Phonetic and affective roles of co-speech gesture in L2 pronunciation

Spencer D. Kelly,[1] Paige Avila,[1] Madeline Chernavsky,[1] Bailey Cooper,[2] Elizabeth Velázquez Fernández,[3] and Yukari Hirata[1]

[1] Colgate University | [2] Georgetown University | [3] The University of Puerto Rico

Second-language (L2) instructors often use hand gestures to teach pronunciation, yet empirical benefits vary in size and scope. Focusing on L2 Mandarin tone production, we explored whether physically exaggerating and emotionally emphasizing tone gestures enhanced pronunciation. Participants imitated videos of a native speaker producing tones in three conditions: Speech Alone (S), Speech + Gesture (SG), and Speech + Gesture + Enthusiasm (SGE). In S, the native speaker spoke tones without arm movements or enthusiastic facial expressions. In SG and SGE, exaggerated hand gestures followed tone contours, and in SGE, the speaker also produced enthusiastic facial expressions. While gesture and enthusiastic expressions yielded only modest pronunciation benefits—slightly higher F0 for Tone 1 and greater Tone 3 lengthening toward native values—self-ratings of motivation, enjoyment, preference, and helpfulness were substantially higher for SG and SGE than S, suggesting that gesture and enthusiastic expressions in L2 may influence affective experience more than correct pronunciation.

Keywords: second language (L2), tone production, pronunciation, hand gesture, emotion, phonetics

Phonetic and affective roles of co-speech gesture in L2 pronunciation

Achieving accurate pronunciation is one of the most challenging parts of second language (L2) learning. This is particularly difficult when non-tonal language speakers are learning tonal languages such as Mandarin (Huang, 2000; Hu, 2010). Mandarin has four major tones—Tone 1 high-level, Tone 2 rising, Tone 3 falling then rising, and Tone 4 falling (Jongman et al., 2006)—and they are phonemic:

https://doi.org/10.1075/jslp.25067.kel | Published online: 23 April 2026
Journal of Second Language Pronunciation ISSN 2215-1931 | E-ISSN 2215-194X

the pitch contour of each syllable distinguishes word meanings (Elliott, 1991). Non-tonal language speakers struggle with these tones for many reasons, including non-tonal languages' use of pitch mainly for pragmatic or syntactic purposes; perceptual bias toward L1 phonetic categories (Lin, 1985); context-dependent tonal change (Xu, 1997); and the communicative stakes of errors (e.g., "wěi dà" = great vs. "wèi dà" = big stomach). To teach proper perception and production, language instructors commonly use hand gestures to visualize tone contours and ask students to imitate them while speaking. Yet evidence for pronunciation benefits of gesture is mixed (Li et al., 2021; Xi et al., 2020; Zheng et al., 2018). The present study tests whether exaggerated gestures, combined with enthusiastic facial expressions, improve tone pronunciation and affective experience.

The body as a teaching tool

The body is an accessible and valuable resource for supporting perception and production of novel L2 speech sounds (Hardison, 2025). Although most research focuses on L2 perception, there is a growing number of studies investigating how observing or producing gesture affects L2 speech production (Amand & Touhami, 2016; Gluhareva & Prieto, 2017; Hoetjes & van Maastricht, 2020; Iizuka et al., 2020; Li et al., 2020; Li et al., 2021; Li et al., 2023; Tseng et al., 2025; Xi et al., 2020; Zhang et al., 2020; Zheng et al., 2018). The results reveal benefits across a variety of gesture types.

For example, focusing on beat gestures, Gluhareva and Prieto (2017) found that native Catalan speakers improved their non-native accent in English, for hard but not easy items, after observing rhythmic beats emphasizing particular stressed L2 syllables. Furthermore, Inceoglu and Ueda (2024) found that when novice learners of French viewed metaphoric gestures linking two words (e.g., connecting "les" and "amis" with a small sweeping motion to indicate that the transition should be pronounced as a /z/ sound), learners produced the linking consonant more correctly compared to observing no gesture. Moving to prosodic gestures, Li and colleagues (2023) had native Catalan speakers watch a French instructor produce (or not) gestures capturing the suprasegmental pitch contours of sentences. Results showed that although gesture instruction did not improve comprehensibility and fluency of L2 speech, it did improve accentedness ratings and acoustic measures of vowel pronunciation. Finally, studies using handclapping as a method of pronunciation training have yielded mixed results. Iizuka et al. (2020) found no benefits for native English speakers producing L2 Japanese segmentals (long vowels, geminates, and moraic nasals), whereas Zhang et al. (2020) reported that Chinese speakers improved vowel length, but not accentedness ratings, when repeating French words while handclapping compared with repetition alone.

Other studies have used iconic gestures to directly represent how speakers should correctly produce mouth and tongue movements during L2 pronunciation (Hoetjes & van Maastricht, 2020; Li et al., 2021). For example, Hoetjes and van Maastricht (2020) found that when L2 Spanish speakers observed an instructor making a round gesture with the thumb and forefinger (like an OK sign) to represent rounding of the lips, learners were better at articulating the /u/ phoneme in Spanish. However, gestures were ineffective when the teacher uttered the much more challenging Spanish /θ/ phoneme using a different gesture: pushing forward one hand with four fingers and thumb pressed together to roughly represent the way the tongue pushes against the teeth. This suggests that gestures have variable effects on L2 pronunciation depending on gesture type and speech difficulty.

Focusing specifically on the L2 production of Mandarin, the benefits of gesture are evident but also variable (Li et al., 2021; Xi et al., 2020; Zheng et al., 2018). For example, Xi et al. (2020) trained native Catalan speakers on Mandarin aspirated plosives (e.g., /pʰ/, /tʰ/, /kʰ/) with and without a corresponding "hand burst gesture" that visually represented the sudden release of air in these sounds. They found that observing gestures during training led to small but statistically significant improvements in the pronunciation of plosives, but only when the gestures accompanied aspirated speech. Moreover, a follow-up study found that when learners were asked to produce the gestures themselves during training, the quality of the gesture mattered: producing gestures *accurately* was necessary for pronouncing the speech in the most native-like fashion (Li et al., 2021).

With regard to tone production in Mandarin, Zheng et al. (2018) found that the effects of producing hand gestures were inconsistent across tones. In the study, native English speakers (with no background in Mandarin) were asked to imitate the speech and gesture of a native Mandarin speaker. When participants produced metaphoric gestures tracing the "shape" of the four tones, they produced more native-like pitch patterns than no gesture (measured by F0) only for Tone 4: the falling gesture slightly dragged the end of the tone downwards towards native levels. In light of Li et al. (2021), one explanation for why there were not more robust effects is that participants may not have gestured in the most effective way. For example, participants in the Zheng et al. (2018) study were sitting, and it is known that gesturing while sitting affects acoustic properties less than gesturing while standing (Pouw et al., 2020). Moreover, the gestures were small and constrained: participants gestured with the right index finger in a space that roughly corresponded to the size of the computer screen where they viewed the native model. This is noteworthy because smaller gestures made with the wrist and hand influence acoustic properties of speech to a lesser degree than larger gestures using the upper body and entire arm (Pouw et al., 2021). In short, gesture production can enhance Mandarin tone pronunciation, but there is room for

improvement. The present study asks whether more robust approaches can boost performance to a greater extent.

Affective factors in L2 learning

Affective engagement is a powerful force in L2 instruction. Naturalistic studies of L2 classrooms increasingly highlight the prevalence and benefits of positive emotions—such as motivation, rapport, enjoyment, and engagement—in language learning contexts (e.g., Dewaele & MacIntyre, 2014; Dörnyei & Ushioda, 2021; Henry & Thorsen, 2018; Krashen, 1982; MacIntyre & Gregersen, 2012). While many affective factors lie outside the control of L2 instructors, one readily accessible tool is how they affectively use their bodies during instruction. In a broad educational context, teachers show enthusiasm and engagement through a range of expressive behaviors: vocal animation, encouraging facial expressions such as smiling, wide-opened eyes, eye contact, demonstrative gestures, and energetic body movements (Collins, 1977; Keller et al., 2016; McCroskey et al., 1995). These bodily actions make a real difference in learner outcomes. In a large-scale review of 120 studies, Keller et al. (2016) found positive correlations between an instructor's bodily enthusiasm and student achievement.

This broader educational pattern is also evident in language learning, where L2 instructors show enthusiasm and engagement through bodily expressions, with similar benefits (Allen, 2000; Dewaele & Li, 2021; Liu, 2021; Megawati & Hartono, 2020; Sime, 2006; Smotrova, 2017; Yuan, 2024). For example, when asked to reflect on their English teacher's effectiveness, Chinese learners of English (EFL) reported a strong association between perceived teacher enthusiasm (through speech and bodily expression) and their own emotional engagement (Dewaele & Li, 2021). With specific regard to L2 pronunciation, which provokes particularly high anxiety in language students (Baran-Łucarz, 2014), Smotrova (2017) observed that instructors who showed engagement and support through their bodily actions created a positive classroom atmosphere and made learning more enjoyable. In the same vein, Megawati and Hartono (2020) suggest that teachers' positive facial expressions, such as smiling, are particularly effective in motivating L2 students.

Moving into the laboratory, only a handful of controlled experiments have attempted to systematically test the role of bodily actions on affective experiences in L2 speakers and learners (Algana & Hardison, 2024; Kamiya, 2025; Hirata et al., 2024; Sueyoshi & Hardison, 2005; Tseng et al., 2025; Zheng et al., 2018). These experiments tested a variety of second languages, including English, Japanese, and Mandarin, and employed diverse dependent measures, such as comprehension questions, discrimination tasks, and pitch production analyses. The

results showed that L2 speakers almost always preferred and/or enjoyed L2 exposure accompanied by hand gestures. Interestingly, despite this almost uniformly positive affective reaction to hand gestures, only half of the studies showed that gestures actually improved L2 perception or production compared to a speech-only baseline: Sueyoshi and Hardison (2005) showed that gestures and facial movements boosted L2 comprehension; Tseng et al. (2025) found that six days of training English with beat gestures improved Mandarin speakers' production of English intonation, rhythm, and stress; and Zheng et al. (2018) showed that gestures helped with the non-native speakers' proper pronunciation of Tone 4 in Mandarin. The other studies in this body of literature showed negligible effects on L2 speech tasks.

The present study

The present study builds on previous work exploring gesture's role in L2 pronunciation in three new ways. First, it uses more robust gestures than previous studies, such as in Zheng et al. (2018) where small, tightly constrained gestures were produced with the index finger while seated. We know from past research that larger movements, involving the whole arm and shoulder, produced while standing have the most robust impacts on speech production (Pouw et al., 2020; Pouw et al., 2021). For this reason, we had participants in the current study observe videos of a model standing and gesturing with maximal arm and hand extensions. Participants, also standing, were asked to imitate those actions as faithfully as possible. By using maximally robust body movements, we wanted to give gestures the best chance of influencing Mandarin tone production. Additionally, to ensure gesture accuracy, a determining factor of how much gesture affects speech production (Li et al., 2021; Xi et al., 2020), the experimenters corrected participants who were incorrectly imitating the native model.

Secondly, given the beneficial role of enthusiastic bodily expressions in L2 instruction, we combined gestures with positive enthusiastic expressions to mutually strengthen their beneficial effects in an L2 context (Del Rio & Alvarez, 2002; Sime, 2006; Smotrova, 2017). This is important because, traditionally, research on co-speech gestures has isolated cognitive functions from emotional functions, and recent theoretical work encourages both to be studied together (Aslan et al., 2024; Kelly & Tran, 2023).

Thirdly, we investigated tone production of L2 Mandarin learners with varying degrees of study experience. Previous research tested participants with no Mandarin experience, and the effects of gestures were modest and isolated to Tone 4 (Zheng et al., 2018). The authors raised the possibility of a floor effect because tone production may have been too challenging for novices. The present

study explored the next step by using the L2 speakers who have prior experience in studying Mandarin.

Participants watched and imitated videos of a native speaker producing Mandarin tones in three conditions: Speech Alone (S), Speech + Gesture (SG), and Speech + Gesture + Enthusiasm (SGE), in which enthusiasm was expressed through the model smiling. There were two sets of dependent measures: acoustic measures (F0 and duration) and affective self-assessments (motivation, helpfulness, enjoyment, and preference). Note that the present study did not involve training with a pre-test and post-test design; the study simply measured how accurately each participant produced tones and how affective self-assessments differed across these three conditions. Thus, any mention of effects of gesture or the native model's enthusiasm refers to conditional differences within each participant, not as the result of training, but in the act of imitating the model.

For the acoustic measures, we predicted that imitating speech and gesture (SG) would produce more native-like pitch patterns than imitating only speech (S), and these effects would be more robust and widespread than in Zheng et al. (2018). This was based on two lines of research. First, biomechanical studies have shown that larger and more exaggerated gestures affect speech production the most (Pouw et al., 2020; Pouw et al., 2021). And second, studies have demonstrated that hand gestures have the biggest influence on speech production when they are produced accurately (Li et al., 2021; Xi et al., 2020). Finally, based on research suggesting that positive facial expressions and hand gestures co-occur in L2 contexts and are mutually enhancing (Keller et al., 2016; Smotrova, 2017), we expected the SGE condition to generate more accurate tonal pronunciation compared to SG and S.

For the affective self-assessments, we predicted that the SG and SGE conditions would both produce higher self-evaluations of enjoyment, motivation, helpfulness, and preference than the S condition (Algana & Hardison, 2024; Hirata et al., 2024; Kamiya, 2025; Sueyoshi & Hardison, 2005; Zheng et al., 2018). In addition, because enthusiastic facial expressions add a layer of positive affect, we expected the SGE condition to produce even higher affective ratings than the SG condition.

Method

Participants

A sample of eighteen right-handed, female college undergraduates were included in the study. Sample size was determined through G*Power with the following parameters: power of 0.95, an effect size of 0.25 and a p-value less than 0.05.

Data from 23 participants were collected; five were excluded (three for not following instructions, two for poor acoustic quality). Participants were all female-identifying to match the naturally higher pitch in females and the pitch of the native model, making tonal differences and acoustic comparisons clearer. Pre-experiment surveys were administered to collect information on participants' self-reported language background and handedness. Right-hand dominance was measured using items from the Edinburgh Handedness Inventory (EHI) (Oldfield, 1971).

We recruited monolingual native English speakers who were L2 Mandarin learners at varying levels from the Chinese language program at Colgate University. Building on Zheng et al. (2018), which showed limited effects of gesture in native English speakers who had never studied Mandarin, the present study examined whether having some experience studying Mandarin yields different results. Participants had formally studied Mandarin for at least one semester (56 hours of formal class time) in college and reported no more than 11 total years of Mandarin study (mean: 5.1 years; range 0.5–11 years).[1] Participants were recruited from Chinese language courses in a four-year language program at our university with 5, 3, 5, and 5 students from the 100-, 200-, 300-, and 400- level courses, respectively (each level totals 112 hours of class time per year). The Chinese language instructors of the program reported that these courses at four levels, when finished, roughly correspond to levels 1, 2, 3, and 4, respectively, of the standardized proficiency test used widely in the U.S. A. (HSK or Hànyǔ Shuǐpíng Kǎoshì Proficiency Test). Sixteen out of 18 participants self-reported experience in studying before college (mean of 3.1 years; range of 0–9 years). The instructors reported that students with approximately 1–4 years of pre-college Mandarin study are placed into the 100-level, those with 5–7 years into the 200- or 300-levels, and those with 8 or more years into the 400-level.

Materials

Tonal stimuli

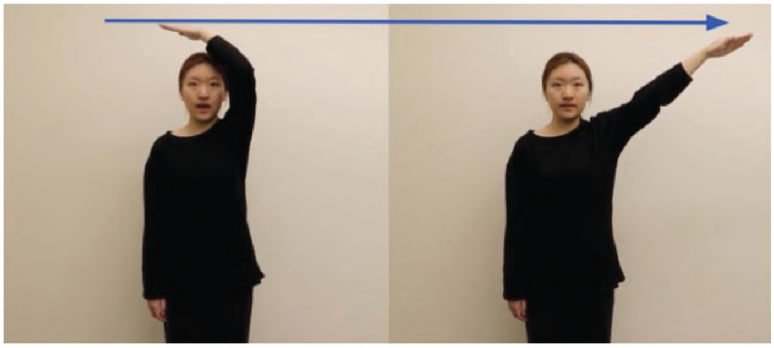
A native female speaker of standard Mandarin (from Guangzhou, China) produced syllables "ma," "mi," and "mu" across four tones. We chose the vowels "a," "i," and "u" for simplicity and cross-linguistic familiarity (Zheng et al., 2018).

---

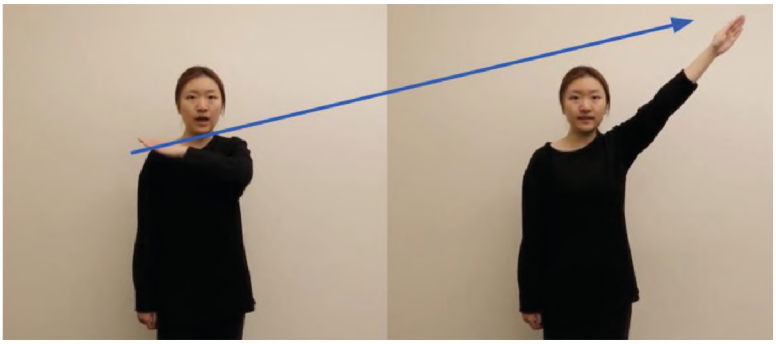
1. We acknowledge that this range is rather broad for L2 learning studies, and the only reason for not selecting a more targeted proficiency was the small participant pool in a rural liberal arts college. Our research goal was to build on results with totally novice learners as in Zheng et al. (2018) and examine whether gesture has a stronger influence on learners with prior experience of Mandarin study. We discuss this limitation in the final discussion section.

Video and audio recording for stimulus sets

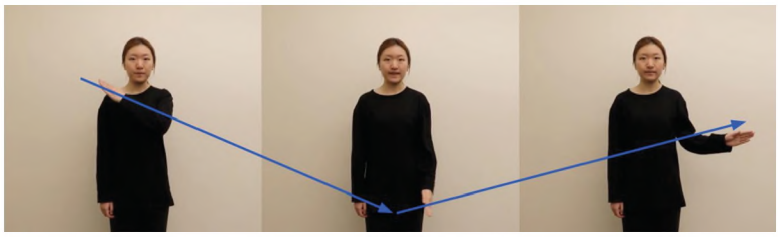
The model produced the four Mandarin tones standing to allow large gestures, which was done to maximize the acoustic effects of gesture during imitation (Pouw et al., 2020; Pouw et al., 2021). The model was framed in the videos to allow participants to see her body from the upper leg to about half a foot above her head. Each of the 12 tokens (4 tones x 3 syllables) lasted about one second. Because producing gestures can affect speech acoustics (Krahmer & Swerts, 2007), the audio was recorded separately without gestures or enthusiastic facial expressions and dubbed afterward to ensure the sound was identical across all three conditions. The dubbed speech was produced as neutrally as possible without gestures and enthusiastic facial expressions. Figure 1 shows the acoustic contours of the four Mandarin tones as visually represented by the metaphoric hand gestures.



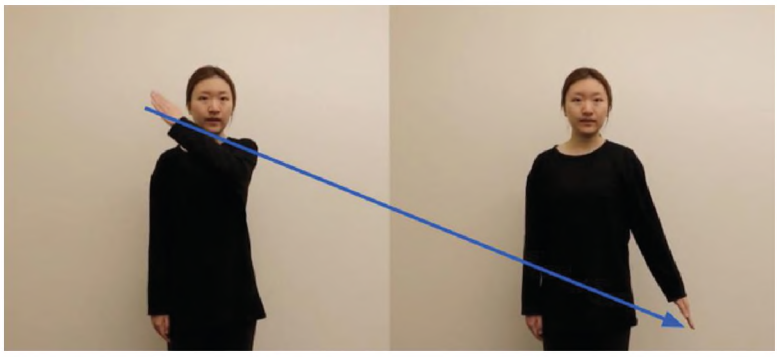
a.



b.



c.



d.

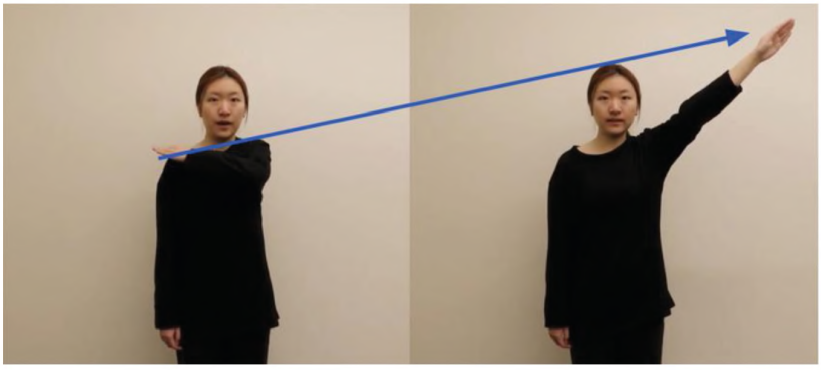
Figure 1. Hand gestures used for the four Mandarin tones

Note. SG gestures for "mi:" (a) Tone 1: horizontal sweep above head, left→right; (b) Tone 2: rising diagonal from left shoulder rightward to above the head; (c) Tone 3: V-shape (left shoulder down to stomach, up to chest); (d) Tone 4: falling diagonal from left of head to right leg. Images cropped and mirrored for left-to-right clarity.

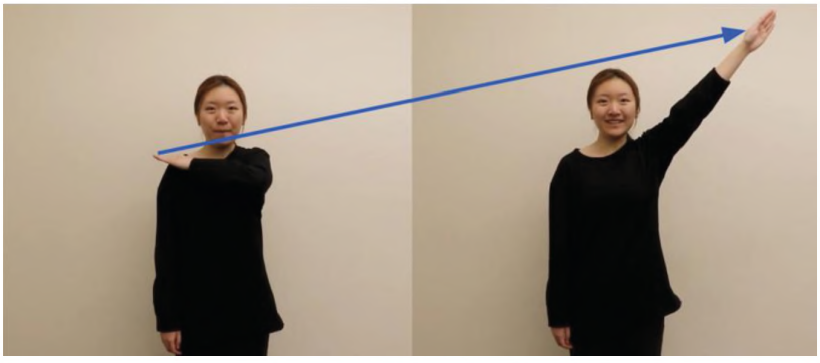
We recorded the three syllables across four tones for each experimental condition: Speech Alone (S), Speech + Gesture (SG), and Speech + Gesture + Enthusiasm (SGE). In the S condition, the model only spoke and did not move her arms or make any enthusiastic facial expressions (Figure 2a). In the SG condition, the model spoke as she produced a gesture that visually mapped onto the tone the model produced, and there were no enthusiastic facial expressions present (Figure 2b). In the SGE condition, the model performed the gestures, but also made facial expressions, including smiling and widening the eyes (Figure 2c). These facial expressions were modelled on enthusiastic expressions that have been shown to motivate students in actual classrooms (Keller et al., 2016).

Construction of stimulus sets. Each tonal pair consisted of two different tones using the same syllable (Appendix). For example, a tonal pair would be the syllable "mi" produced with Tone 4 followed by Tone 1. To avoid "tone sandhi" effects, we strategically chose tonal pairs whose tonal patterns remain consistent across

a.



b.



c.

Figure 2. Comparison of S, SG, and SGE conditions

Note: Tone 2 ("ma") across conditions: (a) S: no gesture, neutral face; (b) SG: speech + gesture, neutral face; (c) SGE: speech + gesture + enthusiastic facial expressions (eyes widening, broad smile).

contexts. We also used tone combinations that reflected tone pairings in real Mandarin words. Each of the four tones was presented 12 times across the pairs in different combinations, resulting in 24 unique tonal pairs per condition.

After creating the tonal pairs, we randomized their order and made additional adjustments to ensure that no consecutive pairs used the same tonal pattern. The same randomized order was used across all three conditions to prevent order effects from becoming a confound (see Appendix for the full sequence). On average, each stimulus set for a condition lasted approximately 5 minutes, including a 5-second pause between each tonal pair.

Post-experiment evaluation

At the end of the experiment, participants completed a self-assessment questionnaire. The questionnaire asked about four affective and subjective dimensions: motivation, helpfulness of gestures, preference, and enjoyment. The "motivation" item asked participants which condition most increased their desire to continue learning Mandarin. "Helpfulness" asked participants to rate the perceived utility of the gestures on a Likert scale (1 = least, 5 = most helpful) and to identify which condition was most intuitive to perform. "Preference" asked participants which condition they would choose to see in the classroom and which condition they would choose to use to teach Mandarin tones to a beginner. "Enjoyment" asked participants to rank how pleasant each of the three conditions was to perform.

Procedure

Instructions

To determine eligibility, participants completed an online self-report language background questionnaire to ensure that they had some experience studying and speaking Mandarin. Participants who qualified were brought to the lab, completed consent forms, and were given a general introduction to the experiment, described as involving gestures and the Mandarin language. After entering the testing room, the experimenter demonstrated the different gestures accompanying each tone and encouraged the participants to practice them. At this time, participants imitated the experimenter, who corrected any mistakes made in the production of the gestures. The experimenter then explained that a model would appear on the screen who would produce different Mandarin tones in monosyllabic pairs, and after a 5-second pause (with a black screen), participants should imitate exactly what they heard in speech and saw in gesture (participants were told to mirror the gestures with their right hands). No instruction was given on the enthusiastic facial expressions because we were most interested in how an

instructor's enthusiasm affects tone production in a speaker. Participants were informed that there would be three stimulus sets, each preceded by a practice video, with a one-minute break in between sets. Participants were then given a chance to ask any clarification questions regarding the experiment.

Following the introduction, participants put on a Logitech Wireless Headset H600, which allowed participants to stand and produce gestures across a wide physical space. After the headset was placed correctly, the first stimulus set (with practice) was played.

Practice trials

The practice video was approximately 40 seconds long and followed the same structure as the experimental videos but only contained three tonal pairs. In total, there were three practice trials, one right before each of the three stimulus sets. The purpose was to familiarize the participants with the condition they would be seeing (S, SG, or SGE) and to make sure they understood the instructions clearly. It also allowed the experimenter to give feedback on the participants' gestures, given previous research showing that producing accurate gestures has the most positive benefits for L2 pronunciation (Li et al., 2021; Xi et al., 2020).

Experimental trials

The stimulus sets began with a countdown ("3-2-1") followed by the first tonal pair. As mentioned above, a black screen followed the pair where participants had to replicate exactly what they saw and heard. This pattern repeated for all of the 24 tonal pairs. At the end of the set, the participant rested for about one minute before the next set began. The other two stimulus sets followed the same pattern as outlined above. The order of the conditions was counterbalanced across participants to offset fatigue effects, resulting in six stimulus set orders.

Post-experiment evaluation

Once the participants finished the three stimulus sets, they filled out the post-experiment survey. In total, the experiment took approximately one hour to complete. At the end, participants were compensated $20 and debriefed on the nature of the study.

Design, labeling, and analysis

We reported two acoustic measures—F0 and duration—and a battery of affective assessment measures. There was also a third acoustic measure, intensity, but it showed no significant effects in facilitating native-like tone pronunciation, so for brevity, these results are not reported.

F0

F0 was analyzed using Phonetic and Acoustic Analysis Toolkit (PRAAT) to determine if there were effects of the three conditions (S, SG, and SGE) on the participants' F0 contours. We divided each vowel portion of the syllable into quarters: 0%, 25%, 50%, 75%, and 100% (based on Wang et al., 2003), taking a data point, in Hz, at each quartile (Figure 3). Thus, each vowel had five data points, totaling 10 data points per tonal pair (5 recordings x 2 tones). For every participant, we recorded 720 data points (10 data points per pair x 24 pairs x 3 conditions). The 0% time point, i.e., the beginning of the vowel, was determined at a point where the vowel formants began (and where the nasal murmur ended for /m/ before the vowel). The 100% time point, i.e., the end of the vowel, was determined at a point where the three vowel formants ceased.

Four investigators labeled the spectrograms for all 18 participants, with each investigator assigned a subset to code independently. To assess inter-labeler reliability, one investigator (A) re-coded portions of the data originally labeled by the other three investigators (B, C, and D). Consistency between investigator A and the others was excellent: Cronbach's α = .954 (N = 671) for A vs. B; .876 (N = 696) for A vs. C; and .945 (N = 735) for A vs. D.

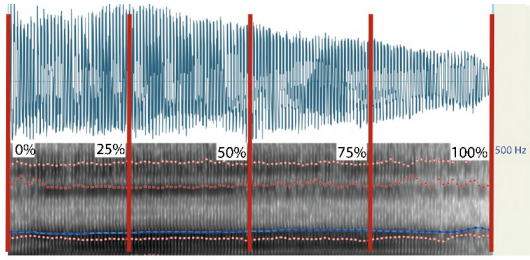


Figure 3. Five time points labeled for F0 analysis for Tone 1

Note. An example spectrogram and waveform displaying how the experimenters labeled each individual vowel for each syllable in the tonal pairs within the stimulus sets. The vowel was divided into quarters, giving five distinct time points per vowel (red vertical lines) and 10 data points per tonal pair. F0 (the blue dotted line) was then extracted in Hz from each time point.

Duration

The duration of each of the four tones varies in native Mandarin (e.g., Tone 3 is typically the longest) and thus, we examined whether the participants' vowel duration was significantly affected by the three conditions. Duration of each vowel for each tone was calculated using a PRAAT script that measured the time (in milliseconds) from the onset to the offset of the vowels (/a/, /i/, /u/). Using the onset

and offset boundaries for each vowel segment, the script computed duration as *endTime minus startTime*.

Post-experiment evaluation: Questionnaire

Given the nature of the ordinal data in the questions related to ordering and rankings, Friedman's and Cochran's tests were used to analyze participants' responses to questions about motivation, enjoyment, preference, and intuitiveness.

Results

Acoustic analyses

F0

Figure 4 shows the mean F0 at five time points for each tone and condition, with the native model plotted for reference. Participants approximated the model's contour for Tone 1 (flat) although the participants' F0 values were consistently lower than the model's. Tone 2 (rising) was not as steeply rising as the model's tone; Tone 3 (dipping) showed a shallower mid-vowel dip; and Tone 4 (falling) was not as steeply falling as the model (higher F0 at the end), replicating Zheng et al. (2018). We examined whether SG and SGE shifted contours toward the native pattern: For Tone 1, since it is a flat tone with F0 almost unchanged across time, a Condition effect might show such a shift. For the other tones, however, a shift towards the native pattern would require Condition × Time interactions since contours change across time. As we describe below, we found a significant main effect of Condition for Tone 1, but no expected interactions emerged for the other tones.

For Tone 1, there was a significant main effect of Condition, $F(2, 3226) = 33.90$, $p < .001$, $\eta^2_p = .102$.[2] Neither the effect of Time nor the interaction of Condition × Time was significant. Mean F0 values collapsed across time were highest in SG (268 Hz), significantly higher than SGE (261 Hz) ($p < .001$, $\eta^2_p = .06$), and SGE was also significantly higher than S (258 Hz) ($p < .001$, $\eta^2_p = .065$). Although the differences were small, SG was closest to the model, SGE was second closest, and S was the furthest.

For Tone 2, there were main effects of Condition, $F(2, 3122) = 8.53$, $p < .001$, $\eta^2_p = .10$ (SGE < S: $p < .001$, $\eta^2_p = .22$; SGE < SG: $p = .02$, $\eta^2_p = .08$), and Time,

2. To more accurately represent subject-level effect sizes, partial eta squared values were calculated based on conditional averages within each subject instead of across items.

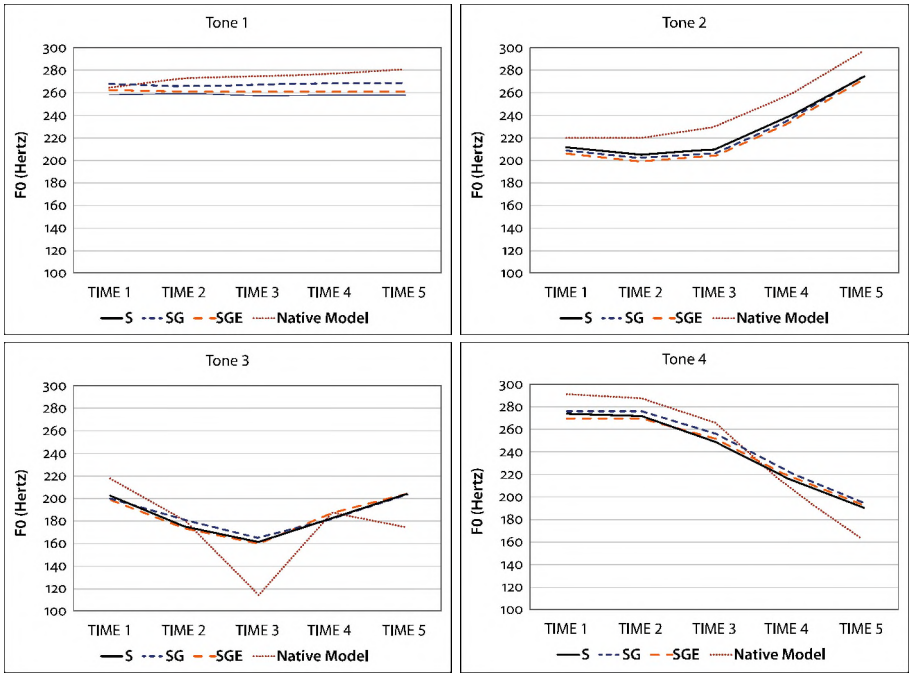


Figure 4. Mean F0 at five time points by condition for each tone

Note. There were no significant Condition x Tone interactions. Dotted line represents the native model.

$F(4, 3122) = 869.89$, $p < .001$, $\eta^2_p = .87$, with an upward linear trend ($p < .001$, $\eta^2_p = .89$). Condition × Time was not significant (ns).

For Tone 3, Condition and Condition × Time were ns (Condition: $F(2, 2768) = 0.14$; interaction: $F(8, 2765) = 1.16$), but there was a Time effect, $F(4, 2765) = 141.46$, $p < .001$, $\eta^2_p = .49$, showing a U-shaped quadratic trend ($p < .001$, $\eta^2_p = .71$).

For Tone 4, there were main effects of Condition, $F(2, 3181) = 10.59$, $p < .001$, $\eta^2_p = .09$ (SG > S: $p < .001$, $\eta^2_p = .15$; SG > SGE: $p = .02$, $\eta^2_p = .10$), and Time, $F(4, 3181) = 905.24$, $p < .001$, $\eta^2_p = .94$, with a downward linear trend ($p < .001$, $\eta^2_p = .95$). Condition × Time was ns.

In summary, there were main effects of Condition in all but Tone 3, and of Time in all but Tone 1, but the Condition x Time interaction was not significant in any of the tones. The main effect of Condition in Tone 1 suggested that SG and SGE slightly raised F0 levels towards native levels (10 Hz and 3 Hz, respectively), but outside of this isolated and modest effect, we did not find robust evidence

that gesture and enthusiasm moved the speakers' F0 contours towards the native model.

Duration

Given reports that native Tone 3 is typically longest (e.g., Jongman et al., 2006; Yang et al., 2017), we asked whether gesture conditions lengthened production. As shown in Figure 5, Tone 3 showed a main effect of Condition, $F(2, 645) = 11.09$, $p < .001$, $\eta^2_p = .24$: SG ($p = .003$, $\eta^2_p = .36$) and SGE ($p < .001$, $\eta^2_p = .31$) were both significantly longer than S, but no different from one another. Tones 1, 2 and 4 showed no Condition effects (all ns). Thus, gesture (with or without enthusiasm) selectively increased the duration of Tone 3 toward a more native-like pattern.

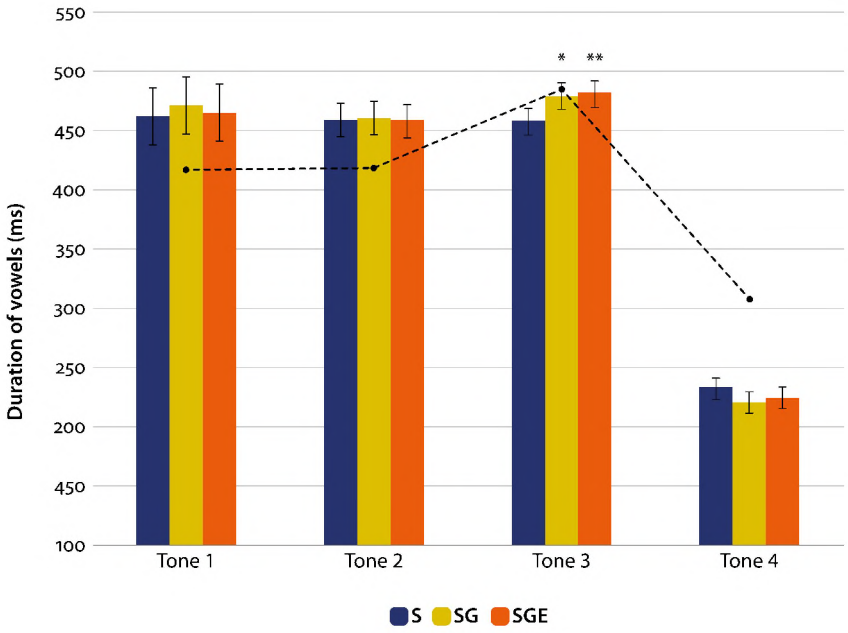


Figure 5. Mean vowel duration (ms) by condition across tones

Note. Tone 3 is longer in SG and SGE than in S. Standard errors are shown. Broken lines show native Mandarin speakers' mean vowel duration ($N = 156$) from Yang et al. (2017).[3]

3. We used Yang et al.'s (2017) data here to represent typical vowel durations: Tone 3 longest, Tone 4 shortest, and the other tones in between. The model native speaker of the present study did not show a typical pattern, with the mean durations of Tones 1, 2, 3, and 4 being 486, 225, 401, and 214 ms. This idiosyncrasy is probably due to the fact that the speaker recorded them in this tone sequence order, with Tone 1 spoken first, with no instruction on consistent speaking rate, and that only 9 tokens (3 vowels x 3 repetitions) from one speaker were averaged per tone, thus not providing a generalized pattern.

Self-assessment questionnaire

Participants evaluated gesture-based conditions—especially when accompanied by enthusiastic facial expressions—more positively than speech alone (Table 1).

Table 1. Conditional differences in self-assessment questions

	Conditions		
	S	**SG**	**SGE**
a. Motivation	3	2	13
b. Helpfulness	2.89	4.22	4.39
c. Intuitiveness	6	6	6
d. Preference (Learn)	2	4	12
e. Preference (Teach)	0	5	13
f. Enjoyment	2	5	11

Note. Numbers in (a), (c), (d), (e), and (f) indicate the total counts of participants' selections. Numbers in (b) indicate the mean of helpfulness in the scale from 1 to 5.

Motivation

Rankings differed by condition (Friedman $\chi^2(2) = 13.37$, $p < .001$); pairwise tests showed SGE > S ($p < .001$) and SGE > SG ($p < .001$) (Table 1a).

Helpfulness and intuitiveness

Helpfulness ratings varied across conditions (Friedman $\chi^2(2) = 13.37$, $p < .001$); S < SG ($p = .025$) and S < SGE ($p = .009$) (Table 1b). Intuitiveness did not differ ($\chi^2(2) = 0.00$, ns) (Table 1c).

Preference

For classroom learning, preferences differed (Cochran's $Q(2) = 14.78$, $p < .001$); SGE was chosen six times more often than S ($p < .001$) and three times more often than SG ($p = .012$) (Table 1d). For teaching a novice, preferences also differed (Cochran's $Q(2) = 14.33$, $p < .001$); S was selected less than SG ($p = .038$) and SGE ($p < .001$), while SG vs. SGE was marginal ($p = .087$) (Table 1e).

Enjoyment

Rankings differed (Friedman $\chi^2(2) = 12.20$, $p = .002$); S < SG ($p = .016$) and S < SGE ($p < .001$) (Table 1f).

Discussion

For the acoustic measures, we found isolated support for our predictions. While there were no significant interactions of Condition x Time on F0 patterns, there were main effects of Condition for tones 1, 2, and 4. Notably for Tone 1, the SG and SGE conditions slightly raised F0 levels towards the native levels. For Duration, the only tone where we found facilitative effects of conditions towards the native model was Tone 3: the tones in the SG and SGE conditions were longer than the S tone, a pattern reaching closer to native Mandarin. In contrast to these modest and uneven acoustic results, the self-assessment questions showed robust and nearly uniform support for our predictions: compared to the S condition, the SGE condition significantly boosted self-ratings of motivation, enjoyment, preference, and helpfulness, while the SG condition led to higher ratings of enjoyment, preference, and helpfulness.

Acoustic Measures

There were only two findings from the acoustic analyses showing that our manipulations moved L2 speech toward native levels: SG and SGE both elevated F0 in Tone 1 (Figure 4) and extended duration in Tone 3 (Figure 5). Focusing first on duration, previous research has suggested that Tone 3 tends to be the longest of the four Mandarin tones (Jongman et al., 2006; Yang et al., 2017), suggesting that gesture helped speakers sound more native-like for this tone. To our knowledge, this gestural lengthening of duration has not been observed in Mandarin tone production, but there is precedent for beat gestures lengthening L1 Dutch syllable duration (Krahmer & Swerts, 2007) and durational gestures elongating L2 vowel duration in Japanese (Li et al., 2020).

The F0 effect for Tone 1 requires more attention. It is well established that one of the limitations of non-native L2 Mandarin speakers is that they do not produce large enough contrasts within and between tones (Chen, 1974; Miracle, 1989), so the fact that both gesture conditions moved L2 speakers closer to native levels suggests that the hands can be useful for expanding F0 range. However, the results suggest that this expansion is quite isolated. For example, note the deviations between non-native speakers and native speakers for Tones 3 and 4. It seems that the non-native F0 patterns with insufficient dipping in Tone 3 and a relatively shallow downward slope in Tone 4 do not easily change, even for L2 Mandarin learners taking university language classes. For these more difficult tones, it appears that hand gestures and enthusiastic facial expressions did not help learners correct these patterns.

One possible explanation for how hand movements affected F0 in Tone 1 comes from research on the biomechanics of co-speech gesture (Pouw et al., 2020; Pouw et al., 2021). Tone 1 showed that both gesture conditions produced higher pitch across all five time points than the speech condition, suggesting that producing gestures high above the head can elevate pitch too. This account resonates with Pouw et al. (2020) who showed that large arm and shoulder movements produced exaggerated spoken phonation. They argue that F0 is determined by air pressure and larynx muscle tonus, with increased air pressure producing higher F0. Because arm movements engage trunk muscles involved in expiration—the phase of respiration during which speech occurs—these movements can directly affect phonation by increasing respiratory force. In our study, the gesture for Tone 1 involved raising the hand and tracing a horizontal line from the farthest point above the left side of the head to the right side of the body, a motion that strongly engages the core. Thus, the observed main effect of condition on Tone 1 may reflect the impact of exaggerated gestures on phonation when coupled with speech, much like how a musician might extend the torso upwards to reach a high note. In the context of Mandarin pronunciation, this extension may help learners move closer to native pitch ranges.

In contrast, the main effects of Condition alone are hard to interpret for Tones 2 and 4 because only a Condition x Time interaction effect is useful for comparison to native speech. For example in the SG condition, producing a higher pitch than S for Tone 4 across *all* time points does not improve pronunciation. To show that SG was helping Mandarin pronunciation towards the native model, it would be necessary to see an interaction between condition and time, with SG starting higher than S at times 1, 2, and 3, but finishing lower than S at time 5, which would mirror the pitch pattern of native speakers (see Figure 4).

Taken together, the acoustic results add modest support to studies showing that producing gestures can aid L2 pronunciation (Gluhareva & Prieto, 2017; Hoetjes & van Maastricht, 2020; Iizuka et al., 2020; Li et al., 2020; Li et al., 2021; Tseng et al., 2025; Xi et al., 2020; Zheng et al., 2018). All of these studies have shown that either observing or producing gestures can produce more native-like acoustic properties in L2 speech production, such as rhythm, stress, intonation, syllable articulation, and fundamental frequency. However, some of these findings are limited in terms of how widespread (Hoetjes & van Maastricht, 2020; Li et al., 2021; Zheng et al., 2018) and robust (Xi et al., 2020) the effects are. Coupled with the present study's isolated acoustic effects, including slight elevation of F0 for Tone 1 and elongation of Tone 3 toward native levels, the most cautious conclusion is that gestures can support L2 pronunciation, but their direct effects on phonetic accuracy are modest and leave room for further improvement.

Affective measures

In contrast to the uneven effects of gesture and facial expressions on tone pronunciation, the affective self-reports revealed a widespread influence of gesture and emotion. For all questions but the "intuitive" measure, the SGE condition had more positive affective evaluations than the S condition. Moreover, in the case of motivation and preference for classroom learning, it also exceeded the SG condition. This suggests that the combination of gesture and enthusiastic facial expressions goes beyond gesture alone in affectively engaging L2 speakers, which is consistent with research suggesting that multiple channels of bodily communication facilitate the learning experience (Collins, 1978; Keller et al., 2016; McCroskey et al., 1995). Of course, it is possible that the increased engagement was driven *exclusively* by the positive facial expression—with no combined benefit of hand gestures—but this is unlikely for at least two reasons. One, previous L2 research has shown robust affective influences of hand gestures alone (Algana & Hardison, 2024; Hirata et al., 2024; Kamiya, 2025; Tseng et al., 2025; Zheng et al., 2018), and two, the SG condition in the present study was rated higher than the S condition in three affective measures: helpfulness, enjoyment, and preference for teaching novices. This suggests that it was the combination of hand gestures and enthusiastic facial expressions that contributed to the positive evaluations in SGE, and future research should attempt to more systematically tease apart the relative contributions of both.

These results are interesting in light of the growing attention on affective factors in L2 teaching and learning contexts (Dewaele & MacIntyre, 2014; Dörnyei & Ushioda, 2021; Henry & Thorsen, 2018; MacIntyre & Gregersen, 2012). For example, Dewaele and MacIntyre (2014) surveyed over 1,700 foreign language (FL) learners from around the world and found that a stimulating classroom environment—featuring engaging activities and highly involved, supportive, and skilled teachers—not only mitigates the anxiety of FL learning but also makes it more motivating and rewarding. This is important because the FL anxiety correlates with the learners' willingness to communicate (Baran-Lucarz, 2014). Given that expressive bodily actions are a particularly effective way for a teacher to create such a stimulating L2 learning environment (Allen, 2000; Henry & Thorsen, 2018), it makes sense that learners would like them.

Finally, it is possible that the affective results simply reflected participants being engaged in a more interesting and novel activity for the SG and SGE conditions compared to the S condition, making the latter relatively unattractive. That may be true for some participants, but a full third of them found the Speech condition to be the most intuitive (33%), and some even ranked it as the most preferable, motivating, or enjoyable. Indeed, in our exit interview, some participants

commented that they liked pronouncing the tones without gesture because they found them to be distracting. Nevertheless, the results clearly indicate that most people did favor the SG and SGE conditions. Regardless of the reasons why, we think this makes our acoustic results even more interesting, as we discuss in the concluding section.

Limitations of present study

There are at least four limitations of the study that warrant attention. First, as we mentioned in the Method section, our participants varied widely in their amount of Mandarin study experience. This wide pool likely added significant variability to the effects of gesture imitation. Future research should examine whether learners at more narrowly defined proficiency levels show more even and robust benefits of gesture imitation. Second, participants had relatively short exposure to the three conditions (about five minutes each), and perhaps more multimodal experience is necessary to more fully shape acoustic properties. Third, our three conditions all used identical dubbed speech produced as neutrally as possible by the Mandarin model. This is good experimental control because it is well known that producing gesture affects speech acoustics (Krahmer & Swerts, 2007), but perhaps including these natural acoustic variations from using gestures would have further improved participants' pronunciation.

Finally, it is possible that the present study (and Zheng et al., 2018) did not provide specific enough instructions on how to produce the gestures. Perhaps our participants were unaware of the weak points of their tone production, and they did not have specific enough focus when they made the gestures. It would be interesting for future studies to give more targeted instruction—with more direct guidance on how to exaggerate gesture production—about specific aspects of nonnative Mandarin speech, such as Tone 3 not dipping low enough at time 3 or Tone 4 not going down steeply enough from the beginning to the very end.

That said, even if methodological changes could improve the acoustic benefits of gesture, there may still be limits. After all, it is possible that gestures may be better suited for helping with other aspects of L2 learning, such as pragmatic processing, semantic comprehension, and vocabulary acquisition (Allen, 1995; García-Gámez & Macizo, 2019; Gullberg, 2006; Huang et al., 2019; Kelly et al., 2009; Macedonia & Klimesch, 2014; Macedonia & Müeller, 2016; Morett, 2018; Sweller et al., 2020; Tellier, 2008). For instance, Sweller et al. (2020) found that English speakers who produced iconic gestures while learning Japanese verbs recalled the words much more effectively, both immediately and after a delay. This facilitative semantic effect is robust and has been observed across a range of languages, including in controlled studies using artificial languages designed to eliminate

phonetic confounds (Macedonia & Klimesch, 2014). This contrasts with the more modest and variable effects of gesture on L2 phonetics. As an explanation for this discrepancy, Kelly (2017) argues that representational hand gestures like iconics and metaphorics are more naturally designed to communicate meaning in a language than they are for communicating sounds in a language. It seems that gestures can be co-opted for phonetic purposes, but their original and most natural function is likely semantic, not phonetic. Indeed, even beat gestures, which often emphasize prosodic elements of speech, do so in the service of clarifying higher-level meaning (McNeill, 1992).

Multiple functions of gesture and enthusiasm

Despite gestures having limited effects on actually pronouncing Mandarin in a more native fashion, participants *felt* as though they helped a lot. Moreover, they quite enjoyed them, and when gestures were accompanied by enthusiastic facial expressions, the combination was seen as especially motivating. What might explain this disconnect between actual performance and subjective impressions?

It is well established in the fields of cognitive and educational psychology that people are not good judges of what is the most effective type of learning (Bjork et al., 2013; Roediger & Karpicke, 2006). When an activity feels right, it is often misattributed to *being* right. This breakdown in metacognition can result in people adopting all sorts of ineffective, unproven, or non-optimal strategies. One possibility is that producing gestures with tones is one such "non-optimal" strategy. Although evidence suggests only uneven and modest benefits for tone pronunciation, teachers and learners in Mandarin classrooms worldwide routinely produce gestures for the four Mandarin tones without questioning their effectiveness.

This perspective, however, describes only part of the picture, as other factors help explain why these gestures remain so widely used in L2 classrooms. Even if tone gestures provide limited *direct* gains, they may offer robust *indirect* benefits by boosting positive affect. Increased motivation and enjoyment may reduce pronunciation anxiety and raise willingness to communicate (WTC), thereby increasing speaking practice and resilience (Baran-Łucarz, 2014; Mercer & Dörnyei, 2020). All this, ultimately, may improve pronunciation in the long run. In this way, gesture and enthusiasm in the L2 classroom may serve two purposes: it may offer a simple and accessible tool for supporting the pronunciation of difficult speech sounds, and it can help create a positive and engaging environment that makes L2 speakers want to keep coming back for more.

Funding

This study is funded by Colgate's Center for Language and Brain.

Acknowledgements

This study was conducted at Colgate University with the support from the Center for Language and Brain. We are grateful to Gloria Liu for being the model for the stimuli and Jordan Shapiro for his assistance in the creation of stimulus sets and analysis of the data. We are also grateful to John Crespi, Jing Wang, and the other instructors of the Chinese Language Program at Colgate University for their help in recruiting participants. The authors thank Amanda Brown, Marianne Gullberg, Masaaki Kamiya, and Laura Morett for their valuable comments, which helped shape this paper.

References

Algana, M., & Hardison, D. M. (2024). Variable effects of speakers' visual cues and accent on L2 listening comprehension: A mixed-methods approach. *Language Teaching Research*.

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.

Allen, L. Q. (2000). Nonverbal accommodation in foreign language. *Applied Language Learning*, *11*(1), 155–176.

Amand, M., & Touhami, Z. (2016). Teaching the pronunciation of sentence final and word boundary stops to French learners of English: Distracted imitation versus audio-visual explanations. *Research in Language*, *14*(4), 377–388.

Aslan, Z., Özer, D., & Göksun, T. (2024). Exploring emotions through co-speech gestures: The caveats and new directions. *Emotion Review*, *16*(4), 265–275.

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. *Canadian Modern Language Review*, *70*(4), 445–473.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.

Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics 2*, 159–171.

Collins, M. L. (1978). Effects of enthusiasm training on preservice elementary teachers. *Journal of Teacher Education*, *29*(1), 53–57.

Del Rio, P. & Álvarez, A. (2002). From activity to directivity: The question of involvement in education. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 59–72). Blackwell.

Dewaele, J. M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, *4*(2), 237–274.

Dewaele, J. M., & Li, C. (2021). Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research, 25*(6), 922–945.

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation* (3rd ed.). Routledge.

Elliott, C. E. (1991). The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai'i Working Papers in ESL*, *10*(2), 177–204.

García-Gámez, A. B., & Macizo, P. (2019). Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics*, *40*(2), 473–507.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609–631.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *International Review of Applied Linguistics*, *44*(2), 103–124.

Hardison, D. M. (2025). *The multimodal context of phonological learning*. University of Toronto Press.

Henry, A., & Thorsen, C. (2018). Teacher–student relationships and L2 motivation. *The Modern Language Journal*, *102*(1), 218–241.

Hirata, Y., Friedman, E., Kaicher, C., & Kelly, S. D. (2024). Multimodal training on L2 Japanese pitch accent: learning outcomes, neural correlates and subjective assessments. *Language and Cognition*, *16*(4), 1718–1755.

Hoetjes, M., & van Maastricht, L. (2020). Using gesture to facilitate L2 phoneme acquisition: The importance of gesture and phoneme complexity. *Frontiers in Psychology*, *11*, 575032.

Hu, B. (2010). The challenges of Chinese: A preliminary study of UK learners' perceptions of difficulty. *The Language Learning Journal*, *38*(1), 99–118.

Huang, J. (2000). Students' major difficulties in learning Mandarin Chinese as an additional language and their coping strategies (ERIC Document Reproduction Service No. ED452736). *ERIC*. https://eric.ed.gov/?id=ED452736

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Language Learning*, *69*(1), 177–197.

Iizuka, T., Nakatsukasa, K., & Braver, A. (2020). The efficacy of gesture on second language pronunciation: An exploratory study of handclapping as a classroom instructional tool. *Language Learning*, *70*(4), 1054–1090.

Inceoglu, S., & Ueda, R. (2024). The effectiveness of hand gestures on the development of L2 French pronunciation. In A. Brown & S. W. Eskildsen (Eds.), *Multimodality across epistemologies in second language research* (pp. 139–152). Routledge.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of East Asian psycholinguistics: Vol. 1. Chinese* (pp. 250–257). Cambridge University Press.

Kamiya, N. (2025). The limited effects of visual and audio modalities on second language listening comprehension. *Language Teaching Research*, *29*(4), 1688–1714.

Keller, M. M., Hoy, A. W., Goetz, T., & Frenzel, A. C. (2016). Teacher enthusiasm: Reviewing and redefining a complex construct. *Educational Psychology Review*, *28*(4), 743–769.

Kelly, S. D. (2017). Exploring the boundaries of gesture–speech integration during language comprehension. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture?: How the hands function in speaking, thinking and communicating* (pp. 243–265). John Benjamins Publishing Company.

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.

Kelly, S. D., & Ngo Tran, Q. A. (2023). Exploring the emotional functions of co-speech hand gesture in language and communication. *Topics in Cognitive Science*, *17*(3), 586–608.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception, and visual perception. *Journal of Memory and Language*, *57*(3), 396–414.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.

Li, P., Baills, F., & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Studies in Second Language Acquisition*, *42*(5), 1015–1039.

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.

Li, P., Baills, F., Baqué, L., & Prieto, P. (2023). The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research, 39*(4), 1077–1105.

Lin, W. C. (1985). Teaching Mandarin tones to adult English speakers: Analysis of difficulties with suggested remedies. *RELC Journal*, *16*(2), 31–47.

Liu, W. (2021). Does teacher immediacy affect students? A systematic review of the association between teacher verbal and non-verbal immediacy and student motivation. *Frontiers in Psychology*, *12,* 713978.

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.

Macedonia, M., & Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in Psychology*, *7*, 953.

MacIntyre, P., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive-broadening power of the imagination. *Studies in Second Language Learning and Teaching*, *2*(2), 193–213.

McCroskey, J. C., Richmond, V. P., Sallinen, A., Fayer, J. M., & Barraclough, R. A. (1995). A cross-cultural and multi-behavioral analysis of the relationship between nonverbal immediacy and teacher evaluation. *Communication Education*, *44*(4), 281–291.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Megawati, W., & Hartono, R. (2020). The impact of teacher's verbal and non-verbal communication on students' motivation in learning English. *English Education Journal*, *10*(4), 436–448.

Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press.

Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of Chinese Language Teachers Association*, *24*, 49–65.

Morett, L. M. (2018). In hand and in mind: Effects of gesture production and viewing on second language word learning. *Applied Psycholinguistics*, *39*(2), 355–381.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Pouw, W., Harrison, S. J., & Dixon, J. A. (2020). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, *149*(2), 391–404.

Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2021). Gesture–speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, *1491*(1), 89–105.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Sime, D. (2006). What do learners make of teachers' gestures in the language classroom? *International Review of Applied Linguistics in Language Teaching*, *44*(2), 211–230.

Smotrova, T. (2017). Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly*, *51*(1), 59–89.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699.

Sweller, N., Shinooka-Phelan, A., & Austin, E. (2020). The effects of observing and producing gestures on Japanese word learning. *Acta Psychologica*, *207*, 103079.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235.

Tseng, G., Liu, Y. T., & Fan, S. Y. C. (2025). Gesture to learn, hum to speak: Promoting L2 pronunciation through non-verbal techniques. *English Teaching & Learning*. Advance online publication.

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

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), 3571–3585.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61–83.

Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the duration of Mandarin tones. *INTERSPEECH 2017*, 1407–1411.

Yuan, L. (2024). EFL teacher–student interaction, teacher immediacy, and students' academic engagement in the Chinese higher learning context. *Acta Psychologica*, *244*, 104185.

Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, *24*(5), 666–689.

Zheng, A., Hirata, Y., & Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 Mandarin tone production. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2179–2195.

Appendix

Stimuli Pairs

1 mì_mī
2 má_mǎ
3 mā_má
4 mí_mǐ
5 mū_mǔ
6 mǔ_mù

7 mǐ_mī
8 mū_mú
9 mǎ_mā
10 mù_mú
11 mī_mǐ
12 mú_mù

13 mǎ_mà
14 mú_mǔ
15 mà_má
16 mù_mū
17 mī_mí
18 má_mà

19 mì_mí
20 mǔ_mū
21 mí_mì
22 mā_mǎ
23 mǐ_mì
24 mà_mā

Address for correspondence

Spencer D. Kelly
Department of Psychological and Brain Sciences
Colgate University
13 Oak Dr.
Hamilton, NY 13346
United States

skelly@colgate.edu
https://orcid.org/0000-0002-4562-8155

Co-author information

Paige Avila
Colgate University
ernestocifuentes30@gmail.com

Madeline Chernavsky
Colgate University
mnchernavsky@gmail.com

Bailey Cooper
Georgetown University
baileyjc@cox.net

Elizabeth Velázquez Fernández
The University of Puerto Rico
elizabeth.velazquez1@upr.edu

Yukari Hirata
Department of East Asian Languages
Colgate University
yhirata@colgate.edu

Publication history

Date received: 11 October 2025
Date accepted: 24 February 2026
Published online: 23 April 2026