Title
Observing Metaphorically Congruent Gestures Enhances Sensitivity to L2 Lexical Tone

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Batten, Rylan
Farran, Bashar
Morett, Laura M.

Publication Date
2026

Observing Metaphorically Congruent Gestures Enhances Sensitivity to L2 Lexical Tone

Rylan N Batten, Bashar M Farran, Laura M Morett

Abstract

This study investigates how observing gestures varying in congruence with pitch contours influences processing of lexical tones during second language (L2) word learning. Adult L1 English speakers with no tonal language experience completed pre- and post-learning tests in which they heard pairs of Mandarin words with correct (matching) or incorrect (mismatching) tones while N400 event-related potentials were recorded. During the learning phase, participants learned Mandarin words differing minimally in lexical tone while observing congruent, incongruent, or no pitch gestures. Participants completed two post-learning tests of lexical tone association and word meaning discrimination immediately after the learning phase and 24 hours later. Distinct N400 responses emerged for L2 words learned with congruent gestures in the lexical tone discrimination task, indicating that observing congruent pitch gestures facilitates neural sensitivity to lexical tones.

Keywords: Lexical tone, pitch gestures, second language acquisition, event-related potentials, N400, phonological processing

Introduction

The acquisition of lexical tones presents special challenges in second language (L2) learning, particularly for individuals whose first language (L1) is atonal. Unlike atonal languages, tonal languages such as Mandarin Chinese rely on pitch variations to distinguish meaning at the word level. In contrast, atonal languages use pitch primarily to convey intonation, which may indicate questioning, emphasis, or emotion rather than differentiating between word meanings. This fundamental difference can make it difficult for atonal L1 speakers to perceive, process, and produce lexical tones, as they may initially interpret pitch changes as carrying paralinguistic rather than lexical information (Gandour, 1983; Wong & Perrachione, 2007). In addition to perceptual challenges, the complexity of lexical tone learning stems from the cognitive load associated with categorizing and recalling unfamiliar pitch patterns (Chandrasekaran et al., 2010; Francis et al., 2008; Wang et al., 2003). Traditional pedagogical methods, which emphasize auditory training through tone discrimination and categorization, may fail to provide sufficient support for learners who lack experience with tonal systems, resulting in persistent difficulties acquiring lexical tone.

Observing pitch gestures improves atonal L1 speakers' ability to perceive, produce, and internalize L2 lexical tones (Baills et al., 2019; Morett & Chang, 2015; Morett et al., 2022; Zhen et al., 2019). This study extends this line of inquiry by investigating the impact of observing pitch gestures on sensitivity to lexical tone and the extent to which this change in sensitivity corresponds with increases or decreases in recall of L2 word meanings.

A growing body of evidence supports the idea that observing gestures enhances L2 acquisition (Goldin-Meadow & Alibali, 2013; Gullberg, 2006; Macedonia & von Kriegstein, 2012). Observing pitch gestures facilitates perception and production of L2 lexical tones via visual depiction of their pitch contours (Baills et al., 2019; Hannah et al., 2017; Morett & Chang, 2015; Morett et al., 2022; So et al., 2010; Zhen et al., 2019). These findings align with cognitive theories emphasizing multimodal learning, including Dual Coding Theory, which asserts that information is processed through interconnected but distinct verbal and nonverbal channels (Paivio, 1991), and theories of embodied cognition, which maintain that language is processed through aligned sensory and motor experiences (Barsalou, 2008).

Morett et al. (2022) coined the vertical conceptual metaphor of pitch, which proposes that high pitch is associated with upward motion and position and low pitch is associated with downward motion and position. The vertical conceptual metaphor of pitch is evident not only in descriptions of sounds (e.g., "high note," "low tone") but also influences perceptual and motor representations of them (Dolscheid et al., 2013). When gestures align with these spatial metaphors, conceptual mapping is reinforced, facilitating the learning process. Conversely, gestures incongruent with these metaphors (e.g., upward gestures paired with low tones) disrupt these conceptual mappings, hindering perception and categorization of lexical tones (Morett et al., 2022).

Studies utilizing event-related potentials (ERPs) have demonstrated that congruence between visual cues and speech enhances processing efficiency during language comprehension. One key neural marker of semantic integration is the N400 component, which indexes the brain's response to semantic mismatches between auditory and visual stimuli. Specifically, N400 amplitude is greater in response to semantically incongruent stimuli than congruent stimuli, reflecting increased processing difficulty and reduced efficiency in semantic integration (Holcomb & Neville, 1990; Holcomb & Anderson, 1993; Wu & Coulson, 2007). In the context of gesture-speech integration, gestures congruent with the meaning of concurrent speech attenuate the N400 response, suggesting

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

that they facilitate semantic processing by providing visual support aligned with auditory input. Conversely, gestures incongruent with the meaning of concurrent speech increase the N400 response, reflecting heightened effort in resolving conflicting semantic information from the visual and auditory modalities (Holle & Gunter, 2007; Kelly et al., 2009; Wu & Coulson, 2007).

In line with prior findings and theories of multimodal integration and the vertical conceptual metaphor of pitch, we hypothesized that congruency of pitch gestures with lexical tones would affect subsequent N400 responses to words differing in them. Specifically, compared to L2 Mandarin words learned with incongruent or no gestures, we predicted that words learned with congruent pitch gestures would elicit increased N400 amplitude differences when they mismatched vs. matched in lexical tone and meaning. In the no gesture condition, we expected N400 amplitudes intermediate between those in the congruent and incongruent pitch gesture conditions for lexical tone and word meaning discrimination. This would reflect reduced facilitation of lexical tone processing in the absence of exposure to gesture but also the absence of visual–auditory conflict during learning. Crucially, this study extends previous work by measuring sensitivity to differences in lexical tone via N400 responses not only before learning and immediately after learning but also during a 24-hour delayed post-test to examine whether the benefits of gesture-speech congruency extend beyond immediate learning. This design allows us to examine changes from before to after learning as well as the persistence of the effects of pitch gesture over extended intervals, offering new insight into the robustness of the effects of multimodal integration of pitch gestures with lexical tones on L2 word learning.

Methods

Participants

29 L1 English speakers ($M = 22.3$ years, $SD = 3.4$) with no prior exposure to tonal languages were recruited. Inclusion criteria were self-reported normal hearing and vision, right-handedness, and no history of neurological, language, or reading disorders, including amusia. One participant from the no gesture group was excluded from ERP analyses due to N400 amplitudes exceeding ±3 standard deviations from the grand mean.

A priori power analysis was conducted using G*Power 3.1 (Faul et al., 2009) to determine the minimum sample size required for a mixed-design ANOVA with three between-participants groups and three measurement points (pre-, immediate post-, and delayed post-test). The analysis was based on an effect size of $f = 0.29$ (Morett, 2023), an alpha level of 0.05, desired power of 0.80, and an assumed correlation among repeated measures of 0.5. Results indicated that a total sample size of 27 participants was sufficient to achieve adequate power (actual power = 0.82) to detect the hypothesized interaction effect.

Materials

Auditory Stimuli 12 pairs of monosyllabic Mandarin words differing minimally in lexical tone (level, rising, falling-rising, and falling) were adopted from Morett & Chang (2015) and Morett (2023). English glosses were high-frequency, concrete words that are highly familiar and easily interpretable (e.g., full, to wave, to pack, to return, to jump, etc.).

Video stimuli Videos featured an L1 Mandarin speaker, shown from head to torso, who used hand gestures to illustrate the pitch contours of Mandarin lexical tones. For example, an upward hand motion depicted the rising tone, while a downward motion depicted the falling tone. The speaker's face was blurred to control for any differences in mouth movements by lexical tone and encourage attention to the hands. Temporal synchronization was verified by visually inspecting the acoustic waveforms and corresponding video frames.

Experimental Design

The learning phase included three between-participants conditions: congruent pitch gesture, incongruent pitch gesture, and no gesture. In the congruent pitch gesture condition, the speaker produced hand gestures with movement trajectories that accurately depicted the pitch contours of the lexical tones in the spoken Mandarin words. In the incongruent pitch gesture condition, gestures conveyed the pitch contour of a different lexical tone than the lexical tone of the presented word. Lastly, in the no gesture condition, a static image of the same speaker featured in the gesture videos was displayed. This controlled for the visual presence of the speaker while removing dynamic visual-prosodic cues, allowing us to isolate the influence of pitch gestures on lexical tone learning. Representative still images of all three learning conditions are shown in Figure 1.

Procedures Prior to beginning the experimental tasks, participants were familiarized with the setup and task instructions. Behavioral response and EEG data were acquired in pre-, immediate post-, and delayed post-tests.

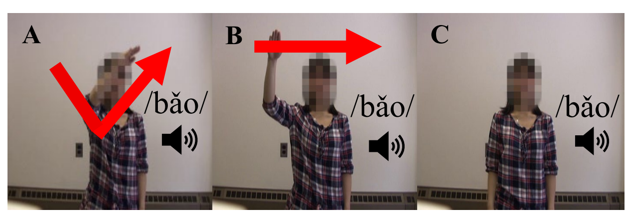

Figure 1: *(A) Congruent gesture depicting the pitch contour of the corresponding word's tone; (B) incongruent gesture depicting the pitch contour of another word's tone; (C) no gesture condition.*

Participants first completed a pre-test lasting approximately 15 minutes, which included a lexical tone discrimination task (see Figure 2). In this task, participants listened to pairs of Mandarin words with either matching or mismatching lexical tones and indicated whether the tones in each pair matched or mismatched by pressing the corresponding button (counterbalanced across participants). Each trial began with a fixation cross, presented for a jittered duration of 450–500 ms, followed by presentation of the prime word. After a 100-ms blank screen, a second fixation cross appeared for another jittered interval of 450–500 ms, preceding presentation of the target word. Upon presentation of the target word, participants indicated whether its lexical tone matched or mismatched that of the prime word. This sequence was repeated until all Mandarin word pairs had been presented 6 times, each in randomized order, resulting in 72 matching and 72 mismatching trials, for a total of 144 trials. Word repetitions were distributed across blocks and intermixed across matching and mismatching trials to minimize habituation to specific word pairings. All participants experienced an identical test-trial structure and repetition schedule, ensuring that any repetition-related attenuation in N400 amplitude was equivalent across groups and did not differentially affect the ERP congruency contrasts.

In the learning phase (see Figure 3), which immediately followed the pre-test, participants were randomly assigned to one of three groups (congruent gesture, incongruent gesture, or no gesture) in a between-participants design. In the congruent pitch gesture group, Mandarin words were paired with gestures depicting the pitch contours of their lexical tones. In the incongruent pitch gesture group, Mandarin words were paired with gestures depicting the pitch contours of the lexical tones of words differing minimally in lexical tone. In the no gesture group, Mandarin words were paired with a still image of the speaker with her hands by her side. Each learning trial began with a fixation cross, jittered at 500 ms, followed by presentation of a Mandarin word in the congruent, incongruent, or no gesture condition, which was followed by its English gloss as text and synthesized speech. Neither behavioral nor ERP data were collected during learning trials. Participants were exposed to 12 Mandarin words repeated 8 times across 12 blocks (i.e., 96 trials total).

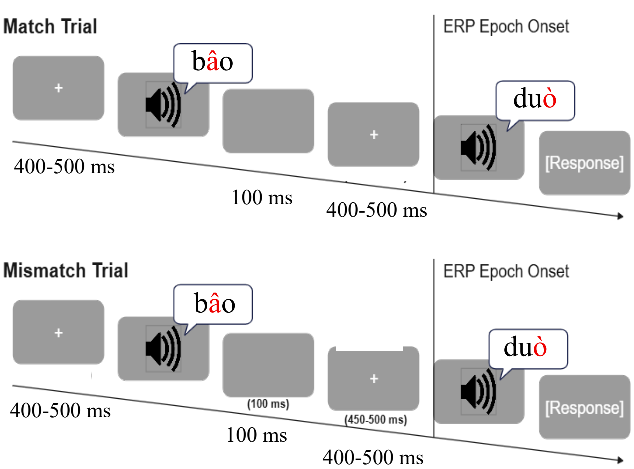

Figure 2: Sample Matching and Mismatching Trials from the Lexical Tone Discrimination Task (Pre-, Immediate Post-, and Delayed Post-Tests)

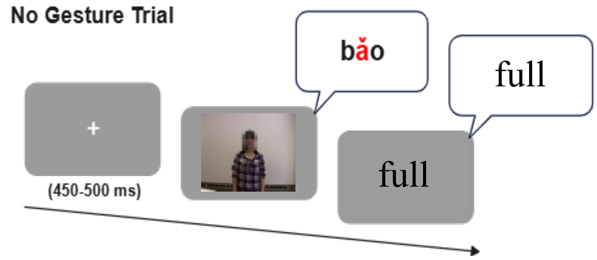

Figure 3: Example Learning Trial Sequence (No Gesture Condition)

Following the learning phase, participants completed an immediate post-test, which contained a lexical tone discrimination task identical to the one administered in the pre-test (see Figure 2). In addition, it contained a word meaning discrimination task. Each trial of this task began with a fixation cross, presented for a jittered duration of 450–500 ms, followed by the audio playback of a prime Mandarin word. A target English word then appeared on the screen as text, and participants indicated whether it matched or mismatched the meaning of the prime Mandarin word. Each Mandarin word was presented 6 times in randomized order, resulting in a total of 144 trials: 72 matching and 72 mismatching pairs. The delayed post-test, approximately 24 hours after the first session, was identical to the immediate post-test.

Data Acquisition and Analysis

Behavioral Data Response accuracy and latency were collected using a Chronos response pad in conjunction with E-Prime 3.0. For latency data, trials with incorrect responses or reaction times below 100 ms or exceeding three standard deviations (SDs) above the mean were excluded to eliminate outliers.

EEG Data EEG data were collected using a Brain Products actiCHamp system with 32 electrodes placed according to the international 10-20 system. Data were sampled at 500 Hz with a band-pass filter of 0.1–30 Hz. During data collection, signals were referenced to the left mastoid, and offline re-referencing was performed using the average of the left and right mastoids. Impedances were kept below 20 kΩ throughout the recording session. Electrode selection for analysis was based on established scalp distributions of the N400 component observed in auditory and cross-modal language processing paradigms. A cluster of electrodes comprising C3, CP5, CP1, Pz, P3, P4, CP6, CP2, and C4 was selected a priori as the region of interest (ROI) for statistical analysis. This approach maximizes sensitivity to expected N400 effects and reduces the likelihood of Type I error associated with exploratory analyses across the full electrode array. The ROI was confirmed by visual inspection of grand-average ERP topographies, which demonstrated maximal negative deflections in the N400 time window over the selected channels for all experimental conditions.

Results

N400 ERP Results

Pooled N400 amplitudes were analyzed for the lexical tone and word meaning discrimination tasks. Linear mixed-effects models for each test (pre-test, immediate post-test, delayed post-test) were used to evaluate fixed effects of learning group (congruent pitch gesture, incongruent pitch gesture, no gesture) and stimulus match (matching, mismatching) on response amplitudes in the N400 time window (300–500 ms). To account for variability across participants and stimuli, random intercepts were specified for both participants and items.

Lexical Tone Discrimination In the pre-test (see Figure 4A), the main effect of stimulus match failed to reach significance (B = 0.65, SE = 0.89, z = 0.73, p = .47), indicating the absence of a measurable N400 mismatch effect prior to learning. Likewise, the main effect of learning group failed to reach significance, indicating no differences between groups prior to learning. Relative to the congruent pitch gesture group, the incongruent pitch gesture group failed to differ in amplitude (B = 1.51, SE = 1.09, z = 1.39, p = .166), whereas the no gesture group showed a marginally lower amplitude (B = -1.80, SE = 1.09, z = -1.66, p = .098). Importantly, the interaction between learning group and stimulus type failed to reach significance for the congruent pitch gesture group (B = 0.70, SE = 0.84, z = 0.84, p = .403), the incongruent pitch gesture group (B = –2.18, SE = 1.63, z = –1.34, p = .180), and the no gesture group (B = –0.41, SE = 1.63, z = -0.25, p = 0.801), indicating that neural responses to words with mismatching versus matching lexical tones did not differ by group prior to learning. These results confirm that comparable neural responses were observed in each learning group and that no pre-existing differences in lexical tone discrimination were present prior to the learning phase.

In the immediate post-test (Figure 4B), we observed a significant main effect of stimulus match (F(1,144) = 29.29, p < .001), with larger N400 amplitudes for words with matching vs. mismatching lexical tones, indicating emergence of sensitivity to lexical tone absent prior to learning in the pre-test. In addition, we observed significant interactions between stimulus match and learning group (F(2,144) = 30.08, p < .001), reflecting greater neural differentiation between matching and mismatching trials in the congruent pitch gesture group. Planned comparisons indicated a robust stimulus match effect in the congruent gesture group (B = 12.84, SE = 1.58, z = 8.13, p < .001). By contrast, the incongruent gesture (B = 1.47, SE = 1.60, z = 0.92, p = .36) and no gesture (B = 0.25, SE = 1.59, z = 0.16, p = .87) groups failed to show a significant effect of stimulus match, indicating an absence of differential neural sensitivity when pitch gesture was either conflicting or absent. Thus, only observing congruent pitch gestures supported neural sensitivity to lexical tone immediately after learning.

In the delayed post-test, the congruent gesture group continued to show a significant stimulus match effect (B = -2.53, SE = 1.28, t = -1.98, p = .048), whereas the incongruent gesture group exhibited a non-significant trend toward a stimulus match effect (B = 0.43, SE = 1.54, t = 0.28, p = .084). Notably, the no gesture group also showed a significant effect of stimulus match (B = 1.43, SE = 0.89, t = 3.82, p = .002). Importantly, within the no gesture group, the mismatching condition elicited greater negativity than the matching condition, and the polarity of the mismatch response in this group contrasted sharply with that of the congruent gesture group. Specifically, the mismatching condition was negative in the no gesture group but positive in the congruent gesture group, yielding a between-group reversal in polarity in the N400 time window. The main effect of learning group failed to reach significance. Compared to the congruent gesture group (B = 1.70, SE = 1.20, z = 1.41, p = .158), neither the incongruent gesture group (B = -2.99, SE = 1.90, z = -1.57,

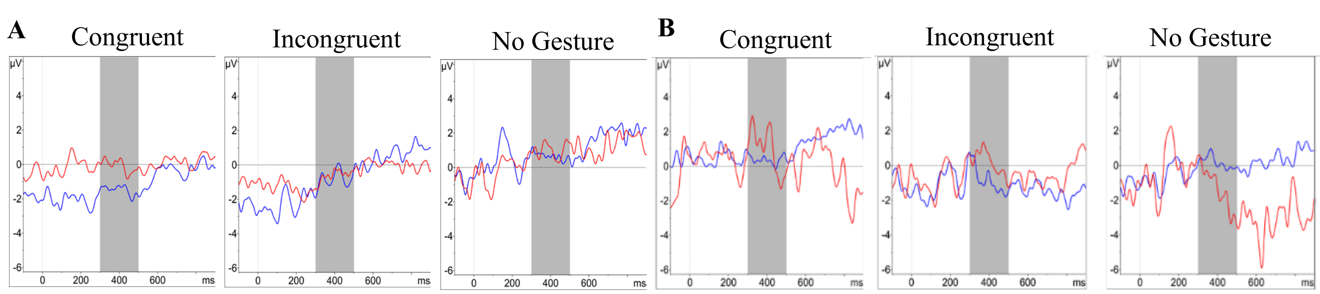

Figure 4: ERP Waveforms from Pooled Channels (C3, CP5, CP1, Pz, P3, P4, CP6, CP2, and C4) for Matching (Blue) and Mismatching (Red) Word Pairs in (A) Immediate Post-Test and (B) Delayed Post-Test of the Lexical Tone Discrimination Task, with Gray Shading Indicating the N400 Time Window.

p = .116) nor the no gesture group (B = -2.08, SE = 1.94, z = -1.07, p = .285) differed significantly in overall N400 amplitude. Together, these results show that although sensitivity to stimulus match persisted in the delayed post-test, its polarity and magnitude differed across learning conditions. While the congruent gesture group maintained a stable response profile, the no gesture group exhibited an opposite-sign response to mismatch, indicating qualitatively different neural representations in the absence of gesture during learning.

Word Meaning Discrimination In the immediate post-test (see Figure 5A), we failed to observe a significant main effect of stimulus match (B = -1.43, SE = 1.10, z = -1.30, p = .193). We observed a marginal main effect of learning group, with participants in the no gesture group exhibiting a non-significant trend towards greater N400 amplitudes compared to those in the congruent group (B = 2.51, SE = 1.35, z = 1.86, p = .063). However, the interaction between learning group and stimulus match failed to reach significance (congruent vs. incongruent: B = –0.38, SE = 1.96, z = –0.19, p = .848; congruent and incongruent vs. no gesture: B = 2.93, SE = 1.96, z = 1.50, p = .134). These results indicate that group differences in neural sensitivity to semantic incongruity failed to emerge immediately after learning, suggesting that word meanings were not robustly encoded.

In the delayed post-test (see Figure 5B), we observed a significant main effect of stimulus match (B = -2.70, SE = 0.84, z = -3.23, p = .002). We failed to observe a significant main effect of learning group, with the no pitch gesture group exhibiting N400 amplitudes comparable to the congruent pitch gesture group (B = -0.04, SE = 2.28, z = -0.02, p = .99). Additionally, the interaction between learning group and stimulus match failed to reach statistical significance (congruent vs. incongruent: B = 4.79, SE = 3.34, t = 1.43, p = .160; congruent and incongruent vs. no gesture: B = 2.70, SE = 3.34, t = 0.81, p = .424). These findings fail to provide conclusive evidence that any group, including the congruent pitch gesture group, maintained differential sensitivity to word-meaning pairings during the delayed post-test.

Discussion

This study investigated how the congruency of pitch gestures observed during learning with the vertical conceptual metaphor of pitch influences neural processing of lexical tones in Mandarin L2 words. The present findings extend previous research by demonstrating that the congruency of pitch gestures with the vertical conceptual metaphor of pitch affects neural processing of lexical tone. This effect was evident from electrophysiological responses, with N400 amplitudes in the post-tests of the lexical tone discrimination task differing according to the alignment of pitch gestures with the vertical conceptual metaphor of pitch. These findings provide the first evidence indicating that cross-modal congruency with this metaphor shapes delayed retention of newly learned tonal contrasts. The current findings indicate that congruent pitch gestures facilitate the learning and retrieval of tone–meaning associations, consistent with prior evidence that semantically aligned gestures enhance multimodal integration and memory encoding (e.g., Kelly et al., 2010; Morett & Chang, 2015). The absence of similar enhancement in the incongruent and no gesture conditions suggests that the observed effects are best interpreted as facilitation through congruency, rather than conflict induced by incongruent cues. Notably, the no gesture condition provided a critical control, confirming that facilitation stems from gesture–tone congruence rather than visual stimuli or dual-modality input.

Although no significant group differences emerged in behavioral accuracy or reaction time, the ERP findings revealed evidence of sensitivity to gesture-speech congruency. Specifically, the congruent gesture group showed an N400 effect consistent with the integration of word–meaning mappings, suggesting that neural indices may capture subtle learning-related changes not reflected in overt performance measures. Additional factors that may have contributed to the divergence between behavioral and neural findings include the relatively brief training exposure, individual variability in learning rates, and the possibility that task demands limited the sensitivity of behavioral responses. Taken together, these results suggest that although behavioral responses provide an important index of learning, neurophysiological measures offer a complementary window into the underlying processes, particularly in early stages of learning.

Lexical Tone Discrimination Task Results from the pre-test confirmed that all learning groups showed similar N400 responses regardless of tonal match, confirming a

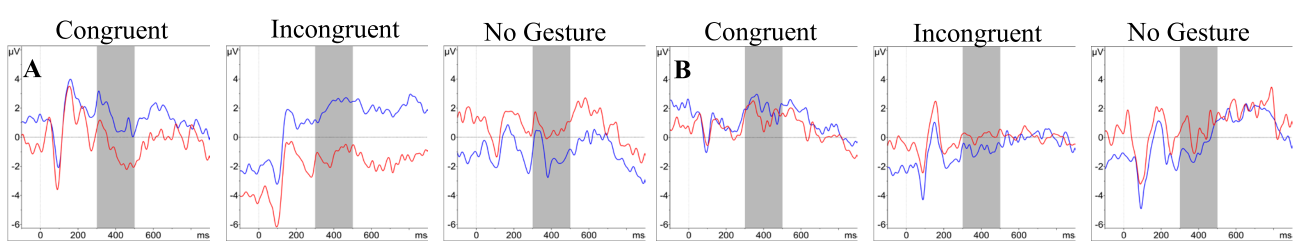

Figure 5: ERP Waveforms and Topographic Plots from Pooled Channels (C3, CP5, CP1, Pz, P3, P4, CP6, CP2, and C4) for Matching (Blue) and Mismatching (Red) Word Pairs in the (A) Immediate Post-Tests and (B) Delayed Post-Tests of the Word Meaning Discrimination Task, with Gray Shading Indicating the N400 Time Window

lack of pre-existing differences prior to learning. This absence of N400 differences prior to learning permits attribution of post-learning changes to observation of pitch gestures when learning L2 Mandarin words differing minimally in lexical tone. Considered in relation to the group-level differences observed in the immediate and delayed post-tests, it also shows that the N400 reflects subtle changes in lexical tone processing over time and across learning conditions.

In the immediate post-test, the congruent pitch gesture group exhibited a stimulus match effect in the N400, aligning with previous work demonstrating that observing gestures conveying the pitch contours of lexical tones of L2 Mandarin words facilitates subsequent sensitivity to lexical tone in these words (Baills et al., 2019; Morett, 2023). The larger N400 differences in the congruent than the incongruent and no gesture groups reflect greater sensitivity to differences between lexical tones, indicating that multimodal congruence during learning promotes stronger associations between pitch contours and lexical tone categories. The effect was widespread across channels, particularly in centroparietal regions, which are commonly implicated in phonological processing (Silva-Pereyra et al., 2003; Kutas & Federmeier, 2011). The no gesture group failed to show consistent N400 differences in post-tests, suggesting limitations in auditory-only tone exposure for facilitating sensitivity to lexical tones in this context. Regarding the reversed-polarity observed in the delayed posttest, we cautiously propose that this may reflect variability in processing strategies across participants, as opposed to being explained by regression of discriminatory skills, or participant confusion.

These results provide support for the Dual Coding Theory (Paivio, 1991) and embodied cognition (Barsalou, 2008) frameworks, which emphasize the roles of aligned multimodal information and the human body in language processing, respectively. Through grounding of lexical tone pitch contours via hand gestures congruent with the vertical conceptual metaphor of pitch, stronger memory traces are formed that are more easily accessed over time. The findings also contribute to our understanding of how abstract phonological constructs such as lexical tone are embodied through sensorimotor experience.

Word Meaning Discrimination Task The findings from the word meaning discrimination task failed to support our hypothesis that pitch gesture congruency during L2 word learning would enhance neural sensitivity to word meaning discrimination, as indexed by the N400. Across the immediate and delayed post-tests, there were no significant group by stimulus match interactions, and the lack of N400 differences between learning groups suggests that sensitivity to semantic congruency failed to develop or persist following training. Although a significant main effect of stimulus match was observed in the delayed post-test, indicating general sensitivity to word meaning discrimination across all groups, this effect was not modulated by learning condition.

In contrast to the lexical tone association task, the word meaning discrimination task required participants to access and evaluate the congruency of L2 Mandarin words with L1 English meanings. Notably, Morett (2023) demonstrated that observing pitch gestures during L2 learning can enhance sensitivity to word meaning discrimination, as reflected in greater N400 effects for mismatching trials. The absence of a comparable effect in the current study may reflect differences in training intensity, stimulus complexity, or the timing of post-learning assessment. It is possible that the consolidation of gesture-related learning effects for word meaning discrimination requires more extensive training or longer retention intervals than those employed in this study (see also Kelly et al., 2009; Macedonia & Klimesch, 2014; Macedonia & Kriegstein, 2012). Additionally, the contrast in effects between the semantic and phonological testing conditions may have emerged because the lexical tone association task directs participants to explicitly evaluate the lexical tones they are presented, which may have increased the salience of lexical tone pitch contours, relative to the word meaning association tests.

Taken together, these results indicate that observing congruent pitch gestures does not robustly enhance word meaning discrimination, as evidenced by the absence of significant interactions between learning group and stimulus match. Although observing congruent pitch gestures facilitated lexical tone discrimination, their impact on semantic integration and retention of word meanings appears limited in this context.

Conclusion

The present study offers evidence that congruency of pitch gestures observed during tonal L2 word learning with the vertical conceptual metaphor of pitch shapes subsequent processing of L2 lexical tone. Congruent pitch gestures elicited sustained differences in N400 amplitudes for lexical tone discrimination at both immediate and delayed intervals. These effects suggest that observing congruent pitch gestures enhances phonological encoding of lexical tone, enabling more efficient category formation and retrieval over time. In contrast, observing incongruent pitch gestures and the absence of gestures interferes with or delays this process, resulting in less robust N400 differences during L2 lexical tone processing. These findings underscore that visual and embodied cues consistent with the vertical conceptual metaphor of pitch support L2 lexical tone acquisition. These cues facilitate learning by linking auditory and visual depictions of lexical tone, thereby promoting richer and more robust learning. This is especially valuable for tonal L2s, for which successful learning requires sensitivity to lexical tone, which is absent from atonal L1s.

References

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*(1), 33–58. https://doi.org/10.1017/S0272263118000074

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: psychophysical evidence for linguistic relativity. *Psychological science*, *24*(5), 613–621. https://doi.org/10.1177/0956797612457374

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175. https://doi.org/10.1016/S0095-4470(19)30813-7

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's Role in Speaking, Learning, and Creating Language. *Annual Review of Psychology*, *64*(Volume 64, 2013), 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Gullberg, M. (2006). *Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon)*. *44*(2), 103–124. https://doi.org/10.1515/IRAL.2006.004

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*. https://doi.org/10.3389/fpsyg.2017.02051

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*(4), 379–411. https://doi.org/10.1080/01690969308407583

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312. https://doi.org/10.1080/01690969008407065

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*(7), 1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

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

Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Kelly, S. D., Bailey, A., & Hirata, Y. (2017). Metaphoric Gestures Facilitate Perception of Intonation More than Length in Auditory Judgments of Non-Native Phonemic Contrasts. *Collabra: Psychology*, *3*(1), 7. https://doi.org/10.1525/collabra.76

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Macedonia, M., & Klimesch, W. (2014). Long-Term Effects of Gestures on Memory for Foreign Language Words Trained in the Classroom. *Mind, Brain, and Education*, *8*(2), 74–88. https://doi.org/10.1111/mbe.12047

Macedonia, M., & Kriegstein, K. von. (2012). Gestures Enhance Foreign Language Learning. *Biolinguistics*, *6*(3–4), Article 3–4. https://doi.org/10.5964/bioling.8931

Morett, L. M. (2023). Observing gesture at learning enhances subsequent phonological and semantic processing of L2 words: An N400 study. *Brain and Language*, *246*, 105327. https://doi.org/10.1016/j.bandl.2023.105327

Morett, L. M., & Chang, L.-Y. (2015). Emphasizing 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. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Silva-Pereyra, J., Rivera-Gaxiola, M., Aubert, E., Bosch, J., Galán, L., & Salazar, A. (2003). N400 during lexical decision tasks: A current source localization study. *Clinical Neurophysiology*, *114*(12), 2469–2486. https://doi.org/10.1016/S1388-2457(03)00248-7

So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI Evidence for Cortical Modification during Learning of Mandarin Lexical Tone. *Journal of Cognitive Neuroscience*, *15*(7), 1019–1027. https://doi.org/10.1162/089892903770007407

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. https://doi.org/10.1017/S0142716407070312

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