



# Talker Normalization in Chinese Bilinguals: A Comparative Study

LU Mingxi<sup>1</sup>, TIAN Yujia<sup>1</sup>, TAO Ran<sup>1</sup>

<sup>1</sup>Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China  
mency.lu@polyu.edu.hk, yogayoga.tian@connect.polyu.hk, ran.tao@polyu.edu.hk

## Abstract

Talker normalization enables listeners to adapt to speaker-specific acoustic variability, thereby facilitating speech perception across different talkers. While previous research suggests that talker normalization can occur in both Mandarin and Cantonese speech contexts, the extent to which it applies to non-linguistic contexts remains unclear. This study examined talker normalization in Mandarin-Cantonese bilinguals, comparing its effects in speech and nonspeech contexts. We recruited 36 bilingual participants to complete a forced-choice word identification task in one Mandarin paradigm and two Cantonese paradigms. The results revealed that under all three paradigms, tone normalization occurred robustly in the speech condition but was absent in the nonspeech condition. These findings provide novel evidence that talker normalization operates primarily within linguistic boundaries, supporting the view that talker normalization is language-specific rather than domain-general.

**Index Terms:** pitch processing, talker normalization, tone normalization, tonal language, Mandarin-Cantonese Bilinguals

## 1. Introduction

Fundamental frequency (F0) is the primary acoustic correlate of pitch perception. Talker normalization is a crucial perceptual mechanism that allows listeners to adapt to speaker-specific acoustic variability, thereby ensuring accurate linguistic interpretation. This adaptation enables listeners to normalize phonetic units across speakers despite variations in acoustic properties. For instance, when a male speaker produces a high-pitched tone and a female speaker produces a mid-pitched tone at acoustically similar frequencies, listeners must adapt to F0 changes between different talkers to recalibrate their perception and correctly categorize phonetic units [1]. Tone normalization investigates how listeners normalize tonal perception in varying F0 contexts, making it a specific manifestation of talker normalization within tonal languages.

The study of tone normalization in Mandarin and Cantonese has significantly advanced our understanding of its underlying mechanisms. Researchers have found that Mandarin listeners can utilize contextual F0 information to perform tone normalization when faced with different talkers. Tone normalization in Mandarin listeners can be achieved effectively by using F0 varieties among different talkers [2]. Compared with Mandarin listeners, Cantonese listeners process more complex pitch contours with greater variations in pitch height, making normalization more challenging [3]. In Cantonese, tone perception relies not only on F0 but also on other acoustic features, such as duration and intensity. Research has shown that Cantonese listeners exhibit higher sensitivity to the

acoustic characteristics in tone perception, which may be attributed to the complexity of the Cantonese tonal system, leading listeners to rely more on acoustic cues for tone normalization [4, 5]. For example, Cantonese listeners are more sensitive to speaker specific F0 variations during the tone normalization process, in contrast to Mandarin listeners [6]. Findings on the effects of speech and nonspeech contexts on talker normalization have led to a debate over whether pitch processing is governed by domain-general mechanisms or language-specific mechanisms. Research adopting Mandarin and Cantonese speech as contexts has consistently found that speech context can elicit reliable context effects. However, studies using Mandarin stimuli have produced mixed findings regarding the effect of the nonspeech context, leading some to favor a domain-general perspective [6, 7]. In contrast, the consistent lack of nonspeech contextual effects in studies of Cantonese stimuli supports a language-specific viewpoint [1, 8, 9, 10, 11, 12].

We found that studies yielding different results regarding talker normalization in nonspeech conditions employed distinct experimental paradigms, and they all used monolingual participants. Prior research has primarily adopted two experimental paradigms: the Mandarin paradigm [7] and the Cantonese paradigm [1]. Notable methodological differences exist between these approaches. Mandarin research often involves the identification of ambiguous artificial tone contours from a continuum, whereas Cantonese studies typically utilize naturally produced tones. For example, a recent study on Mandarin adults and children used a continuum generated from a Mandarin level tone to a rising tone (e.g., /i55/ to /i35/) [6]. In contrast, researchers studying Cantonese take advantage of the three level tones in naturally produced speech: high-level, mid-level, and low-level tones [13]. These level tones have similar contours but differ in pitch height.

Given these methodological differences, our study first aimed to determine whether the discrepancies in nonspeech tone normalization results arise from differences in experimental paradigms or participants' linguistic backgrounds. To address this question, we recruited Mandarin-Cantonese bilingual participants to complete both experimental paradigms, bridging the existing research gap. We analyzed their tone normalization responses across two context types: speech and nonspeech, specifically examining (1) whether tone normalization occurs consistently in speech contexts across paradigms, and (2) whether nonspeech contexts fail to elicit tone normalization across paradigms. Additionally, an innovation of this study is the use of contour tones (Tones 1 and 2) in the Cantonese paradigm that differ in both F0 height and F0 direction to further investigate tone normalization mechanisms in a more complex tonal structure, which differs from previous Cantonese studies that predominantly investigated level tones (Tones 1, 3, and 6). This methodological expansion provides additional

evidence for tone normalization effects in Cantonese, offering a more comprehensive understanding of how tonal structure influences talker normalization across different linguistic contexts.

## 2. Methodology

This study adopted stimuli and experimental designs comparable to those utilized in previous research [10, 11, 12]. As a component of a larger project investigating non-linguistic musical context effects, the present report focused specifically on the comparison between speech and nonspeech contexts. The stimuli preparation and experimental procedure are briefly described in the following sections.

### 2.1. Participants

We recruited 36 Mandarin-Cantonese bilinguals (15 males, mean age = 21.47 yrs, SD = 1.29) who had normal hearing and were right-handed. All participants provided written informed consent prior to the experiment. The experimental protocol was approved by the Human Subjects Ethics Sub-Committee of The Hong Kong Polytechnic University (PolyU, Reference Number: HSEARS20241029003).

### 2.2. Stimuli

The stimuli in this study included targets and contexts across two conditions, categorized into two types: speech and nonspeech. All speech contexts and target stimuli were produced by four native Mandarin-Cantonese bilingual speakers (two males). These speakers included one female with a high pitch range, one female with a low pitch range, one male with a high pitch range, and one male with a low pitch range. The speech context is a four-syllable meaningful sentence, including Mandarin phrase '这个字是' (This word is), as well as the Cantonese phrase '呢個字係' (This word is). Additional phrases, acting as fillers, including the Chinese phrases '我现在读' (Now I am reading) and '请留心听' (Please listen carefully), and their Cantonese counterparts '我依家讀' (Now I am reading) and '請留心聽' (Please listen carefully) were also recorded. After recording the natural production of sentences from the four speakers, the F0 trajectories of the contexts were lowered and raised by three semitones. In summary, three sets of speech contexts were formed: F0 lowered, F0 unshifted, and F0 raised. These F0-manipulated contexts were used to trigger contrastive context effect [13] during participants' perception of ambiguous targets. Specifically, more high-level tone responses were expected in the lowered context condition, and more low-level tone and low-raising tone responses were expected in the raised context condition. The nonspeech contexts were generated by mapping the F0 trajectory and intensity profile of the speech contexts onto triangle waves. Mandarin syllable /i/ and Cantonese syllable /ji/ were selected as the target syllables, corresponding to Mandarin characters '衣' and '姨' (e.g., tonal syllables /i55/ and /i35/), and Cantonese characters '医', '倚', '意', and '二' (e.g., /ji55/, /ji25/, /ji33/, and /ji22/).

The target characters were manipulated to create continua. The original recording was used to generate the target continuum by changing the pitch contours. For example, an 11-step pitch continuum transitioning from Mandarin /i55/ to /i35/ was synthesized in Praat through interpolation. The original pitch trajectory of recording /i55/ was replaced by 11 pitch trajectories using the built-in overlap-add synthesis method in Praat [14]. For the identification task, we used only the middle

step, which is the most ambiguous step of the continuum, while the endpoints were used as fillers. Cantonese /ji55/ and /ji25/ were manipulated in a similar manner to create a Cantonese tone continuum. All targets were adjusted to an intensity of 55 dB and duration of 450 ms. All contexts were adjusted to an intensity of 55 dB and duration of 1000 ms.

### 2.3. Procedure

The participants completed three practice blocks followed by six experimental blocks. The experimental blocks were designed as word identification tasks. Participants were asked to make a forced choice in three tasks. Task 1 is the Cantonese word identification task with natural stimuli (CIDn) among '医' /ji55/, '意' /ji33/, and '二' /ji22/. Task 2 and Task 3 are the identification tasks with artificial tones in continua of Mandarin (MIDc) and Cantonese (CIDc), e.g., '衣' /i55/ or '姨' /i35/ and '医' /ji55/ or '倚' /ji25/, respectively. In each trial, the participants saw a fixation (+) appearing on the screen, followed by the context stimulus played through the inserted earphones. After hearing the context and jittering silence (ranging 300-500 ms), a target syllable was presented. A question mark was presented 500 ms after the onset of the target and remained on the screen for up to 1500 ms. Participants were instructed to press a designated key for the characters they heard after seeing the question mark to ensure the whole target syllable was heard before a response is made. With this setting, reaction times were not meaningful to indicate mental process and thus not analyzed. We focused on participants' judgments of the targets, consistent with the standard procedure used in previous research.

In the present study, each task session had the following structures: for Task 1, the design similarly included three pitch shifts from four speakers, and each condition was repeated four times, yielding 48 experimental trials. For Tasks 2 and 3, filler trials consisted of two steps from each of the four speakers, adding eight trials, for a total of 56 trials per task. As in Task 1, the design also included three pitch shifts from four speakers. Each trial contained T1, T2, and an intermediate tone step between T1 and T2, which was repeated twice, resulting in 48 experimental trials. Additionally, filler trials consisted of two steps from each of the four speakers, contributing 8 trials, bringing the total to 56 trials for this task.

The three tasks were repeated twice to form six experimental blocks. To control order effects, the six experimental blocks were counterbalanced across participants.

### 2.4. Analysis

In our analysis, we examined tone normalization across three experimental paradigms, e.g., MIDc, CIDc, and CIDn. Since the primary aim of this study was to investigate the effects of speech and nonspeech contexts on talker normalization in Mandarin-Cantonese bilinguals across different paradigms. Following previous research [11, 12, 13], we calculated the mean Perceptual Height (PH) for participants across the three experimental paradigms to examine patterns of tone normalization.

In the CIDc and MIDc paradigms, PH was specifically coded to facilitate a more intuitive interpretation of the talker normalization effects: high-level tones were coded as 0, whereas low-level tones were coded as 1. Under the Lowered F0 condition, a PH mean closer to 0 indicated that participants exhibited talker normalization, whereas under the Raised F0 condition, a PH mean closer to 1 suggested the presence of

talker normalization. In the CIDn paradigm, PH coding was expanded to accommodate a more complex tonal structure: high-level tones were coded as 6, mid-level tones as 3, and low-level tones as 1. Under the Lowered F0 condition, a PH mean closer to 6 indicates that talker normalization occurred, whereas under the Raised F0 condition, a PH mean closer to 1 suggests the presence of talker normalization. This coding scheme allows for a unified analysis of PH across different paradigms, providing a clearer interpretation of how F0 shifts influence talker normalization.

To systematically examine the effects of speech and nonspeech contexts on tone normalization, we conducted a three-way ANOVA on target tone perception, analyzing the main effects and interactions of Paradigm (experimental paradigm), Context (speech vs. nonspeech), and Shift (F0 shift condition). Additionally, to obtain more detailed statistical insights, we calculated the effect sizes and performed post hoc tests. Specifically, we first compared paradigms to assess the consistency of talker normalization across the Mandarin and Cantonese paradigms. Next, we examined Shift conditions to determine whether pitch shifts influenced normalization effects. Finally, given our primary focus on the contrast between speech and nonspeech contexts, we analyzed the Context  $\times$  Shift interaction to assess whether nonspeech contexts failed to elicit talker normalization. Furthermore, considering the methodological novelty of this study—specifically, the inclusion of both MIDc and CIDn paradigms, as well as the use of contour tones in the CIDc paradigm—we conducted independent two-way ANOVAs (Context  $\times$  Shift). These independent analyses allowed us to systematically examine tone normalization effects in Cantonese and Mandarin contour tones, as well as Cantonese level tones, thereby providing further insights into how different tonal categories (high, mid, and low) modulate talker normalization.

Bonferroni-corrected post-hoc tests were conducted within each paradigm to explore the specific effects of Context and Shift under different experimental conditions.

### 3. Result

To examine the effects of Context and Shift on participants' responses, separate analyses were conducted for each experimental paradigm (CIDc, MIDc, and CIDn). Across all three tasks, a consistent Context  $\times$  Shift interaction emerged, with speech contexts exhibiting sensitivity to pitch manipulations, while nonspeech contexts showed no significant sensitivity to pitch manipulations. Notably, the CIDn perceptual height task revealed the strongest effect sizes ( $\eta^2 = .575$ ), indicating that explicit perceptual height ratings are highly sensitive to pitch shift and suggesting that explicit pitch perception measures may provide a more sensitive index of talker normalization than categorical tone identification tasks.

A consistent pattern of results emerged in both the CIDc and MIDc paradigms. Two-way ANOVA revealed a significant main effect of Shift in CIDc,  $F(1, 35) = 28.147, p < .001, \eta^2 = .068$ , and in MIDc,  $F(1, 35) = 23.740, p < .001, \eta^2 = .061$ , both indicating medium effect sizes. Additionally, a significant Context  $\times$  Shift interaction was observed in CIDc,  $F(1, 35) = 22.618, p < .001, \eta^2 = .105$ , and in MIDc,  $F(1, 140) = 12.34, p < .001, \eta^2 = .081$ , both reflecting large effect sizes. However, the main effect of Context was not significant in either paradigm CIDc:  $F(1, 35) = 4.253, p = .047, \eta^2 = .021$  and not significant in MIDc ( $F(1, 35) = 0.912, p = .346, \eta^2 = .004$ ). Post-hoc analyses further confirmed that Shift effects were significant in the speech context (both  $ps < .001$ ) but not in the

nonspeech context (CIDc:  $p = .441$ ; MIDc:  $p = .689$ ). Moreover, Context effects were significant for both lowered ( $p = .001$ ) and raised ( $p = .021$ ) conditions in CIDc, whereas in MIDc, they were only significant for the raised condition ( $p = .002$ ) but not for the lowered condition ( $p = .074$ ).

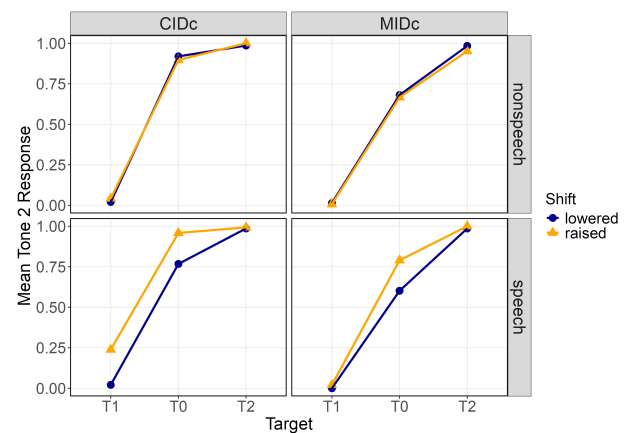


Figure 1. Mean percentage of Tone2 response for the two context conditions under the CIDc and MIDc paradigms. Responses to the raised and lowered F0 conditions are shown as green and red lines, respectively.

The CIDn paradigm had the strongest effect. A two-way ANOVA revealed a highly significant main effect of Shift,  $F(1, 35) = 430.473, p < .001, \eta^2 = .575$  (a very large effect size), and a robust Context  $\times$  Shift interaction,  $F(1, 35) = 445.728, p < .001, \eta^2 = .557$ . The main effect of Context was not significant,  $F(1, 35) = 1.051, p = .256, \eta^2 = .009$ . Post-hoc analyses confirmed the same pattern observed in the other tasks but with stronger effect sizes: Shift effects were highly significant in the speech context ( $p < .001$ ) but not in the nonspeech context ( $p = .773$ ). Moreover, both lowered and raised conditions showed significant differences between speech and nonspeech contexts (both  $ps < .001$ ).

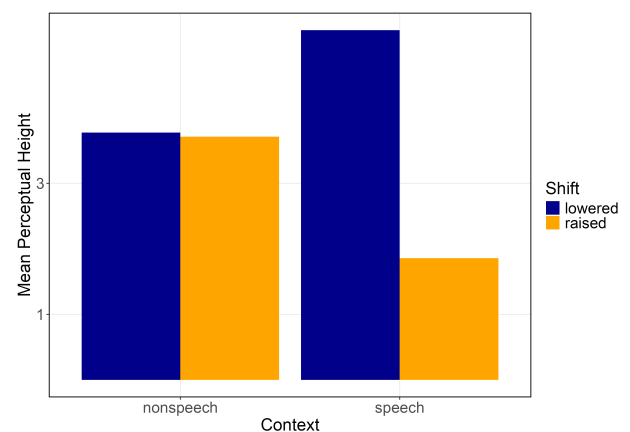


Figure 2. The perceptual height of the targets in CIDn paradigm.

A cross-task comparison further reinforces this pattern. Across all three tasks, the speech context consistently showed sensitivity to pitch manipulations (Shift effects), whereas the nonspeech context did not. This effect was most pronounced in the CIDn paradigm ( $\eta^2 = .557$ ), indicating that explicit perceptual height ratings offer a more fine-grained assessment

of pitch normalization than categorical tone identification tasks (CIDc:  $\eta^2 = .105$ , MIDc:  $\eta^2 = .081$ ). Notably, the effect of F0 Shift on Tone 2 responses in the speech context aligns with expectations from talker normalization: in the Lowered F0 condition, participants were more likely to categorize the stimulus as Tone1, indicating that they adjusted their perception based on the preceding F0 context. In the Raised F0 condition, Tone1 identification decreased, consistent with a shift in the perceptual boundary due to talker normalization. This shift pattern was absent in the nonspeech context, suggesting that talker normalization primarily operates within linguistic boundaries. Furthermore, the similarity between the CIDc and MIDc paradigms suggests that talker normalization mechanisms are shared across Mandarin and Cantonese bilinguals, while the larger effects observed in the CIDn paradigm indicate that explicit perceptual height tasks capture finer-grained pitch normalization effects compared to categorical identification tasks.

## 4. Discussion

By examining pitch processing across different experimental paradigms, we aimed to clarify whether talker normalization operates as a language-specific mechanism or as a more generalized auditory process. Our findings provide strong evidence in favor of a language-specific account, as speech contexts consistently exhibited sensitivity to pitch manipulations, whereas nonspeech contexts showed no significant effects.

### 4.1. Context Effects on Talker Normalization

A significant context  $\times$  shift interaction was observed across all three experimental paradigms (CIDc, MIDc, and CIDn). Specifically, the CIDc and MIDc tasks. Participants' responses were significantly modulated by pitch shifts within speech contexts, whereas these effects were absent in nonspeech contexts. These results suggest that talker normalization is primarily engaged in linguistic environments, where listeners primarily rely on phonetic and prosodic cues to adjust their perception of incoming speech signals.

### 4.2. Differential Sensitivity Across Paradigms

Notably, the CIDn perceptual height task revealed the strongest effects, with the largest effect size for the Shift factor ( $\eta^2 = .575$ ). This suggests that explicit pitch perception measures serve as a more sensitive measure of talker normalization compared to categorical tone identification tasks. One possible explanation for the large effect in CIDn is related to the data coding: we used 1, 3, and 6 as the dependent variable, which are 10 times larger than the percentage values used in the CIDc and MIDc tasks. This difference in data scaling may have amplified the observed effect. Future analyses could address this potential confound. Additionally, categorical tasks may involve phonological encoding, which could partially obscure the effects of talker-related acoustic variability. In contrast, perceptual height judgments rely more directly on acoustic properties, thus amplifying the observed normalization effects. This finding underscores the importance of using diverse methodological approaches to capture different aspects of pitch processing, particularly in tone perception studies.

### 4.3. Implications for Level vs. Contour Tone Normalization

Our results also suggest that talker normalization effects may be more pronounced for level tones than for contour tones. The strong effects observed in the CIDn paradigm, which primarily involved level tone judgments, indicated that listeners adjusted their perception of pitch height more readily in these cases. This aligns with prior findings that level tones rely more on absolute pitch cues, whereas contour tones incorporate dynamic pitch movement, which may exhibit reduced susceptibility to talker-induced shifts [14]. Future studies could further investigate whether similar patterns hold across other tonal languages, such as Thai or Vietnamese, to determine whether these effects generalize to other tonal languages.

### 4.4. Possible Explanations for Prior Contradictory Findings

An important consideration in interpreting our results is the methodological differences between our study and prior research. Previous studies examining talker normalization effects in Mandarin paradigms have reported varying results regarding the role of nonspeech contexts. One potential explanation for these discrepancies is how nonspeech stimuli were manipulated and presented.

In our study, we used a triangular wave to construct nonspeech materials. A triangular wave is a type of non-sinusoidal waveform named for its characteristic triangular shape. In contrast, Holt's study employed a single sinusoidal wave at F0 as the nonspeech stimulus, which consisted of a series of pure tones at four times the F0 of the speech stimuli [15]. Although both types of nonspeech stimuli contained complete F0 information, the triangular wave had a richer harmonic structure than the sinusoidal wave. As a result, the nonspeech context effect observed in Holt's study [15] cannot be generalized to all types of nonspeech contexts, underscoring the necessity of considering stimulus properties when evaluating talker normalization effects.

## 5. Conclusions

In summary, our study provides compelling evidence that talker normalization is primarily a language-specific process, with significant pitch adaptation effects observed only in speech contexts. The strongest effects emerged in the CIDn paradigm, suggesting that explicit pitch perception tasks may offer a more sensitive measure of normalization than categorical tone identification tasks. These findings advance our understanding of speech perception mechanisms and have important implications for linguistic theory, speech technology, and auditory training programs. Future research should continue to explore the broader implications of talker normalization across different populations and methodological frameworks.

## 6. Acknowledgements

This study was supported by two internal grants from Hong Kong Polytechnic University awarded to TAO Ran (Project IDs: P0048115 and P0056421).

Miss LU Mingxi thanks her supervisor Prof. TAO Ran for his guidance and Miss Tian Yujia for her invaluable support and feedback throughout this project. LU Mingxi is grateful for the learning experience this challenging project provided.

We extend our sincere appreciation to all participants who came to our lab and trusted us with their participation. Without their contribution and cooperation, this project would not have been

possible. We also thank the anonymous reviewers for their constructive comments.

## 7. References

- [1] G. Peng, C. Zhang, H. Y. Zheng, J. W. Minett, and W. S. Y. Wang, "The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems," *J. Speech Lang. Hear. Res.*, vol. 55, no. 2, pp. 579–595, 2012, doi: 10.1044/1092-4388(2011/11-0025).
- [2] C. Zhang and S. Chen, "Toward an integrative model of talker normalization," *J. Exp. Psychol.*, vol. 42, no. 8, pp. 1252–1268, 2016, doi: 10.1037/xhp0000216.
- [3] P. Wong, S. T. Cheng, and F. Chen, "Cantonese tone identification in three temporal cues in quiet, speech-shaped noise, and two-talker babble," *Front. Psychol.*, vol. 9, 2018, doi: 10.3389/fpsyg.2018.01604.
- [4] X. Tong, S. M. K. Lee, M. M. L. Lee, D. Burnham, and J. Snyder, "A tale of two features: Perception of Cantonese lexical tone and English lexical stress in Cantonese-English bilinguals," *PLOS ONE*, vol. 10, no. 11, 2015, doi: 10.1371/journal.pone.0142896.
- [5] P. Wong and H.-Y. Chan, "Acoustic characteristics of highly distinguishable Cantonese entering and non-entering tones," *J. Acoust. Soc. Am.*, vol. 143, no. 2, pp. 765–779, 2018, doi: 10.1121/1.5021251.
- [6] F. Chen, K. Zhang, Q. Guo, and J. Lv, "Development of achieving constancy in lexical tone identification with contextual cues," *J. Speech Lang. Hear. Res.*, vol. 66, no. 4, pp. 1148–1164, 2023, doi: 10.1044/2022\_JSLHR-22-00257.
- [7] Y.-C. Kuo, S. Rosen, and A. Faulkner, "Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 2815–2824, 2008, doi: 10.1121/1.2896755.
- [8] R. Soo and M. Babel, "Perceptual effects of lexical competition on Cantonese tone categories," *Lab. Phonol.*, vol. 14, no. 1, 2023.
- [9] R. Tao and G. Peng, "Music and speech are distinct in lexical tone normalization processing," in *Proc. 34th Pacific Asia Conf. Lang., Inf. Comput.*, 2020.
- [10] R. Tao, K. Zhang, and G. Peng, "Music does not facilitate lexical tone normalization: A speech-specific perceptual process," *Front. Psychol.*, vol. 12, pp. 1–14, 2021, doi: 10.3389/fpsyg.2021.717110.
- [11] C. Zhang, G. Peng, and W. S. Y. Wang, "Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1088–1099, 2012, doi: 10.1121/1.4731470.
- [12] C. Zhang, G. Peng, and W. S. Y. Wang, "Achieving constancy in spoken word identification: Time course of talker normalization," *Brain Lang.*, vol. 126, no. 2, pp. 193–202, 2013, doi: 10.1016/j.bandl.2013.05.010.
- [13] P. C. M. Wong and R. L. Diehl, "Perceptual normalization for inter- and intratalker variation in Cantonese level tones," *J. Speech Lang. Hear. Res.*, vol. 46, no. 2, pp. 413–421, 2003, doi: 10.1044/1092-4388(2003/034).
- [14] F. Chen and G. Peng, "Context effect in the categorical perception of Mandarin tones," *Journal of Signal Processing Systems*, vol. 82, pp. 253–261, 2016.
- [15] J. Huang and L. L. Holt, "General perceptual contributions to lexical tone normalization," *J. Acoust. Soc. Am.*, vol. 125, no. 6, pp. 3983–3994, 2009, doi: 10.1121/1.3125342.
- [16] L. Cabrera, F.-M. Tsao, D. Gnansia, J. Bertoni, and C. Lorenzi, "The role of spectro-temporal fine structure cues in lexical-tone discrimination for French and Mandarin listeners," *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 877–882, 2014, doi: 10.1121/1.4887444.
- [17] F. Cao, R. Tao, L. Liu, C. A. Perfetti, and J. R. Booth, "High proficiency in a second language is characterized by greater involvement of the first language network: Evidence from Chinese learners of English," *J. Cogn. Neurosci.*, vol. 25, no. 10, pp. 1649–1663, 2013, doi: 10.1162/jocn\_a\_00414.
- [18] F. Chen and G. Peng, "聲學密度假說——基於普通話和粵語母語者聲調歸一化的對比研究," in *高山仰止—王士元教授九十歲賀壽文集*, G. Peng, J. Kong, Z. Shen, and F. Wang, Eds., Hong Kong: City University of Hong Kong Press, 2023, pp. 201–212.
- [19] L. L. Holt, "The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2801–2817, 2006, doi: 10.1121/1.2354071.
- [20] Y. Li, C. Tang, J. Lu, J. Wu, and E. F. Chang, "Human cortical encoding of pitch in tonal and non-tonal languages," *Nat. Commun.*, vol. 12, no. 1, pp. 1–12, 2021, doi: 10.1038/s41467-021-21430-x.
- [21] C. B. Moore and A. Jongman, "Speaker normalization in the perception of Mandarin Chinese tones," *J. Acoust. Soc. Am.*, vol. 102, no. 3, pp. 1864–1877, 1997, doi: 10.1121/1.420092.
- [22] G. Peng, "Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese," *J. Chin. Linguist.*, vol. 34, no. 1, pp. 134–154, 2006.
- [23] F. M. Tsao, "Perceptual improvement of lexical tones in infants: Effects of tone language experience," *Front. Psychol.*, vol. 8, p. 558, 2017.