



OPEN ACCESS

EDITED BY

Xin Wang,
Macquarie University, Australia

REVIEWED BY

Haiquan Huang,
Hubei University of Technology, China
Debra Hardison,
Michigan State University, United States

*CORRESPONDENCE

Bashar M. Farran
✉ bfarran@health.missouri.edu

RECEIVED 01 April 2024

ACCEPTED 08 July 2024

PUBLISHED 24 July 2024

CITATION

Farran BM and Morett LM (2024) Multimodal cues in L2 lexical tone acquisition: current research and future directions. *Front. Educ.* 9:1410795. doi: 10.3389/feduc.2024.1410795

COPYRIGHT

© 2024 Farran and Morett. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multimodal cues in L2 lexical tone acquisition: current research and future directions

Bashar M. Farran* and Laura M. Morett

Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, United States

This review discusses the effectiveness of visual and haptic cues for second language (L2) lexical tone acquisition, with a special focus on observation and production of hand gestures. It explains how these cues can facilitate initial acquisition of L2 lexical tones via multimodal depictions of pitch. In doing so, it provides recommendations for incorporation of multimodal cues into L2 lexical tone pedagogy.

KEYWORDS

lexical tone, second language acquisition, multimodality, gesture, tonal languages

1 Introduction

Imagine a language where the meaning of a word hinges on its pitch. This is the reality in tonal languages, where pitches, not just phonemes, determine word meaning. Most world languages, including Mandarin Chinese, Vietnamese, Thai, Yorùbá and various African languages, are tonal (Maddieson, 2013). While mastery of tonal first languages (L1s) comes naturally, second language (L2) learning of tonal languages entails a unique challenge, particularly for learners whose first language is atonal (Wang et al., 2006, 2020).

L2 acquisition of lexical tones encompasses both perception and production. Although perception often precedes production in L2 lexical tone acquisition (Wang et al., 1999), the relationship between them is not always straightforward, and improvements in perception do not necessarily entail improvements in production, and vice versa (Leather, 2011). L2 lexical tone acquisition involves perception of not only auditory cues, but also visual and haptic cues such as hand gestures (Gullberg, 2006). The importance of these multimodal cues in facilitating L2 lexical tone perception and production has increasingly gained recognition (McCafferty, 2004; Hostetter, 2011; Lewis and Kirkhart, 2022; Zhang et al., 2023). Multisensory learning, which integrates multiple sensory modalities, is more effective than unisensory approaches due to optimization of the brain for multisensory environments, suggesting that L2 lexical tone pedagogy could be enhanced by incorporating such approaches (Shams and Seitz, 2008). Macedonia and Kepler (2013) argue that use of pedagogical approaches informed by neuroscience findings into L2 instruction can significantly enhance learning via a three-pronged approach: (1) utilizing multisensory experiences for vocabulary acquisition, (2) incorporating imitation exercises to leverage mirror neurons for pronunciation training, and (3) tailoring instruction to brain development stages for optimal grammar and pronunciation outcomes. Moreover, multisensory cues enhance learning outcomes by supporting content comprehension (Dick et al., 2009). Understanding how nonverbal cues enhance auditory representations can shed light on how multimodal approaches can be leveraged to facilitate acquisition of an unfamiliar tonal L2 (Yip, 2002; Liu et al., 2022).

2 Auditory training methods

Cognitively, tonal languages require awareness of pitch, which permits discrimination, identification, and manipulation of lexical tones. In the intricate acoustic signal of speech, multiple cues such as formant frequencies, amplitude, and temporal information coexist with pitch contours. Thus, tonal language comprehension entails selective attention to pitch cues in conjunction with suppression of other acoustic information (Huang and Johnson, 2011). This selective attention to pitch cues is shaped by experience with lexical tone. Moreover, pitch perception in tonal languages goes beyond recognizing static pitch levels as it entails tracking rapid pitch movements and complex tonal contours over time (Gandour, 1983; Xie and Myers, 2015). Thus, processing of pitch within the speech stream is critical to L2 lexical tone acquisition (Jasmin et al., 2020).

Neurologically, the ability to selectively focus on pitch involves specialized mechanisms shaped by tonal language experience (Gandour et al., 2003; Xu et al., 2006). Lexical tone processing involves both subcortical and cortical structures (Gandour and Krishnan, 2016). Initially, L2 lexical tone processing is predominantly handled by the right hemisphere or bilaterally, but with increased exposure, it becomes more left lateralized and akin to L1 processing (Gandour et al., 2004; Wang et al., 2004; Gandour, 2006; Xi et al., 2010; Kaan et al., 2013).

Considering the cognitive and neurological complexities of lexical tone processing, auditory methods have been developed to facilitate L2 lexical tone learning. These methods include discrimination training, categorization training, and auditory corrective feedback.

Discrimination training involves exposure to contrasting pairs of tones and subsequent testing via determination of whether trained tones are the same or different. For example, *má* and *mà* could be presented consecutively in training, and discrimination between the rising and falling tones could then be tested by determining whether *chó* and *chò* are perceived as the same or different. Discrimination tasks are perceptual, involving the discernment of differences in pitch contours and other acoustic cues. Discrimination training leads to significant improvements in perception of differences between lexical tones (Wang et al., 1999; Wayland and Guion, 2004; Hao, 2012).

Categorization training involves exposure to labeled tones and subsequent testing via labeling of unlabeled tones. For example, the tones in *má* and *mà* could be labeled as rising and falling in training, and categorization could then be tested by labeling *má* as rising and *mà* as falling. Thus, identification tasks draw on memory as well as perception because they require mapping acoustic features of lexical tones onto their representations. Categorization training improves L2 lexical tone identification, particularly in the early stages of acquisition, but may not be sufficient for accurate production (Leather, 1990; Wang et al., 2003; Duanmu, 2007; Ladefoged and Johnson, 2015).

The distinction between discrimination and categorization is significant because discrimination can precede categorization in L2 lexical tone acquisition. However, discrimination and categorization are related; thus, they can support one another. Understanding the relationship between discrimination and categorization is essential for designing effective language learning materials, speech recognition systems, and other natural language processing applications for tonal languages.

Discrimination and categorization training based on a small set of stimuli in experimental tasks may not fully capture the natural variations of lexical tones in everyday speech. This limitation helped lead to the emergence of High Variability Perception Training (HVPT)

in lexical tone learning tasks. This training entails exposure to lexical tones within varying linguistic contexts or produced by multiple speakers in the interest of more closely approximating the natural variability encountered in real-life tonal language processing (Lively et al., 1994; Pisoni and Lively, 1995). HVPT improves both perception and production of L2 lexical tones as it enhances generalization across different contexts and speakers (Guion et al., 2000; Wang et al., 2003). This approach emphasizes the importance of exposure to diverse linguistic input to achieve more robust language learning outcomes.

Auditory corrective feedback may consist of recasts, in which the correct tone is heard in response to incorrect tone production; contrastive feedback, which highlights the difference between attempted and correct pronunciation; and explicit feedback, which provides verbal explanations of errors and correction techniques (Lee and Lyster, 2016; Saito, 2021). The effectiveness of auditory corrective feedback relies upon perception as well as memory because differences between incorrect and correct tones must be perceived and remembered to produce them correctly. Auditory corrective feedback improves L2 lexical tone production accuracy by highlighting errors and modeling correct pronunciation (Bryfonski and Ma, 2020).

While auditory methods have been a mainstay in L2 lexical tone acquisition, they have limitations stemming from challenges inherent in relying solely on auditory input and feedback. Furthermore, L1 background and the L2 tone system may limit the effectiveness of auditory methods.

3 Visual cues

Visual cues can be powerful tools for enhancing L2 lexical tone acquisition. One approach utilizes static visual depictions of lexical tone pitch contours (Figure 1). These depictions, which may consist of lines, graphs, or color-coded charts, visually represent fundamental frequency (F0) variations characterizing tones (Godfroid et al., 2017). Such visual depictions facilitate understanding of lexical tone contours (Zhou and Olson, 2023), as evidenced by enhanced perception of lexical tones cross-linguistically (Burnham et al., 2022). Moreover, visual depictions of pitch contours improve categorization of L2 lexical tones compared to auditory input (Chun et al., 2012).

Building upon the benefits of visual depictions of pitch contours, another approach leverages pitch gestures to enhance L2 lexical tone learning. Also known as tone gestures or tone-bearing gestures, pitch gestures are hand or body movements that visually convey pitch patterns of words or syllables via fundamental frequency (Morett and Chang, 2015; Figure 2). Pitch gestures spontaneously occur in conjunction with tonal languages (Krahmer and Swerts, 2007) and are often produced with the hands or head but may also include eyebrow movements or body posture changes corresponding with tones (Antonioni and Chin, 2018; Lacombe et al., 2022).

Observing pitch gestures enhances perception and production of L2 lexical tones. Observing eye movements, head movements, and hand gestures conveying pitch contours enhances understanding and pronunciation of L2 Mandarin tones (Chen and Massaro, 2008). Additionally, observing pitch gestures positively impacts discrimination between L2 Mandarin words differing in lexical tone (Morett and Chang, 2015; Morett, 2023).

Visual cues such as observed pitch gestures provide tangible depictions of lexical tones that strengthen mental representations of

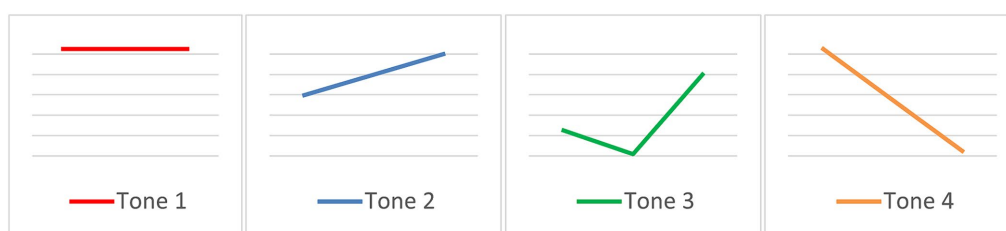


FIGURE 1
Images of pitch contours of Mandarin lexical tones.

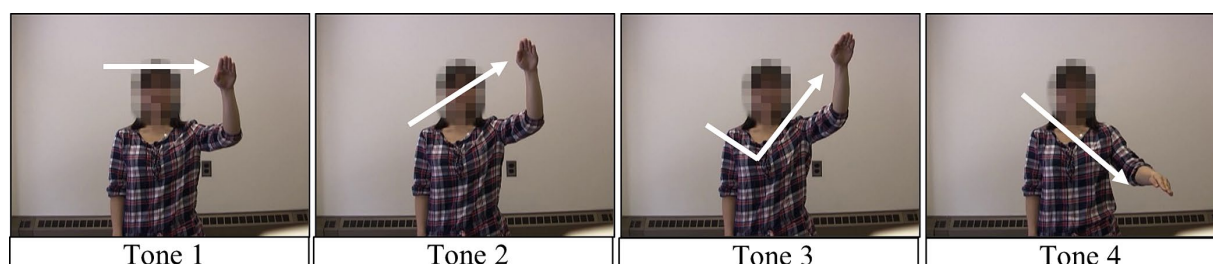


FIGURE 2
Pitch gestures for Mandarin lexical tones.

them via encoding and retrieval and enhance their perception and memory. In addition, visual cues offer additional support when auditory processing is impaired or exposure to tonal languages is limited.

While observing pitch gestures supports L2 lexical tone perception and production, relying solely on visual input may entail limitations. Visual depictions alone may not fully capture the richness and complexity of tonal variation, leading to incomplete or oversimplified learning outcomes. Additionally, visual depictions of lexical tones may encourage dependence on visual cues, neglecting development of auditory perception skills necessary for real-world communication. For example, use only of visual input for L2 Mandarin tone learning results in lower perception accuracy compared to use of both visual and auditory input (Jiang, 2017). Therefore, integrating visual cues with input from audition and other modalities may yield superior learning outcomes.

Theories providing explanations for the effects of visual cues on L2 lexical tone acquisition include dual coding theory and multimedia learning theory. Dual coding theory posits that information can be processed via both auditory (verbal) and visual (non-verbal) channels (Paivio, 1991, 2014a), each of which has strengths and weaknesses. Visual cues excel at conveying spatial information and relationships, while verbal cues are better suited for conveying linear sequences and abstract concepts. When visual and verbal cues occur together, tones can be processed via both the auditory and visual channels simultaneously. The resulting multimodal representations enhance encoding, storage, and retrieval of L2 lexical tones, improving their acquisition (Paivio, 2014b).

Multimedia learning theory emphasizes the importance of using multiple modes of representation to facilitate learning. This theory emphasizes combining different modalities (e.g., auditory, visual) to optimize learning outcomes and improve comprehension and retention of material (Mayer, 2005, 2009; Gullberg, 2022). It posits that learning is an active process that entails building connections between information presented in different modalities. Like dual coding theory, multimedia

learning theory maintains that presenting corresponding verbal and visual information simultaneously can enhance learning. This process leads to deeper understanding, improved retention, and enhanced knowledge transfer and real-world application (Mayer and Moreno, 1998; Mayer, 2005, 2014). For L2 lexical tone acquisition, multimodal methods that combine auditory verbal input with visual representations of pitch contours are consistent with multimedia learning theory.

4 Haptic cues

Haptic approaches to L2 lexical tone learning involve the use of bodily movements to facilitate and reinforce production and perception of lexical tones. Haptic approaches posit that physical interaction with lexical tone can enhance its cognitive processing and memory retention. Examples of haptic approaches may include hand movements conveying tonal contours or tactile feedback corresponding to pitch changes. One promising haptic approach is gesture production, which entails enactment of specific hand or arm movements to convey lexical tones. This approach capitalizes on the close connection between speech production and bodily movements, as well as the benefit of haptic cues for language learning.

Pitch gesture production improves discrimination and production of L2 lexical tone (Hannah et al., 2017). More specifically, producing pitch gestures, rather than merely observing them, leads to better learning outcomes (Baills et al., 2019). Producing hand gestures in conjunction with lexical tone not only enhances production of lexical tone but also improves discernment of subtle tonal differences (Zheng et al., 2018; Li et al., 2020; Yu et al., 2024). This suggests that producing hand movements results in deeper understanding of tonal contrasts, enhancing L2 tone acquisition. From a neurological perspective, speech perception and production involve distributed neural networks

that encompass not only auditory and motor cortices but also somatosensory and premotor areas (Guenther and Vladusich, 2012). This overlap suggests that haptic cues may recruit additional neural resources, resulting in enriched representations of lexical tones.

Despite their potential benefits, haptic approaches to L2 lexical tone acquisition may entail challenges. Firstly, the design and implementation of activities involving haptic cues requires careful consideration. Appropriate gestures or movements must be selected and consistently mapped to lexical tones, ensuring that associations are intuitive and easy to remember. Secondly, explicit instruction and feedback may be necessary to ensure that lexical tones are conveyed accurately via haptic cues. Thirdly, cultural and contextual factors may influence the acceptability and effectiveness of learning approaches involving haptic cues.

Multimodal methods incorporating haptic cues align with the principles of embodied cognition, providing evidence that cognitive processes are grounded in sensorimotor experiences and interactions with the physical world (Lakoff and Johnson, 2017; Shapiro, 2019). Embodied cognition proposes that recruitment of multiple sensory modalities facilitates acquisition and representation of abstract concepts by activating relevant physical experiences via mental simulation. Mental simulation leads to a stronger connection between acoustic features of tone and embodied experience, fostering more accurate production and perception.

5 Integrated multimodal cues

Research has increasingly explored integration of multimodal cues in the auditory, visual, and haptic modalities to enhance perception and production of L2 lexical tone. This approach focuses on the synergistic effects of engaging multiple sensory channels via complementary sources of information and its reinforcement of the mapping between lexical tones and their depictions. Integration of multiple modalities engages a broad range of cognitive and sensory processes, resulting in effective learning. This enhances attention, memory, and engagement with content, leading to improved acquisition and retention of L2 lexical tone. Thus, integration of visual and haptic cues should enrich representations of lexical tone, enhancing categorization and differentiation of lexical tones. Visual and haptic cues should be consistent with the vertical conceptual metaphor of pitch, which posits that high pitch is associated with upward positions and motion and that low pitch is associated with downward positions and motion. Visual–auditory mappings aligned with this metaphor result in accurate and robust representations of L2 lexical tones (Morett et al., 2022).

Multimodal approaches may help overcome the challenges associated with learning L2 Mandarin tones (Pelzl et al., 2022). Moreover, methods integrating visual and haptic cues are more effective than unimodal methods, highlighting the benefits of multimodality in facilitating L2 lexical tone acquisition (Godfroid et al., 2017). However, the effectiveness of multimodality may depend on several factors, such as the specific combination of modalities employed, the design and implementation of instructional materials, and prior tonal language experience. Although the factors discussed here provide explanations for the effectiveness of multimodal approaches, further research is needed to fully understand the underlying mechanisms and to optimize the design and implementation of multimodal instructional approaches to L2 lexical tone acquisition.

6 Discussion

Moving forward, insights from this review can inform development of strategies to enhance L2 tone acquisition. One strategy is to incorporate multimodal cues into existing curricula, leveraging techniques such as pitch gesture observation, pitch gesture production, and images of pitch contours to enhance L2 lexical tone acquisition. However, it is essential to critically evaluate existing instructional methods to determine their efficacy for both teachers and learners. To ensure maximum effectiveness, activities should convey lexical tone intuitively via the vertical conceptual metaphor of pitch.

Although existing research provides insight into how multimodal learning benefits L2 lexical tone acquisition, several topics warrant further investigation. Future research should determine the optimal combination of cues in different modalities by comparing their impacts on L2 lexical tone learning, as assessed via multiple measures. Additionally, research on the cognitive and neural correlates of lexical tone learning is needed to better understand the mechanisms enabling enrichment of representations via multimodal input. Furthermore, development and evaluation of technology-based tools presents opportunities to leverage digital technologies to enhance L2 tone instruction via multimodal learning. Addressing these research gaps will advance the understanding of multimodal learning and its implications for L2 lexical tone acquisition, informing development of practices that facilitate L2 lexical tone learning.

In summary, research illuminating the impact of multimodal cues on L2 lexical tone acquisition presents compelling evidence supporting their efficacy, particularly with respect to observation and production of hand gestures. Incorporating visual and haptic cues from gestures alongside auditory cues provides an enriched learning experience, enhancing perception and production of L2 lexical tone. The research reviewed here underscores the benefits of multimodal approaches, highlighting how visual depictions such as observed pitch gestures and haptic approaches such as gesture production can complement auditory input, resulting in enriched mental representations of L2 lexical tones. Taken together, this work demonstrates that multimodality enriches mental representations of L2 lexical tone, leading to improved learning outcomes.

Author contributions

BF: Writing – original draft. LM: Writing – review & editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. LM was funded by US National Science Foundation CAREER award #2140073.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Antoniou, M., and Chin, J. L. L. (2018). What can lexical tone training studies in adults tell us about tone processing in children? *Front. Psychol.* 9:1. doi: 10.3389/fpsyg.2018.00001
- Baills, F., Suárez-González, N., González-Fuente, S., and Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Stud. Second. Lang. Acquis.* 41, 33–58. doi: 10.1017/S0272263118000074
- Bryfonski, L., and Ma, X. (2020). Effects of implicit versus explicit corrective feedback on Mandarin tone acquisition in a SCMC learning environment. *Stud. Second. Lang. Acquis.* 42, 61–88. doi: 10.1017/S0272263119000317
- Burnham, D., Vatikiotis-Bateson, E., Vilela Barbosa, A., Menezes, J. V., Yehia, H. C., Morris, R. H., et al. (2022). Seeing lexical tone: head and face motion in production and perception of Cantonese lexical tones. *Speech Comm.* 141, 40–55. doi: 10.1016/j.specom.2022.03.011
- Chen, T. H., and Massaro, D. W. (2008). Seeing pitch: visual information for lexical tones of Mandarin-Chinese. *J. Acoust. Soc. Am.* 123, 2356–2366. doi: 10.1121/1.2839004
- Chun, D., Jiang, Y., and Ávila Reyes, N. (2012). Visualization of tone for learning Mandarin Chinese. In: *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, (Eds.), J. Levis and K. LeVelle (IA: Iowa State University). 77–89.
- Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., and Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain Mapp.* 30, 3509–3526. doi: 10.1002/hbm.20774
- Duanmu, S. (2007). The phonology of standard Chinese. 2nd Edn: Oxford University Press.
- Gandour, J. T. (1983). Tone perception in far eastern languages. *J. Phon.* 11, 149–175. doi: 10.1016/S0095-4470(19)30813-7
- Gandour, J. T. (2006). "Brain mapping of Chinese speech prosody" In: *The handbook of East Asian psycholinguistics: volume 1: Chinese*. eds. E. Bates, L. H. Tan, O. J. L. Tzeng and P. Li (Cambridge: Cambridge University Press), 308–319. doi: 10.1017/CBO9780511550751.030
- Gandour, J. T., Dziedzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang.* 84, 318–336. doi: 10.1016/S0093-934X(02)00505-9
- Gandour, J. T., and Krishnan, A. (2016). "Processing tone languages" in *Neurobiology of language* (San Diego: Elsevier), 1095–1107. doi: 10.1016/B978-0-12-407794-2.00087-0
- Gandour, J. T., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357. doi: 10.1016/j.neuroimage.2004.06.004
- Godfroid, A., Lin, C.-H., and Ryu, C. (2017). Hearing and seeing tone through color: an efficacy study of web-based, multimodal Chinese tone perception training. *Lang. Learn.* 67, 819–857. doi: 10.1111/lang.12246
- Guenther, F. H., and Vladusich, T. (2012). A neural theory of speech acquisition and production. *J. Neurolinguistics* 25, 408–422. doi: 10.1016/j.jneuroling.2009.08.006
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am.* 107, 2711–2724. doi: 10.1121/1.428657
- Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *Int. Rev. Appl. Linguist. Lang. Teach.* 44, 103–124. doi: 10.1515/iral.2006.004
- Gullberg, M. (2022). "Studying multimodal language processing" In: *The Routledge handbook of second language acquisition and psycholinguistics*. eds. A. Godfroid and H. Hopp. 1st ed (New York: Routledge), 137–149. doi: 10.4324/9781003018872-14
- Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J., and Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers. *Front. Psychol.* 8, 1–15. doi: 10.3389/fpsyg.2017.02051
- Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *J. Phon.* 40, 269–279. doi: 10.1016/j.wocn.2011.11.001
- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137, 297–315. doi: 10.1037/a0022128
- Huang, T., and Johnson, K. (2011). Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. *Phonetica* 67, 243–267. doi: 10.1159/000327392
- Jasmin, K., Sun, H., and Tierney, A. T. (2020). Effects of language experience on domain-general perceptual strategies bioRxiv. doi: 10.1101/2020.01.02.892943
- Jiang, Y. (2017). Examining the auditory approach: lexical effects in the perceptual judgment of Chinese L2 tone production. *Chin. a Sec. Lang. Res.* 6, 225–250. doi: 10.1515/caslar-2017-0010
- Kaan, E., Wayland, R., and Keil, A. (2013). Changes in oscillatory brain networks after lexical tone training. *Brain Sci.* 3:2. doi: 10.3390/brainsci3020757
- Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005
- Lacombe, N., Dias, T., and Petitpierre, G. (2022). Can gestures give us access to thought? A systematic literature review on the role of co-thought and co-speech gestures in children with intellectual disabilities. *J. Nonverbal Behav.* 46, 119–136. doi: 10.1007/s10919-022-00396-4
- Ladefoged, P., and Johnson, K. (2015). A course in phonetics. 7th Edn. Stamford, CT: Cengage Learning.
- Lakoff, G., and Johnson, M. (2017). Metaphors we live by. Chicago: University of Chicago Press.
- Leather, J. (1990). "Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers" In: *New Sounds 90: Proceedings of the Amsterdam symposium on the Acquisition of Second Language Speech*. eds. J. Leather and A. James. (Amsterdam: University of Amsterdam), 305–341.
- Leather, J. (2011). *Interrelation of perceptual and productive learning in the initial acquisition of second-language tone*. New York: De Gruyter Mouton. 75–102. doi: 10.1515/9783110882933.75
- Lee, A. H., and Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Stud. Second. Lang. Acquis.* 38, 35–64. doi: 10.1017/S0272263115000194
- Lewis, T. N., and Kirkhart, M. W. (2022). "Researching the effect of gestures on the learning and retention of vocabulary in a naturalistic setting" In: *Gesture and multimodality in second language acquisition: a research guide*. eds. G. Stam and K. Urbanski. 1st ed (New York: Routledge). doi: 10.4324/9781003100683
- Li, P., Baills, F., and Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Stud. Second. Lang. Acquis.* 42, 1015–1039. doi: 10.1017/S0272263120000054
- Liu, L., Lai, R., Singh, L., Kalashnikova, M., Wong, P. C. M., Kasisopa, B., et al. (2022). The tone atlas of perceptual discriminability and perceptual distance: four tone languages and five language groups. *Brain Lang.* 229:105106. doi: 10.1016/j.bandl.2022.105106
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am.* 96, 2076–2087. doi: 10.1121/1.410149
- Macedonia, M., and Kepler, J. (2013). Three good reasons why foreign language instructors need neuroscience. *J. Stud. Educ.* 3:1. doi: 10.5296/jse.v3i4.4168
- Maddieson, I. (2013). "Tone" in *The world atlas of language structures online*. WALS online (v2020.3) [data set]. eds. M. Dryer and M. Haspelmath. Available at: <https://zenodo.org/record/7385533>
- Mayer, R. E. (2005). "Cognitive theory of multimedia learning" In: *The Cambridge handbook of multimedia learning*. ed. R. Mayer. (Cambridge: Cambridge University Press), 31–48. doi: 10.1017/CBO9780511816819.004
- Mayer, R. E. (2009). *Multimedia learning*. New York, NY: Cambridge University Press.
- Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learn. Instr.* 29, 171–173. doi: 10.1016/j.learninstruc.2013.04.003
- Mayer, R. E., and Moreno, R. (1998). A cognitive theory of multimedia learning: implications for design principles. *J. Educ. Psychol.* 91, 358–368.
- McCafferty, S. G. (2004). Space for cognition: gesture and second language learning. *Int. J. Appl. Linguist.* 14, 148–165. doi: 10.1111/j.1473-4192.2004.0057m.x
- Morett, L. M. (2023). Observing gesture at learning enhances subsequent phonological and semantic processing of L2 words: an N400 study. *Brain Lang.* 246:105327. doi: 10.1016/j.bandl.2023.105327
- Morett, L. M., and Chang, L.-Y. (2015). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Lang. Cogn. Neurosci.* 30, 347–353. doi: 10.1080/23273798.2014.923105

- Morett, L. M., Feiler, J. B., and Getz, L. M. (2022). Elucidating the influences of embodiment and conceptual metaphor on lexical and non-speech tone learning. *Cognition* 222:105014. doi: 10.1016/j.cognition.2022.105014
- Paivio, A. (1991). Dual coding theory: retrospect and current status. *Can. J. Psychol.* 45, 255–287. doi: 10.1037/h0084295
- Paivio, A. (2014a). “Bilingual dual coding theory and memory” in *Foundations of bilingual memory*. eds. R. R. Heredia and J. Altarriba (New York: Springer), 41–62.
- Paivio, A. (2014b). Intelligence, dual coding theory, and the brain. *Intelligence* 47, 141–158. doi: 10.1016/j.intell.2014.09.002
- Pelzl, E., Liu, J., and Qi, C. (2022). Native language experience with tones influences both phonetic and lexical processes when acquiring a second tonal language. *J. Phon.* 95:101197. doi: 10.1016/j.wocn.2022.101197
- Pisoni, D. B., and Lively, S. E. (1995). “Variability and invariance in speech perception: a new look at old problems in perceptual learning” In: *Speech perception and linguistic experience: issues in cross-language research*. ed. W. Strange. (Timonium, MD: York Press), 433–459.
- Saito, K. (2021). “Effects of corrective feedback on second language pronunciation development” In: *The Cambridge handbook of corrective feedback in second language learning and teaching*. eds. E. Kartchava and H. Nassaji (Cambridge: Cambridge University Press), 407–428.
- Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006
- Shapiro, L. (2019). *Embodied cognition*. 2nd Edn. . New York, NY: Routledge.
- Wang, Y., Behne, D. M., Jongman, A., and Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Appl. Psycholinguist.* 25, 449–466. doi: 10.1017/S0142716404001213
- Wang, Y., Jongman, A., and Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc. Am.* 113, 1033–1043. doi: 10.1121/1.1531176
- Wang, T., Potter, C. E., and Saffran, J. R. (2020). Plasticity in second language learning: the case of Mandarin tones. *Lang. Learn. Dev.* 16, 231–243. doi: 10.1080/15475441.2020.1737072
- Wang, Y., Sereno, J. A., and Jongman, A. (2006). “L2 acquisition and processing of Mandarin tones” in *The handbook of east Asian psycholinguistics: volume 1: Chinese*. eds. P. Li, L. H. Tan, E. Bates and O. J. L. Tzeng (Cambridge: Cambridge University Press), 250–256.
- Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am.* 106, 3649–3658. doi: 10.1121/1.428217
- Wayland, R. P., and Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: a preliminary report. *Lang. Learn.* 54, 681–712. doi: 10.1111/j.1467-9922.2004.00283.x
- Xi, J., Zhang, L., Shu, H., Zhang, Y., and Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience* 170, 223–231. doi: 10.1016/j.neuroscience.2010.06.077
- Xie, X., and Myers, E. (2015). The impact of musical training and tone language experience on talker identification. *J. Acoust. Soc. Am.* 137, 419–432. doi: 10.1121/1.4904699
- Xu, Y., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Tong, Y., et al. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum. Brain Mapp.* 27, 173–183. doi: 10.1002/hbm.20176
- Yip, M. (2002). “Introduction” in *Tone* (Cambridge: Cambridge University Press), 1–16.
- Yu, K., Zhang, J., Li, Z., Zhang, X., Cai, H., Li, L., et al. (2024). Production rather than observation: comparison between the roles of embodiment and conceptual metaphor in L2 lexical tone learning. *Learn. Instr.* 92:101905. doi: 10.1016/j.learninstruc.2024.101905
- Zhang, Y., Ding, R., Frassinelli, D., Tuomainen, J., Klavinskis-Whiting, S., and Vigliocco, G. (2023). The role of multimodal cues in second language comprehension. *Sci. Rep.* 13:20824. doi: 10.1038/s41598-023-47643-2
- Zheng, A., Hirata, Y., and Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 Mandarin tone production. *J. Speech Lang. Hear. Res.* 61, 2179–2195. doi: 10.1044/2018_JSLHR-S-17-0481
- Zhou, A., and Olson, D. (2023). The use of visual feedback to train L2 lexical tone: evidence from Mandarin phonetic acquisition. *Pronun. Sec. Lang. Learn. Teach. Proc.* 13:1. doi: 10.31274/psllt.15715