post-tests were created by recording a male native Mandarin speaker saying each word. A speaker of a different sex than the speakers featured in videos was featured in audio recordings to ensure that participants could generalize lexical tone across speakers.

Table 1: Pairs of Mandarin words differing minimally in lexical tone with English translations.

Word 1		Word 2	
Pinyin	English	Pinyin	English
hui1	to wave	hui2	to return
bao1	to pack	bao3	full
chou1	to pump	chou4	to stink
xiang2	to surrender	xiang3	to think
tiao2	to shift	tiao4	to jump
duo3	to hide	duo4	to chop

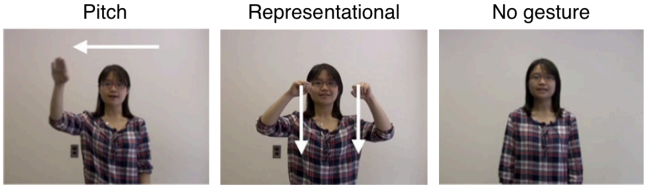

Figure 1: Screenshots of videos from each of the three learning conditions (arrows represent hand motion).

Procedures

Before completing the experimental task, participants provided informed consent and completed a demographic questionnaire and a brief non-speech tone perception test (Mehr et al., 2017). This test, which consists of pairs of pure tones differing in pitch for which the direction of the pitch change is identified (up vs. down), has been validated against the Montreal Battery of Amusia, a more extensive test of non-speech tone perception (Peretz et al., 2008).

The experimental task consisted of three phases: pre-test, learning, and post-test. In the pre-test, primes consisted of silent pitch gesture videos, and targets consisted of audio recordings of Mandarin words. Participants responded by pressing one of two buttons (counterbalanced across participants) to indicate whether the lexical tones of Mandarin word targets matched (k = 72) or mismatched (k = 72) pitch gesture primes. In the learning phase, participants learned Mandarin words by watching videos in which they were presented in pairs (order counterbalanced across participants) and were accompanied by either pitch gesture (n = 10), representational gesture (n = 15), or no gesture (n = 17). Participants were instructed to learn the meanings of words as they would be subsequently tested on them, and no mention of the tonal properties of words was made. All 12 words were presented randomized in order in 3 blocks, such that each word was presented 3 times and a total of 36 trials were presented in the learning phase. In the post-test, a prime and a different target word were selected from among the set of learned Mandarin words. Prime and target words had either the same (k = 72) or different (k = 72) lexical tones, and participants pressed one of two buttons (counterbalanced across participants) to indicate whether their lexical tones were the same or different.

Results

Effects of gesture observation on lexical tone discrimination accuracy

Signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2004) was used to decompose responses on lexical tone discrimination tasks into two conceptually and statistically distinct parameters: *Discrimination* or *sensitivity* (*d'*), which captures how well participants successfully discriminated prime-target pairs differing in lexical tone from prime-target pairs with the same lexical tone (*d'*), and *response criterion* (*c*) or *response bias,* which captures the criterial level at which participants judged lexical tones to be different, regardless of whether they actually differed.

To determine whether lexical tone discrimination accuracy differed by test and learning condition, response data were analyzed using mixed effects probability unit (*probit*) models, which operate on trial-level data and account for participant- and item-level variability within the same model. Probit mixed effect models allow responses (1 = same; 0 = different) rather than *d'* values to be used as the dependent variable (DV), with measures of sensitivity expressed as *d'* values. In these models, congruency of lexical tone between prime and target words (same vs. different), test (pre-test vs. post-test), and learning condition (no gesture vs. pitch gesture vs. representational gesture) were included as fixed effects using weighted mean centered (Helmert) contrast coding. The intercept represents overall response bias (*c*), and the main effect of congruency represents overall discrimination performance (*d'*), with an alpha level < .05 indicating that overall response bias and/or discrimination performance exceeds chance. The main effect of learning condition represents its effect on response bias (*c*), and the interaction of learning condition with congruency represents the effect of learning condition on discrimination accuracy (*d'*), with an alpha level < .05 indicating that the effect of learning condition on response bias and discrimination performance exceeded chance. Random slopes were included with the maximal random effect structure permitted to achieve model convergence (Barr et al., 2013).

The main model (k = 10,790) revealed that response bias did not differ significantly by test (B=0.06, SE=0.06, z=0.95, p=.34), learning condition (B=-0.12, SE=0.08, z=-1.48, p=.14), or the interaction between them (B=-0.11, SE=0.12, z=-0.89, p=.37). By contrast, it revealed that discrimination accuracy differed significantly by the interaction between test and learning condition (B=0.83, SE=0.13, z=6.16, p<.001; see Figure 2). One sample t-tests revealed that accuracy was significantly below chance in groups assigned to all three learning conditions at pre-test (t_P=-7.55, p_P<.001; t_R=-4.15, p_R<.001; t_N=-7.28, p_N<.001), whereas it significantly exceeded chance in groups assigned to all three learning conditions at post-test (t_P=5.15, p_P<.001; t_R=3.82, p_R=.002; t_N=3.66, p_N=.002). Simple effect analyses by learning condition (k = 3,109 – 4,225) revealed that discrimination

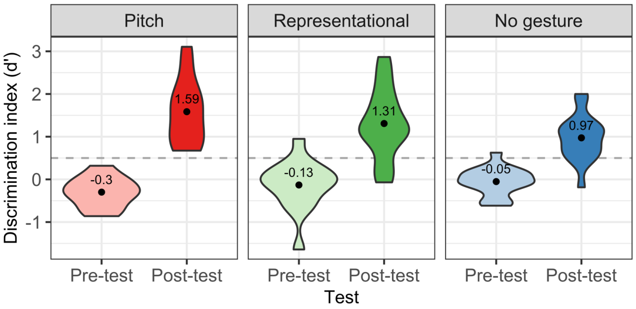

Figure 2: Violin plot of lexical tone discrimination accuracy by learning condition and test (dots and values represent cell means; dashed lines represent chance).

accuracy increased significantly from pre-test than to post-test across learning conditions (B_P=1.87, SE_P=0.10, z_P=19.12, p_P<.001; B_R=1.34, SE_R=0.09, z_R=14.68, p_R<.001; B_N=1.05, SE_N=0.08, z_N=13.16, p_N<.001). Simple effect analyses by test (*k* = 4,806 – 5,984) revealed that, at pre-test, discrimination accuracy was significantly *lower* in the groups assigned to the pitch and representational gesture conditions than the group assigned to the no gesture condition (*B*=-0.25, *SE*=0.09, *z*=-2.77, *p*=.006) and marginally *lower* in the group assigned to the pitch gesture condition than the groups assigned to the no gesture and representational gesture conditions (*B*=-0.15, *SE*=0.08, *z*=- 1.87, *p*=.06). At post-test, by contrast, discrimination accuracy was significantly *higher* in the groups assigned to the pitch and representational gesture conditions than the group assigned to the no gesture condition (*B*=0.31, *SE*=0.09, *z*=3.61, *p*<.001) and significantly *higher* in the group assigned to the pitch gesture condition than the groups assigned to the no gesture and representational gesture conditions (*B*=0.58, *SE*=0.09, *z*=6.53, *p*<.001). The pre-test to post-test increase in discrimination accuracy was greatest in the pitch gesture condition (1.29), followed by the representational gesture condition (1.18), followed by the no gesture condition (0.92).

Effects of individual differences in speech and non-speech tone perception on gesture's impact on lexical tone discrimination accuracy

To examine the relationship between audiovisual lexical tone discrimination (pre-test) and auditory lexical tone discrimination (post-test) and determine whether it differed between groups assigned to each learning condition, a probit mixed effect model with prime-target congruency (same vs. different), pre-test response (same vs. different), and learning condition (no gesture vs. pitch gesture vs. representational gesture) as fixed effects using weighted mean centered (Helmert) contrast coding was implemented using post-test response as the dependent variable. This model (*k* = 10,790) revealed that neither response bias nor discrimination accuracy differed significantly by pre-test response or its interaction with learning condition.

To examine the relationships between non-speech tone perception and audiovisual lexical tone discrimination (pre-test) as well as auditory lexical tone discrimination (post-test) in individual participants, *d'* for the pre-test and post-test was computed on a per-participant basis. Pearson product-moment correlations were then computed between scores on the non-speech tone perception test and *d'* for both the pre-test and post-test. These correlations revealed that, prior to lexical tone learning, non-speech tone perception was not significantly correlated with audiovisual lexical tone discrimination (*r*(39) = -.03; *t* = -0.20; *p* = .84). By contrast, non-speech tone perception prior to lexical tone learning was significantly positively correlated with auditory lexical tone discrimination following lexical tone learning (*r*(39) = .46; *t* = 3.22; *p* = .003). Further examination revealed that non-speech tone perception and auditory tone discrimination were significantly positively correlated in participants assigned to the pitch gesture learning condition (*r*(8) = .64; *t* = 2.34; *p* = .047) and the no gesture learning condition (*r*(14) = .50; *t* = 2.16; *p* = .048), but not the representational gesture learning condition (*r*(13) = .33; *t* = 1.28; *p* = .22; see Figure 3).

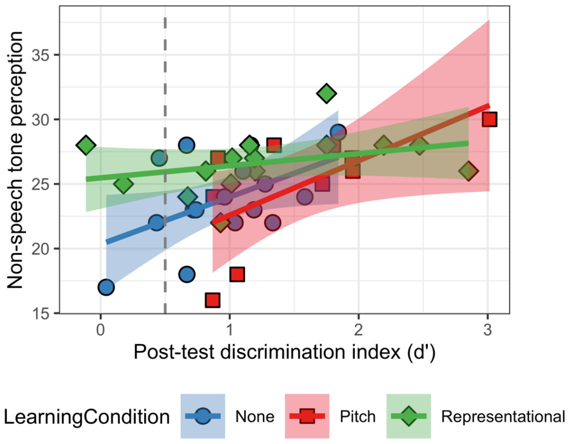

Figure 3: Scatter plot of non-speech tone perception test score by post-test discrimination accuracy (dashed line represents chance) by learning condition.

Discussion

The primary goal of the current study was to examine the extent to which the effects of observing pitch and representational gesture on lexical tone acquisition by atonal language speakers generalize to a lexical tone discrimination paradigm. The results indicate that observing pitch gestures enhanced lexical tone learning to a greater extent than no gesture, consistent with previous work using a lexical tone identification paradigm (Baills et al., 2019; Hannah et al., 2017; Morett et al., 2022; Morett & Chang, 2015; Zhen et al., 2019). However, the results also indicate that observing representational gestures conveying word referents enhanced, rather than hindered, lexical tone discrimination. Although this finding is inconsistent with previous evidence that observing these gestures hinders lexical tone identification (Morett & Chang, 2015), it is consistent with previous evidence that it did not hinder Japanese vowel length differentiation, although it did not facilitate it, either (Kelly & Lee, 2012). Thus, taken together, these results suggest that the effect of observing representational gestures conveying word meanings on the learning of novel L2 speech sounds such as lexical tones may vary based on task difficulty, such that it is detrimental or neutral in more challenging tasks such as identification paradigms and facilitatory in easier tasks such as discrimination paradigms. By contrast, observing pitch gestures conveying lexical tones may facilitate their learning regardless of task difficulty. Given that it wasn't possible to manipulate task difficulty in the current study, it will be important to manipulate it in future research to confirm this explanation.

A secondary goal of the current study was to examine the extent to which individual differences in lexical and non-speech tone perception affect the impact of observing gestures on L2 lexical tone learning. The results suggest that individual differences in non-speech tone perception prior to lexical tone learning contributed to post-test discrimination of lexical tones learned by observing pitch gesture and no gesture, but not representational gesture. This finding differs from previous work (Morett et al., 2022), which found no such effect of these individual differences on identification of lexical tones learned by observing pitch gesture or no gesture. One possible reason for this discrepancy may be the similarity in the difficulty of the tone discrimination task used in the post-test of the current study, which entailed same-different tone judgments for different words, and the non-speech tone perception test, which entailed identification of the difference between two pure tones as either upwards or downwards. In both the post-test and non-speech tone perception test, the stimuli were auditory, which could explain why individual differences in non-speech tone perception prior to lexical tone learning failed to contribute to pre-test lexical tone discrimination, which was tested using audiovisual stimuli. This difference in modality between the pre-test and post-test stimuli also provides a possible explanation for why pre-test responses failed to predict post-test discrimination accuracy. The significant effects of observing pitch and representational gestures on post-test lexical tone discrimination in this and other studies suggest that audiovisual depictions of lexical tone can affect how it is processed in the auditory modality, however. With that said, the results of this analysis should be interpreted with caution due to the limited sample size of the current study, and replication will be necessary for confidence in this finding.

In conclusion, the results of the current study indicate that observing pitch gestures enhances lexical tone learning regardless of how it is assessed, whereas the effect of observing representational gestures on lexical tone learning may depend on task difficulty, including the difficulty of the way in which it is assessed. More specifically, observing representational gestures may facilitate perception of lexical tones—and novel L2 speech sounds more generally—when learning and assessment tasks are easier, whereas they may hinder perception of these sounds when learning and assessment tasks are more difficult. Furthermore, the results suggest that individual differences in non-speech tone perception prior to learning may predict how well lexical tones can be learned by observing pitch and no gesture, but not how well they can be learned by observing representational gesture. Together, these findings suggest that task difficulty as well as individual differences in sensitivity to non-speech sounds influence the effects of observing gesture on novel L2 speech sound learning.

References

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, *79*(4), 521–529. https://doi.org/10.1111/j.1540-4781.1995.tb05454.x

Baills, F., Suárez-González, N., González-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*, 33–58. https://doi.org/10.1017/S0272263118000074

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Casasanto, D., & Boroditsky, L. (2003). Do we think about time in terms of space? *Proceedings of the Annual Meeting of the Cognitive Science Society*.

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

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. https://doi.org/10.1016/j.bandc.2012.09.005

Garcia-Gamez, A. B., & Macizo, P. (2019). Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics*, *40*(2), 473–507. https://doi.org/10.1017/S0142716418000656

Godfroid, A., Lin, C.-H., & Ryu, C. (2017). Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training. *Language Learning*, *67*(4), 819–857. https://doi.org/10.1111/lang.12246

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J., & Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers. *Frontiers in Psychology*, *8*, 2051. https://doi.org/10.3389/fpsyg.2017.02051

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. https://doi.org/10.1044/2014_JSLHR-S-14-0049

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, *33*(5), 353–367.

Hoetjes, M., & Van Maastricht, L. (2020). Using gesture to facilitate L2 phoneme acquisition: The importance of gesture and phoneme complexity. *Frontiers in Psychology*, *11*.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30*(3), 129–148.

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

Frontiers in Psychology, *5*. https://doi.org/10.3389/fpsyg.2014.00673

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. https://doi.org/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*(2), 313–334. https://doi.org/10.1080/01690960802365567

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. https://doi.org/10.1111/j.1467-9922.2011.00673.x

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998. https://doi.org/10.1002/hbm.21084

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Lawrence Erlbaum Associates.

Maddieson, I. (2013). Tone. *The World Atlas of Language Structures Online*.

Mehr, S. A., Kotler, J., Howard, R. M., Haig, D., & Krasnow, M. M. (2017). Genomic imprinting is implicated in the psychology of music. *Psychological Science*, *28*(10), 1455–1467. https://doi.org/10.1177/0956797617711456

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. https://doi.org/10.1080/23273798.2014.923105

Morett, L. M., Feiler, J. B., & Getz, L. M. (2022). Elucidating the influences of embodiment and conceptual metaphor on lexical and non-speech tone learning. *Cognition*, *222*, 105014. https://doi.org/10.1016/j.cognition.2022.105014

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

Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a Second Language*, *54*(1), 51–78. https://doi.org/10.1075/csl.18009.pel

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., & Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception*, *25*(4), 331–343. https://doi.org/10.1525/mp.2008.25.4.331

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227. https://doi.org/10.3389/fpsyg.2010.00227

Porter, A. (2016). A helping hand with language learning: Teaching French vocabulary with gesture. *The Language Learning Journal*, *44*(2), 236–256. https://doi.org/10.1080/09571736.2012.750681

Shapiro, L. (2019). *Embodied cognition*. Routledge.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043. https://doi.org/10.1121/1.1531176

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. https://doi.org/10.1121/1.428217

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

Xi, X., Li, P., Baills, F., & Prieto, P. (2020). Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research*, *63*(11), 1–15. https://doi.org/10.1044/2020_JSLHR-20-00084

Yip, M. (2002). *Tone*. Cambridge University Press.

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. https://doi.org/10.1016/j.cognition.2019.03.004