Journal of Linguistics & Cultural Studies

Review

Practical Dilemmas and Breakthrough Paths in Teaching Chinese Tones in International Chinese Language Education

Yijie Wang[1,*]

[1] School of Humanities, Jianghan University, Wuhan, China
* Correspondence: Yijie Wang, School of Humanities, Jianghan University, Wuhan, China

Abstract: Tone sits at the core of the Chinese phonological system, and how to teach it has long been one of the central challenges in international Chinese language instruction. Research on tone acquisition now stretches back more than three decades, yet classroom practice continues to grapple with a cluster of difficulties: teachers' limited professional knowledge, the absence of connected-speech training in textbooks, learners' negative L1 transfer together with emotional anxiety, restricted instructional hours, and an underuse of technology. This paper reviews these practical dilemmas in a systematic way and, drawing on recent work in perceptual training, multimodal teaching, visual feedback, and contextualized instruction, sets out six paths forward-perception first, multimodal coordination, technological support, contextual integration, autonomous extension, and emotional support. The aim is to narrow the divide between tone acquisition research and classroom practice, and to offer frontline teachers a reference that combines systematic grounding with practical applicability.

Keywords: tone teaching; practical dilemmas; multimodal teaching; perceptual training; international Chinese language education

1. Introduction

Tone is the most distinctive suprasegmental element in the Chinese phonological system and the feature that most clearly sets Chinese apart from non-tonal languages, including those of the Indo-European family [1]. For learners, tonal accuracy matters in two ways. It bears directly on meaning: bēizi (cup) and bèizi (quilt), or shuìjiào (to sleep) and shuǐjiǎo (dumplings), differ by only one tone yet mean quite different things. Beyond that, it heavily shapes how native speakers judge their Chinese proficiency, since the familiar "foreign accent" stems largely from tonal errors [2]. Tone teaching has consequently long been treated as both a key and a difficult component of international Chinese pronunciation instruction [3].

A closer look at the literature, though, reveals a persistent gap. Over more than three decades, the field has built up substantial findings on the description of tonal errors, the order in which they are acquired, and the causes behind them; even so, much of this work centers on static phenomena such as citation tones and leans toward experimental analysis rather than classroom application, leaving genuinely usable methods-ones that teachers can take directly into daily instruction-still in short supply [4] [5],. Learners face a parallel problem outside class: without well-founded practice routes and timely feedback, further progress is hard to come by. Once tonal errors take hold, they fossilize readily, weakening communicative efficiency and producing real discouragement. Starting from frontline practice, then, this paper examines the concrete dilemmas tone teaching presents across four levels-teachers, textbooks, learners, and the instructional environment-and, integrating findings on perceptual training, multimodal teaching, and visual feedback, looks for workable paths forward. Its purpose is to bridge the gap between tone acquisition research and classroom teaching, and to provide a reference for tone instruction within international Chinese language education.

Received: 16 August 2026
Revised: 23 September 2026
Accepted: 07 October 2026
Published: 13 October 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

2. Practical Dilemmas in Tone Teaching

Why tone teaching is difficult cannot be attributed to the complexity of tones alone. Teachers, textbooks, learners, and the instructional environment all contribute, and the difficulties they create are closely interwoven. Working through these practical dilemmas is a necessary first step before any path forward can be identified.

2.1. Teachers: Insufficient Professional Knowledge and a Uniform Teaching Approach

Teachers carry the main responsibility for transmitting tonal knowledge. Survey data indicate that learners pick up tones almost entirely from their teachers rather than from textbooks [6], a finding that both highlights how central teachers are and places considerable demands on their professional competence. In practice, though, many teachers have had little systematic phonetic training and lack a firm conceptual grasp of how tones are articulated, what their contours look like, and how pitch range works. One survey reports that as many as 95.3% of learners could not identify the speech organs involved in tone production, while 96.9% did not understand how the articulation of the four tones differs [6] [2],. Because tones can be neither seen nor touched, describing their articulatory essentials in words is often difficult, and instruction tends to settle into a single "model and imitate" pattern, with feedback that stays general rather than becoming specific. Tone work is also concentrated largely in the initial pronunciation stage; once vocabulary and grammar take over, many teachers push tones to the margins and fail to sustain tone awareness across the course, leaving learners' persistent problems with little chance of correction.

2.2. Textbooks: Scattered Organization and the Absence of Connected-Speech Training

Textbooks provide an important basis for instruction, yet their organization of tonal content has clear shortcomings. For one thing, tonal knowledge is largely confined to a handful of chapters in the pronunciation stage, presented sporadically and revisited only rarely. Later textbooks generally contain no dedicated sections for tone review or transition, which makes it difficult to support the shift of instructional focus from citation tones toward tones in connected speech. Having reviewed the Chinese textbooks currently in wide use, the author found that most provide dedicated chapters and training for monosyllabic tones, whereas for disyllabic tones they offer neither independent sections nor systematic explanations of errors or corrective design. For another, textbook exercises are dominated by mechanical training such as read-aloud and imitation, with insufficient appeal and communicative value. They mostly remain at the level of citation tones and disyllabic words and pay inadequate attention to tones and intonation at the sentence and discourse levels; tone sandhi rules, such as the sandhi of yī (one) and bù (not) and third-tone sandhi, are also insufficiently recycled. A study of textbooks found that intonational content is generally absent from many textbooks, even though learners' "foreign accent" is closely related to inadequate intonation training [7].

2.3. Learners: Negative L1 Transfer, Weak Perception, and Emotional Anxiety

Learners' own difficulties first show up as negative first-language transfer. Those whose mother tongue is a non-tonal language generally lack pitch awareness, find tonal categories hard to establish, and tend to interpret and replace tones through stress and intonation [8]. Learners whose mother tongue is a tonal language, although equipped with some tonal experience, remain susceptible to interference from their native tone system and produce systematic errors [9]. Research indicates that, whatever the language background, perceptual confusion between Tone 2 and Tone 3 is a difficulty learners share [10]. Many cannot perceive tonal differences accurately and, understandably, then struggle to correct themselves. After-class practice often lacks a well-founded pathway and timely feedback: learners "cannot hear the difference, cannot correct the error, and cannot keep practicing." Errors left uncorrected tend to entrench and, at the advanced stage, may even move toward fossilization [11]. Repeated tonal mistakes also readily give rise to apprehension and frustration; learners may hold back from speaking for fear of reading tones incorrectly, which makes the learning difficulty still greater.

2.4. Environment: Limited Instructional Hours and Inadequate Use of Technology

From the standpoint of the instructional environment, tone training is constrained to a considerable degree by class hours. Tone teaching is usually compressed into the first one or two months of Chinese study-a short period carrying a heavy load-and is seldom arranged systematically afterward, which makes continued development of tonal ability difficult to ensure. The spread of online teaching has added new challenges: limited by network sound quality and modes of interaction, online tone correction often falls short of offline classrooms in immediacy and precision. Visual speech analysis software such as Praat has been shown to benefit tone teaching, yet its use in frontline classrooms remains limited. Constrained by technical barriers and equipment conditions, most teachers cannot use it effectively, and the support technology might offer tone teaching has yet to be fully realized.

The dilemmas of tone teaching run across teaching, learning, materials, and environment, and they are interrelated as well as mutually reinforcing. Overcoming them therefore cannot rest on improving any single link; what is needed instead is a comprehensive solution arrived at through systematic analysis.

3. Recent Advances in Tone Acquisition Research and Their Pedagogical Implications

Any serious attempt to resolve these dilemmas has to return to tone acquisition research for its theoretical grounding. Recent years, driven by developments in experimental phonetics, cognitive psychology, and educational technology, have produced a steady stream of new findings in perceptual training, multimodal teaching, visual feedback, and contextualized instruction, and these carry real implications for how teaching might be improved.

3.1. Perceptual Training: Driving Production through Auditory Discrimination

The relation between perception and production runs through the center of this research. Whether one precedes the other remains an open question, but the field largely agrees that perceptual accuracy feeds directly into production: learners who cannot hear tonal differences rarely manage to produce them correctly [12]. Studies of categorical perception add a further point, showing that native speakers and second-language learners build tonal categories along systematically different lines. Learners from non-tonal backgrounds tend to rely on cues such as mean pitch and are distinctly less sensitive to the slope of the pitch contour [13]. The implication for teaching is that the traditional priority given to production over perception ought to be reversed, with perceptual training moved forward [14]. Minimal pairs and AXB discrimination provide useful paradigms for this, helping learners form tonal categories that remain stable and allowing listening, in a genuine sense, to drive speaking.

3.2. Multimodal Teaching: Multi-Sensory Collaborative Intervention

Because tones are abstract and resist verbal description, multimodal teaching calls several channels-visual, kinesthetic, and auditory-into coordinated input. Abstract pitch can be made concrete visually through color coding, moving contours, and staff-notation diagrams; kinesthetically, the TPR gesture method traces tonal change through the rise and fall of hand movements [15]; and on the auditory side, exaggerated models together with correct–incorrect contrasts reinforce perception. Empirical work suggests that multimodal scaffolding lowers cognitive load while strengthening both motivation and confidence in pronunciation. Li [16] conducted a 12-week comparative experiment with primary-school learners in Thailand. The experimental class, taught through a three-dimensional model combining visual, kinesthetic, and auditory modes, raised its overall tone-test scores by 66.2%, clearly outperforming the control class in both discrimination and production accuracy, and recurring errors such as a low Tone 1 and an incompletely realized Tone 4 were also substantially reduced. Taken together, these results attest to the effectiveness of multi-sensory collaborative intervention.

3.3. Visual Feedback: Technology-Assisted Error Correction

For the problem that tones can be neither seen nor touched, visual feedback offers a technological route. Speech analysis software such as Praat presents the F0 contour in real time, letting learners see directly how their own tones differ from the standard model and then adjust themselves in targeted ways [2]. Mobile speech applications, portable and immediate, make independent after-class practice easier to sustain. Multimodal articulatory research based on ultrasound tongue imaging goes further still, rendering in motion physiological mechanisms such as tongue position during pronunciation [17]. At present this work serves mainly research and teacher training, but it also opens up new possibilities for demonstrating the mechanics of tone articulation more deeply in the classroom.

3.4. Contextualized Instruction: From Citation Tones to Connected Speech

Where citation-tone teaching falls short is in handling real communication, and researchers have accordingly proposed a shift from the static to the dynamic and from isolated syllables to connected speech. Xu [18] notes that citation-tone teaching cannot accommodate the phonetic changes that arise in actual speech, and that overemphasizing such tones yields half the result for twice the effort; she argues for moving the teaching focus toward the chunk tones of disyllabic words and phrases. Tiered instruction makes a closely related point, stressing a gradual sequence from citation syllable through disyllabic word and sentence to discourse, with tones consolidated through genuine communicative tasks [2]. Teaching on this basis must do more than train syllable tones: tone sandhi, neutral tones, and intonation, as prosodic phenomena of connected speech, call for comparable attention if tone learning is genuinely to align with communication. Across the dimensions of listening, seeing, moving, and using, then, recent research opens several lines for tone teaching, and they converge on a single orientation-perception as the foundation, multimodality as the means, technology as the support, and context as the destination.

4. Breakthrough Paths for Tone Teaching

The dilemmas set out above call for a coordinated response, and recent work on tone acquisition suggests the direction it might take. What is needed is an approach organized around six elements-perception first, multimodal coordination, technological support, contextual integration, autonomous extension, and emotional support-and built into a pedagogical path that teachers can both put into practice and sustain over time.

4.1. Perception First: Building a Tone Perception Training System

The long-standing emphasis on read-aloud practice at the expense of auditory discrimination ought to be reversed at its root, so that perception becomes the starting point of tone teaching. Minimal-pair drills built around the four tones offer a workable means: sets such as mā (mom)-má (hemp)-mǎ (horse)-mà (to scold), or bēizi (cup)-bèizi (quilt), lead learners to track differences in contour and pitch range. Experimental paradigms such as AXB discrimination can be adapted for classroom use, supporting activities in which learners identify a tone from a recording or choose between alternatives after judging sameness and difference; through these, tonal categories that are clear and stable gradually take shape. The widespread confusion between Tone 2 and Tone 3 calls for particular attention. Contrasting the position of the turning point and the starting pitch of the two lets learners first hear the distinction and then produce it, so that listening genuinely carries speaking forward.

4.2. Multimodal Coordination: Engaging Multiple Senses

The visual, kinesthetic, and auditory channels all deserve full use if abstract pitch is to be turned into forms learners can perceive. Simplified five-level diagrams, color coding, and dynamic contours can trace the trajectory of each of the four tones. Gestures support kinesthetic memory in a comparable way: the hand moves horizontally for Tone 1, rises toward the upper right for Tone 2, dips and then rises to mark a turning point for Tone 3, and swings downward for Tone 4. Exaggerated reading can accompany these movements

to widen the pitch range, which is then drawn back to a natural level once muscular memory has formed. Teacher modeling and correct–incorrect contrasts reinforce perception on the auditory side. Games can then build on this groundwork-tone matching, tone Bingo, and tone chains-moving the work away from mechanical imitation toward multi-sensory participation, lowering cognitive load as enjoyment rises.

4.3. Visual Feedback: Establishing an Immediate Error-Correction Mechanism

Technology has a genuine contribution to make to the problem of tone feedback. In the classroom, software such as Praat can display the F0 contour of learners' pronunciation in real time, so that deviations from the standard model in contour and pitch range show up directly and adjustments can be aimed at the point that is off. Where technological resources are limited, gestures and curves drawn on the blackboard serve a similar visual purpose without any software. After class, learners can be directed toward speech applications with tone recognition and scoring functions; these return immediate feedback on the tone of individual syllables, and some add spaced repetition aimed at the syllables that cause difficulty, doing much to compensate for feedback the classroom cannot supply.

4.4. Contextual Integration: Moving from Citation Tones to Real Communication

Citation tones, treated in isolation, will carry learners only so far, and tone teaching needs to reach past them-from the static toward the dynamic and from individual words toward discourse. The chunk tones of disyllabic words and phrases make a productive focus, since they let learners feel how tones combine within a larger unit. The gradual sequence from citation syllable through disyllabic word and sentence to discourse still holds, but rules are best trained within specific words and sentences rather than memorized on their own: the sandhi of yī (one) and bù (not), third-tone sandhi, and neutral tones all work this way. Tone and intonation, meanwhile, ought to be taught in coordination. Role play, situational dialogue, and story retelling provide real tasks in which learners put tones to meaningful use, bringing tone learning genuinely into line with communication.

4.5. Autonomous Extension: Building a Sustainable Practice Mechanism

The problem that learners often do not know how to practice after class, and then find it hard to persist, merits attention in its own right. Tiered tasks offer one way to address it. At the basic level comes read-aloud imitation and tone discrimination; at a higher level, learners record themselves and compare the recording with the standard model; and at the challenge level, they express themselves freely around a given topic. Feedback is most effective when several sources are combined-automated application scoring, peer evaluation, and periodic teacher checks-so that the loop remains varied rather than fixed. Recording check-ins and learning portfolios document the process, while the display of incremental progress and the organization of tone showcase activities help sustain motivation, extending practice beyond the classroom and turning short-term training into a long-term habit.

4.6. Emotional Support: Reducing Anxiety and Building Confidence

Learners' emotional states deserve a place alongside the tones themselves. Teachers can cultivate a relaxed classroom atmosphere in which errors are allowed and treated rationally, and offer feedback that is specific and positive rather than simply negative. Showing learners their incremental progress in tone work and sharing cases of successful learning help build confidence in pronunciation and ease the anxiety that holds speech back for fear of a mistake. Once emotional security is joined to the sense of achievement that progress brings, both the initiative to practice and the capacity to keep practicing tend to grow.

The six paths, it is worth stressing, do not stand apart from one another; they offer mutual support and take effect together. Perception and multimodality address whether learners understand what they are studying; visual feedback and contextualization address whether they can practice it accurately and put it to use; and the autonomous

mechanism together with emotional support address whether they can continue. From classroom work into out-of-class practice, from cognition into affect, the six form a single integrated system-and a realistic, feasible basis on which the dilemmas of tone teaching can be overcome.

5. Conclusion

That tone teaching is difficult stems from the way factors involving teachers, textbooks, learners, and the instructional environment interweave, and from this it follows that no single area of improvement, taken on its own, will be enough. Against a systematic review of the practical dilemmas that tone teaching presents, and drawing on recent work in tone acquisition-across perceptual training, multimodal teaching, visual feedback, and contextualized instruction-this paper has put forward a coordinated set of breakthrough paths: perception first, multimodal coordination, technological support, contextual integration, autonomous extension, and emotional support, all aimed at narrowing the divide between research on tone acquisition and what happens in the classroom. It is worth noting, though, that the discussion here rests largely on literature review and teaching observation; the paths proposed still await empirical testing across a wider range of teaching settings. Looking ahead, future research might turn to long-term longitudinal studies of tone acquisition, the development of tone resources and toolkits tailored to frontline teaching, and closer attention to the training of teachers' competence in pronunciation instruction-steps that, taken together, can help raise the quality of tone teaching in international Chinese language education on a continuing basis.

References

1. J. M. Zhao, "From tones in some tone languages to Chinese tones," in *Selected Papers from the Second International Symposium on Chinese Language Teaching*, Beijing, China: Beijing Language Institute Press, 1987, pp. 171–181.
2. Y. D. Song, "Strategies for teaching Chinese tones as a foreign language," *Language Teaching and Linguistic Studies*, no. 3, pp. 48–53, 2009.
3. Q. H. Du, "On tone teaching in modern Chinese," *Language Teaching and Linguistic Studies*, no. 1, pp. 75–81, 1992.
4. Y. Q. Chi, "A review of research on teaching Chinese tones as a foreign language," *Journal of PLA University of Foreign Languages*, vol. 28, no. 1, pp. 55–58, 2005.
5. B. Yi and J. Liang, "A review of research on the acquisition of Chinese tones as a second language," *Journal of Tianjin Normal University (Social Science Edition)*, no. 2, pp. 77–80, 2010.
6. J. Guan, "A preliminary inquiry into the reform of tone teaching," *Language Teaching and Linguistic Studies*, no. 4, pp. 51–54, 2000.
7. X. H. Zhu, "On the construction and implementation of a Chinese pronunciation teaching system," 2019. [Online]. Available: https://www.linglab.cn/news/312
8. L. J. Zhang, "Phonological and acoustic processing of Chinese tones by American learners," *Chinese Teaching in the World*, vol. 25, no. 2, pp. 268–275, 2011, doi: 10.13724/j.cnki.ctiw.2011.02.010.
9. H. J. Zhang, "A review of research on the acquisition of Chinese tones as a foreign language," *Foreign Language and Literature*, vol. 28, no. S1, pp. 66–68, 2012.
10. X. Y. Ge, X. X. Wei, S. Abudouhelili, et al., "Research progress on Mandarin tone perception by L2 Chinese learners across regions," *Overseas English*, no. 11, pp. 48–50, 2026.
11. C. N. Lê and T. K. L. Trần, "Difficulties in tone teaching for Vietnamese learners of Chinese and the application of visualization tools," *HCMUE Journal of Science*, vol. 22, no. 8, pp. 1507–1517, 2025, doi: 10.54607/hcmue.js.22.8.5009(2025).
12. Y. Lu, "Research methods, level classification, and future prospects of categorical perception of Mandarin tones," *Journal of Jilin University (Social Sciences Edition)*, vol. 66, no. 4, pp. 235–244, 2026.
13. Y. M. Ci, Y. X. Ge, Q. Zheng, et al., "Individual differences in the sub-lexical perception and production of L2 Mandarin tones," *Chinese Journal of Applied Linguistics*, vol. 49, no. 3, pp. 336–365, 2026.
14. R. Hao, "On musical perception and tone teaching in Chinese as a foreign language," *Forum on Chinese Culture*, no. 2, pp. 185–190, 2015.
15. J. Yu, "A new teaching plan for tones," *Language Teaching and Linguistic Studies*, no. 1, pp. 77–81, 2007.
16. Z. Y. Li, "The application of multimodal scaffolding in tone acquisition for beginning Chinese learners: A case study of Assumption College Primary Section, Thailand," M.A. thesis, Guangxi Normal University, Guilin, China, 2026, doi: 10.27036/d.cnki.ggxsu.2026.002324.
17. J. X. Zhang, H. Z. Yu, and J. Zhang, "A multimodal articulatory physiological dataset of Mandarin Chinese based on ultrasound tongue imaging," *China Scientific Data*, vol. 11, no. 2, pp. 571–590, 2026.

18. D. Xu, "Some reflections on Chinese tones," *Nankai Linguistics*, no. 2, pp. 11–23, 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.