Brain & Language 246 (2023) 105327

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Observing gesture at learning enhances subsequent phonological and semantic processing of L2 words: An N400 study

Laura M. Morett

Department of Speech, Language and Hearing Sciences, University of Missouri, 421 Lewis Hall, Columbia, MO 65211, United States
Department of Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama, United States

ARTICLE INFO

Keywords:
Gesture
Lexical tone
L2 word learning
N400

ABSTRACT

This study employed the N400 event-related potential (ERP) to investigate how observing different types of gestures at learning affects the subsequent processing of L2 Mandarin words differing in lexical tone by L1 English speakers. The effects of pitch gestures conveying lexical tones (e.g., upwards diagonal movements for rising tone), semantic gestures conveying word meanings (e.g., waving goodbye for to wave), and no gesture were compared. In a lexical tone discrimination task, larger N400s for Mandarin target words mismatching vs. matching Mandarin prime words in lexical tone were observed for words learned with pitch gesture. In a meaning discrimination task, larger N400s for English target words mismatching vs. matching Mandarin prime words in meaning were observed for words learned with pitch and semantic gesture. These findings provide the first neural evidence that observing gestures during L2 word learning enhances subsequent phonological and semantic processing of learned L2 words.

1. Introduction

Learning words in a second language (L2) entails mapping their phonological forms onto their meanings. When the phonological forms of L2 words differ minimally in an unfamiliar speech sound, discrimination between them based on this speech sound is necessary to map them onto their meanings. Observing pitch gestures—hand gestures conveying pitch height and contour—when learning Mandarin words differing minimally in lexical tone facilitates atonal first language (L1) speakers' differentiation between their phonological forms and meanings (Baills et al., 2019; Morett et al., 2022; Morett & Chang, 2015; Zhen et al., 2019). By contrast, observing iconic gestures conveying the referents of words fails to do so (Morett & Chang, 2015). The current study employs the N400, an event-related potential (ERP) indexing prediction and semantic integration effort, to determine how observing pitch gestures conveying lexical tone and semantic gestures conveying word meanings at learning influences subsequent processing of Mandarin words differing minimally in lexical tone. In doing so, it provides insight into how information about the phonology and semantics of L2 words conveyed via gesture at learning affects subsequent phonological and semantic processing of these words.

1.1. L2 lexical tone acquisition in the context of word learning

Lexical tone is a speech sound present in many world languages that consists of pitches (fundamental frequencies; *f0*) differing in duration (e.g., short, long), height (e.g., low, high) and/or contour (e.g., rising, falling) that is used to distinguish between word meanings or inflections (Gussenhoven, 2004; Maddieson, 2013; Yip, 2002). Mandarin, the most widely-spoken tonal language, has four principal lexical tones: (1) high-flat; (2) rising; (3) low or low-dipping; and (4) falling (Chao, 1965; Ho, 1976; Howie, 1974). These tones often differ minimally between words; for example, as in mā (*mother*), má (*hemp*), mǎ (*horse*), and mà (*scold*). For words with similar forms such as these, lexical tone is the sole phonological cue by which meaning can be differentiated. For words differing in form, lexical tone influences semantic access, albeit at a later point in time than initial consonants and vowels (C.-Y. Lee, 2007; Shuai et al., 2012; Wiener & Turnbull, 2016). Thus, L2 learners of tonal languages, such as Mandarin, who are L1 speakers of atonal languages, such as English, must learn to discriminate between the lexical tones of such words to discriminate between their meanings.

Many adult English speakers, even with advanced L2 Mandarin proficiency, have difficulty acquiring Mandarin lexical tone and reliably differentiating between words differing minimally in it (Pelzl, 2019). Despite this, many adult English speakers unfamiliar with Mandarin can

E-mail address: lmorett@health.missouri.edu.

https://doi.org/10.1016/j.bandl.2023.105327
Received 28 April 2023; Received in revised form 26 September 2023; Accepted 29 September 2023
Available online 5 October 2023
0093-934X/

learn to successfully differentiate between the tones and meanings of such words given brief training (Wang et al., 1999; Wong & Perrachione, 2007), although success varies based on individual differences (Chandrasekaran et al., 2010; Cooper & Wang, 2012; Perrachione et al., 2011). Such increases in differentiation accuracy are associated with a right-to-left shift in functional brain activity (R. R.-W. Lee et al., 2017; Wang et al., 2003; Wong et al., 2007), approximating the left-lateralized functional brain activity characterizing lexical tone processing in L1 Mandarin speakers (Wang et al., 2001, 2004). In Thai, another tonal language, the mismatch negativity (MMN), an ERP sensitive to auditory deviance, develops in English speakers following lexical tone training in response to deviant tones differing from standard tones in initial pitch. By contrast, the MMN develops in Mandarin speakers following such training for all types of deviant tones (Kaan et al., 2007, 2008). These results suggest that speakers of atonal languages, such as English, are sensitive primarily to initial pitch differences in L2 lexical tones, whereas speakers of other tonal languages, such as Mandarin, are sensitive to later differences in pitch contours in L2 lexical tones. Together, these findings suggest that brief training based on a limited set of L2 stimuli differing minimally in lexical tone enhances subsequent identification of lexical tones and meanings of words differing minimally in it and changes the neural signatures of their processing in adult English speakers.

1.2. Gesture observation and L2 word learning

Behavioral response data provides evidence that observing hand gestures while learning L2 words affects how these words are subsequently processed in accordance with the information conveyed via gesture in relation to speech. A substantial body of research has shown that observing gestures conveying L2 word meanings via their form and motion while learning L2 words improves subsequent memory for these words (Allen, 1995; Bergmann & Macedonia, 2013; Garcia-Gamez & Macizo, 2019; Kelly et al., 2009; Macedonia et al., 2011; Macedonia & Knösche, 2011; Porter, 2016; Sweller et al., 2020; Tellier, 2008). This may be due to gestures' visual conveyance of motor imagery associated with the meanings of these words, which enriches semantic representations of them via dual coding through the visual and verbal channels, resulting in robust multimodal traces (Paivio, 1990). It is important to note, however, that behavioral response data suggests that the beneficial effect of observing gestures conveying word meanings may only apply to phonologically dissimilar L2 words. By contrast, phonologically similar L2 words, including words differing minimally in lexical tone, are *less* likely to be remembered when they are learned by observing gestures conveying their meanings than without observing these gestures at learning (Kelly & Lee, 2012; Morett & Chang, 2015). This difference based on phonological similarity may be due to gestures conveying word meanings interfering with attention to the phonological forms of L2 words, which is crucial to distinguishing between phonologically similar L2 words. Thus, to promote differentiation between the meanings of phonologically similar L2 words, gestures may need to direct attention to differences in phonology rather than meaning.

A growing number of studies provide evidence that observing gestures conveying unfamiliar L2 speech sounds, including lexical tone, via their form and motion facilitates auditory discrimination between these speech sounds (Baills et al., 2019; Hannah et al., 2017; Hoetjes & Van Maastricht, 2020; Morett et al., 2022; Xi et al., 2020; Zhen et al., 2019). Successful discrimination between these speech sounds is necessary to differentiate between the meanings of L2 words differing minimally in them; therefore, observing gestures conveying these speech sounds visually should enhance differentiation between them, in turn enhancing differentiation between their meanings. Pitch gestures conveying the pitch contours of lexical tones are often used to teach Mandarin lexical tone formally and are sometimes spontaneously produced by L1 Mandarin speakers to emphasize differences in lexical tone between words similar in form. Longitudinal work has demonstrated superior perception and production of Mandarin lexical tones in children and L2 learners whose teachers employ pitch gestures relative to those whose teachers do not employ pitch gestures (Chen, 2013; Jia & Wang, 2013a, 2013b). Likewise, in lab settings, training incorporating observation of pitch gesture facilitates English speakers' differentiation between the meanings of Mandarin words differing minimally in lexical tone (Baills et al., 2019; Morett & Chang, 2015). Notably, observing pitch gestures enhances differentiation between L2 lexical tones more effectively than observing gestures conveying other unfamiliar L2 phonological contrasts (e.g., vowel length; Hirata et al., 2014; Hirata & Kelly, 2010; Kelly et al., 2014, 2017). This finding suggests that the vertical conceptual metaphor that pitch gestures are based on is particularly conducive to conveying lexical tone (Morett et al., 2022), as is the case for musical tone (Casasanto et al., 2003; Connell et al., 2013). Together, these findings suggest that pitch gestures direct L1 atonal language speakers' attention to differences between the pitch contours of lexical tones, enhancing their discrimination between L2 words differing in lexical tone.

1.3. The N400 as a neural signature of L2 word learning and gesture-speech integration

The N400 is a late posterior negativity reflecting semantic integration and predictive processing effort. Differences in the N400 can be elicited via sentence comprehension as well as priming, making it ideal for comparing semantic and phonological processing in these contexts. Although N400 differences often track with differences in behavioral measures of accuracy and latency, they can nevertheless provide important insight into implicit processing and individual differences in language comprehension that are not always evident from behavioral measures. For example, N400 differences between L2 words and pseudowords are correlated with hours of L2 instruction, whereas discrimination accuracy is not (McLaughlin et al., 2004). Moreover, individual differences in N400 responses during L2 sentence processing show much greater variation than behavioral acceptability judgments and are related to age of acquisition and motivation (Tanner et al., 2014). These findings suggest that N400 differences may provide insight into semantic and phonological processing during early stage L2 acquisition that may not be discernible from behavioral response data.

With respect to semantic integration effort, N400 differences for semantically congruent and incongruent stimuli have been observed following L2 word learning as well as during processing of heard L1 words in conjunction with observed gestures conveying their meanings. Despite an initial absence of N400 differences, following brief (<4–14 hrs.) exposure to an unfamiliar L2, significant N400 differences are evident between L2 pseudowords and words (McLaughlin et al., 2004) as well as incorrect and correct L1 translations of L2 words (Pu et al., 2016). These findings indicate that the N400 reflects mapping of the phonological forms of L2 words onto their meanings in early stage L2 acquisition. With respect to lexical tone, L1 Mandarin speakers demonstrate N400 differences for Mandarin target words with lexical tones mismatching vs. matching pictorial primes (e.g., picture: *hua1* "flower", sound: *gua1* "melon" vs. *hua1* "flower"; Malins & Joanisse, 2012) as well as sentential contexts (e.g., My mother's chicken is very ill, she doesn't lay gallbladder (dǎn) vs. eggs (dàn); Brown-Schmidt & Canseco-Gonzalez, 2004). In L1 sentential and priming contexts, the N400 for words accompanied by gestures mismatching them in meaning is larger than for words accompanied by gestures matching them in meaning (Bernardis et al., 2008; Holle & Gunter, 2007; Kelly et al., 2004; Wu & Coulson, 2005, 2007). These N400 effects parallel those for words matching vs. mismatching sentential contexts in meaning in L1 (Özyürek et al., 2007), indicating that gestures conveying meaning affect semantic processing of words in priming and sentential contexts similarly to how words affect one another's semantic processing in these contexts. During recognition of L2 words previously learned with or without gestures conveying their meanings, no N400 differences were observed. However, the late positive component (LPC), an ERP reflecting recognition, was larger for L2 words learned with than without gestures conveying their meanings, suggesting that such gestures facilitate L2 word recognition (Kelly et al., 2009). This difference in the LPC rather than the N400 may have been due to the type of paradigm used (recognition of learned vs. unlearned words vs. priming or sentential processing) rather than the presence of gesture at learning. At present, however, the sensitivity of the N400 to semantic integration of spoken words with gestures conveying their meanings remains unexplored in early stage L2 learning. Thus, the extent to which observing gestures conveying the meanings of L2 words at learning affects the N400 during subsequent association of these L2 words with their meanings is currently unclear.

In addition to semantic integration effort, the N400 also reflects predictive processing in language more generally. This effect extends to speech sounds, with non-rhyming words eliciting a larger N400 than rhyming words during auditory rhyme judgment tasks (Coch et al., 2005; Noordenbos et al., 2013; Perrin & Garcıa-Larrea, 2003; Praamstra & Stegeman, 1993; Rugg, 1984; Rugg & Barrett, 1987). Phonological neighborhood size also affects the N400, with some work showing larger N400s for words with more phonological neighbors (Carrasco-Ortiz et al., 2017) and other work showing smaller N400s for such words (Yates et al., 2020), providing further evidence that the N400 indexes speech sound prediction. In L2, the N400 reflects the phonological-semantic interface, as evidenced by larger N400s in L2 learners of English for words differing from expected English words in one phoneme than for expected English words in auditory sentential contexts (Heidlmayr et al., 2021). With respect to lexical tone, N400 effects can be elusive in advanced L2 learners who are L1 atonal language speakers. Although larger N400s have been observed in atonal L1 speakers for Mandarin words inconsistent than consistent with sentential contexts in vowels, no significant differences in N400s have been observed in atonal L1 speakers for Mandarin words inconsistent vs. consistent with sentential contexts in lexical tone (Pelzl et al., 2019). A similar pattern of N400 results has been observed in atonal L1 speakers in a Mandarin nonword decision task in which nonwords differ minimally from words in vowels vs. lexical tones (Pelzl et al., 2021), suggesting that these results cannot be explained by prediction alone. To date, no published research has examined whether the N400 reflects differences in lexical tone and their impact on the meanings of L2 words in early-stage acquisition by atonal L1 speakers, so it is currently unclear whether these findings extend to it.

1.4. Current study

The current study was the first to investigate the neural substrates of the impact of gesture observation on L2 lexical tone acquisition by L1 atonal language speakers. More specifically, it investigated the influence of information conveyed via gesture about the lexical tones and meanings of L2 Mandarin words differing minimally in lexical tone in relation to their subsequent phonological and semantic processing. To do so, all words were learned while observing either pitch gestures conveying the pitch contours of lexical tones, semantic gestures conveying word meanings, or no gestures. Subsequently, N400 differences were examined using a lexical tone discrimination task, in which a target Mandarin word had either the same or a different lexical tone than a preceding prime Mandarin word differing in form; and a meaning discrimination task, in which the meaning of a target English word either matched or mismatched a preceding prime Mandarin word. Examining the N400 in the same way in these two priming tasks permitted direct comparison of semantic and phonological processing of Mandarin words learned by observing pitch gesture, semantic gesture, and no gesture. Additionally, collection of response accuracy and latency data permitted comparison of behavioral and neural measures. Thus, this study provided insight into the N400 as an index of phonological and semantic processing of L2 words differing minimally in an unfamiliar phoneme, revealing whether it is more sensitive than response accuracy and latency.

Based on previous work demonstrating that observing pitch gestures at learning facilitates lexical tone identification and word-meaning association for Mandarin words differing minimally in lexical tone by atonal L1 speakers (Baills et al., 2019; Morett et al., 2022; Morett & Chang, 2015; Zhen et al., 2019), we predicted that observing pitch gestures at learning would increase N400 differences in both the lexical tone and meaning discrimination tasks to a greater extent than observing semantic gestures and no gestures. If confirmed, these findings would indicate that observing pitch gestures when learning L2 words differing minimally in lexical tone facilitates subsequent processing of both lexical tone and meaning in such words by atonal L1 speakers. In doing so, they would provide neural evidence that observing pitch gestures enhances L2 word-meaning association via lexical tone differentiation. Alternatively, observing pitch gestures at learning may increase N400 differences in the lexical tone discrimination task to a greater extent than observing semantic gestures or no gestures, whereas observing semantic gestures at learning may increase N400 differences in the meaning discrimination task to a greater extent than observing pitch gestures or no gestures. These findings would indicate that information about the lexical tones of L2 words conveyed via pitch gestures influences subsequent processing of the lexical tones of these words, whereas information about the meanings of L2 words conveyed via semantic gestures influences subsequent processing of the meanings of these words.

2. Methods

2.1. Participants

44 L1 English speakers with no tonal language knowledge (age range: 18–31 yrs.; 29 females, 15 males) participated in this experiment on a volunteer basis or in return for partial course credit. This sample size exceeds by eight the minimum sample size of 36 participants computed via a post-hoc power analysis including main effects and interactions with 85% power to detect $d = 0.547$ with $\alpha = 0.05$. All participants were right-handed and had normal hearing and normal or corrected-to-normal vision, and none had any 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. Informed consent was provided by all participants prior to participation.

2.2. 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 gesture. All Mandarin words were monosyllabic and consisted of phonemes present in English.

Vid***eo recordings of Mandarin words used in the word learning task were derived from vid***eos used in Morett and Chang (2015) and

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

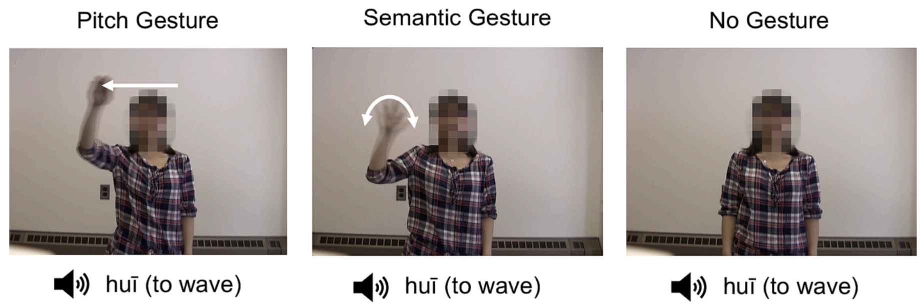

Fig. 1. Screenshots of Vid***eos from Each Learning Condition with Mandarin Word and English Translation, with Arrows Representing Hand Motion (Face Blurred in Figure Only).

are available via the following link: https://osf.io/nufkd/?view_only=34c5065fcf4c405da0a0e4cdf78e267b. The frame size of these vid***eos was reduced from 100% to 25% to ensure that they could be viewed comfortably from a normal seating distance (~40–45 cm) to minimize eye movements. These vid***eos featured a female L1 Mandarin speaker from the torso up saying each Mandarin word. While saying each Mandarin word, the speaker either produced a pitch gesture conveying the pitch contour of the word's lexical tone, a semantic gesture conveying the word's meaning, or kept her hands still and did not produce a gesture (see Fig. 1). All gestures were produced relative to the speaker's frame of reference. Table 2 provides descriptions of pitch gestures used to convey lexical tones, and Table 3 provides descriptions of semantic gestures used to convey word meanings. To ensure that Mandarin words were acoustically identical across conditions, soundtracks of the pitch and semantic gesture conditions were replaced with that of the no gesture condition. Additionally, to ensure that aud***io-visual synchrony was comparable across conditions, Mandarin word onsets were aligned with gesture stroke onsets and the speaker's face was blurred to eliminate any discrepancies between lip movements and speech.

In contrast to training vid***eos, which featured a female speaker's voice in conjunction with her body, aud***io stimuli used in the lexical tone and meaning discrimination tasks featured a male voice. This was done to ensure that participants could generalize lexical tone across speakers, as in Morett and Chang (2015). These recordings were segmented into individual files with 0.1 s of silence preceding and following words using Audacity (Version 2.3.0). To eliminate any differences in volume across individual recordings, the intensity of these aud***io files was normalized to 71.44 dB using Praat (Version 6.1.06; Boersma & Weenink, 2016).

2.3. Procedure

Each participant sat facing a computer screen located approximately 40–45 cm away. Vid***eo stimuli were presented on the computer screen at a 1920 × 1080 resolution with masking, such that they were 7.5 cm in height and 10 cm in width, subtending a 79.61 – 86.30 degree visual angle. Aud***io stimuli were presented through loudspeakers placed beside the screen at a pre-specified volume level.

Participants first completed the word learning task (see Fig. 2 for sample trial), in which they were instructed to learn the meanings of Mandarin words as they would subsequently be tested on them. To discourage participants from attending exclusively to lexical tone at the expense of meaning, no mention of the tonal properties of words was made. Participants were randomly assigned to one of three learning conditions: pitch gesture (n = 13), semantic 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 via the vid***eo from the corresponding learning condition. Following a 1000 ms interstimulus interval, the English translation of the preceding Mandarin word was presented for 1000 ms via text. Following another 1000 ms interstimulus interval, this sequence was presented again. 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

Table 2
Pitch Gesture Descriptions.

Lexical tone	Pitch gesture
Tone 1 (high flat)	Moves right hand from left to right above eye level
Tone 2 (rising)	Moves right hand diagonally from left of chest level to right above head
Tone 3 (low dipping)	Moves right hand from left above eye level to center at stomach level, then to right above eye level
Tone 4 (falling)	Moves right hand from left above eye level to right at stomach level

Table 3
Semantic Gesture Descriptions.

Mandarin word	English translation	Semantic gesture
hui1	to wave	Waves right hand from side to side repeatedly
hui2	to return	Moves right hand in circle horizontally with index finger extended twice
bao1	to pack	Moves partially clasped right hand in arcing motion twice towards palm of upright left hand
bao3	full	Strokes both hands with fingers spread downward simultaneously over stomach three times
chou1	to pump	Moves both clasped hands downward and then back upward from eye level to hip level three times
chou4	to stink	Turns to face palm of upright left hand and fans right hand in front of nose three times
xiang2	to surrender	Raises both hands with palms out and fingers spread to eye level on either side of face simultaneously
xiang3	to think	Turns to right and taps extended right index finger against right cheek four times
tiao2	to shift	Moves both extended clasped hands across one another in an × shape horizontally and then returns them to original position
tiao4	to jump	Jumps while moving both hands up and then back down simultaneously
duo3	to hide	Crouches down and moves quickly to right while moving clasped hands close to one another near face
duo4	to chop	Brings right hand with fingers together down four times while extending left hand with palm up

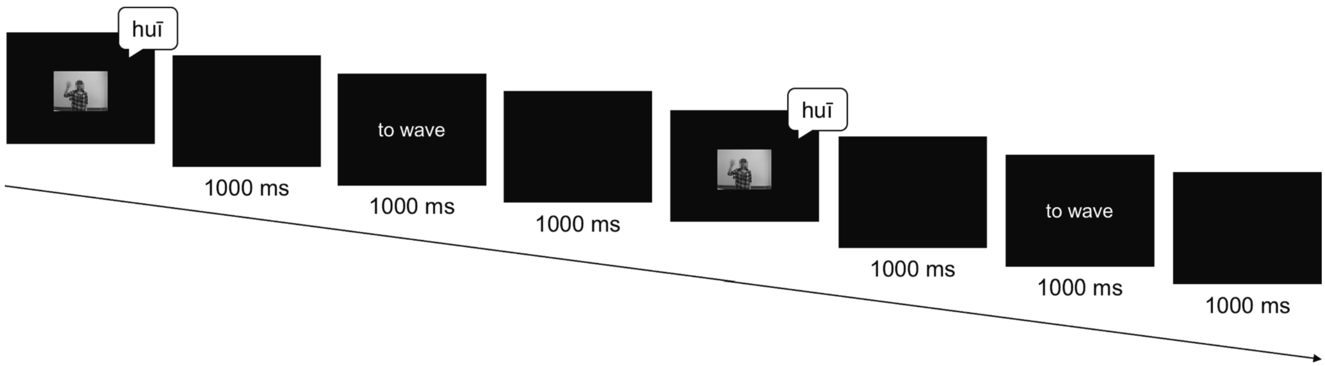

Fig. 2. Sample Trial from Word Learning Task (Semantic Gesture Condition).

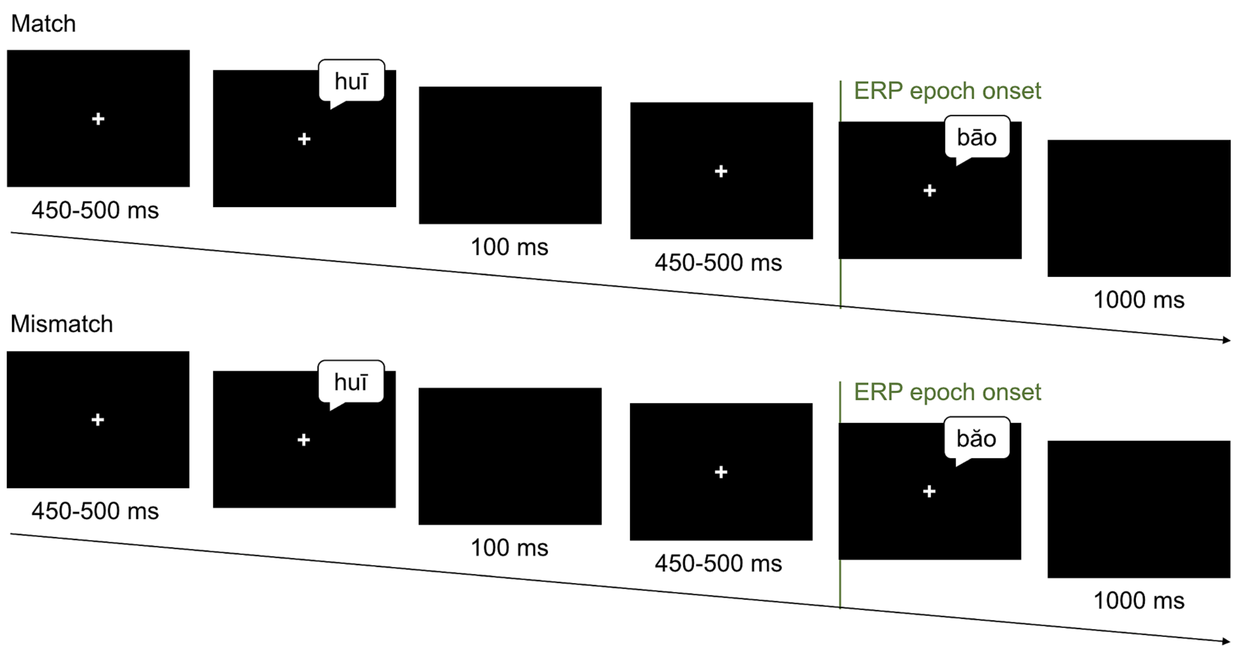

Fig. 3. Sample Matching and Mismatching Trials from Lexical Tone Discrimination Task.

presentation counterbalanced across participants), following Morett and Chang (2015). All six pairs of Mandarin words were presented in this way in random order across three blocks, such that each word was presented three times and a total of 36 trials were presented in the word learning task. The word learning task lasted approximately 7 min.

Following a brief break, participants completed either the lexical tone or meaning discrimination task (order counterbalanced across participants). In each trial of the lexical tone discrimination task (see Fig. 3 for sample trials), a fixation cross appeared for a duration jittered between 450 and 500 ms, followed by an aud***io recording of a learned Mandarin prime word accompanied by a fixation cross. Following a 100 ms interstimulus interval, this sequence was repeated for a learned Mandarin target word different in form with either the same or a different lexical tone as the prime word. Participants indicated whether the lexical tone of the target word matched or mismatched the lexical tone of the prime word by pressing one of two buttons with either their left or right index finger (counterbalanced across participants). No feedback on response correctness was provided. Once participants responded, a 1000 ms intertrial interval occurred, followed by the next trial. Half of the trials consisted of prime and target words with matching lexical tones, and half of the trials consisted of words with mismatching lexical tones, with trial types randomly interleaved. A total of 144 trials was presented, with each possible combination of words and lexical tones presented an equal number of times within each trial type (matching vs. mismatching). The lexical tone discrimination task lasted approximately 12 min.

In each trial of the meaning discrimination task (see Fig. 4 for sample trials), a fixation cross appeared for a duration jittered between 450 and 500 ms, followed by an aud***io recording of a learned Mandarin prime word accompanied by a fixation cross. Following a 100 ms interstimulus interval and a fixation cross jittered in the same manner, an English target word that was the translation of either the Mandarin prime word or the other Mandarin word in the pair differing minimally from it in lexical tone was presented as text. As in other similar studies examining the N400 (Pu et al., 2016), participants indicated whether the meaning of the target word matched or mismatched that of the prime word by pressing one of two buttons with either their left or right index finger (counterbalanced across participants). No feedback on response correctness was provided. Once participants responded, a 1000 ms intertrial interval occurred, followed by the next trial. Half of the trials consisted of prime and target words with matching meanings, and half of trials consisted of prime and target words with mismatching meanings, with trial types randomly interleaved. A total of 144 trials was presented, with each prime-target pair presented an equal number of times within each trial type (matching vs. mismatching meanings). The meaning discrimination task lasted approximately 12 min.

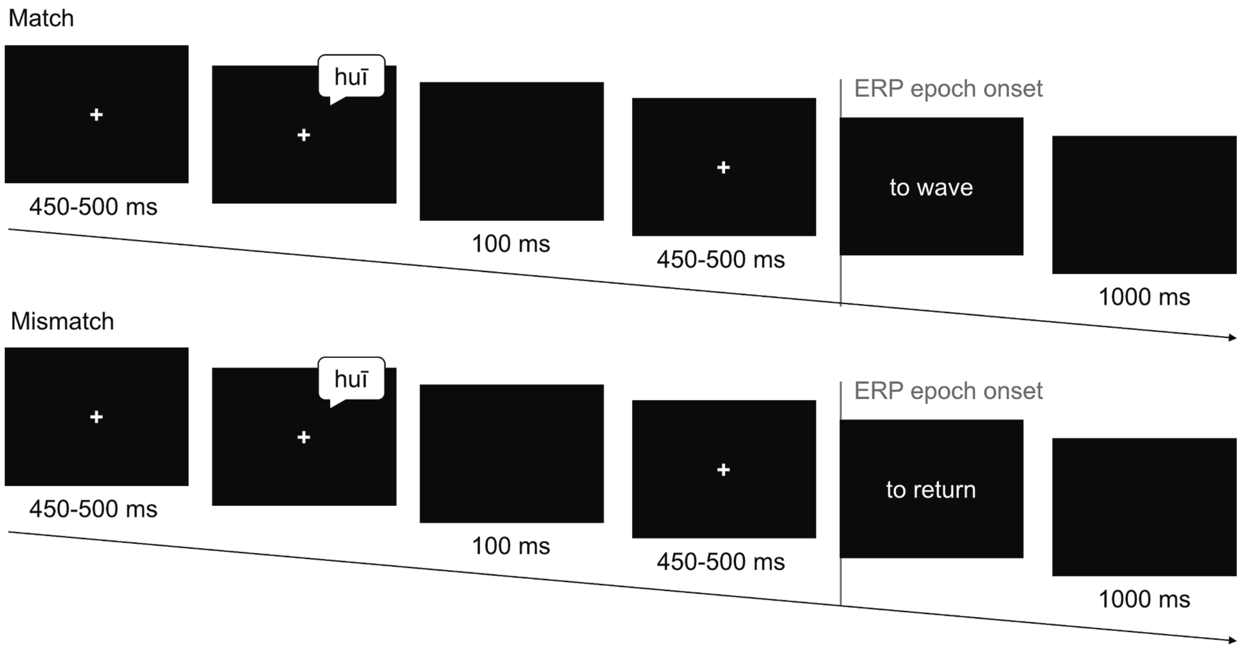

Fig. 4. Sample Matching and Mismatching Trials from Meaning Discrimination Task.

2.4. Behavioral response data analysis

De-identified data and analysis scripts are available via the following link: https://osf.io/nufkd/?view_only=34c5065fcf4c405da0a0e4cdf78e267b. Due to a glitch in the experimental script, behavioral response data was not collected during the meaning discrimination task; thus, only behavioral response data collected during the lexical tone discrimination task was collected and analyzed. Signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2004) was used to decompose responses into two conceptually and statistically distinct parameters: Discrimination or sensitivity (*d'*), which captures how well participants successfully discriminated prime-target pairs with mismatching lexical tones from prime-target pairs with matching lexical tones, and response criterion (*c*) or response bias, which captures the criterial level at which participants judged lexical tones as mismatching, regardless of whether they were actually mismatching.

To determine whether lexical tone discrimination accuracy differs by learning condition, response data were analyzed using a mixed effects probability unit (probit) model, which operates on trial-level data and accounts for participant- and item-level variability within the same model. Probit mixed effect models allow responses (1 = matching; 0 = mismatching) rather than *d'* values to be used as the dependent variable (DV), with measures of sensitivity expressed as *d'* values. In these models, lexical tone match between prime and target words (match vs. mismatch) and learning condition (no gesture vs. pitch gesture vs. semantic 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 lexical tone match represents overall discrimination performance (*d'*), with an alpha level < 0.05 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 lexical tone match represents the effect of learning condition on discrimination accuracy (*d'*), with an alpha level < 0.05 indicating that the effect of learning condition on response bias and discrimination performance exceeded chance.

To determine whether lexical tone discrimination latency differs by learning condition and lexical tone match, latency of correct responses in milliseconds was analyzed using a linear mixed effects model. Learning condition (no gesture vs. pitch gesture vs. semantic gesture) and lexical tone match (match vs. mismatch) were included as fixed effects using weighted mean centered (Helmert) contrast coding, and reaction time served as the outcome variable. For both the accuracy and latency models, random slopes were included with the maximal random effect structure permitted to achieve model convergence (Barr et al., 2013). For all effects of learning condition reaching significance, Tukey HSD post-hoc tests were conducted using the *emmeans* package to test for differences between levels.

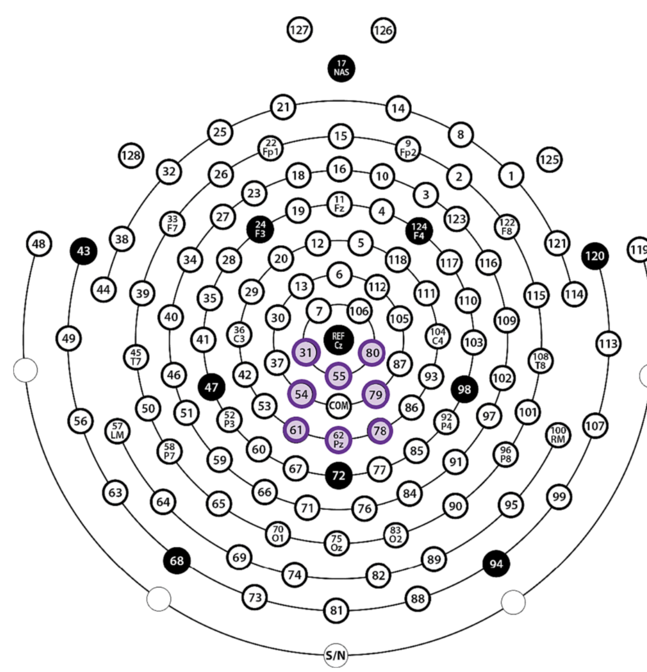

Fig. 5. 128 Channel Montage Used for EEG Recording with Channels in Central Posterior ROI Included in Analysis Highlighted (Purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Mean Accuracy and Latency by Learning Condition and Match for Lexical Tone Discrimination Task (Standard Deviations in Parentheses).

	Congruency	Pitch gesture	Semantic gesture	No gesture
Accuracy	Match	0.79 (0.41)0.75	0.78 (0.42)0.68	0.70 (0.46)0.67
	Mismatch	(0.43)	(0.47)	(0.47)
Latency (ms)	Match	1102.76 (348.98)1108.24	1177.74 (363.69)1193.85	1168.11 (375.98)1178.04
	Mismatch	(356.64)	(402.58)	(385.72)

Table 5
Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of Mean N400 Amplitudes for Lexical Tone Discrimination Task (Observations = 2296).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
InterceptLearning condition 1	0.39	1.66	0.24	0.81
(no gesture vs. pitch gesture + semantic gesture)Learning condition 2	−0.82	3.86	−0.21	0.83
(no gesture + pitch gesture vs. semantic gesture)				
	−0.39	3.29	−0.12	0.91
Match (matching vs. mismatching)	−3.93	1.23	−3.20	0.003^{**}
Learning condition 1 × Match	−14.40	1.89	−7.61	$< 0.001^{***}$
Learning condition 2 × Match	9.82	1.56	6.28	$< 0.001^{***}$

Random effect	s^2
Participant	9.53
Channel	3.39
Channel x Learning condition 1	1.13
Channel x Learning condition 2	4.31
Channel x Match	5.14

2.5. EEG recording and data analysis

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 1000 Hz with an anti-aliasing low-pass filter of 4000 Hz.

EEG data were pre-processed and analyzed offline using EEGLab (Delorme & Makeig, 2004) and ERPLab (Lopez-Calderon & Luck, 2014). 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 from −100 to 800 ms relative to target word onset, with −100 to 0 serving as a baseline. 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 match and averaged across participants for ERP analyses.

Following several other studies examining the N400 for gesture-speech integration (He et al., 2020; Holle & Gunter, 2007; Obermeier et al., 2011), the standard N400 time window ranging from 300 to 500 ms was used for statistical analysis. Mean amplitudes recorded during this time window for each condition were averaged across a central posterior “region of interest” (ROI) consisting of channels 31, 54, 55, 61, 62, 78, 79, and 80 based on inspection of scalp voltage topographies and previous research (Drijvers & Özyürek, 2018; He et al., 2020; Morett et al., 2020; see Fig. 5).

ERP data were analyzed using linear mixed effect models with learning condition and match as fixed factors,[1] participant and channel as random factors, and mean amplitude as the outcome variable for both the lexical tone discrimination task and meaning discrimination task. Prior to entry into these models, fixed effects were coded using weighted mean centered (Helmert) contrast coding in order of the levels mentioned above. Random slopes were included with the maximal random effect structure permitted to achieve model convergence. For all effects of learning condition reaching significance, Tukey HSD post-hoc tests were conducted using the *emmeans* package to test for differences between levels.

3. Results

3.1. Behavioral response accuracy and latency

Table 4 contains descriptive statistics for *d’* and latency (ms) from the lexical tone discrimination task by learning condition. Response bias did not differ significantly between the no gesture vs. pitch gesture and semantic gesture conditions ($B = 0.01$, $SE = 0.11$, $z = 0.05$, $p = .96$) or the no gesture and pitch gesture vs. semantic gesture conditions ($B = 0.08$, $SE = 0.10$, $z = 0.85$, $p = .40$). Although discrimination accuracy differed significantly between the no gesture vs. pitch gesture and semantic gesture conditions ($B = -0.59$, $SE = 0.09$, $z = -6.56$, $p < .001$), it did not differ significantly between the no gesture and pitch gesture vs. semantic gesture conditions ($B = -0.02$, $SE = 0.08$, $z = -0.24$, $p = .81$). Tukey-corrected post-hoc tests failed to reveal significant differences in discrimination accuracy between any of the learning conditions. Although response latency was significantly lower for words with matching than mismatching lexical tones ($B = 21.31$, $SE = 10.47$, $z = 2.04$, $p = .04$), it did not differ significantly between the no gesture vs. pitch gesture and semantic gesture conditions ($B = -50.59$, $SE = 55.95$, $z = -0.90$, $p = .37$) or the no gesture and pitch gesture vs. semantic gesture

[1] To determine whether the effects of learning condition and match varied by task order, task order was included as a fixed factor in initial models. No significant main effects or interactions of task order with learning condition or match were observed, so task order was not included as a fixed effect in the final models.

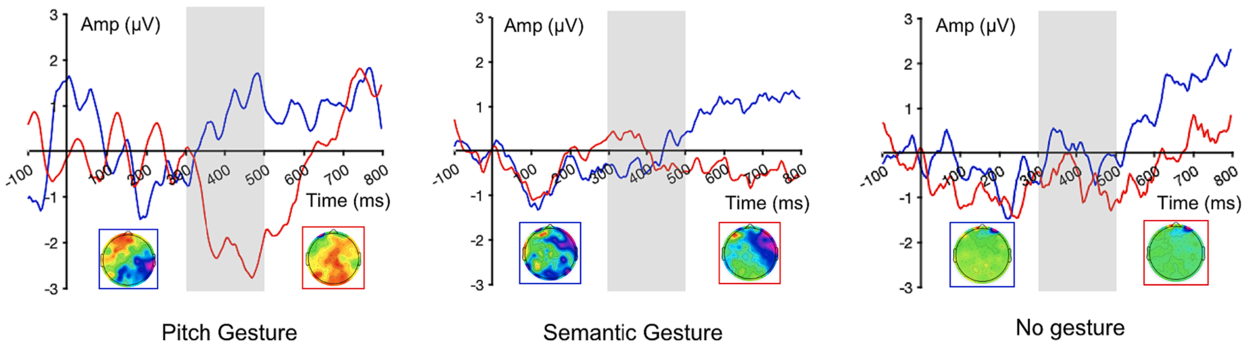

Fig. 6. Central-Posterior ERP Waveforms and Topographic Plots for Matching (Blue) and Mismatching (Red) Word Pairs Learned with Pitch Gesture, Semantic Gesture, and No Gesture in Lexical Tone Discrimination Task (Gray Shading Indicates N400 Time Window). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conditions (B = 45.78, SE = 48.24, z = 0.95, p =.35). Additionally, no significant interactions were observed between match and the no gesture vs. pitch gesture and semantic gesture conditions (B = 0.85, SE = 25.42, z = 0.03, p =.97) or the no gesture and pitch gesture vs. semantic gesture conditions (B = 25.41, SE = 22.10, z = 1.15, p =.25).

3.2. N400: Lexical tone discrimination task

To ensure that any differences by learning condition and match observed in the N400 time window were not driven by baseline differences, it was assessed whether data in the baseline window differed by condition. No significant main effects or interactions were found, indicating that baseline amplitudes did not differ by condition. Additionally, to ensure that any differences by learning condition were not driven by outliers, it was assessed whether participants' mean amplitudes in the N400 time window were greater or less than 3 standard deviations from the grand mean, and data from one participant in the no gesture condition was excluded on this basis.

As can be seen from Table 5, which contains parameter estimates for the model of N400 amplitudes for the lexical tone discrimination task, there was a significant main effect of match as well as significant interactions between match and both learning condition contrasts (no gesture vs. pitch gesture and semantic gesture; no gesture and pitch gesture vs. semantic gesture). Tukey HSD-adjusted planned comparisons indicated that N400 amplitudes differed significantly for Mandarin target words learned with pitch gestures with lexical tones mismatching than matching the lexical tones of Mandarin prime words (B = 14.72, SE = 1.65, z = 8.90, p <.001), whereas N400 amplitudes did not differ significantly for target words learned with semantic gestures (B = -2.30, SE = 1.58, z = -1.46, p =.15) or no gesture (B = 0.32, SE = 1.65, z = 0.19, p =.85; see Fig. 6) with lexical tones mismatching vs. matching the lexical tones of prime words.

3.3. N400: Meaning discrimination task

To ensure that any differences by learning condition and match observed in the N400 time window were not driven by baseline differences, it was assessed whether mean amplitudes in the baseline window differed by condition. No significant main effects or interactions were found, indicating that baseline amplitudes did not differ by condition. Additionally, to ensure that any differences by learning condition were not driven by outliers, it was assessed whether participants' mean amplitudes in the N400 time window were greater or less than 3 standard deviations from the grand mean, and no data was excluded on this basis.

As can be seen from Table 6, which contains parameter estimates for the model of N400 amplitudes for the meaning discrimination task, there was a significant main effect of match as well as significant interactions between match and both learning condition contrasts (no gesture vs. pitch gesture and semantic gesture; no gesture and pitch gesture vs. semantic gesture). Tukey HSD-adjusted planned comparisons indicated that N400 amplitudes differed significantly for English target words with meanings mismatching and matching prime Mandarin words learned with pitch gestures (B = 21.25, SE = 1.78, z = 11.92, p <.001) and semantic gestures (B = 5.51, SE = 1.66, z = 3.32, p =.001), whereas N400 amplitudes did not differ significantly for English target words with meanings mismatching vs matching Mandarin prime words learned with no gesture (B = 1.18, SE = 1.72, z = 0.69, p =.49; see Fig. 7.).

3.4. Relationship between behavioral response accuracy and latency and N400 differences

To determine the extent to which behavioral and neural measures were related in the lexical tone discrimination task, Pearson correlations between d' and response latency and N400 differences between words with matching and mismatching lexical tones were computed for

Table 6
Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of Mean N400 Amplitudes for Meaning Discrimination Task (Observations = 2352).

Fixed effect	Coefficient	SE	Wald z	p
InterceptLearning condition 1	−1.62	0.98	−1.65	0.11
(no gesture vs. pitch gesture + semantic gesture)Learning condition 2	−0.92	2.06	−0.45	0.66
(no gesture + pitch gesture vs. semantic gesture)				
	1.01	1.73	0.59	0.56
Match (matching vs. mismatching)	−8.94	0.99	−9.02	< 0.001***
Learning condition 1 × Match	−20.07	2.48	−8.11	< 0.001***
Learning condition 2 × Match	5.71	2.07	2.76	0.006**

Random effect	s^2
Participant	4.29
Channel	2.83

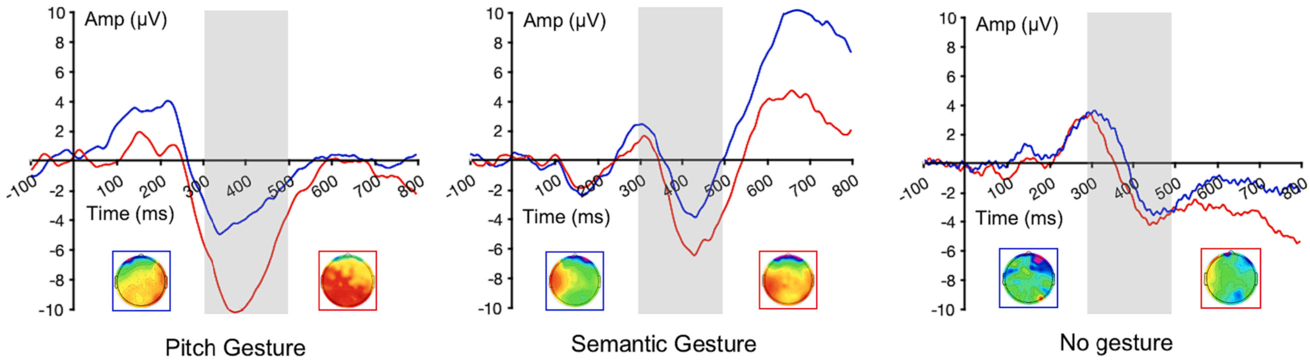

Fig. 7. Central-Posterior ERP Waveforms and Topographic Plots for Matching (Blue) and Mismatching (Red) Word Pairs Learned with Pitch Gesture, Semantic Gesture, and No Gesture in Meaning Discrimination Task (Gray Shading Indicates N400 Time Window). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Pearson Correlations Between d' and Latency and N400 Differences for Words with Matching and Mismatching Lexical Tones by Learning Condition in Lexical Tone Discrimination Task.

	Statistic	Pitch gesture	Semantic gesture	No gesture
d', N400	*r*	0.51	−0.44	−0.42
Latency (ms)	*p*	0.14	0.15	0.16
, N400	*r*	−0.68	0.09	0.23
	p	0.03*	0.79	0.45

participants in each learning condition. As can be seen from Table 7, the only correlation that reached statistical significance was between response latency and N400 differences in the pitch gesture condition. Thus, for words learned by observing pitch gestures, response latencies decreased as N400 differences increased.

4. Discussion

This study investigated how observing gestures conveying pitch contours and word meanings when learning L2 Mandarin words differing minimally in lexical tone affects subsequent phonological and semantic processing of these words by L1 English speakers. In particular, it examined whether such words learned with such gestures were associated with similar or varying N400 differences across two priming tasks: a meaning discrimination task, in which their meanings matched or mismatched those of English target words, and a lexical tone discrimination task, in which their lexical tones matched or mismatched those of Mandarin prime words differing in form. In the lexical tone discrimination task, a significantly larger N400 was observed for target words with lexical tones mismatching than matching those of prime words learned with pitch gestures, whereas the N400 did not differ significantly by match for target words learned with semantic or no gestures. In the meaning discrimination task, a significantly larger N400 was observed for English target words with meanings mismatching than matching the meanings of Mandarin prime words learned with pitch and semantic gestures, whereas the N400 did not differ significantly by match for words learned with no gestures. These findings reveal that observing gestures conveying phonological information differentiating between L2 words at learning enhances subsequent phonological and semantic processing of these L2 words. Moreover, they reveal that, by comparison, observing gestures conveying semantic information differentiating between L2 words at learning enhances subsequent semantic, but not phonological, processing of these L2 words.

The finding that observing pitch gestures at learning is associated with a significant N400 difference for matching and mismatching trials in the lexical tone discrimination task provides the first neural evidence that observing pitch gestures enhances L1 English speakers' differentiation between lexical tones. This N400 difference reflects sensitivity to phonological similarity between the lexical tones of learned Mandarin prime and target words via prediction, similar to the N400 difference elicited by rhyming and non-rhyming words (Coch et al., 2005; Noordenbos et al., 2013; Perrin & Garcıa-Larrea, 2003; Praamstra & Stegeman, 1993; Rugg, 1984; Rugg & Barrett, 1987). This focus on phonological similarity provides a potential explanation for why an N400 difference was observed in the lexical tone discrimination task for L1 English speakers briefly exposed to L2 lexical tone, whereas Pelzl and colleagues (2019, 2021), whose tasks probed the phonological-semantic interface, failed to observe an N400 difference in L1 English speakers who were advanced Mandarin L2 learners. Another potential explanation is that the Mandarin words tested in the current study are all monosyllabic, which may elicit greater N400 differences following brief exposure than multisyllabic Mandarin words (Pelzl et al., 2019). In contrast to the significant N400 difference observed for matching and mismatching Mandarin words learned with pitch gestures but not semantic or no gestures, response accuracy and latency did not differ significantly across learning conditions, contrary to previous studies (Baills et al., 2019; Morett et al., 2022; Morett & Chang, 2015; Zhen et al., 2019). Some possible reasons for this discrepancy in behavioral response data between the current study and previous studies may include the use of a discrimination rather than a categorization task, which may be more demanding and may therefore reflect more individual and group level variation in lexical tone perception, as well as analysis of data collected after learning only rather than both before and after learning, which may be more sensitive to changes in lexical tone perception due to training. Notably, the only behavioral response variable significantly correlated with N400 differences was latency for correct trials in the pitch gesture condition. This finding suggests that only the largest N400 differences may be evident in online behavioral measures such as response latency. Thus, in many cases, the N400 may provide a more sensitive index of semantic and phonological processing of newly learned L2 words than response accuracy and latency.

Likewise, the finding that observing pitch gestures at learning is associated with a significant N400 difference for matching and mismatching trials in the meaning discrimination task provides the first neural evidence that observing pitch gestures enhances L1 English speakers' differentiation between the meanings of Mandarin words differing in lexical tone. For Mandarin words differing minimally in lexical tone, distinguishing between lexical tones is necessary to distinguish between meanings, and observing pitch gestures conveying lexical tones facilitates lexical tone discrimination. Thus, the larger N400 observed for L1 English target words with meanings mismatching than matching L2 Mandarin prime words learned in the pitch gesture condition reflects greater semantic integration effort. This N400 difference is analogous to the larger N400 observed for L1 Mandarin target words with lexical tones mismatching than matching the meanings of pictorial primes and sentential contexts (Brown-Schmidt & Canseco-

Gonzalez, 2004; Malins & Joanisse, 2012). Because behavioral response data was unavailable for the meaning discrimination task in the current study, it is unclear whether it differed for words learned with pitch gesture and no gesture, as was the case for the N400. However, two previous studies (Baills et al., 2019; Morett & Chang, 2015) show that, relative to no gestures, observing pitch gestures when learning L2 Mandarin words differing minimally in lexical tone improves subsequent association of these words with their L1 English translations. Thus, the N400 difference observed for L1 English target words mismatching vs. matching L2 Mandarin prime words learned with pitch gesture in the current study may reflect explicit mapping of the phonological forms of L2 Mandarin words differing in lexical tone onto their L1 meanings, which may strengthen word-meaning associations.

The finding that observing semantic gestures is associated with a significant N400 difference for matching and mismatching trials in the meaning discrimination task provides the first neural evidence that observing gestures conveying word meanings enhances L1 English speakers' differentiation between the meanings of Mandarin words differing minimally in lexical tone. Because behavioral response data was unavailable for the meaning discrimination task in the current study, it is unclear whether it differed for words learned with semantic gesture and no gesture, as was the case for the N400. However, this N400 difference is inconsistent with behavioral evidence from previous studies showing a lack of enhancement of word-meaning association accuracy for phonologically similar L2 words learned with semantic gestures over no gesture in previous studies (Kelly & Lee, 2012; Morett & Chang, 2015). This discrepancy between the current neural and previous behavioral findings suggests that the N400 may reflect beneficial effects of observing semantic gestures conveying the meanings of Mandarin L2 words differing minimally in lexical tone on subsequent semantic processing of these words that are not evident in behavioral response data. Although only differences within the N400 time window were tested to address the research hypotheses, it is worth noting that there appears to be a difference beginning at around 600 ms for words learned with semantic gesture that does not appear for words learned with pitch gesture or no gesture in the meaning as well as the lexical tone discrimination task. This difference may be due to the late positive component (LPC), which represents recognition and is larger for L2 words learned with than without semantic gestures conveying their meanings (Kelly et al., 2009). To confirm that this is the case, future research can include unlearned words, which should elicit smaller LPCs than learned words, in the test phase.

Although this study provides insight into the N400 as a neural signature reflecting the impacts of pitch and semantic gestures on the learning and subsequent processing of Mandarin L2 words by L1 English speakers, it has some limitations. One limitation is the sample size, which constrained the analyses to detecting medium to large effects. A second limitation is the lack of response accuracy and latency data for the meaning discrimination task due to a technical glitch. In future work, it will be important to collect behavioral data whenever ERP data are collected to determine the extent to which they are consistent or differ across all tasks. A third limitation is the lack of a delayed follow-up test, which would reveal the robustness of the impacts of observing pitch and semantic gestures on semantic and phonological processing of learned L2 words over time. In light of previous behavioral work showing delayed effects of gesture on production but not perception of L2 phonological contrasts (Li et al., 2021, 2022), the sensitivity of the N400 may help to reveal delayed effects of gesture observation on phonological and semantic processing of L2 words. A fourth limitation is the relatively brief duration of the word learning phase. Although the duration of the word learning phase was based on previous work showing effects of gesture observation on L2 lexical tone categorization and word-meaning association by L1 English speakers (Morett et al., 2022; Morett & Chang, 2015), a longer word learning phase might produce more robust effects, particularly if a delayed follow-up test is implemented. A fifth limitation is the greater noise distributed throughout the ERP data collected during the lexical tone discrimination task relative to the meaning discrimination task, which may be due to their elicitation by speech stimuli, which take time to unfold unlike text stimuli. A final limitation is that, although this study helps to shed light on the mechanisms underlying the impact of observing gesture on phonological and semantic processing of newly learned L2 words, the extent to which the results generalize to naturalistic acquisition of L2 vocabulary by learners is unclear and should be investigated in future research.

In conclusion, this study provides the first neural evidence that observing gestures conveying the semantics and phonology of L2 words at learning influences subsequent semantic and phonological processing of these words. Thus, it supports multimodal and embodied theories of language acquisition and processing by demonstrating that information conveyed via gesture is integrated into representations of language, affecting its subsequent processing. Notably, it demonstrates that the N400 reflects the impacts of observed gestures even in cases in which they are not evident from behavioral response accuracy and latency. These findings have important implications for how gesture can be used to support acquisition of phonologically similar L2 words differing in meaning, providing evidence-based recommendations that learners and their instructors can leverage to enhance vocabulary acquisition in unfamiliar L2s.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A link to the stimuli, data, and analysis code on OSF is provided in the manuscript.

Acknowledgements

This work was supported by the US National Science Foundation [grant number 2140073] and a Level 1 Research Grant from the University of Alabama Research Grants Council. All stimuli, data, and analysis scripts are publicly available via the Open Science Framework at https://osf.io/nufkd/?view_only=34c5065fcf4c405da0a0e4cdf78e267b. The author thanks Jin Zhao and Wendy Chang for assistance with recording audio and video stimuli, respectively, and Maddy Armstrong and Carlee Brandon for assistance with data collection. In addition, the author thanks the reviewers and attendees of the 2022 and 2023 annual meetings of the Cognitive Science Society and the 2023 annual meeting of the American Association for Applied Linguistics for feedback on portions of this research.

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

Bergmann, K., & Macedonia, M. (2013). A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In R. Ayelett, B. Krenn, C. Pelachaud, & H. Shomodaira (Eds.), *Lecture Notes in Computer Science* (Vol. 8108, pp. 139–148). Springer. https://doi.org/10.1023/B:JOPR.0000017223.98667.10.

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology, 25*(7–8), 1114–1128. https://doi.org/10.1080/02643290801921707

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (6.0.39).* [Computer software].

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.

Carrasco-Ortiz, H., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2017). interactions in the neighborhood: Effects of orthographic and phonological neighbors on N400 amplitude. *Journal of Neurolinguistics, 41*, 1–10. https://doi.org/10.1016/j.jneuroling.2016.06.007

Casasanto, D., Phillips, W., & Boroditsky, L. (2003). Do we think about music in terms of space? Metaphoric representation of musical pitch. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128*(1), 456–465. https://doi.org/10.1121/1.3445785

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

Chen, C.-M. (2013). Gestures as tone markers in multilingual communication. *Research in Chinese as a Second Language, 9*, 143.

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. https://doi.org/10.1162/0898929052880020

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

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on cantonese word learning. *The Journal of the Acoustical Society of America, 131*(6), 4756–4769. https://doi.org/10.1121/1.4714355

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language, 177*, 7–17. https://doi.org/10.1016/j.bandl.2018.01.003

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

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

He, Y., Luell, S., Muralikrishnan, R., Straube, B., & Nagels, A. (2020). Gesture's body orientation modulates the N400 for visual sentences primed by gestures. *Human Brain Mapping, 41*(17), 4901–4911. https://doi.org/10.1002/hbm.25166

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. https://doi.org/10.1016/j.neuropsychologia.2021.107831

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

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. https://doi.org/10.1159/000259792

Hoetjes, M., & Van Maastricht, L. (2020). Using gesture to facilitate L2 phoneme acquisition: the importance of gesture and phoneme complexity. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.575032

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

Howie, J. M. (1974). On the domain of tone in mandarin. *Phonetica, 30*(3), 129–148. https://doi.org/10.1159/000259484

Jia, L., & Wang, J. (2013a). on the effects of visual processing on tone production by english-speaking learners of chinese. *TCSOL Studies, 52*(4), 30–34.

Jia, L., & Wang, J. Q. (2013b). The effects of visual processing on tone perception by native English-Speaking learners of Chinese. *chinese Teaching in the World, 27*, 548–557.

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and mandarin Chinese: An event-related potentials training study. *BMC Neuroscience, 9*(1), 53. https://doi.org/10.1186/1471-2202-9-53

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research, 1148*, 113–122. https://doi.org/10.1016/j.brainres.2007.02.019

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). https://doi.org/10.1525/collabra.76

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., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253–260. https://doi.org/10.1016/S0093-934X(03)00335-3

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

Lee, C.-Y. (2007). Does horse activate mother? processing lexical tone in form priming. *Language and Speech, 50*(1), 101–123.

Lee, R.-R.-W., Hsu, C.-H., Lin, S.-K., Wu, D. H., & Tzeng, O.-J.-L. (2017). learning transforms functional organization for mandarin lexical tone discrimination in the brain: Evidence from a MEG experiment on second language learning. *Journal of Neurolinguistics, 42*, 124–139. https://doi.org/10.1016/j.jneuroling.2016.12.004

Li, P., Baills, F., Baqué, L., & Prieto, P. (2022). The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research, 02676583221124075*. https://doi.org/10.1177/02676583221124075

Li, P., Xi, X., Baills, F., & Prieto, P. (2021). Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters. *Language, Cognition and Neuroscience, 36*(10), 1313–1328.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00213

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education, 5*, 196–211. https://doi.org/10.1111/j.1751-228X.2011.01129.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

I. Maddieson Tone. The World Atlas of Language Structures Online 2013.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in mandarin Chinese. *Neuropsychologia, 50*, 2032–2043. https://doi.org/10.1016/j.neuropsychologia.2012.05.002

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*(7), 703–704. https://doi.org/10.1038/nn1264

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*, Article 105014. https://doi.org/10.1016/j.cognition.2022.105014

Morett, L. M., Landi, N., Irwin, J., & McPartland, J. C. (2020). N400 amplitude, latency, and variability reflect temporal integration of beat gesture and pitch accent during language processing. *Brain Research, 1747*, Article 147059. https://doi.org/10.1016/j.brainres.2020.147059

Noordenbos, M. W., Segers, E., Wagensveld, B., & Verhoeven, L. (2013). Aberrant N400 responses to phonological overlap during rhyme judgements in children at risk for dyslexia. *Brain Research, 1537*, 233–243. https://doi.org/10.1016/j.brainres.2013.09.018

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture–speech integration: When synchrony is lost, memory can help. *Journal of Cognitive Neuroscience, 23*(7), 1648–1663. https://doi.org/10.1162/jocn.2010.21498

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. https://doi.org/10.1162/jocn.2007.19.4.605

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

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. https://doi.org/10.1017/S027226312000039X

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. https://doi.org/10.1017/S0272263117000444

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461–472. https://doi.org/10.1121/1.3593366

Perrin, F., & García-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research, 17*(1), 36–47. https://doi.org/10.1016/S0926-6410(03)00078-8

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

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research, 1*(2), 73–86. https://doi.org/10.1016/0926-6410(93)90013-U

Pu, H., Holcomb, P. J., & Midgley, K. J. (2016). Neural changes underlying early stages of L2 vocabulary acquisition. *Journal of Neurolinguistics, 40*, 55–65. https://doi.org/10.1016/j.jneuroling.2016.05.002

Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language, 23*(2), 225–240. https://doi.org/10.1016/0093-934X(84)90065-8

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. https://doi.org/10.1016/0093-934X(87)90132-5

Shuai, L., Li, B., & Gong, T. (2012). *2012*. Speech Prosody: Priming effects of tones and segments in lexical processing in mandarin.

Sweller, N., Shinooka-Phelan, A., & Austin, E. (2020). The effects of observing and producing gestures on japanese word learning. *Acta Psychologica, 207*, Article 103079. https://doi.org/10.1016/j.actpsy.2020.103079

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition, 17*(2), 277–293.

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., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). the role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics, 25* (03), 449–466. https://doi.org/10.1017/S0142716404001213

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of mandarin tones by chinese and american listeners. *Brain and Language, 78*, 332–348. https://doi.org/10.1006/brln.2001.2474

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*, 1019–1027. https://doi.org/10.1162/089892903770007407

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

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in mandarin chinese. *Language and Speech, 59*(1), 59–82.

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

Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping, 28*, 995–1006. https://doi.org/10.1002/hbm.20330

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654–667. https://doi.org/10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*(3), 234–245. https://doi.org/10.1016/j.bandl.2006.12.003

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

Yates, M., Shelley-Tremblay, J., & Knapp, D. L. (2020). Measuring the influence of phonological neighborhood on visual word recognition with the N400: Evidence for semantic scaffolding. *Brain and Language, 211*, Article 104866. https://doi.org/10.1016/j.bandl.2020.104866

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