Title
Observing Gestures During L2 Word Learning Facilitates Differentiation Between Unfamiliar Speech Sounds and Word Meanings

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Author
Morett, Laura M.

Publication Date
2023

Observing Gestures During L2 Word Learning Facilitates Differentiation Between Unfamiliar Speech Sounds and Word Meanings

Laura M. Morett[1] (lmorett@ua.edu)
Department of Educational Studies, University of Alabama, Tuscaloosa, AL 35487

Abstract

This study investigated how observing pitch gestures conveying lexical tones and representational gestures conveying word meanings when learning L2 Mandarin words differing in lexical tone affects their subsequent semantic and phonological processing in L1 English speakers using the N400 event-related potential (ERP). Larger N400s for English target words mismatching vs. matching Mandarin prime words in meaning were observed for words learned with pitch and representational gesture, but not no gesture. Additionally, larger N400s for Mandarin target words mismatching vs. matching Mandarin prime words in lexical tone were observed for words learned with pitch gesture, but not representational or no gesture. These findings provide the first ERP evidence that observing gestures conveying phonological and semantic information during L2 word learning enhances subsequent phonological and semantic processing of learned L2 words.

Keywords: gesture; L2 word learning; N400

Introduction

Lexical tone refers to the use of pitch in language to differentiate between word meanings or inflections (Yip, 2002). Lexical tone differs from intonation, which refers to the use of pitch in language to convey emotion and emphasis (Cruttenden, 1997). When learning a tonal second language (L2), such as Mandarin, speakers of atonal first languages (L1s), such as English, must learn to differentiate between lexical tones, which can be quite challenging (Pelzl, 2019). In Mandarin, there are four lexical tones, each with a distinct pitch contour: Tone 1 (high-flat), Tone 2 (rising), Tone 3 (low or low-dipping), and Tone 4 (falling; Chao, 1965; Ho, 1976; Howie, 1974). Because many words in Mandarin differ minimally in lexical tone, learning to differentiate between these pitch contours is critical to differentiating between words when L1 speakers of English learn Mandarin as an L2.

One way to enhance L2 acquisition of Mandarin lexical tones by L1 English speakers is by showing visual depictions of the pitch contours of lexical tones as they are heard at learning (Bluhme & Burr, 1971; Godfroid et al., 2017; Liu et al., 2011). The combination of these visual and acoustic cues results in robust multimodal representations of lexical tone, as postulated by dual coding theory (Paivio, 1990). Moreover, observing pitch gestures, which convey pitch contours haptically as well as visually, also facilitates L2 acquisition of Mandarin lexical tones by L1 English speakers (Baills et al., 2019; Hannah et al., 2017; Morett et al., 2022; Zhen et al., 2019). This finding is in line with other work showing that observing gestures conveying unfamiliar phonological contrasts via their form and motion facilitate their acquisition in an unfamiliar L2 (Hirata et al., 2014b, 2014a; Hoetjes & Van Maastricht, 2020; Xi et al., 2020). Thus, observing gestures conveying L2 speech sounds may enrich representations via mental simulation of embodied action, as postulated by theories of embodied cognition (Barsalou, 2008; Gibbs, 2006; Shapiro, 2019). Moreover, pitch gestures and visual depictions of pitch contours are grounded in in the vertical conceptual metaphor of pitch, in 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). Indeed, observing visual depictions and pitch gestures incongruent with the pitch contours of lexical tones hinders their L2 acquisition (Morett et al., 2022), suggesting that visual-auditory mappings conflicting with the vertical conceptual metaphor of pitch may result in less robust representations of L2 lexical tones.

In addition to enhancing acquisition of L2 Mandarin lexical tones, observing congruent pitch gestures at learning enhances differentiation between the meanings of L2 Mandarin words differing minimally in lexical tone (Baills et al., 2019; Morett & Chang, 2015). Although the visual and haptic features of pitch gestures do not map directly onto word meanings, their mapping onto pitch contours facilitates differentiation between lexical tones, which is necessary to differentiate between word meanings. In contrast to pitch gestures, representational gestures have visual and haptic features that map directly onto word meanings. A large body of previous research indicates that observing representational gestures conveying the meanings of L2 words from atonal languages at learning enhances subsequent association of these words with their meanings (Allen, 1995; Garcia-Gamez & Macizo, 2019; Kelly et al., 2009; Macedonia et al., 2011; Porter, 2016; Tellier, 2008). Conversely, some research also indicates that observing representational gestures incongruent with the meanings of L2 words from atonal languages at learning interferes with subsequent association of these words with their meanings (Garcia-Gamez & Macizo, 2019; Kelly et al., 2009). Importantly, however, such effects have not been observed for phonologically similar L2 words learned with representational gestures in either atonal or tonal languages (Kelly & Lee, 2012; Morett & Chang, 2015), suggesting that observing representational gestures at learning may distract attention from key phonological distinctions between these L2 words.

One way to gain further insight into how observing pitch and representational gestures affects L2 acquisition of Mandarin words differing minimally in lexical tone is by examining the N400 event related potential (ERP). The N400 is a late posterior negativity reflecting semantic integration,

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

with larger N400s indicating lower semantic relatedness and smaller N400s indicating higher semantic relatedness. In L1, target words accompanied by representational gestures with incongruent meanings elicit a larger N400 than target words accompanied by representational gestures with congruent meanings (Bernardis et al., 2008; Holle & Gunter, 2007; Kelly et al., 2004; Wu & Coulson, 2005, 2007). These differences in N400s are similar to those elicited by words with meanings incongruent vs. congruent with sentential contexts (Özyürek et al., 2007), providing evidence that representational gestures are integrated semantically with words similarly to how words are integrated semantically with one another. In L2, larger N400s have been observed for L1 target words that are incorrect vs. correct translations of L2 prime words following brief L2 exposure (Pu et al., 2016), indicating that the N400 reflects mapping of the phonological forms of L2 words onto their meanings. However, L2 words learned with and without congruent representational gestures elicit similar N400s during subsequent recognition in an old-new paradigm despite eliciting differences in the late positive component, an ERP reflecting recognition (Kelly et al., 2009). To date, no research has examined whether the N400 differs for L2 words learned with representational gestures congruent vs. incongruent with their meanings.

In addition to semantic integration, the N400 also reflects prediction in a linguistic context, with larger N400s indicating lower predictability and smaller N400s indicating higher predictability. Although predictability is often conceptualized semantically, it applies to phonological processing, as well. In L1, larger N400s are elicited by non-rhyming than rhyming words in rhyme judgment tasks (Coch et al., 2005; Noordenbos et al., 2013; Perrin & Garcıa-Larrea, 2003; Praamstra & Stegeman, 1993; Rugg, 1984; Rugg & Barrett, 1987). In L2, words differing from an expected word in a phoneme elicit larger N400s than the expected word (Heidlmayr et al., 2021). Although Mandarin words differing from expected words in a vowel elicit larger N400s than expected words in L1 speakers of atonal languages who are highly proficient in L2 Mandarin, there is no significant difference in N400s elicited by expected words and words differing from them in lexical tone in this population (Pelzl et al., 2019, 2021). Given that words differing from expected words in lexical tone elicit a larger N400 than expected words in L1 Mandarin speakers (Brown-Schmidt & Canseco-Gonzalez, 2004; Li et al., 2008), this finding suggests that differences in lexical tone may be insufficient to disrupt predictive processing of L2 Mandarin in L1 atonal language speakers. Thus, reinforcing the acoustic features of L2 lexical tone via pitch gesture at learning may be necessary to induce subsequent prediction of lexical tone in this population.

The current study employed the N400 to investigate the impact of observing pitch and representational gestures at learning on subsequent semantic and phonological processing of L2 Mandarin words differing minimally in lexical tone in L1 English speakers. In light of the research discussed above, it was predicted that observing pitch gestures at learning would elicit differences in the N400 for learned L2 Mandarin target words with meanings and lexical tones matching vs. mismatching prime words. These results would provide evidence that observing pitch gestures when learning L2 Mandarin words differing in lexical tone enhances word-meaning association via lexical tone differentiation.

Methods

Participants

44 adult native English speakers without tonal language knowledge (age range: 18-32 years; 29 females, 13 males) participated on a volunteer basis or in return for partial course credit. All participants were right-handed and had normal hearing and normal or corrected-to-normal vision. Additionally, participants had no documented sensory, speech, language, learning, or neurological disorders. Data from 2 participants were excluded due to the presence of artifacts in more than 50% of trials. Thus, the final sample consisted of 42 participants.

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

Audio recordings for use during the test phase were created by recording a male native Mandarin speaker saying each word. A speaker of a different sex than the speaker 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

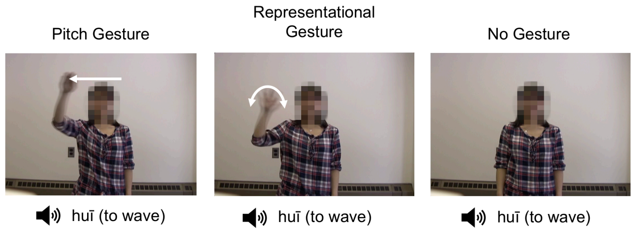

Figure 1: Screenshots of videos from each of the three learning conditions (arrows represent hand motion).

Procedures

In the learning phase, participants were instructed to learn the meanings of Mandarin words as they would subsequently be tested on them, and no mention of the tonal properties of words was made. Participants were randomly assigned to one of three learning conditions: pitch gesture (n = 13), iconic gesture (n = 15), or no gesture (n = 14). In each trial of this task, one of the Mandarin words listed in Table 1 was presented in the video from the corresponding learning condition. Following a 1000 ms interstimulus interval, the English gloss of the preceding Mandarin word was presented for 1000 ms via text. Following a 1000 ms intertrial interval, the trial was repeated with the other Mandarin word in the pair to emphasize the difference in lexical tone between them (order of presentation counterbalanced across participants). All 12 words were presented in this manner in random order in 3 blocks, such that each word was presented 3 times and a total of 36 trials were presented in the learning phase.

The test phase consisted of two tasks: a meaning discrimination task and a lexical tone discrimination task. In both tasks, a prime word was presented with a fixation cross preceding it for a duration jittered between 450 – 500 ms interval. Following a 100 ms interstimulus interval, this sequence was repeated for a target word, which was followed by a 1000 ms intertrial interval. In the meaning discrimination task, English L1 target words either had the same meaning as the preceding Mandarin L2 prime word (k = 72) or a meaning that instead corresponded to the L2 Mandarin word paired with the Mandarin prime word (k = 72). In the lexical tone discrimination task, a prime and a different target word were selected from among the set of learned L2 Mandarin words. Prime and target words had either the same (k = 72) or different (k = 72) lexical tones.

EEG Recording and Data Analysis

Electroencephalographic (EEG) data were recorded via a 128-channel Hydrocel Geodesic sensor net (Electrical Geodesics, Inc., Eugene, OR, USA) with electrodes placed according to the international 10/20 standard. EEG signals were recorded using NetStation 5.4.2 with a NetAmps 300 Amplifier. The online reference electrode was Cz and the ground electrode had a centroparietal location. EEG data were sampled at 1,000 Hz with an anti-aliasing low-pass filter of 4000 Hz.

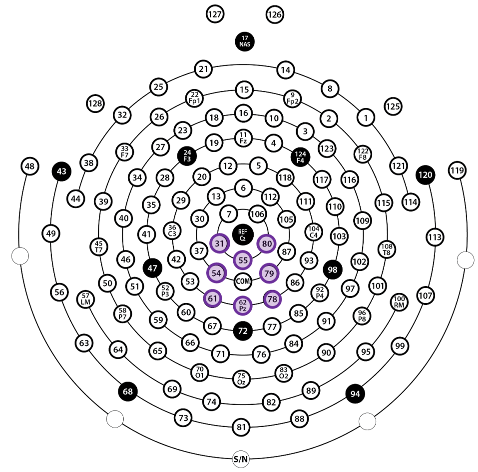

Figure 2: 128-channel montage used for EEG recording with channels in central posterior ROI included in analysis highlighted (purple).

EEG data were pre-processed and analyzed offline using EEGLab and ERPLab. Continuous EEG data were high pass filtered at 0.1 Hz to minimize drift and re-referenced to the online average of all electrodes. Subsequently, excessively noisy or flat channels and data from between-block breaks were removed. Continuous data were then downsampled to 250 Hz, low pass filtered at 30 Hz, and segmented into epochs relative to target word onset. Epoched data were screened for artifacts and abnormalities using a simple voltage threshold of 100 μv and a moving-window peak-to-peak threshold with 500 ms windows, a 100 ms step function, and a 120 μv threshold. Across included participants, 11.5% of trials were rejected, with rejections equally distributed across conditions (F < 1). Finally, trials were classified by condition and congruency and averaged across participants for ERP analyses.

Following other studies examining the N400 for gesture-speech integration, the 300-500 ms time window was selected for statistical analysis. Mean amplitudes recorded during each condition were averaged across a central posterior "region of interest" (ROI) based on inspection of scalp voltage topographies and previous research (see Figure 2).

ERP data were analyzed using linear mixed effect models with condition and congruency as fixed factors, participant and channel as random factors, and mean amplitude in the N400 window as the outcome variable for both the meaning and lexical tone discrimination tasks. Prior to entry into these models, all fixed effects were coded using weighted mean centered (Helmert) contrast coding in order of the levels mentioned. Random slopes were included with the maximal random effect structure permitted to achieve model convergence. For all effects reaching significance for factors with more than two levels, Tukey HSD post-hoc tests were conducted using the *emmeans* package to test for differences between levels.

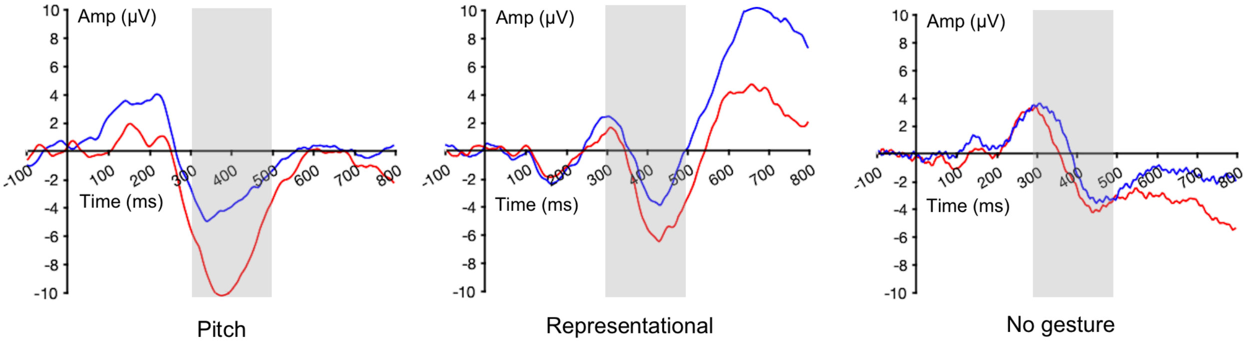

Figure 3: ERP waveforms for congruent (blue) and incongruent (red) word pairs learned with pitch, representational, and no gestures in meaning discrimination task (gray shading indicates N400 time window).

Results

Meaning Discrimination Task

We observed a main effect of congruency (B=-3.31, SE=0.97, z=-3.40, p<.001) as well as interactions of congruency with the no gesture vs. pitch and representational gesture conditions (B=-24.13, SE=2.00, z=-12.05, p<.001) and the no gesture and pitch vs. representational gesture conditions (B=-8.78, SE=2.11, z=-4.17, p<.001). Estimated marginal means revealed that N400 amplitude was larger (i.e., more negative) for target English words with meanings incongruent than congruent with prime Mandarin words learned with pitch gestures (B=9.40, SE=1.75, z=5.37, p<.001) and representational gestures (B=8.15, SE=2.54, z=3.21, p=.02), whereas N400 amplitude did not differ for English target words with meanings incongruent vs. congruent with Mandarin prime words learned with no gesture (B=1.85, SE=1.69, z=1.10, p=.88; see Figure 3.)

Lexical Tone Discrimination Task

We observed a main effect of congruency (B=-2.66, SE=0.80, z=3.32, p<.001) as well as interactions of congruency with the no gesture vs. pitch and representational gesture conditions (B=-24.13, SE=2.00, z=-12.05, p<.001) and the no gesture and representational vs. pitch gesture conditions (B=-7.86, SE=1.68, z=-4.69, p<.001). Estimated marginal means revealed that N400 amplitude was larger (i.e., more negative) for target Mandarin words with lexical tones incongruent than congruent with prime Mandarin words learned with pitch gestures (B=6.88, SE=1.44, z=4.77, p<.001), whereas N400 amplitude did not differ for target Mandarin words with lexical tones incongruent vs. congruent with prime Mandarin words learned with representational gestures (B=-2.68, SE=1.34, z=-1.99, p=.35) or no gesture (B=0.91, SE=2.77, z=0.33, p=.99; see Figure 4).

Discussion

The current study examined how observing pitch and representational gestures when learning L2 Mandarin words differing minimally in lexical tone affects L1 English speakers' semantic and phonological processing via the N400. With respect to semantic processing, larger N400s were observed for English target words mismatching than matching the meanings of Mandarin prime words learned with pitch and representational gesture, but not no gesture. These results build on previous behavioral findings indicating that L1 English speakers are more likely to differentiate between the meanings of L2 Mandarin words differing minimally in lexical tone when they are learned with congruent pitch gestures than without gestures (Baills et al., 2019; Morett & Chang, 2015), providing the first evidence that the N400 reflects this increase in word-meaning association accuracy. Moreover, they provide the first

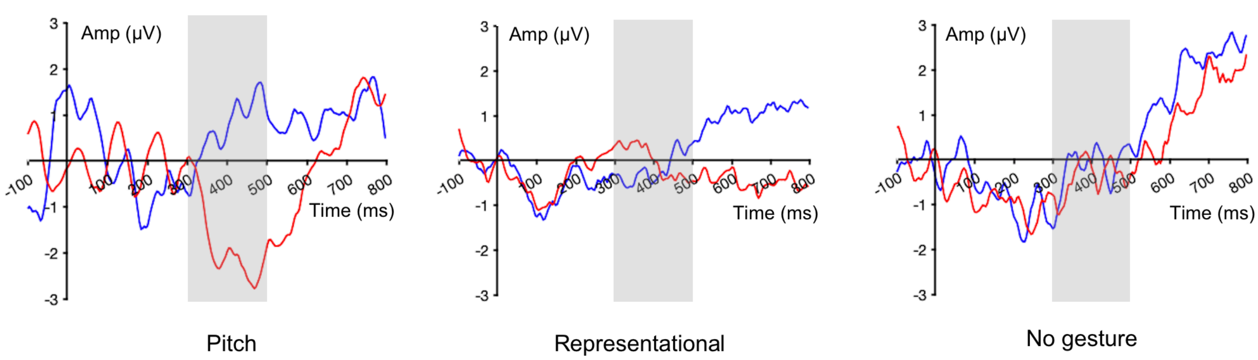

Figure 4: ERP waveforms for congruent (blue) and incongruent (red) word pairs learned with pitch, representational, and no gestures in lexical tone discrimination task (gray shading indicates N400 time window).

evidence that observing congruent representational gestures when learning L2 Mandarin words differing minimally in lexical tone enhances L1 English speakers' differentiation between their meanings relative to no gesture. The lack of such evidence in previous research may be due to employment of behavioral measures (Kelly & Lee, 2012; Morett & Chang, 2015) and the old-new paradigm, which is less conducive to the N400 than the semantic priming translation paradigm employed in the current study, for ERP data collection (Kelly et al., 2009). In light of the novelty of the N400 differences in semantic processing of L2 Mandarin words differing in lexical tone learned with pitch and representational gesture observed in the current study, future research should reveal the extent to which these results can be replicated.

With respect to phonological processing, larger N400s were observed for Mandarin target words mismatching than matching the lexical tones of Mandarin prime words learned with pitch gesture, but not representational or no gesture. These results build on previous findings indicating that N400s are similar for expected and unexpected Mandarin L2 words differing from expected words in lexical tone in L1 English speakers (Pelzl et al., 2019, 2021), and that they differ in L1 Mandarin speakers (Brown-Schmidt & Canseco-Gonzalez, 2004; Li et al., 2008), suggesting that observing pitch gestures conveying the lexical tones of L2 Mandarin words at learning results in more L1-like differentiation between lexical tones. Thus, these results support behavioral findings demonstrating that observation of such pitch gestures when learning L2 Mandarin words differing in lexical tone facilitates L1 English speakers' lexical tone categorization (Baills et al., 2019; Hannah et al., 2017; Morett et al., 2022; Zhen et al., 2019). Notably, the ERP data collected during the lexical tone discrimination task is noisier and the N400 effect isn't as large as it is in the meaning discrimination task. Thus, it is crucial to determine the extent to which differences in the N400 for L2 Mandarin words with matching vs. mismatching lexical tones learned with congruent pitch gestures can be replicated in future work.

In conclusion, the results of the current study reveal that observing pitch gestures when learning L2 Mandarin words differing in lexical tone enhances L1 English speakers' differentiation between their lexical tones and meanings. Furthermore, they reveal that observing representational gestures when learning such L2 Mandarin words also enhances L1 English speakers' differentiation between their meanings, albeit not to the same extent as observing pitch gestures. In addition to providing insight into replicability, future research should further explore the relationship between the N400 and behavioral measures. Nevertheless, the current study provides the first ERP evidence that observing gestures conveying phonological and semantic information during L2 word learning enhances subsequent phonological and semantic processing of learned L2 words.

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.

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.

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

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, *25*(7), 1114–1128.

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

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research*, *33*(2), 103–135.

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.

Coch, D., Grossi, G., Skendzel, W., & Neville, H. (2005). ERP nonword rhyming effects in children and adults. *Journal of Cognitive Neuroscience*, *17*(1), 168–182.

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.

Cruttenden, A. (1997). *Intonation*. Cambridge U. Press.

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.

Gibbs, R. W. (2006). *Embodiment and cognitive science*. Cambridge University Press.

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.

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.

Heidlmayr, K., Ferragne, E., & Isel, F. (2021). Neuroplasticity in the phonological system: The PMN and the N400 as markers for the perception of non-native phonemic contrasts by late second language learners. *Neuropsychologia*, *156*, 107831.

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014a). Effects of hand gestures on auditory learning of L2 vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2090–2101.

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014b). Effects of observing or producing hand gestures on non-native speakers' auditory learning of Japanese short and long vowels. *The Journal of the Acoustical Society of America*, *135*(4), 2353–2353.

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*.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175–1192.

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

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*(1), 253–260.

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

Kelly, S. D., 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.

Li, X., Yang, Y., & Hagoort, P. (2008). Pitch accent and lexical tone processing in Chinese discourse comprehension. *Brain Research*, *1222*, 192–200.

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

Macedonia, M., 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.

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.

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.

Noordenbos, M. W., Segers, E., Wagensveld, B., & Verhoeven, L. (2013). Aberrant N400 responses to phonological overlap during rhyme judgements in children. *Brain Research*, *1537*, 233–243.

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616.

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.

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2021). Even in the best case scenario L2 learners have difficulty perceiving and utilizing tones in Mandarin: Findings from behavioral and event-related potential experiments. *Studies in Second Language Acquisition*, *43*(2), 268–296.

Pelzl, E., Lau, E., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, *41*, 59–86.

Perrin, F., & Garcıa-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research*, *17*(1), 36–47.

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

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, *1*(2), 73–86.

Pu, H., Holcomb, P. J., & Midgley, K. J. (2016). Neural changes underlying early stages of L2 vocabulary acquisition. *Journal of Neurolinguistics*, *40*, 55–65.

Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language*, *23*(2), 225–240.

Rugg, M. D., & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language*, *32*(2), 336–361.

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.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*(6), 654–667.

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*(3), 234–245.

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.

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.