

Effect of Rap Music Context on Lexical Tone Normalization

TIAN Yujia¹, YE Yanyuan¹, LU Mingxi¹, JIA Fanlu², TAO Ran¹

¹ Research Centre for Language Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University ²Department of Psychology, Jinan University

yogayoga.tian@connect.polyu.hk, yanyuan.ye@connect.polyu.hk,
mency.lu@connect.polyu.hk, spe_jiafl@ujn.edu.cn, ran.tao@polyu.edu.hk

Abstract

This study explores the role of rap music in lexical tone normalization among Mandarin speakers, addressing a gap in existing research that primarily focuses on speech contexts. While previous studies suggest that tone normalization is largely speech-specific, rap music, which combines elements of both speech and music, may provide unique insights. We examined the effects of rap, cello, and drum contexts compared to typical speech contexts. Our findings reveal that rap music, unlike purely instrumental music, elicited tone normalization effects similar to those of speech. This suggests that the pitch information in rap music may activate cognitive mechanisms akin to those used in speech processing.

The presence of human voices in rap creates a frame of reference, enabling listeners to normalize tones based on contextual pitch information. This challenges the notion that lexical tone normalization is exclusive to speech, highlighting the potential for speech-like elements in music to influence auditory perception. Our research underscores the importance of vocal elements in tone normalization and suggests that rhythm alone is not a critical factor.

Future research should investigate other speech-like materials and include participants with diverse linguistic and musical backgrounds to deepen our understanding of these mechanisms. By expanding the scope of contexts and participant diversity, we aim to further elucidate the cognitive processes underlying tone normalization and its broader implications for language education, rehabilitation, and AI technologies.

1 Introduction

1.1 Research Background

Speech normalization focuses on how phonologically identical utterances can exhibit significant acoustic variation among different speakers, yet listeners still recognize words across these variations. This phenomenon has been thoroughly investigated in vowel perception, where the "frame of reference" theory is pivotal. This theory suggests that listeners utilize contextual cues to form a cognitive framework, aiding in the accurate interpretation of vowel sounds. Ladefoged and Broadbent (1957) showed that vowels assessed within a precursor phrase with altered vowel formant frequencies were perceived differently compared to those in a phrase with lower formant frequencies. This concept has been reinforced by numerous studies (Ainsworth, 1974; Dechovitz, 1977; Nearey, 1978, 1989; Remez et al., 1987; Johnson, 1990).

Similarly, lexical tone normalization, another aspect of speech normalization, refers to the process by which individuals interpret tone information in various perceptual environments. Leather (1983) observed normalization effects with Mandarin tones, indicating that the pitch range of a contextual utterance affects the perception of test tones with varying F0 values.

Many studies have explored how context influences the perception of level tones in Cantonese (Francis, 2006; Wong & Diehl, 2003; Zhang, 2012). There are also some studies using Mandarin as subjects (Chen, F. & Peng, G., 2016). These studies indicate that target tones are often perceived as lower when the contextual pitch is higher, and vice versa. Zhang et al. (2012) found a disparity in the perception of Cantonese level tones between speech and non-speech contexts. Musical context, as an important non-speech context, has also been partially studied (Zhang et al., 2013; Tao

and Peng, 2020). These studies indicate that musical context does not have the same effect on tone perception as speech context.

Previous research suggests that lexical tone normalization is largely speech-specific, influenced mainly by speech context rather than non-speech or music contexts. These findings imply that music and speech contexts activate different cognitive mechanisms in processing tonal information. However, these conclusions are primarily based on typical musical styles, overlooking special genres like rap music, which shares some acoustic features with speech. To better understand the role of contexts that blend music and speech in lexical tone normalization, it is essential to examine rap music. If rap music does not evoke tonal normalization, it would support the idea that normalization is specific to speech. On the other hand, if rap music does evoke tonal normalization, it would suggest that contexts with speech-like elements can also trigger this process. Furthermore, as rap combines elements of both language and music, it does not necessarily imply that language and music share common cognitive features.

Rap is defined as a style of popular music, developed by African Americans in New York in the 1970s, where words are spoken rhythmically and often in rhyming sentences over an instrumental backing.¹ As a genre of hip-hop music, rap features strong rhythm and a lack of melody. Therefore, we believe that rap occupies a space between speech and music, potentially playing a unique role in the normalization of lexical tones among Mandarin speakers.

Previous research suggests that lexical tone normalization is largely speech-specific, influenced mainly by speech rather than non-speech contexts. However, these conclusions often overlook genres like rap music, which shares acoustic features with speech. This study aims to explore whether rap music can evoke lexical tone normalization, challenging the speech-specific hypothesis. By examining rap, cello, and drum contexts, we seek to understand the role of different contexts in Mandarin tone normalization. This exploration is crucial for determining whether

elements of rap music can bridge the gap between speech and non-speech contexts.

This research fills a gap in the study of rap contexts in lexical tone normalization, providing new insights into the speech-specific hypothesis. By focusing on Mandarin, we aim to offer additional evidence in a field where results from Cantonese and Mandarin have sometimes conflicted. This study contributes to a deeper understanding of how different auditory contexts influence tone perception, potentially reshaping theories of speech processing.

The findings have implications for tone language education, particularly Mandarin teaching. They could inform speech and music-related rehabilitation therapies, offering new strategies for auditory training and cognitive rehabilitation. Additionally, the study's insights can advance AI technologies, such as speech-to-text systems and voice recognition, by improving algorithms that process tonal languages. Understanding how different contexts affect tone normalization can enhance language learning tools and improve AI's speech processing capabilities, making technology more accessible and effective for tone language speakers.

Furthermore, this research could impact the development of educational curricula by integrating musical elements into language teaching, thereby enriching the learning experience. In therapeutic settings, the findings might lead to innovative approaches that utilize musical contexts to aid in speech recovery and cognitive development. The potential applications of this research extend to various fields, highlighting its broad relevance and utility.

Our primary question is whether the rap context can facilitate the normalization of Tone 1 and Tone 2 in Mandarin. We hypothesize that rap, as a unique non-speech context, may trigger tone normalization processes distinct from those in speech and instrumental music contexts. By comparing these contexts, we aim to determine whether vocal form has a different contrastive context effect than instrumental music. This investigation will provide insights into the cognitive processes underlying tone perception and

¹ The Oxford English Dictionary defines "Rap" as "a style of popular music (developed by New York Blacks in the 1970s) in which words (usually improvised) are spoken rhythmically and often in rhyming sentences over an

instrumental backing". In the online version of the Encyclopedia Britannica, Rap is defined as "the competitive use of rhyming lines spoken over an ever-more-challenging rhythmic base".

the potential for cross-modal influences between music and language.

In conclusion, this study explores the boundary between music and language, providing evidence for speech normalization. By examining contexts that combine elements of both, we aim to uncover whether lexical tone normalization involves a unique frame of reference, similar to vowel perception. This exploration is crucial for understanding the broader cognitive mechanisms underlying speech perception and normalization. Ultimately, the research seeks to bridge gaps in current knowledge, offering a comprehensive view of how diverse auditory contexts influence language processing. Through this work, we hope to contribute to the ongoing dialogue in linguistics and cognitive science, paving the way for future studies that further unravel the complexities of human auditory perception.

2 Methodology

We utilized a similar experimental design and stimuli as in previous research (Tao et al., 2021). Below is a brief overview of the stimuli preparation and experimental procedure; for more detailed information, refer to (Zhang et al., 2013; Zhang et al., 2017).

2.1 Participants

The study involved 24 native Mandarin speakers [12 females, average age (mean±SD) = 23.5±2.2], all of whom were right-handed and had no hearing impairments. This number was determined to ensure the four context type conditions were counterbalanced among participants. According to Peng et al. (2010), tone inventories can influence categorical perception without context. Therefore, participants were selected exclusively from Northern China, speaking only Mandarin and no other dialects. None had formal musical training or professional experience in music. All participants provided written consent prior to the study, which was approved by the Human Subjects Ethics Subcommittee of The Hong Kong Polytechnic University.

2.2 Stimuli

Stimuli of this study comprised targets and contexts across four conditions: rap, cello, drum, and speech. These were designed to explore the effects of different rhythmic and melodic contexts

on lexical tone normalization, categorized into two main groups: vocal and instrumental, each with distinct characteristics.

In the vocal group, the rap and speech contexts were produced by four native Mandarin speakers (two males and two females), each with over five years of amateur rap experience and approximately seven hours of weekly exposure to rap music. The content for both speech and rap contexts consisted of a six-syllable meaningful sentence: "下面你会听到 (Below you will hear)" and "现在我我说的是 (Now what I say is)." The rap context was derived from a track by Chinese rapper Pharaoh, specifically the song "百变酒精," chosen for its strong rhythmic elements and clear articulation, ensuring it was representative of Chinese rap music. The rap materials were recorded with a speaker performing alongside a background music (BGM) track, with only the vocal component recorded. This approach aimed to make the rap material accessible while retaining distinctive rap characteristics. In contrast, the speech context maintained a natural conversational rhythm, typical of spontaneous speech, with variations in tempo and intonation.

Meanwhile, the instrumental group consisted of cello and drum contexts, both purely instrumental and devoid of vocal elements. These contexts were designed to match the pitch and rhythm of the rap context, providing a basis for comparison. The cello context emphasized melodic continuity, with musical notes aligned to the pitch of each syllable in the rap context, and its rhythm was consistent with the rap context, following the pattern: "X X X X | X X". The drum context utilized the sound of a woodfish to represent rhythmically strong instruments and featured strong, syncopated beats with complex rhythmic patterns, creating a rhythmically dominant context compared to the more melodic cello context.

The selection of these four sound types aimed to isolate the effects of rhythm and melody on lexical tone normalization. By comparing the rhythmically strong drum and rap contexts with the rhythmically weaker cello and speech contexts, we sought to determine whether rhythm alone could account for any observed effects. The instrumental group allowed us to examine the impact of rhythm without vocal influence, while the vocal group provided insight into the role of vocal melody and rhythm.

The Mandarin syllable /i/ was chosen as the target syllable, derived from natural recordings by the same speakers. An 11-step tone continuum was constructed, ranging from Mandarin Tone 1 (high-level tone) to Tone 2 (mid-rising tone). In Mandarin, /i/ with a high-level tone means "clothes" (coded as stimulus Number 1), while with a mid-rising tone, it means "aunt" (coded as stimulus Number 11). After recording the sentences, the F0 trajectories were adjusted by three semitones to create F0-lowered and F0-raised contexts. The music contexts were similarly produced. All targets were set to an intensity of 55 dB and a duration of 450 ms, while all contexts were adjusted to an intensity of 55 dB and a duration of 1000 ms.

2.3 Experiment Procedure

Participants attended a practice block, four experimental blocks, and a subjective choice block. The practice block consisted of two repetitions and two phrases with two tones, totaling eight trials. This block included only one talker (different from the four talkers in the subsequent experimental blocks) and one context (speech). It was designed to familiarize participants with the experimental procedure. The study comprised four experimental blocks, each representing a distinct context condition: cello, drum, rap, and speech. These conditions were systematically counterbalanced among participants to ensure each appeared equally across different positions. Each participant was randomly assigned to one of several counterbalancing sequences, ensuring that each context condition appeared in each position an equal number of times. The subjective choice block was a judgment task where participants decided whether the heard sound belonged to rap or speech.

The experimental blocks were designed as tone identification tasks. Participants were asked to judge the target syllable after attentively listening to the preceding context. Specifically, participants were instructed to focus on the entire utterance and press "1" for Tone 1 or "2" for Tone 2 using their right hand. During each trial (see Fig. 1), a forward mask (+) was displayed for 500 ms, followed by the context stimulus played through inserted earphones. After hearing the context and a jittering silence (ranging from 300 ms to 500 ms), the target syllable was presented, ranging from Mandarin Tone 1 to Tone 2. Participants then identified the target tone.

To mitigate order effects, participants pressed the corresponding keys when a question mark appeared 500 ms after the target onset, remaining for up to 1500 ms. Reaction times were not analyzed, as they did not provide meaningful insights into psycholinguistic properties. The focus was on participants' tone judgments, aligning with standard research procedures.

Each context condition included two F0 frequency shifts, two content types from four speakers, and 11 target steps, totaling 176 trials per condition. The experimental blocks were counterbalanced to avoid order effects.

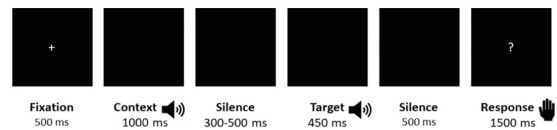


Figure 1: The trial procedure of the Mandarin word identification task.

2.4 Analysis

In line with previous studies (Chen, et al., 2016; Tao, et al., 2021; Zhang et al., 2023), we analyzed the Tone 2 identification rate to assess lexical tone normalization.

For the tone identification task, we performed a Probit analysis of the recognition responses to calculate the position and boundary width of the category boundaries (Finney, 1971), comparing between different types of contexts.

We performed one-way repeated measures ANOVAs on identification rates (IR) with Context as the main factor, followed by three-way repeated measures ANOVAs on Tone 2 identification rates. Two main factors were Context (cello, drum, rap, speech) and F0 shift of context frequency (low, high).

3 Results

The data analysis procedure largely followed previous research (e.g., Chen, et al., 2016; Tao, et al., 2021; Zhang et al., 2023).

For each stimulus, the identification score was calculated as the percentage of responses where participants identified the stimulus as either 'Tone 1' or 'Tone 2'. Figure 2 shows the average percentage of Tone 2 responses under two F0 shift conditions (high vs. low) across four contexts. The red line indicates responses to target stimuli in the

high-F0 context, while the green line represents those in the low-F0 context.

Consistent with previous studies, the results in Figure 2 were reanalyzed using Probit analysis to provide a clearer understanding of the data and to estimate the identification boundaries and shifts based on the preceding context (Chen & Peng, 2016). The boundary of categorical tone perception was defined as the onset F0 of the target stimuli corresponding to 50% on the lines. Detailed results are presented in Table 1 and Figure 3.

A repeated measures ANOVA was conducted to examine the categorical boundary, considering context type (cello, drum, rap, speech) and context frequency shifts (high, low) as within-subject factors. The analysis indicated a significant main effect of context type, $F(3, 188) = 14.139, p < 0.05$, while no significant effect was found for context frequency, $F(1, 46) = 0.432, p = 0.730$.

In a separate analysis, another repeated measures ANOVA was performed with the same factors. This time, a significant main effect of context frequency was observed, $F(3, 188) = 14.954, p < 0.01$, but no significant effect for context type, $F(1, 46) = 0.493, p = 0.688$.

Post hoc analyses were conducted for each context type individually. In the speech condition, variations in context frequency (high vs. low) significantly affected the results, $F(1, 46) = 15.195, p < 0.01$, leading to an approximate 1.016 shift in categorical boundary positions (refer to Table 1). Similarly, the rap condition showed a significant effect, $F(1, 46) = 13.338, p < 0.01$, with a boundary shift of approximately 1.012 (see Table 1). However, no significant effects were found for the cello condition, $F(1, 46) = 0.001, p = 0.977$, or the drum condition, $F(1, 46) = 0.002, p = 0.967$. These results suggest that the primary effects were driven by the speech and rap contexts, likely due to the perception of ambiguous tone stimuli in the middle range of the continuum (see Fig. 2, stimuli No. 5 to No. 7).

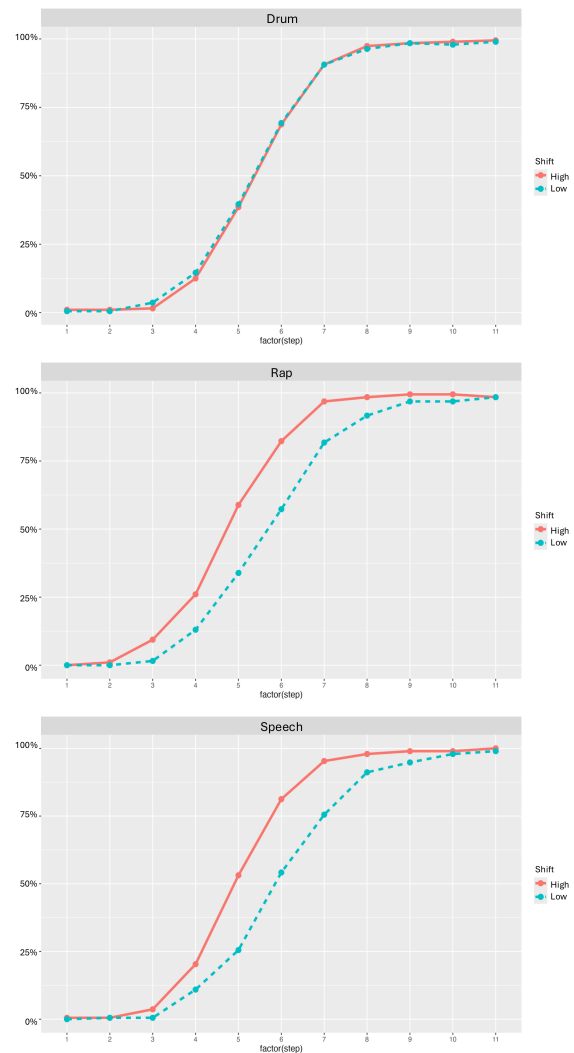


Figure 2 Average Tone 2 Response by Step and Shift for Each Context (from top to bottom: Cello, Drum, Rap and Speech)

Contexts	High	Low	Difference
Cello	5.357	5.349	0.008
Drum	5.429	5.419	0.009
Rap	4.805	5.817	1.012
Speech	5.012	6.028	1.016

Table 1 Derived categorical boundary positions for each type of context with high and low mean F0.

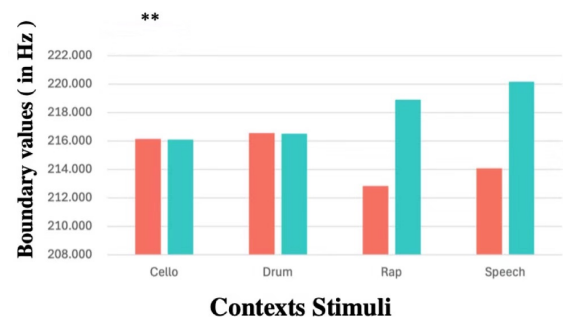
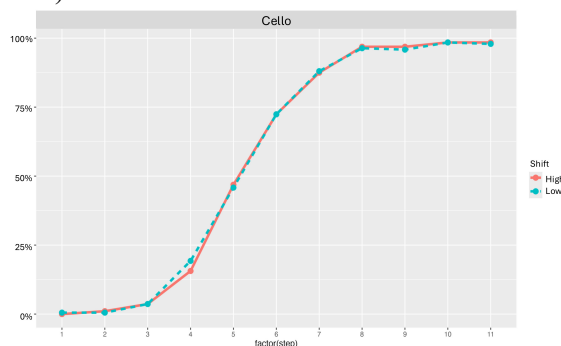


Figure 3 Category boundary values of each context types and frequencies. Red bars represent higher mean F0 context; blue bars represent lower mean F0 context. Higher category boundary values indicate more Tone 2 responses. **: $p < 0.01$.

4 Discussion

Our study reveals that typical instrumental music, such as cello and drum, does not induce lexical tone normalization. This finding aligns with previous research using piano as a musical stimulus (Tao et al., 2021), suggesting that instrumental music lacks the necessary elements for this process. Therefore, it appears that the absence of human vocal elements in instrumental music may be a critical factor. This absence might prevent listeners from accessing the pitch cues essential for normalization.

When considering rhythm, it is important to note that drum and rap are rhythmically strong, while cello and speech are rhythmically weak. Despite this classification, rhythm does not appear to significantly impact lexical tone normalization. Instead, rap and speech, despite their rhythmic differences, exhibit similar effects. This indicates that rhythm is not a critical factor in lexical tone normalization, suggesting that other elements may play a more significant role. The consistency of effects across different rhythmic contexts implies that the mechanism underlying tone normalization is robust to rhythmic variations.

Furthermore, the ability of rap music to induce lexical tone normalization challenges the notion that this process is exclusive to speech. This suggests that normalization may not be solely attributable to speech-like elocution but can also be triggered by rap, a distinct musical genre. Consequently, the shared elements between rap and speech might be key factors in this process. It is possible that these shared elements include pitch information and other vocal characteristics inherent in human speech. This finding opens new avenues for exploring how different types of vocalizations can influence cognitive processes related to language.

Moreover, the lack of significant effects from instrumental contexts highlights the importance of human vocal elements. The presence of vocal elements, regardless of classification as music or speech, seems necessary for lexical tone normalization. This reliance on human voices, particularly pitch information, suggests a

mechanism similar to vowel perception, where vocal characteristics establish a frame of reference. Thus, it is crucial to explore what specific elements in human voices contribute to this process. Understanding these elements could provide insights into how the brain processes complex auditory information.

Additionally, the absence of normalization in instrumental contexts may be due to differences from speech. In our experiment, the cello context was created by adjusting the average pitch of each word, which may have been perceived as distracting. Participants might have subjectively excluded these contexts during the task. Therefore, future experiments will explore whether including complete frequency information can induce normalization by better replicating tonal direction and ensuring participant focus. By refining the experimental design, we aim to determine whether more naturalistic instrumental stimuli can elicit normalization effects.

In light of these findings, it is evident that lexical tone normalization may not be limited to speech alone. We aim to identify the key factors involved, hypothesizing that these exist in the common elements between rap and speech stimuli. Future research will explore diverse rap genres and incorporate EEG experiments to distinguish the roles of speech information and human voice elements. This approach will help clarify the components that induce lexical tone normalization and explore the mechanism from multiple perspectives. By employing advanced neuroimaging techniques, we can gain a deeper understanding of the neural correlates of this phenomenon.

However, one limitation of our study is the relatively small number of participants (24 native Mandarin listeners), which may impact the generalizability of our findings. While our results suggest that lexical tone normalization is not exclusive to speech, the limited sample size means that these findings should be interpreted with caution. The small participant pool may not fully capture the diversity of perceptual abilities and linguistic backgrounds present in the broader population. Consequently, this limitation prevents us from definitively challenging the speech-specific hypothesis of lexical tone normalization. Future studies with larger and more diverse participant groups are necessary to validate our

findings and provide a more comprehensive understanding of the mechanisms involved.

Furthermore, we plan to use professional rappers and recruit participants with varied linguistic and musical backgrounds. Previous studies indicate that musical ability affects tone normalization (Zhang et al., 2023). Including Cantonese speakers or Mandarin-Cantonese bilinguals may provide further insights into perceptual abilities and their influence on results. This diversity in participants will allow us to examine how different backgrounds impact the perception of pitch information and contribute to lexical tone normalization. By considering individual differences, we can better understand the variability in normalization effects across populations.

In conclusion, our study expands the understanding of lexical tone normalization, suggesting it may be influenced by factors beyond speech. By identifying the shared elements between rap and speech, we can better understand the underlying mechanisms. This research has significant implications for future studies, as it highlights the need to explore various sound contexts and their potential to induce lexical tone normalization. Through continued investigation, we aim to uncover the fundamental reasons behind this phenomenon and its broader applications. Ultimately, this work contributes to a more comprehensive understanding of how humans process complex auditory stimuli and adapt to diverse linguistic environments.

Building on these findings, we propose a talker-specific hypothesis. The study found that neither of the instrumental music contexts (cello and drum) significantly affected lexical tone normalization, whereas speech contexts did, supporting the speech-specific hypothesis. Interestingly, rap contexts produced effects similar to speech, suggesting that rap functions more like speech than instrumental music in lexical tone normalization. This finding challenges the speech-specific hypothesis by demonstrating that rap music can induce lexical tone normalization in listeners similarly to typical speech.

Therefore, we suggest that the critical factor may not be speech per se, but rather the presence of human-produced vocal sounds. This implies that different types of vocalizations, under certain conditions, can trigger lexical tone normalization. It may be necessary for listeners to subjectively

recognize and interpret the sounds as human-produced. We aim to further investigate the mechanisms and conditions under which this occurs, proposing that the talker-specific hypothesis could provide a broader framework for understanding these effects. Future research will focus on exploring the underlying processes and constraints of this hypothesis, contributing to a deeper understanding of auditory processing and linguistic adaptation.

5 Conclusion

The study found that neither of the instrumental music contexts (cello and drum) significantly affected lexical tone normalization, whereas speech contexts did, supporting the speech-specific hypothesis.

Interestingly, rap contexts produced effects similar to speech, suggesting that rap functions more like speech than instrumental music in lexical tone normalization. This finding challenges the speech-specific hypothesis by demonstrating that rap music can induce lexical tone normalization in listeners similarly to typical speech.

The lack of significant effects from instrumental music, contrasted with the effects from contexts containing human voices, implies that lexical tone normalization may depend on the presence of human vocal elements, particularly pitch information. Additionally, the similar effects observed across different rhythm types suggest that tone normalization may not be sensitive to variations in rhythm.

Instrumental music contexts did not yield significant effects, whereas materials containing human voices did. This suggests that lexical tone normalization may rely on the presence of human voices, especially the pitch information within those voices.

A type of rap music context can induce lexical tone normalization in listeners similarly to typical speech. This finding may conflict with the speech-specific hypothesis.

Acknowledgments

This study has been supported by an internal grant from The Hong Kong Polytechnic University (Project No P0051041).

References

- Ainsworth, W. A. (1974). The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech*, 17, 103–9.
- Chen, F., & Peng, G. (2016). Context Effect in the Categorical Perception of Mandarin Tones. *Journal of Signal Processing Systems*, 82(2), 253–261. <https://doi.org/10.1007/s11265-015-1008-2>
- Dechovitz, D. (1977). Information conveyed by vowels: A confirmation. *Haskins Laboratory Status Report on Speech Research*, SR-53/54, 213–19.
- Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *Journal of the Acoustical Society of America*, 125(6), 3983–94.
- Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America*, 88, 642–54.
- Ladefoged, & Peter. (2005). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98–104.
- Leather, J. (1983). Speaker normalization in the perception of lexical tone. *Journal of Phonetics*, 11, 373–82.
- Nearey, T. M. (1978). Phonetic feature systems for vowels. Indiana University Linguistics Club, Bloomington, IN.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85, 2088–113.
- Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 38(4), 616–624.
- Peng, G., Zhang, C., Zheng, H. Y., Minett, J. W., & Wang, W. S. Y. (2012). The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research*, 55(2), 579–595. [https://doi.org/10.1044/1092-4388\(2011/11-0025\)](https://doi.org/10.1044/1092-4388(2011/11-0025))
- Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 40–61.
- Tao, R., & Peng, G. (2020). Music and speech are distinct in lexical tone normalization processing. *The 34th Pacific Asia Conference on Language, Information and Computation*.
- Tao, R., Zhang, K., & Peng, G. (2021). Music Does Not Facilitate Lexical Tone Normalization: A Speech-Specific Perceptual Process. *Frontiers in Psychology*, 12(October), 1–14. <https://doi.org/10.3389/fpsyg.2021.717110>
- Zhang, C., Peng, G., & Wang, W. S. (2012). Unequal effects of speech and non-speech contexts on the perceptual normalization of Cantonese level tones. *The Journal of the Acoustical Society of America*, 132(2), 1088–1099. <https://doi.org/10.1121/1.4731470>
- Zhang, C., Peng, G., & Wang, W. S. Y. (2013). Achieving constancy in spoken word identification: Time course of talker normalization. *Brain and Language*, 126(2), 193–202. <https://doi.org/10.1016/j.bandl.2013.05.010>
- Zhang, C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., Peng, G., & Wang, W. S. Y. (2016). Functionally integrated neural processing of linguistic and talker information: An event-related fMRI and ERP study. *NeuroImage*, 124, 536–549. <https://doi.org/10.1016/j.neuroimage.2015.08.064>
- Zhang, K., Sjerps, M. J., & Peng, G. (2021). Integral perception, but separate processing: The perceptual normalization of lexical tones and vowels. *Neuropsychologia*, 156(April 2020), 107839. <https://doi.org/10.1016/j.neuropsychologia.2021.107839>
- Zhang, K., Tao, R., & Peng, G. (2023). The advantage of the music-enabled brain in accommodating lexical tone variabilities. *Brain and Language*, 247, 105348. <https://doi.org/10.1016/j.bandl.2023.105348>