Perceptual Cue Weighting of Mandarin High Vowels by Native Japanese Speakers

Wenhui Zhu[1] and Sun-Hee Lee[2,*]

[1]School of Humanity and Science, Southern University of Science and Technology, Shenzhen, China
[2]College of Chinese Studies, Cyber Hankuk University of Foreign Studies, Seoul, Republic of Korea
Email: zhuwh@sustech.edu.cn (W.H.Z.); lishanxi@cufs.ac.kr (S.H.L.)
*Corresponding author
Manuscript received August 20, 2025; revised December 10, 2025; accepted February 27, 2026.

***Abstract*—This study investigates native Japanese speakers' perception of Mandarin high vowels, demonstrating that learners exhibit weighting of the second formant and third formant frequency cues in the perception. Twenty-seven native Japanese participants performed an XAB perceptual categorization task on synthesized stimuli. The results showed that all participants significantly relied on the F2 frequency, which is associated with the backness feature, in their perception of the stimuli; whereas they did not exhibit a significant reliance on the F3 frequency, related to the roundness feature. The result suggests that the backness feature, not roundness feature, is salient in learners' perception due to the characteristics of Japanese vowels system. The study provides new data for the field of perceptual cue weighting by examining the weighting of phonetic cues and distinctive features in Mandarin high vowels by native Japanese learners. The outcomes are also educationally significant, underscoring the necessity for tailored perception training for learners from various linguistic backgrounds to improve their acquisition of target second language sounds.**

***Keywords*—perceptual cue weighting, distinctive feature, vowel formant, Japanese, mandarin Chinese**

I. Introduction

Asymmetry is a universal phenomenon, and the origin of language is closely linked to the asymmetry of the brain [1]. In the field of speech acquisition, research on asymmetry has long been a focal point. Human infants exhibit categorical perception of speech sounds and gradually adjust to a language-specific perceptual pattern [2] [3], indicating that the innate neural mechanisms of speech acquisition retain plasticity under the influence of linguistic experience. Although the challenges of L2 speech acquisition in adults are well-known, this neural plasticity provides critical theoretical support for successful L2 speech acquisition. Two highly influential theories in L2 speech acquisition: the Speech Learning Model (SLM) [4, 5], and the Perceptual Assimilation Model (PAM) [6, 7], both acknowledge that adult L2 learners retain the ability to acquire cues, offering a theoretical foundation for applying cue weighting theory to L2 phonetic learning.

Perceptual cue weighting is a type of the asymmetry in speech acquisition. Broadly defined, a cue refers to any factor that systematically influences speech perception. Among these, research on acoustic cue weighting has garnered significant attention due to the experimental methods. The acoustic cue weighting theory posits that learners weigh and integrate different acoustic cues when acquiring speech sounds, and this ability is crucial for successful speech acquisition [8]. The process of L1 speech acquisition involves precisely this weighing and integration of acoustic cues. As described by [9]'s 'desensitization' and [10]'s 'Native Language Magnet' theories, infants transition from a general auditory processing mode to a language-specific phonetic pattern, becoming more reliant on cues present in their native language while gradually losing sensitivity to cues absent in it, ultimately forming their L1-specific phonetic patterns. In contrast, L2 learners exhibit distorted phonetic patterns shaped by prior linguistic experience, making it difficult for them to attend to L2-specific cues that do not exist in their L1s. This leads to significant challenges in distinguishing non-native phonemic contrasts.

II. Literature Review

A. Acoustic Cue Weighting

Numerous studies have documented asymmetries in the acoustic cues that L2 learners and native speakers rely on when acquiring L2 target sounds. A classic example is the asymmetry in acoustic cues (duration vs. spectral information) for the English vowel contrast /i/-/ɪ/ [11–15]. It is found that L2 learners tend to rely more on temporal cues than on spectral cues to distinguish the /i/-/ɪ/ contrast, whereas native English-speaking listeners place more weight on spectral cues. For instance, [13] found native Japanese speakers of English categorized /i/ and /ɪ/ based on temporal cues through a cue-weighting task. [15] trained Chinese speakers, who have a tendency to rely on duration cue for the /i/-/ɪ/ distinction to more native-like cue weighting in both perception and production of the /i/-/ɪ/ contrast. In this two research, native Japanese speakers' weight on temporal cues was explained as vowel duration is used phonologically in Japanese. The influence of the L1 is crucial in explaining perceptual cue weighting. [16] also argued that because Spanish has phonemic length contrasts, learners transfer this reliance on duration to their perception of English tense-lax vowel distinctions. However, not all perceptual weighting patterns can be attributed to L1 transfer. Mandarin Chinese lacks vowel length contrasts, but Mandarin speakers often rely more on duration when perceiving many L2 vowels. As a result, duration has been interpreted as a more universally salient feature [17].

Overall, research on acoustic cue weighting has been systematic and comprehensive, covering target languages such as English, European, Asian, African languages, and some Chinese dialects. However, there remains a lack of research on Mandarin Chinese acoustic cue weighting. Most research on Chinese L2 speech acquisition still centers on error description and explanation, highlighting an urgent

need for studies that address L2 learners' phonetic difficulties from the perspective of acquisition mechanisms.

B. Study on Chinese Mandarin and Japanese Vowels

In Mandarin Chinese, there are five to six[1] monophthong vowels: three high vowels /i/, /y/, /u/; one mid vowel /ɤ/; and one low vowel /a/. Among the three high vowels, /i/ and /y/ are front high vowels, distinguished by lip rounding (/i/ is unrounded, while /y/ is rounded). Meanwhile, /u/ and /y/ are both high rounded vowels but differ in backness (/y/ is front, whereas /u/ is back). Please refer to Table 1 for details.

In contrast, Japanese has five vowel phonemes: /a/, /i/, /u/, /e/, and /o/. Among these, /i/ and /u/ are high vowels, /e/ and /o/ are mid vowels, and /a/ is a low vowel. Japanese vowel phonemes are distinguished by two primary distinctive features: height (high / mid / low) and backness (front / central / back) [18–20].

Table 1. Mandarin and Japanese high vowels

	Front		**Back**	
	Unrounded	Rounded	Unrounded	Rounded
Mandarin	i	y		u
Japanese	i イ		ɯウ	

For Japanese native speakers, the greatest challenge in acquiring Mandarin Chinese vowels lies in mastering the vowel /y/. [21] found that Japanese learners exhibited the highest error rate in perceiving the Mandarin vowel /y/, primarily misidentifying it as /i/ (error rate: 44%), followed by /u/ (error rate: 28%). In Teaching Chinese as Foreign Language (TCFL) classrooms, Japanese learners frequently struggle with distinguishing and producing /y/, often confusing it with /i/ and failing to achieve sufficient lip rounding in articulation. While existing research on L2 speech acquisition is abundant, limited studies have explored these errors from the perspective of perceptual cue weighting. Investigating Japanese learners' acquisition of Mandarin /y/ will provide new linguistic data to address this research gap and contribute to a deeper understanding of cross-linguistic perceptual strategies in vowel acquisition.

III. Materials and Methods

A. Stimuli

Following [12] and [17], the experimental stimuli consisted of synthesized vowels (refer to Fig. 1). The top-left and bottom-right corners of the stimulus matrix shown in Fig. 1 represent the most /u/-like and /y/-like instances, respectively. These two stimulus values were based on the average formant frequencies of Mandarin vowels /y/ and /u/ produced by male speakers [22]. Please refer to Table 2. As illustrated in Fig. 1, these stimuli varied along the second formant (F2) (x-axis) and the third formant (F3) (y-axis) frequencies to create a two-dimensional stimulus grid with equidistant frequency steps on the mel scale. The 24 stimuli were synthesized using Praat [23]. The six F2 steps, between the seven different F2 values, were equal along a mel scale (1 step = 127 mel). The six F3 steps, between the seven different F3 values, were equal along a mel scale (1 step = 27 mel). This design allowed for systematic manipulation of spectral cues to examine how listeners weigh F2 and F3 variations in perceiving the Mandarin /y/–/u/ contrast. The mel-scaled steps ensured perceptual equidistance, a crucial feature for investigating potential cue weighting across different learners.

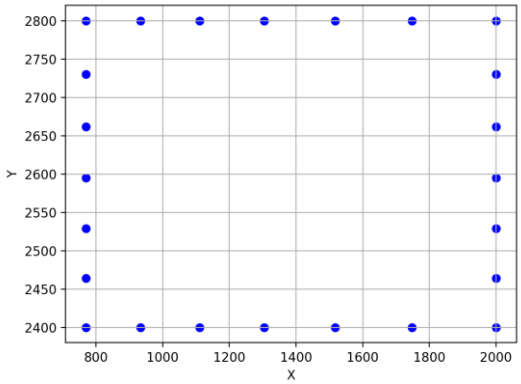

Fig. 1. F2 and F3 values of stimuli (Hz).

Table 2. Values of first three formants of mandarin vowels /u/ and /y/ (Hz) (Bao, 2004)

Vowel	**F1**	**F2**	**F3**
u	465	770	2800
y	340	2000	2400

B. Participant

A total of 27 native Japanese speakers participated in this experiment. All participants were beginner-level Chinese learners with average 90 hours of Mandarin learning experience. Their age was 21.5 on average. They were monolingual speakers with Japanese as their sole native language. All participants reported having normal hearing.

C. Procedure

The experiment adopted an XAB discrimination task where participants heard three vowel stimuli per trial and had to determine whether the first stimulus (X) sounded more like the second (A, representing /y/) or third (B, representing /u/) stimulus. The X stimulus was one of 24 target stimuli, while A and B corresponded to the most /y/-like (top-left corner in Fig. 1) and most /u/-like (bottom-right corner in Fig. 1) endpoints of the stimulus matrix respectively, requiring listeners to categorize each of the 24 stimuli as one of these two endpoints. A critical 1.2 millisecond inter-stimulus interval was implemented between the three stimuli to ensure language-specific phonological processing [17, 26].

During the experiment, each of the 24 target stimuli was presented as the "X" stimulus 10 times, with the presentation order of the second and third stimuli counterbalanced: for each X stimulus's 10 repetitions, the most /y/-like stimulus appeared as the first response option in 5 trials. This balancing of A and B stimuli resulted in 48 distinct XAB triplets, each presented five times for a total of 240 stimulus presentations. Participants responded by selecting either "A" (/y/) or "B" (/u/) while being tested in a quiet room. To ensure participants entered the desired monolingual mode, the

[1] There may be also a retrofex vowel phoneme /ɚ/ in Mandarin, but it has limited distribution and lacks of a clear phonetic description [24][25]

experimenter conducted a 5-minute conversation in Japanese before the perception test. Following this, the experimenter explained the procedure and conducted two oral practice trials of the XAB task using vowels other than /y/ and /u/. The XAB categorization task lasted approximately 30 minutes.

Following [17]'s study, we also measured participants' accuracy in identifying X stimuli representing prototypical Mandarin /u/ or /y/ (the stimuli at the top-left and bottom-right corners of Fig. 1's stimulus grid) firstly. Since each stimulus was repeated 10 times, the maximum correct responses per endpoint was 10, with each participant's accuracy for each endpoint calculated as a percentage score. Participants who failed to meet the 80% correct response criterion for classifying typical /u/ and /y/ stimuli were excluded from all analyses. It should be noted that the endpoint markers also served as response categories in the XAB task, meaning participants were expected to easily select the correct category when the X stimulus was identical to either stimulus A or B. Participants who couldn't reliably classify any vowel endpoint markers might have been fundamentally incapable of performing the vowel categorization task. All the participants passed the test.

This approach follows established psycholinguistic protocols that prioritize pure auditory processing in speech perception research, especially when investigating non-native sound contrasts that may not have direct orthographic equivalents in learners' L1. The auditory-only design ensures that participants' responses reflect their acoustic-phonetic processing rather than being influenced by potentially misleading orthographic representations.

IV. Result and Discussion

A. Result

We use logistic regression analysis to investigate participants' relative weighting of F2 and F3 spectral cues in perceiving the Mandarin /u/-/y/ vowel contrast by examining the influence of independent variables (stimulus F2 and F3 frequencies) on the binary dependent variable (vowel categorization as /u/ or /y/). Analysis of 27 participants' data revealed that F2 value significantly predicted vowel categorization probability ($p<0.0001$), while F3 value showed no statistically significant predictive effect. The goodness-of-fit analysis for univariate models demonstrated that the F2 model effectively explained outcome variance. In the F2 prediction model, the regression coefficient was 0.0055 (SE=0.00016), with Wald test confirming extremely strong statistical association ($z=34.17$). The corresponding odds ratio (OR=1.006, 95% CI[1.005,1.006]) indicated that each 1Hz increase in F2 value systematically increased the probability of perceiving the stimulus as /y/ by 0.6%. The model showed significantly better fit than the null model (residual deviance: 1257.0 vs. 4479.9; AIC=1261), explaining approximately 72% of variance (pseudo $R^2=0.72$). In contrast, the F3 model showed marginal significance ($p=0.060$) with minimal explanatory power (pseudo $R^2=0.0008$; AIC=4480.4). The intercept term in the F3 model was non-significant ($p=0.077$), while the F2 model's intercept was highly significant ($p<0.0001$), further supporting F2 model robustness. Please refer to Fig. 2.

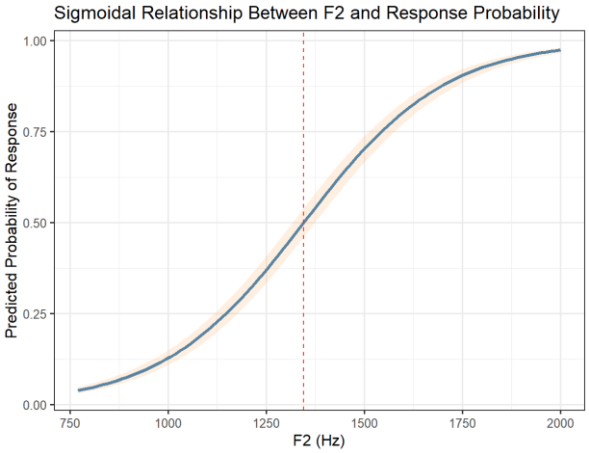

Fig. 2. Effect of F2 on the probability of selecting vowel /y/ stimuli.

Analysis of individual participants' data similarly demonstrated that F2 values significantly predicted the probability of selecting /y/ stimuli ($p < 0.05$), while F3 frequencies showed no statistically significant predictive effect. For detailed experimental data of each individual participant, please refer to Appendix 1.

B. Discussion

This study investigated native Japanese speakers' perception of Mandarin high vowels using synthesized stimuli. The results demonstrated that the 27 Japanese learners showed significant reliance on F2 frequency rather than F3 frequency when perceiving the vowel stimuli. Acoustic studies indicate that F2 frequency is closely related to the backness (front-back distinctive feature), while F3 frequency correlates with the lip rounding feature [27]. This suggests that Japanese native speakers exhibit higher sensitivity to the backness feature than to the lip rounding feature. These findings align with previous research by [28] and [29], which similarly found that Japanese native speakers pay greater attention to F2 frequency than F3 frequency, showing stronger sensitivity to the backness than to the roundness distinctive feature.

The experimental results can be explained by the phonological characteristics of Japanese, where vowel contrasts are not distinguished by lip rounding (see Table 1). Consequently, when perceiving Mandarin vowels, Japanese learners demonstrate lower sensitivity to the lip rounding feature and rely more heavily on the backness cue. This leads to relatively better performance in distinguishing the /y-u/ contrast (which differs primarily in backness) compared to the /y-i/ contrast (which differs primarily in roundness). This pattern of perceptual cue weighting stands in contrast to that of native English speakers [30], highlighting the significant influence of L1 phonology on L2 speech perception.

The study provides empirical evidence for how L1 phonological systems shape the perceptual strategies employed by L2 learners. The findings have important implications for understanding cross-linguistic speech perception and for developing more effective pronunciation training methods tailored to specific L1–L2 combinations. In current TCFL class, instructors typically begin by teaching the relatively easier vowel /i/, subsequently adding lip-rounding to produce /y/. This pedagogical approach decomposes the complex articulatory gesture of "tongue

frontness + lip rounding" while emphasizing the roundness feature. However, as Japanese learners demonstrate limited sensitivity to lip rounding features, they frequently fail to achieve adequate lip protrusion when producing /y/, often substituting it with an insufficiently rounded /i/. Based on our findings, we recommend modifying this instructional sequence for Japanese learners. Teachers can first introduce the back rounded vowel /u/, then guide learners to produce /y/ by advancing the backness while maintaining roundness. Since Japanese learners show greater sensitivity to the backness than to the roundness feature, this alternative progression: moving from /u/ to /y/ via tongue advancement, better aligns with their perceptual strengths. This approach leverages their native sensitivity to backness cue, facilitating more accurate perception and production of the /y/ vowel through backness modification from the established /u/ articulation. The proposed method offers two key advantages: 1) it builds upon Japanese learners' existing perceptual bias toward the backness cue, and 2) it maintains rounding continuity between /u/ and /y/, avoiding the rounding acquisition challenge posed by the traditional /i/-to-/y/ transition. Such L1-specific pedagogical adjustment may improve Japanese learners' mastery of Mandarin vowel /y/.

V. Conclusion

In summary, this study reveals a weighting mechanism between F3 value / lip rounding and F2 value / backness distinctive features in Mandarin high vowel perception among Japanese native speakers, demonstrating their predominant reliance on the F2 / backness cue. These findings carry significant theoretical implications for understanding the perceptual mechanisms of non-native speech contrasts and developmental pathways in SLA. Cross-linguistic studies targeting Mandarin and other languages can provide deeper insights into the role of various acoustic cues in L2 learning. The discovery of this weighting mechanism enriches theoretical frameworks of perceptual cue weighting in language acquisition research. Practically, it highlights the necessity of language-specific perceptual studies for learners from different L1 backgrounds and underscores the importance of adaptive teaching principles in TCFL pronunciation instruction. Specifically, Mandarin phonetics teaching should tailor its approaches according to learners' native language perceptual biases, optimizing instruction by aligning with their inherent cue-weighting strategies. This research thus bridges theoretical psycholinguistic findings with practical pedagogical applications, advocating for L1-informed approaches in L2 listening and production teaching.

Future research should expand the participant pool to include higher-proficiency L2 Mandarin learners to investigate how language proficiency influences perceptual cue weighting, as the current study's exclusive focus on beginners limits generalizability to advanced acquisition stages. Additionally, the pedagogical implications require further exploration through a more granular analysis of training methodologies, such as incorporating real-time visual feedback for lip rounding to enhance practical applications, given that the present design did not systematically assess instructional efficacy. The interaction between perception and production—particularly how F2 reliance in perception may correlate with articulation errors—warrants deeper examination to clarify L2 acquisition mechanisms. Finally, the neurocognitive basis of cue weighting remains underexplored; advancing our understanding of the neural mechanisms underlying observed phoneme perception biases—potentially through EEG, fMRI, or other neuroimaging techniques—would significantly strengthen theoretical frameworks in SLA.

Appendix

Participant	F2 regression coefficient	F3 regression coefficient
1	0.007 (SE=0.001, z=5.121, p<0.001)	0.0004 (SE=0.001, z=0.293，p=0.769)
2	0.007 (SE=0.001, z=5.12, p<0.001）	0.0003 (SE=0.001, z=0.293，p=0.769)
3	0.005 (SE=0.001, z=6.815, p<0.001)	0.0008 (SE=0.001, z=0.70, p=0.484)
4	0.006 (SE=0.001, z=6.354, p<0.001）	-0.0002 (SE=0.001, z=−0.176, p=0.860）
5	0.006 (SE=0.001, z=6.35, p<0.001）	−0.0002 (SE=0.001, z=−0.176, p=0.860）
6	0.0116, (SE=0.003，z=3.562, p<0.001）	0.0006 (SE=0.001, z=0.461, p=0.645）
7	0.0040, (SE=0.001, z=6.896, p<0.001）	0.0002 (SE=0.001, z=0.144, p=0.886)
8	0.0074 (SE=0.001, z=5.592, p<0.001）	0.0002 (SE=0.001, z=0.149, p=0.882)
9	0.0061 (SE=0.001, z=5.442, p<0.001）	0.0005 (SE=0.001, z=0.401, p=0.688)
10	0.0028 (SE=0.00045, z=6.244, p<0.001）	-0.0003 (SE=0.001, z=-0.261, p=0.794
11	0.0028 (SE=0.00045, z=6.244, p<0.001）	-0.0003 (SE=0.001, z=-0.261, p=0.794)
12	0.0028 (SE=0.00045, z=6.244, p<0.001）	-0.0003 (SE=0.001, z=-0.261, p=0.794)
13	0.004 (SE=0.0006, z=6.859, p<0.001）	0.0003 (SE=0.001, z=0.245, p=0.806)
14	0.004 (SE=0.0006, z=7.035, p<0.001）	0.0003 (SE=0.001, z=0.218, p=0.827)
15	0.004 (SE=0.0006, z=7.035, p<0.001）	0.0003 (SE=0.001, z=0.218, p=0.827)
16	0.005 (SE=0.0007, z=6.768, p<0.001）	0.0001 (SE=0.001, z=0.053, p=0.958)
17	0.005 (SE=0.0007, z=6.768, p<0.001）	0.0001 (SE=0.001, z=0.053, p=0.958)
18	0.005 (SE=0.0007, z=6.678, p<0.001）	0.0002 (SE=0.001, z=0.205, p=0.838)
19	0.007 (SE=0.0012, z=5.837, p<0.001）	0.0003 (SE=0.001, z=0.220, p=0.826)
20	0.006 (SE=0.0009, z=6.634, p<0.001）	0.0013 (SE=0.001, z=1.097, p=0.273)
21	0.006 (SE=0.0011, z=6.002, p<0.001）	0.0002 (SE=0.001, z=0.143, p=0.887)
22	0.007 (SE=0.0012, z=5.661, p<0.001）	0.0009 (SE=0.001, z=0.787, p=0.431)
23	0.007 (SE=0.0012, z=5.661, p<0.001）	0.0010 (SE=0.001, z=0.866, p=0.387)
24	0.007 (SE=0.0012, z=5.661, p<0.001）	0.0010 (SE=0.001, z=0.866, p=0.387)
25	0.007 (SE=0.0011, z=6.053, p<0.001）	0.0003 (SE=0.001, z=0.226, p=0.821)
26	0.007 (SE=0.0011, z=6.053, p<0.001）	0.0003 (SE=0.001, z=0.226, p=0.821)
27	0.007 (SE=0.0012, z=5.837, p<0.001）	0.0003 (SE=0.001, z=0.220, p=0.826)

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Wen-Hui Zhu and Sun-Hee Lee are in charge of conceptualization, methodology, and editing; Wen-Hui Zhu wrote the main manuscript text; both authors reviewed the manuscript and approved the final version.

FUNDING

This research was funded by the 2022 International Chinese Language Education Research Topic Youth Project Funding, grant number 22YH61D; Guangdong Philosophy and Social Sciences Planning Project 2025 (General Program and Special Research Grant Category III) Funding, grand number GD25CZY06; and the Undergraduate Teaching Quality and Teaching Reform Project of Southern University of Science and Technology, grant numbers SJZLGC202448.

ACKNOWLEDGMENT

The authors wish to thank Professor Tianming Zhang at the Language Center, Otaru University of Commerce, Japan, for his assistance in the collection of speech data.

REFERENCES

[1] T. Crow, "Directional asymmetry is the key to the origin of modern Homo sapiens," *Laterality,* vol. 9, no. 2, pp. 233–242, 2004. doi:10.1080/13576500342000374

[2] P. K. Kuhl *et al.*, "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science,* vol. 255, no. 5044, pp. 606–608, 1992. doi:10.1126/science.1736364

[3] J. F. Werker and S. Curtin, "PRIMIR: A developmental framework of infant speech processing," *Lang. Learn. Dev.,* vol. 1, no. 2, pp. 197–234, 2005.

[4] J. E. Flege, "Second language speech learning," *in Speech Perception and Linguistic Experience,* W. Strange (Ed.), Timonium: York Press, 1995, pp. 233–277.

[5] J. E. Flege, "Interactions between the native and second-language phonetic systems," *in An Integrated View of Language Development,* P. Burmeister et al. (Eds.), Trier: Wissenschaftlicher Verlag, 2002, pp. 217–243.

[6] C. T. Best, "The emergence of native-language phonological influences in infants," *in The Development of Speech Perception,* J. C. Goodman and H. C. Nusbaum (Eds.), Cambridge: MIT Press, 1994, pp. 167–224.

[7] C. T. Best et al., "Nonnative and second-language speech perception," *in Language Experience in Second Language Speech Learning,* O. Bohn and M. Munro (Eds.), Amsterdam: John Benjamins, 2007, pp. 13–34.

[8] A. Holt and A. J. Lotto, "Cue weighting in auditory categorization," *J. Acoust. Soc. Am.,* vol. 119, no. 5, pp. 3059–3071, 2006. doi:10.1121/1.2188377

[9] O. Bohn, "Cross-language speech perception in adults," *in Speech Perception and Linguistic Experience,* W. Strange (Ed.), Timonium: York Press, 1995, pp. 279–304.

[10] P. K. Kuhl *et al.*, "Phonetic learning as a pathway to language," Philos. Trans. R. Soc. B, vol. 363, no. 1493, pp. 979–1000, 2008.

[11] J. E. Flege *et al.*, "Effects of experience on non-native speakers' production and perception of English vowels," *J. Phon.,* vol. 25, no. 4, pp. 437–470, 1997.

[12] P. Escudero and P. Boersma, "Bridging the gap between L2 speech perception research and phonological theory," *Studies. Second Lang. Acquis.,* vol. 26, no. 4, pp. 551–585, 2004.

[13] I. Grenon *et al.*, "The creation of a new vowel category by adult learners after adaptive phonetic training," *J. Phon.,* vol. 72, pp. 17–34, 2019.

[14] B. Cheng *et al.*, "The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study," *Front. Psychol.,* vol. 10, p. 1178, 2019. doi:10.3389/fpsyg.2019.01178

[15] X. Zhang et al., "Cognitive factors in nonnative phonetic learning," *J. Phon.,* vol. 100, p. 101266, 2023.

[16] M. V. Kondaurova and A. L. Francis, "The relationship between native allophonic experience with vowel duration," *J. Acoust. Soc. Am.,* vol. 124, no. 6, pp. 3959–3971, 2008.

[17] P. Escudero *et al.,* "Native, non-native and L2 perceptual cue weighting for Dutch vowels," *J. Phon.,* vol. 37, no. 4, pp. 452–465, 2009. doi:10.1016/j.wocn.2009.07.006

[18] T. Akamatsu, *Japanese Phonetics: Theory and Practice.* München: Lincom Europa, 1997.

[19] L. Labrune, *The Phonology of Japanese*. Oxford: Oxford University Press, 2012.

[20] H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (Vol. 2). Berlin: Walter de Gruyter, 2015.

[21] Y. Wang, "A preliminary study of the perception of high vowels in Mandarin by Korean and Japanese learners," *Lang. Teach. Linguist. Stud.,* no. 6, 2001.

[22] H. Bao, "Introduction to physiological and acoustic analysis of Mandarin speech (part 1 continued)," *J. Audiol. Speech Pathol.,* vol. 4, pp. 285–286, 2004.

[23] P. Boersma and D. Weenink, Praat: Doing Phonetics by Computer (Version 5.1.05) [Computer Program], 2009. Available: http://www.praat.org

[24] B. Huang and X. Liao, *Modern Mandarin.* Lanzhou: Gansu People's Publishing House, 2024.

[25] Y. Lin, *The Sounds of Chinese.* Cambridge: Cambridge University Press, 2007.

[26] J. F. Werker and J. S. Logan, "Cross-language evidence for three factors in speech perception," *Percept. Psychophys.,* vol. 37, no. 1, pp. 35–44, 1985.

[27] H. Bao and M. Lin, *Essentials of Experimental Phonetics (Revised).* Beijing: Peking University Press, 2014.

[28] R. A. Yamada and Y. Tohkura, "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," *Percept. Psychophys.,* vol. 52, no. 4, pp. 376–392, 1992.

[29] P. Iverson *et al*., "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition,* vol. 87, no. 1, pp. B47–B57, 2003.

[30] W. Zhu et al., "Language-dependent cue weighting in distinctive feature," *Asian-Pac. J. Second Foreign Lang. Educ.,* vol. 8, no. 1, p. 31, 2023. doi:10.1186/s40862-023-00204-6