



## OPEN ACCESS

## EDITED BY

Caicai Zhang,  
The Hong Kong Polytechnic University,  
Hong Kong SAR, China

## REVIEWED BY

Jiaqiang Zhu,  
Hong Kong Polytechnic University,  
Hong Kong SAR, China  
Aspasia Eleni Paltoglou,  
Manchester Metropolitan University,  
United Kingdom

## \*CORRESPONDENCE

Maoping Zheng  
✉ zhengswu@126.com

RECEIVED 08 August 2023

ACCEPTED 07 November 2023

PUBLISHED 27 November 2023

## CITATION

Hao J, Zhong Y, Li H, Li J and Zheng M (2023)  
The effect of instrumental and vocal musical  
experience on tone categorical perception in  
individuals in a Chinese dialect area.  
*Front. Educ.* 8:1274441.  
doi: 10.3389/feduc.2023.1274441

## COPYRIGHT

© 2023 Hao, Zhong, Li, Li and Zheng. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# The effect of instrumental and vocal musical experience on tone categorical perception in individuals in a Chinese dialect area

Jiayi Hao<sup>1,2</sup>, Yuhuan Zhong<sup>3</sup>, Hong Li<sup>4</sup>, Jianbo Li<sup>5</sup> and  
Maoping Zheng<sup>6\*</sup>

<sup>1</sup>Key Laboratory of Cognition and Personality (Ministry of Education), Southwest University, Chongqing, China, <sup>2</sup>Chongqing Vocational College of Culture and Arts, Chongqing, China, <sup>3</sup>Chongqing Vocational and Technical University of Mechatronics, Chongqing, China, <sup>4</sup>Chongqing Foreign Language School, Chongqing, China, <sup>5</sup>Chongqing Municipal Educational Examinations Authority, Chongqing, China, <sup>6</sup>School of Music, Southwest University, Chongqing, China

**Introduction:** Previous studies have shown that music training modulates adults' categorical perception of Mandarin tones. However, the effect of music training on tone categorical perception ability in individuals in Chinese dialect areas remains unclear.

**Methods:** We recruited three groups of college students majoring in music in Chongqing, a dialect area in southwestern China. These groups included an instrumental music group ( $n = 31$ ), a vocal music group ( $n = 32$ ), and an instrumental-vocalist music group ( $n = 30$ ). Additionally, we recruited a group of college students who did not receive any music training ( $n = 32$ ). The accuracy and reaction time of the four groups were measured using the oddball task and compared to explore the differences in their tone categorical perception ability.

**Results:** Considering both between-tone category and within-tone category, the three music groups showed both greater accuracy and shorter reaction times than the non-music group. Regarding the three music groups, no significant differences in accuracy and reaction time were detected between the instrumental and vocal groups. However, the instrumental vocalist group outperformed both the instrumental and vocal groups in terms of accuracy and reaction time.

**Discussion:** Results showed that music training seems to have a positive effect on the categorical perception of Mandarin tone categories in Chinese dialect areas, and demonstrated that the combination of instrumental and vocal music training can further enhance tone categorical perception. To some extent, these findings provide a theoretical basis for the improvement of individual tone perception ability in dialect areas, and provided theoretical support for music and language education.

## KEYWORDS

mandarin, instrumental training, vocal training, tone categorical perception, music experience, dialect area

## 1 Introduction

Commonly used by Chinese people, Mandarin is the standard language in China, with Beijing pronunciation as its standard, the northern dialect its basic dialect, and the typical vernacular its grammatical norm (Li, 2006). Speaking Mandarin can eliminate social estrangement between distinct dialect areas, not only promoting social communication internally but also spreading Chinese culture externally (Wu, 2010).

China is comprised of a large number of dialect and ethnic minority autonomous regions. Each dialect region has a particular dialect that is spoken by local residents. For example, Cantonese is spoken by local Cantonese speakers, while the Minnan dialect is spoken by Taiwanese and Fujian speakers. Noticeably, residents in the southwest region of China use local dialects in their daily social communications. Some studies have found that the linguistic environment of a dialect area affects the level of Mandarin proficiency among the local residents (Wang and Cao, 2002; Xia, 2007). Using sampling inspection, Xia (2007) examined the composition papers of second grade students in the Sichuan dialect area and found that the use of incorrect characters was caused by differences in pronunciation between the local dialect and Mandarin. Han (2017) conducted a statistical analysis on the results of teachers who had taken the National Mandarin Proficiency Test since 2015, and found that teachers in Beijing had the highest level of Mandarin proficiency, followed by Heilongjiang; Hebei and Jilin; Shandong, Shanxi, Jiangxi, Inner Mongolia, and Xinjiang; and Anhui, Shanghai, Fujian, Guangdong, Guangxi, Hainan, Gansu, and Guizhou. The results of the study found that individuals in the North had higher Mandarin scores than those in the South. Thus, dialect regions are an important factor affecting Mandarin proficiency, with vocal tone pronunciation being a key aspect of this (Wang and Cao, 2002).

The division of modern Chinese dialects is based on phonetic standards, and the tonal systems of different dialects vary greatly. People from different dialect backgrounds speak Mandarin with strong regional tones. Some have low actual tonal values for level tone (T1), while others cannot distinguish between level and falling tones (T4). Some studies have found that vocal tone perception ability varies among individuals from different dialect areas (Yu and Huang, 2019). Wang et al. (2016) conducted a comparative study on the Chongqing and Beijing dialects and found that native speakers in Chongqing had weaker categorical perception of the continuum of Yinping–Yangping (rising–falling) in Mandarin than their Beijing counterparts. Yu and Huang (2019) compared residents in the Beijing dialect area with those in three other dialect areas (Boshan, Nanchang, and Cantonese), and found that people speaking different dialects demonstrated different lexical tone categorical perception, mainly in terms of boundary width, an index reflecting the degree of lexical tone perception ability; the narrower the width, the stronger the categorical perception ability. However, the boundary width of the third and fourth tones (T3–T4) in the Guangdong dialect area is smaller than that in the Beijing dialect area. Additionally, individuals' boundary width in the other two dialect areas was significantly larger than that of those in the Beijing dialect area. These findings indicate that different dialect areas exhibit varying tone perception abilities. As stated, Cantonese-speaking listeners showed smaller boundary width than those from Beijing. In particular, tone perception was closely related to tone pronunciation. Meanwhile, Liang (2017) studied the

causes of confusion in Cantonese pronunciation, indicating that tone pronunciation confusion is initially caused by tone perception confusion. Consequently, to improve Mandarin pronunciation, tone perception ability should be addressed.

Mandarin is a tonal language comprising four tones: high-level (T1), mid-rising (T2), low-dipping (T3), and high-falling (T4). These four tones have their own categories and can be categorically perceived. Mandarin tones are characterized by between- and within-category perception. Between-tone category perception refers to the ability to distinguish between the four different tones (e.g., the pitch difference between T1, T2, T3, and T4), while within-tone perception is the ability to perceive internal variations in the fundamental frequency value of each tone (e.g., 210–300 Hz and 230–270 Hz both belong to T2). Changes between tonal categories are usually easily perceived, whereas changes within the tonal categories are not (Studdert-Kennedy and Shankweiler, 1970). Therefore, in this study, we anticipated that there may be some differences in the between and within-tone categorical perception of our participants.

In recent years, an increasing number of scholars have begun to explore the influence of music training on language processing (Schön et al., 2004; Marques et al., 2007; Marie et al., 2011; Zatorre, 2013; Tang et al., 2016; Besson et al., 2018; Nan et al., 2018; Barbaroux et al., 2021), with some investigations finding that music training has an effect on individuals' categorical perception of Mandarin tone (Wu et al., 2015; Tang et al., 2016; Yao and Chen, 2020). Therefore, music training may have an effect on the categorical perception of Mandarin tone.

The influence of music training on the categorical perception of tone is mainly based on the commonality between music and language. Although music and language belong to distinct fields and have traditionally been regarded as different psychological abilities (Jäncke, 2012), they are closely associated with one another and share numerous common characteristics. For example, Patel (2012a) believes that music and language are abilities that are unique to human beings, distinguishing them from other species. Both abilities pertain to the category of hearing. Considering their interior structure, both music and language involve complex and meaningful sequences and can express emotions through changes in pitch and speed. Many researchers have found that music and language also display certain overlapping neural mechanisms (Patel et al., 1998; Maess et al., 2001; Tillmann, 2005; Koelsch, 2006). For example, Broca's area, which is located in the left inferior frontal region of the brain, is believed to be involved in language production. Likewise, through magnetoencephalogram and source analysis, Maess et al. (2001) found that peoples' perception of out-of-tune music and sound sequences also activated Broca's area, as well as the corresponding brain area in the right hemisphere. A similar result was obtained in an event-related potential study: The P600 component, which is associated with language syntactic violations, was also evoked by out-of-tune music chords (Patel et al. (1998). Previous studies using functional magnetic resonance imaging (fMRI) have demonstrated that irregular chords in music can also activate the "language regions" of the brain, including Broca's area, the superior temporal gyrus, transverse temporal gyrus, temporal plane, Wernicke's area, and anterior superior insular cortex (Koelsch et al., 2002; Tillmann et al., 2003; Tillmann, 2005; Koelsch, 2006; Chen et al., 2023). A meta-analysis of music and language revealed that in the low-level processing stage, 50% of the left hemisphere demonstrated an overlapping of language and music (Lai et al., 2014).

In fact, there are indeed theoretical studies and findings supporting the influence of music training on speech processing. On the basis that music and language processing may share similar cognitive resources and neural processes, [Patel \(2012b\)](#) proposed the resource-sharing framework theory, which holds that although music and language have different neural representation networks, they share specific neural mechanisms. Following this, Patel proposed the OPERA theory ([Patel, 2011](#); [Patel, 2014](#)), in which the acronym OPERA represents five key elements: overlap, precision, emotion, repetition, and attention. Overlap indicates that language and music share many perceptual and cognitive processing processes. Compared to language, music learning can require more precise vocal representation, as well as stronger emotional rewards, multiple repetitions, and a high concentration of attention. The theory stipulated that (1) the sensory or cognitive processes shared by music and language (e.g., waveform periodic coding and auditory working memory) are regulated by overlapping brain networks; (2) music requires more precision than language in this process; and (3) music combines emotion, repetition, and attention with sensory and cognitive processes shared by language, spurring the formation of neural plasticity in language processing.

It is also worth considering both the top-down and bottom-up hypotheses. The top-down hypothesis is based on the idea that music training promotes language processing by improving one's general cognitive abilities (e.g., attention, memory, executive functions, etc.) ([Besson et al., 2011, 2018](#)). The bottom-up hypothesis is based on the findings that music training contributes to the enhancement of one's sensitivity to the common acoustic characteristics (pitch, timbre, intensity, and length) of music and language. Another explanation lies in the characteristics of the multi-sensory channels involved in music training. In long-term music training, musicians relate their internal motion and auditory representation to sound more accurately. Therefore, musicians' perceptual ability is concurrently supported by the absorption of imperceptible information through other channels (vision, touch, etc.), confirming that they had better perception than that of non-musicians ([Lee and Noppeney, 2011](#)).

The hypotheses reviewed above provide a theoretical basis for the positive effects of music training on language processing. Nonetheless, there are few studies that have examined the influence of music training on tone categorical perception. [Wu et al. \(2015\)](#) measured Chinese musicians' and non-musicians' categorical perception of a lexical tone continuum (from the high-level tone to the high-falling tone). They found that music majors were more sensitive to the subtle acoustic differences of Mandarin tones, which was mainly reflected in higher accuracy of within-tone categorical perception. Another study provided 12 months of music training to children aged four and five and experimentally examined three metrics of tone categorical perception in these children: boundary position, boundary width, and discrimination accuracy. Boundary position refers to the point at which the two discriminant curves of the two recognition functions (i.e., the intersection of the two curves) correspond to the stimulus value at a 50% recognition rate. Boundary width is the linear distance between the 25 and 75% recognition rates, calculated using means and standard deviations in probability analysis ([Finney, 1972](#)). Generally, a narrower boundary width indicates a faster transition rate from one tonal category to another and a higher level of categorization. This study found that the 12 months of music training enhanced the children's tone categorical perception, as evidenced by significantly

smaller categorical boundary widths in the music group than in the control group. In addition, the music training also enhanced children's perceptual sensitivity for differentiating between within-tone categories ([Yao and Chen, 2020](#)). Previous studies have shown that music training benefits tone categorical perception, which is mainly manifested that the boundary width of tone categorical perception of musically trained individuals is smaller than that of non-musically trained individuals, while the accuracy of within-tone category differentiation is higher in musically trained individuals than in non-musically trained individuals.

Based on this short review, it appears that more experiments are needed to investigate the effect of music training on tone categorical perception in individuals with a tonal language background. In addition, whether music training has the same impact on individuals in dialect areas remains unclear. Therefore, the primary goal of this study was to explore whether music training fosters individuals' categorical perception of tone in dialect areas.

Music training, generally involving vocal and instrumental music training, is a long-term multi-system learning process ([Wang et al., 2015](#)). Instrumental music refers to the type of music played only with instruments, without any vocal sounds or a subordinate vocal voice. In contrast, based on human vocal cords, with the mouth, tongue, and nasal cavity acting on the breath, vocal music issues a rather pleasant, continuous, and rhythmic sound ([Nikjeh et al., 2009](#)). Many scholars are uncertain about which type of training has a larger influence. Accordingly, a few scholars have conducted a series of studies to compare the effects of vocal and instrumental music training on speech perception and production ([Nikjeh et al., 2008, 2009](#); [Kirkham et al., 2011](#)). Initially, they predicted that vocal musicians would perform better than instrumental musicians in all respects, because compared with instrumental musicians, vocalists' musical "instrument" is endogenous to the body (i.e., the larynx and vocal tract). As early as infancy, an individual's larynx receptor system is continuously affected by their auditory system. Thus, this influence begins earlier than that of instrumental music training, so the vocalist's ability to control and perceive the pitch should be stronger. However, the results showed no differences in pitch perception and phonological production between instrumentalists and vocalists. Previous psychoacoustic and electrophysiological studies have shown that the auditory skills of musicians of different music types may vary ([Kishon-Rabin et al., 2001](#); [Nager et al., 2003](#); [Tervaniemi et al., 2006](#); [Seppänen et al., 2007](#)). Therefore, determining which type of music training method has a stronger effect on Mandarin tone categorical perception is the second objective of this study.

To answer these questions, we chose tone categorical perception as the experimental paradigm for this study. There were two main reasons for this. First, previous studies have shown that 3-year-old children can better master the tones of Mandarin, and their accuracy in perceiving the four tones of Mandarin can reach 90% ([Wong et al., 2005](#)). Therefore, when tonal perception is used as the experimental material, ceiling effects may appear. Second, to date, there is little research using Mandarin tone continuum materials to explore the impacts of different music training methods on tone perception. Most of the experimental materials adopted in previous studies were syllables or non-speech stimulus materials. For example, [Kirkham et al. \(2011\)](#) used six syllables generated from four tones as the experimental materials in their experiments on the processing of Mandarin tones by instrumentalists and vocalists. [Nikjeh et al. \(2008\)](#) also studied pitch perception in

instrumentalists and vocalists, and the materials they used were homophonic compounds with non-speech stimuli. Therefore, the tone category perception mode with a more refined perception of tone can be used to better explore the effect of music training on the Mandarin tone perception ability of individuals in dialect areas.

Previous research has shown that music training can promote tone categorical perception of individual tone languages (Marie et al., 2011; Yao and Chen, 2020), but whether music training has the same effect on individual tone language dialect areas has not yet been examined. Perception of individual tone categories varies across different dialect areas (Yu and Huang, 2019). Therefore, in the current study, we first used the oddball paradigm to compare differences in the tone categorical perception of musicians and non-musicians in dialect areas. The oddball paradigm focuses on the random presentation of two stimuli—standard and deviant stimuli—in the auditory channel of the experiment. The probability of occurrence of the standard stimulus was 90% and that of the deviant stimulus was 10%. Participants differentiated between the two stimuli and pressed a key in response.

Furthermore, it is very important to use specific musical training methods to improve the tone categorical perception ability of individuals in dialect areas, but there is no unified approach to doing so in the existing literature. Hence, we also explored whether different music training methods (i.e., instrumental and vocal music training) display different effects on the Mandarin tone categorical perception ability of individuals in dialect areas. There are many ways to train in instrumental music, including keyboard, string, wind, percussion, and so on, and previous studies have found that different types of music training have different effects on musicians' auditory perception (Kishon-Rabin et al., 2001; Nager et al., 2003; Tervaniemi et al., 2006; Seppänen et al., 2007). In this study, we selected the piano as the instrumental training method, because it has a strong universality and wide range. Moreover, some studies have found that piano training can promote tone pitch perception even better than percussion training (Nan et al., 2018; Yao, 2020).

## 1.1 Our hypotheses are as follows

*H1:* Music training will have a beneficial impact on the tone categorical perception of individuals in dialect areas. This was mainly reflected in the fact that the music group had a lower error rate than the non-music group, while the music group responded faster than the non-music group.

*H2:* Considering that vocal musicians obtain early auditory and kinaesthetic feedback from their bodily instrument (i.e., the larynx and vocal tract) when first uttering words in infancy, we propose that the accuracy of the vocal group will be higher than that of the instrumental group, and that the response time of the vocal group will be faster than that of the instrumental group.

*H3:* One study found that vocal musicians with instrumental training appear to have an auditory neural advantage over musicians with only instrumental or vocal training (Nikjeh et al., 2008). Therefore, we propose that individuals who received combined instrumental and vocal training will have lower error rates and faster reaction times than individuals who learned only instrumental or vocal music.

## 2 Materials and methods

### 2.1 Participants

We determined the sample size by utilizing G\* power software (version 3.1) (Faul et al., 2007). This study utilized a two-factor between-subjects design and statistical tests using two-factor repeated measures ANOVA. Based on previous research, to obtain power of 0.80 with alpha level set to 0.05 to observe a medium effect of  $|p| = 0.3$ , the projected sample size was at least 68 people in total. One hundred and twenty-five students from Southwest University were recruited to participate in the study. All participants were from Chongqing dialect area (located in southwest China), and were divided into four groups: the instrumental music group, the vocal music group, the instrumental vocal music group (a vocalist with instrumental music experience), and the non-music group (Table 1). The instrumental group included the participants who had received education and practice of music for at least 8 years and had started daily practice for about 1–3 h before the age of 8. In addition, the subjects in the instrumental music group had never received systematic vocal training. The instruments practiced by the participants mainly included piano (18), violin (6), trumpet (5) and guzheng string (2). The vocal music group practiced by participants included Folk singers (19) and Bel canto singers (13). The students in vocal group have never taken systematic instrumental training. In contrast to that, the instrumental vocal music group had acquired piano training. Both piano and vocal music group have studied for at least 6 years, with an average of 1 to 3 h of practice every day. As the non-musicians had never received formal music training in the past five years and to prevent amusia, all non-musicians

TABLE 1 The demographic characteristics of the musicians and non-musicians and describing statistical data.

	Instrument group (M ± SD)	Vocal group (M ± SD)	Instrument vocalist group (M ± SD)	Non-musician group (M ± SD)
	n = 31	n = 32	n = 30	n = 32
Mean age (range)	20.6 (18–23)	21.7 (18–24)	21.3 (18–24)	21.2 (18–24)
Male/female	12/19	15/17	14/16	15/17
acc (between)	0.96 (0.26)	0.95 (0.06)	0.98 (0.02)	0.84 (0.14)
acc (within)	0.91 (0.09)	0.95 (0.05)	0.97 (0.03)	0.72 (0.21)
rt (between)	315.12 (140.79)	340.10 (98.14)	319.81 (96.17)	485.26 (163.86)
rt (within)	30.8.28 (125.72)	336.50 (130.59)	245.99 (74.29)	392.32 (139.79)

participants were required to complete the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003). The average score of each non-musician participant is 70%. All participants were right-handed (Oldfield, 1971) without hearing impairment. All signed an informed consent in compliance with a protocol approved by the research ethics committee of Southwest University, and were paid for their participation. The study was approved by the Ethics Committee of Southwest University (No. H23089).

This study used a mixed experimental design of 4(subjects: vocal group vs. instrumental group vs. instrumental vocalist group vs. non-musician) x 2 (conditions: between-category vs. within-category) was used. Subjects were the between-group variable and stimuli were the within-group variable.

## 2.2 Stimuli

Two Chinese syllables, /fa2/ and /fa4/, were constructed for the lexical tone condition, which shared the same phoneme but carried differences in their lexical tone contours (2= rising tone and 4= falling tone; Figure 1). The original stimuli were produced by a female native Mandarin speaker and recorded at a sampling rate 44.1 kHz. Then the two speech stimuli are processed in Praat. Except for the pitch contour difference, other acoustic parameters are the same (duration, strength, envelope, etc.). The duration of stimulation is 350 ms, including 10 ms rise and fall time, and the amount of stimulation is 70 dB. The above two stimuli are taken as both ends of the tone continuum and adopt PSOLA synthesis technology to generate 11 stimuli as the test material for tone category perception,

and the stimulated tone curve is shown in Figure 1. Subsequently, before the formal experiment, 15 native speakers of Mandarin made an evaluation of these 11 sound stimuli, and were asked to judge if each stimulus that randomly appeared was two or four tones, and then pressed the key to respond. Each stimulus was presented at random for 6 times with a time interval of 1,000 ms. According to the results of behavioural experiments (Figure 2; Table 2), S3

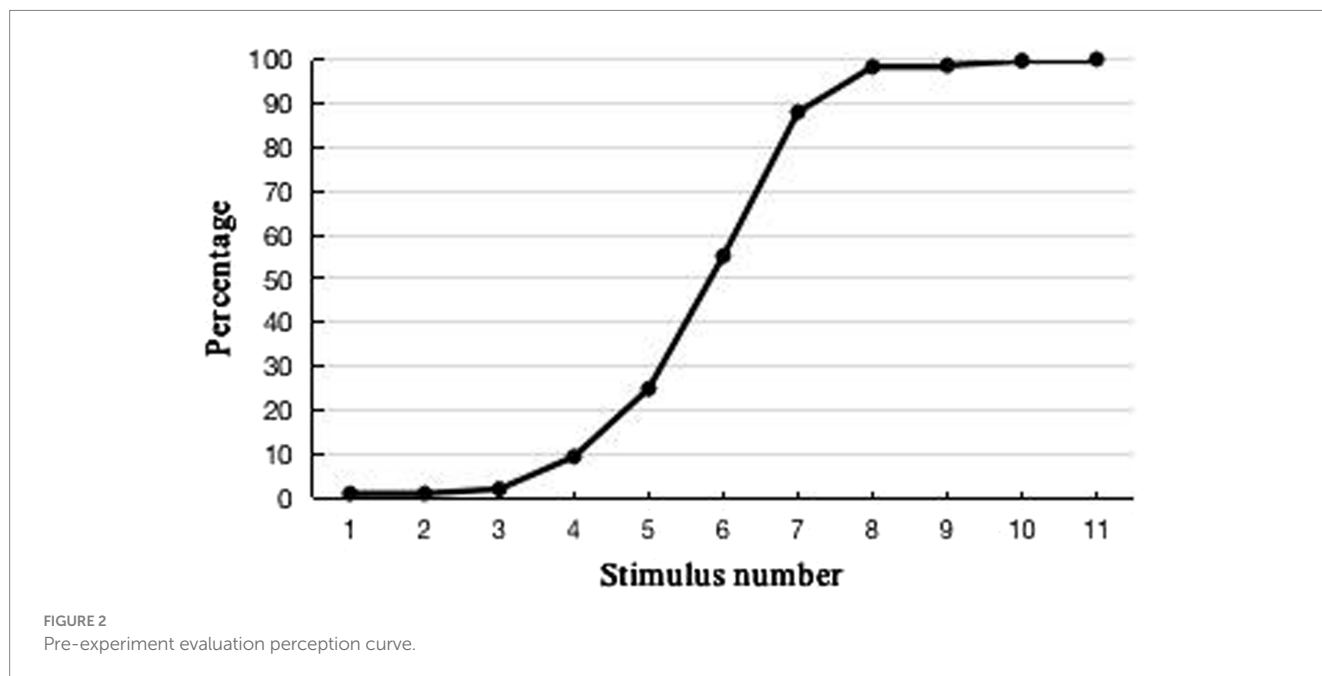
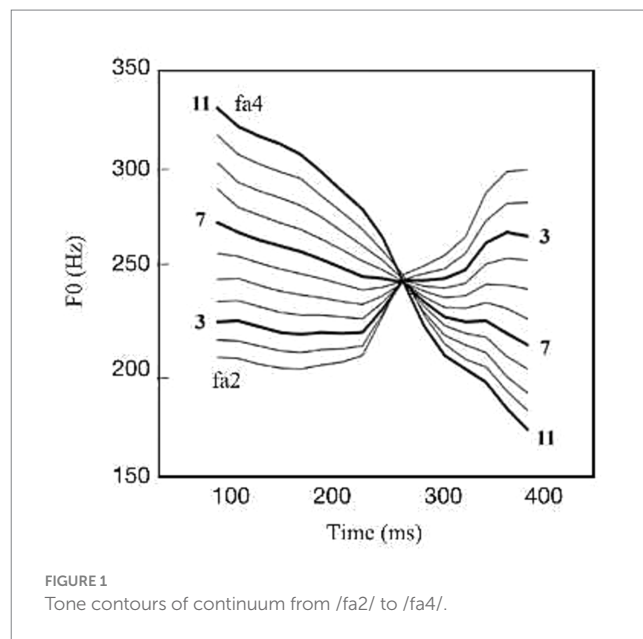


TABLE 2 The behavioural results.

Stimulus number	1	2	3	4	5	6	7	8	9	10	11
/Fa4/accuracy	1.04	1.04	2.07	9.45	24.93	55.17	87.95	98.28	98.57	99.66	100

(Stimulus 3) and S7 (Stimulus 7) are the inflection points of tone categorical perception. The tone probability before S3 and S3 is recognized as the rising tone, and the tone probability after S7 and S7 is recognized as the falling tone. Therefore, S7 is regarded as the standard stimulus, S3 as between-category deviation stimulus, and S11 (Stimulus 11) as within-category deviation stimulus. Hence, S3 and S7 were used as between-category stimulus pairs, and S7 and S11 were adopted as within-category stimulus pairs as experimental materials.

## 2.3 Procedure

The classic oddball paradigm used in the study was referenced from Tang et al. (2016). There are two blocks in the experiment (Figure 3). One block includes standard stimuli and between-category stimuli, and the other includes standard stimuli and within-category stimuli. Each block includes 500 trials that appeared in a pseudo-randomized manner with at least two successive standards between deviants. The probability of standards is 90% and of deviants is 10%. The 15 stimuli at the beginning of each block are standard stimuli. Two blocks are presented randomly. The duration of each block is 15 min, and there is a 5-min rest time between the two blocks. For each trial a fixation cross “+” would first appear at the centre of the screen for 500 ms, subsequent presentation of 350 ms of speech stimulation, followed by a blank screen of 650 ms, completing a single trial. Participants were required to wear headphones and watch silent movies, press the “F” key when hearing the standard stimulus, and press the “J” key when hearing the deviant stimulus. The distance between the movie video and the subjects is 70 cm. The stimulus were presented through binaural headphones with a volume of 70 dB, and the subjects subsequently began to complete the oddball task.

## 2.4 Data analysis

We investigate the perception ability of four groups (instrumental musicians, vocal musicians, instrumental vocalist, non-musicians) and two stimulus types (standard and deviant) between and within-tone categories. Repeated-measures ANOVA on the accuracy and reaction time of between and within-category of tone were conducted, with group as between-subjects factor, and stimulus as within-subjects factor. The accuracy was calculated as the number of correct trials divided by the total number of trials, and reaction time was calculated as the time that the participant responded correctly. All of the analyses were conducted by using SPSS 22.0. The *p*-values were adjusted for

sphericity using the Greenhouse–Geisser method. Post-hoc *t*-tests with Bonferroni adjustments were used for multiple comparisons. Based on the  $\pm 3$  standard deviations, 3 participants were excluded. Ultimately, 122 valid samples were obtained, including 30 in the instrumental group, 31 in the vocal group, 30 in the instrumental group, and 31 in the non-musical group.

## 3 Results

### 3.1 Accuracy

Results on the ACC showed a significant main effect of stimulus type,  $F(1, 121) = 12.81, p < 0.001, \eta^2 = 0.10$  (Figure 4). The accuracy in between-tone category was greater than that in within-tone category (Table 1). The results also showed a significant main effect of the group,  $F(1, 121) = 47.87, p < 0.001, \eta^2 = 0.54$ ; *post hoc t* test showed that the accuracy in instrumental group, vocal group and instrumental vocalist group in the tone categorical perception is higher than that of non-music group (Table 1). The accuracy in instrumental vocalist group was higher than that in instrumental music group ( $p = 0.03$ ), but there was no significant difference between instrumental music group and vocal group,  $p = 0.29$ , or between vocal group and instrumental vocalist group ( $p = 0.24$ ).

The results also showed an interaction between groups and stimulus types,  $F(1, 121) = 5.00, p = 0.003, \eta^2 = 0.11$ . Simple effect analysis showed that, for between-tone categorical perception, accuracy was higher in the instrumental group, the vocal group and the instrumental vocalist group than in the non-musical group ( $p < 0.001$ ) with no difference between the three music groups (instrumental group vs. vocal group:  $p = 0.84$ ; instrumental group vs. instrumental vocalist group:  $p = 0.25$ , and vocal group vs. instrumental vocalist group:  $p = 0.17$ ). For within-tone categorical perception, accuracy was higher in the instrumental group, the vocal group and the instrumental vocalist group than in the non-musical group ( $p < 0.001$ ), and the accuracy in instrumental vocalist group was higher than that in instrumental group ( $p = 0.05$ ), with no difference between the instrumental group and vocal group ( $p = 0.15$ ), or between vocal group and instrumental vocalist group ( $p = 0.56$ ).

### 3.2 Reaction time

Results on the RT showed a significant main effect of stimulus type,  $F(1, 121) = 13.44, p < 0.001, \eta^2 = 0.10$  (Figure 5). The reaction time in between-tone category was longer than that in within-tone

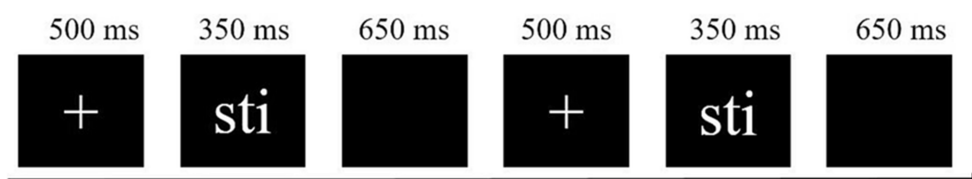


FIGURE 3  
Oddball paradigm flow chart.

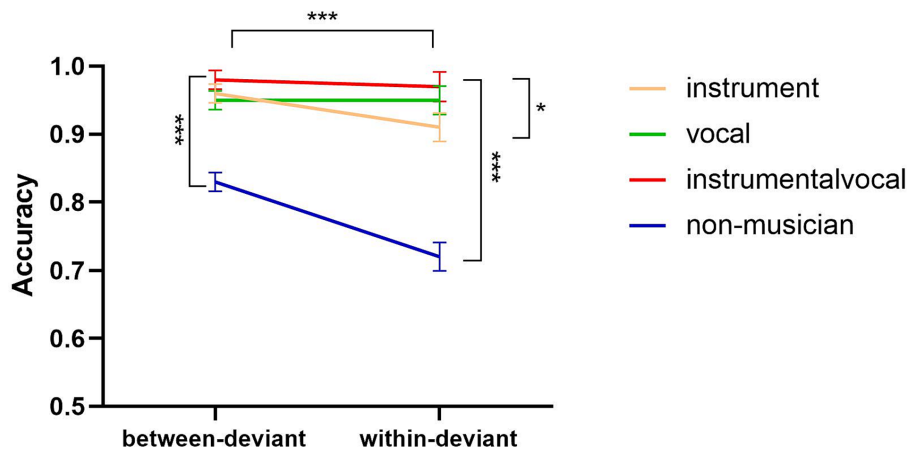


FIGURE 4 Accuracy of instrumental group, vocal group, instrumental and non-musical group in between-category stimulation and within-category stimulation. Error bar indicates standard error (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

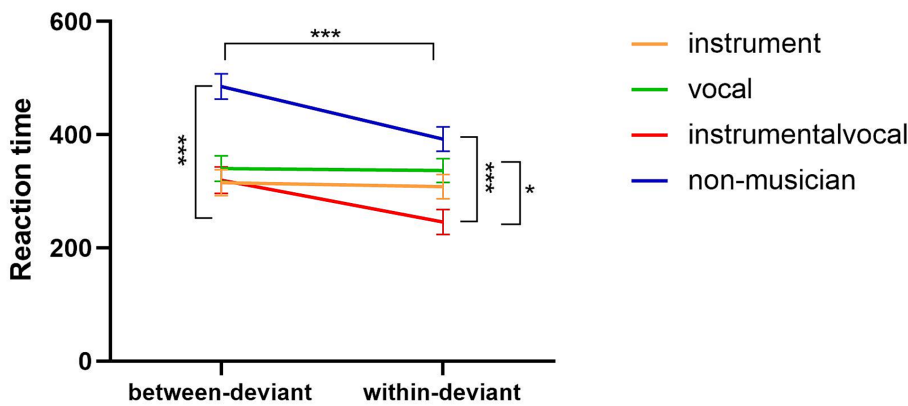


FIGURE 5 Reaction time of instrumental group, vocal group and non-musical group in between-category stimulation and within-category stimulation. Error bar indicates standard error (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

category (Table 1). The results also revealed a significant main effect of the group,  $F(1, 121) = 13.13, p < 0.001, \eta^2 = 0.25$ ; *post hoc t* test showed that the reaction time in instrumental music, the vocal group and the instrumental vocalist group in the perception of tone category remained shorter than that of non-music group (Table 1). The reaction time in instrumental vocal music group was faster than that of vocal group,  $p = 0.04$ , but there was no significant difference between instrumental music group and vocal music group,  $p = 0.32$ , or between instrumental group and instrumental vocalist group ( $p = 0.29$ ).

The results also presented an interaction between groups and stimulus types,  $F(1, 121) = 3.63, p = 0.02, \eta^2 = 0.08$ . Simple effect analysis showed that, for between-tone categorical perception, reaction time was shorter in the instrumental group, the vocal group and the instrumental vocalist group than in the non-musical group ( $p < 0.01$ ) with no difference between the three music groups (instrumental group vs. vocal group:  $p = 0.44$ ; instrumental group vs. instrumental vocalist group:  $p = 0.89$ ; and vocal group vs.

instrumental vocalist group:  $p = 0.54$ ) For within-tone categorical perception, reaction time was shorter in the instrumental group and instrumental vocalist group than in the non-musical group (instrumental group vs. non-musical group:  $p = 0.007$ ; instrumental vocalist group vs. non-musical group:  $p < 0.001$ ), but vocal group and non-musical group had borderline significant ( $p = 0.07$ ), and the reaction in instrumental vocalist group was shorter than that in vocal group ( $p = 0.004$ ). Similarly, there were no difference between the instrumental group and vocal group ( $p = 0.36$ ) or between instrumental group and instrumental vocalist group ( $p = 0.08$ ).

## 4 Discussion

This study used the oddball paradigm to explore the effect of instrumental, vocal, instrumental vocalist, and non-musical experiences on individuals' tone category perception of Mandarin.

The results showed that, compared with the non-music group, the music groups (comprising instrumental, vocal, and instrumental vocalist groups) had a stronger categorical perception of Mandarin tone. In particular, the music groups had better accuracy scores and reaction times than the non-music group in both between and within-tone categorical perception.

#### 4.1 Differences in tone categorical perception accuracy between the music and non-music groups

As stated above, the accuracy rates of the instrumental, vocal, and instrumental vocalist groups were significantly higher than those of the non-music group in both between and within-tone categorical perception. This indicates that receiving music training enhances subtle tone processing. This finding is similar to that of [Tang et al. \(2016\)](#), who also used the oddball paradigm and found that the music group performed better than the non-music group. However, this result differs from the outcomes of [Mok and Zuo \(2012\)](#) and [Wu et al. \(2015\)](#), who did not identify any differences in the between-tone categorical perception of musicians and non-musicians. Instead, musicians only outperformed non-musicians in terms of their within-tone perception ability. There are two possible reasons for these contrasting results. First, our experimental paradigm differs from those used in [Mok and Zuo \(2012\)](#) and [Wu et al. \(2015\)](#). This study used a dual-task singularity experimental paradigm in which the participants were asked to perceive tone categories while watching a silent movie. Such an experimental paradigm increases the difficulty of identifying between-tone and within-tone categorical perception. Second, musicians may have better inhibitory control and be less subjected to distraction than non-musicians ([2015](#); [Slevc et al., 2016](#); [Chen et al., 2017](#)), which may explain why their level of performance was higher than that of non-musicians in our study.

#### 4.2 Differences in tone categorical perception reaction times between the music and non-music groups

Our results showed that the reaction time of the music groups was faster than that of the non-music group in both between and within-tone categorical perception. This may be because music training can improve the pitch sensitivity of the auditory cortex ([Nan et al., 2018](#)) and promote cognitive flexibility. Cognitive flexibility is one of the three core components of executive function, which reflects individuals' response and transformation ability. Previous studies have found that individuals with music training had lower local and overall conversion costs ([Moradzadeh et al., 2015](#)) and faster reaction times ([Hanna-Pladdy and MacKay, 2011](#); [Janus et al., 2016](#)) than those without music training. [Moreno and Farzan \(2015\)](#) believe that executive function represents a potential mechanism that can explain the transfer between musical training and non-musical cognitive ability. Accordingly, there may be a mediating effect on the impact of cognitive flexibility on tone categorical perception reaction times.

#### 4.3 Differences in tone categorical perception accuracy between the music groups

When comparing accuracy between the music groups, we identified that the instrumental and vocal groups showed no differences in the ability to perceive between and within-tone category changes. The initial assumption was that vocalists would outperform instrumentalists in accuracy. A singer's musical instrument is their own throat; thus, they begin practicing music in infancy, earlier than the period in which instrumental music learning begins. Moreover, the singer's pitch discrimination ability is crucial to their performance. However, our hypothesis was not supported by the data; the accuracy rate of tone categorical perception in the instrumental and vocal groups did not differ. This is consistent with the findings of most similar studies ([Nikjeh et al., 2008, 2009](#); [Kirkham et al., 2011](#)). [Nikjeh \(2006\)](#) explored auditory perception and vocal production from individuals' behavior and EEG levels and identified no differences between instrumentalists and vocalists, although there was a unique correlation between pitch perception and voice production in the instrumentalist group. However, some scholars have found that different musicians have different hearing abilities ([Spiegel and Watson, 1984](#); [Kishon-Rabin et al., 2001](#); [Tervaniemi et al., 2006](#)), but these studies examined different instrumental musicians and music types, rather than directly comparing vocalists and instrumentalists.

Although there were no differences between the instrumental vocalist group and the vocal group, the accuracy rate of the instrumental vocalist group was higher than that of the instrumental group. These results differ from those of [Nikjeh et al. \(2008\)](#), whose behavioural study identified no differences between the vocal, instrumental, and instrumental vocalist music groups. There may be two reasons for our results. First, our study used different experimental materials. The auditory stimuli used by [Nikjeh et al. \(2008\)](#) consisted of harmonic polyphony, which approximates the physical characteristics of piano tones, while the experimental material used a continuous unified tone synthesized from original speech. The tonal continuum is a part of speech, and instrumental vocalists are involved in the perception and processing of lyrics during the singing process; therefore, they may have a higher sensitivity to tonal perception than instrumental musicians.

Second, the differences between instrumentalists, vocalists, and instrumental vocalists may be significant at the brain level. Some non-behavioural EEG experiments on the effect of music training on speech perception have identified similar differences. [Yao and Chen \(2020\)](#) conducted a 12-month music training experiment on children aged 4–5 years to explore its impact on the children's tone categorical perception ability. There were no differences between the outcomes of the six-month test and previous music training. Even after 12 months of training, there was no change in the position of the category boundaries. [Nikjeh et al. \(2008\)](#) also compared the pitch perception of instrumentalists, vocalists, and instrumental vocalists, and found no differences between the three groups at the behavioural level. However, at the EEG level, the instrumental vocalists' latency in the P3a component was shorter. [Moreno et al. \(2009\)](#) found that six-month music training can improve the pitch perception level of children aged 4–5 years. [Marie et al. \(2011\)](#) used the EEG technique to compare the differences between French musicians and



non-musicians in terms of tonal and segmental (consonant, vowel) recognition in Mandarin. Their experimental results showed that musicians had higher tone and segment recognition than non-musicians. This was mainly reflected in their N2 and N3 components, which occurred 100 ms earlier than in non-musicians, and in their P3b component, which had a higher amplitude compared to that of non-musicians. In addition, there were no differences between the instrumental vocalist group and vocal group, which may have been because most vocal songs include lyrics, so singing itself involves speech processing, allowing vocalists with vocal music experience to show a higher tone perception accuracy rate than musicians with only instrumental music training. Thus, there were no differences between the instrumental vocalist group and vocal group.

#### 4.4 Differences in tone categorical perception reaction times between the music groups

The experimental results showed that there were no significant differences in the reaction times of the instrumental and vocal groups; however, the reaction time of the instrumental vocalist group was faster than that of the vocal group. The main reason for this is that the instrumental vocalist group mainly recruits vocalists with piano training experience. During piano training, individuals must read the score while playing the keys quickly, which greatly enhances their cognitive conversion ability and finger flexibility. According to the multi-sensory channel features, long-term music training makes the connection between sound-related movement representation and the sound's auditory representation more accurate (Lee and Noppeney, 2011). Vocalists have developed action channel representation through piano training, which supports perceptual processing, thereby enhancing their efficiency.

However, under stimulus conditions, the reaction times for within-tone categorical perception were faster than those for between-tone categorical perception, which is an interesting result. In the oddball task, we found that although individuals had longer reaction times in the between-tone tasks, their accuracy on these tasks was higher than on the within-tone tasks. This indicates that individuals maintained the accuracy of their task responses via a longer delay in their reaction times (Chen et al., 2020).

#### 4.5 Limitations

This study had several limitations. First, it used a dual-task experimental paradigm that comprised the classical oddball paradigm and watching a silent movie; this resulted in some differences between the experimental results and those of the single oddball task. The dual-task paradigm increases the difficulty of the task and places higher demands on the participants (especially in the music group). For example, if the participants in the music group have fewer years of music training and less intense music training, there may be no trial data results for the music and non-music groups. In addition, we acknowledge that the use of the dual-task paradigm may have introduced interference during the experiment, which could have impacted the experimental results. However, according to such a paradigm, the main effect of the experimental results is indeed

significant (i.e., the music group had a lower error rate and responded faster than the non-music group). In future studies, we intend to conduct separate experiments with single-task (oddball paradigm) and dual-task modes to compare the experimental results and verify the accuracy and scientific validity of the dual-task experimental mode.

Second, there was a lack data to conduct a baseline comparison of the tone category perception of individuals in dialect areas. Future studies should measure the tone categorical perception of individuals from the Beijing dialect area (Beijing pronunciation is considered the standard pronunciation of Mandarin) to obtain baseline data, and then compare it with that of those in local dialect areas to better explore the impact of music training on individuals in these areas.

Third, we only selected individuals from the Chongqing dialect area (located in southwestern China) for our experiments. Previous studies have found that there are differences in the perception of tones in different dialect areas (Yu and Huang, 2019); therefore, in the future we will select different dialect areas for further comparisons and research.

Fourth, we only used behavioural research methods in this study, and were not able to provide temporal and spatial evidence of brain activity. Future studies should use EEG or fMRI to analyse both the temporal and spatial neural responses of the tone categorical perception of individuals in dialect areas.

Finally, there are many music training methods. In addition to vocal and instrumental music, music training also includes different instruments (e.g., keyboards, string, or percussion instruments), music genres (e.g., jazz, folk, and rock), music elements (e.g., interval, chord, and rhythm), and so on. In the future, we will continue to explore the effects of music training on the Mandarin tone categorical perception of individuals in dialect areas using different training methods.

## 5 Conclusion

This study supports the OPERA hypothesis that music training can enhance tone categorical perception, even in a dialect environment. Furthermore, music training combining instrumental and vocal music will have a better effect on Mandarin tone categorical perception, even in a dialect environment. To some extent, these findings provide a theoretical basis for the improvement of the tone categorical perception of individuals in dialect areas, as well as theoretical support for music and language education.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving humans were approved by Ethics Committee of Southwest University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

JH: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. YZ: Data curation, Methodology. HL: Writing – review & editing. JL: Conceptualization, Investigation. MZ: Conceptualization, Funding acquisition, Investigation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Elite program of Chongqing (2022YC052).

## References

- Barbaroux, M., Norena, A., Rasamimanana, M., Castet, E., and Besson, M. (2021). From psychoacoustics to brain waves: a longitudinal approach to novel word learning. *J. Cogn. Neurosci.* 33, 8–27. doi: 10.1162/jocn\_a\_01629
- Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094
- Besson, M., Dittinger, E., and Barbaroux, M. (2018). How music training influences language processing: evidence against information encapsulation. *L'Année Psychol.* 118, 273–288. doi: 10.3917/anpsy1.183.0273
- Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., et al. (2023). The human language system, including its inferior frontal component in "Broca's area," does not support music perception. *Cereb. Cortex* 33, 7904–7929. doi: 10.1093/cercor/bhad087
- Chen, J., Chen, J. J., Chen, J., Li, H., Li, X. Y., and Wu, K. (2020). The effect of music training on executive functions in adults. *J. Psychol. Sci.* 43, 629–636. doi: 10.16719/j.cnki.1671-6981.20200317
- Chen, J., Liu, L., Wang, R., and Haizhou, S. (2017). The effect of musical training on executive functions. *Adv. Psychol. Sci.* 25, 1854–1864. doi: 10.3724/SP.J.1042.2017.01854
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G\*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Finney, D. J. (1972). In probit analysis. *J. R. Stat. Soc. Ser. D* 21:222. doi: 10.2307/2986688
- Han, Y. H. (2017). "Putonghua proficiency test development report" in *Language life paper - report on language life in China* (Beijing: Commercial Press), 60–64.
- Hanna-Pladdy, B., and MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology* 25, 378–386. doi: 10.1037/a0021895
- Jäncke, L. (2012). The relationship between music and language. *Front. Psychol.* 3:123. doi: 10.3389/fpsyg.2012.00123
- Janus, M., Lee, Y., Moreno, S., and Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *J. Exp. Child Psychol.* 144, 84–97. doi: 10.1016/j.jecp.2015.11.009
- Kirkham, J., Lu, S., Wayland, R., and Kaan, E. (2011). "Comparison of vocalists and instrumentalists on lexical tone perception and production tasks" in *Proceedings of the 17th international congress of phonetic sciences* (Hong Kong: City University of Hong Kong), 1098–1101.
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12, 125–144. doi: 10.1515/JBCPP.2001.12.2.125
- Koelsch, S. (2006). Significance of broca's area and ventral premotor cortex for music-syntactic processing. *Cortex* 42, 518–520. doi: 10.1016/S0010-9452(08)70390-3
- Koelsch, S., Gunter, T. C., Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *NeuroImage* 17, 956–966. doi: 10.1006/nimg.2002.1154
- Lai, H., Xu, M., Song, Y., and Liu, J. (2014). Distinct and shared neural basis underlying music and language: a perspective from meta-analysis. *Acta Psychol. Sin.* 46, 285–297. doi: 10.3724/SP.J.1041.2014.00285
- Lee, H., and Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci.* 108, E1441–E1450. doi: 10.1073/pnas.1115267108

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, D. C. S. (2006). Chinese as a lingua franca in greater China. *Ann. N. Y. Acad. Sci.* 26, 149–176. doi: 10.1017/S0267190506000080
- Liang, Y. (2017). The production-perception mechanism in tonal shift: the case of Hong Kong Cantonese. *Chinese Lang.* 6, 723–732. doi: 10.3969/j.issn.1003-0751.2002.04.032
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in mandarin Chinese. *J. Cogn. Neurosci.* 23, 2701–2715. doi: 10.1162/jocn.2010.21585
- Marques, C., Moreno, S., Luis Castro, S., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453
- Mok, P. K. P., and Zuo, D. (2012). The separation between music and speech: evidence from the perception of Cantonese tones. *J. Acoust. Soc. Am.* 132, 2711–2720. doi: 10.1121/1.4747010
- Moradzadeh, L., Blumenthal, G., and Wiseheart, M. (2015). Musical training, bilingualism, and executive function: a closer look at task switching and dual-task performance. *Cogn. Sci.* 39, 992–1020. doi: 10.1111/cogs.12183
- Moreno, S., and Farzan, F. (2015). Music training and inhibitory control: a multidimensional model. *Ann. N. Y. Acad. Sci.* 1337, 147–152. doi: 10.1111/nyas.12674
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120
- Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., and Münte, T. F. (2003). The fate of sounds in conductors' brains: an ERP study. *Cogn. Brain Res.* 17, 83–93. doi: 10.1016/S0926-6410(03)00083-1
- Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., et al. (2018). Piano training enhances the neural processing of pitch and improves speech perception in mandarin-speaking children. *Proc. Natl. Acad. Sci.* 115, E6630–E6639. doi: 10.1073/pnas.1808412115
- Nikjeh, D. A. (2006). *Vocal and instrumental musicians: Electrophysiologic and psychoacoustic analysis of pitch discrimination and production*. Florida: University of South Florida.
- Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2008). Hearing of note: an electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology* 45, 994–1007. doi: 10.1111/j.1469-8986.2008.00689.x
- Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: comparison of vocal and instrumental musicians. *J. Acoust. Soc. Am.* 125, 328–338. doi: 10.1121/1.3021309
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? *The OPERA hypothesis*. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142

- Patel, A. D. (2012a). *Music, language, and the brain*. Oxford: Oxford University Press.
- Patel, A. D. (2012b). The OPERA hypothesis: assumptions and clarifications. *Ann. N. Y. Acad. Sci.* 1252, 124–128. doi: 10.1111/j.1749-6632.2011.06426.x
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi: 10.1016/j.heares.2013.08.011
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121
- Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. *Ann. N. Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x
- Seppänen, M., Brattico, E., and Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiol. Learn. Mem.* 87, 236–247. doi: 10.1016/j.nlm.2006.08.011
- Slevc, L. R., Davey, N. S., Buschkuhl, M., and Jaeggi, S. M. (2016). Tuning the mind: exploring the connections between musical ability and executive functions. *Cognition* 152, 199–211. doi: 10.1016/j.cognition.2016.03.017
- Spiegel, M. F., and Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *J. Acoust. Soc. Am.* 76, 1690–1695. doi: 10.1121/1.391605
- Studdert-Kennedy, M., and Shankweiler, D. (1970). Hemispheric specialization for speech perception. *J. Acoust. Soc. Am.* 48, 579–594. doi: 10.1121/1.1912174
- Tang, W., Xiong, W., Zhang, Y.-X., Dong, Q., and Nan, Y. (2016). Musical experience facilitates lexical tone processing among mandarin speakers: behavioral and neural evidence. *Neuropsychologia* 91, 247–253. doi: 10.1016/j.neuropsychologia.2016.08.003
- Tervaniemi, M., Castaneda, A., Knoll, M., and Uther, M. (2006). Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *Neuroreport* 17, 1225–1228. doi: 10.1097/01.wnr.0000230510.55596.8b
- Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusician listeners. *Ann. N. Y. Acad. Sci.* 1060, 100–110. doi: 10.1196/annals.1360.007
- Tillmann, B., Janata, P., and Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Ann. N. Y. Acad. Sci.* 999, 209–211. doi: 10.1196/annals.1284.031
- Wang, Q. S., and Cao, Y. L. (2002). Brief discussion on "the status of tone in putonghua proficiency test." *Zhongzhou J.* 4, 114–118.
- Wang, Y. J., Liu, S. W., and Qing, W. (2016). Tonal patterns and categorical perception of Yinping and Yangping in Chongqing mandarin: implications to historical Chongqing of Yinping and Shangsheng. *Chinese J. Phonetics* 7, 18–27.
- Wang, M., Ning, R., and Zhang, X. (2015). Musical experience alleviates the aging in the speech perception in noise. *Adv. Psychol. Sci.* 23, 22–29. doi: 10.3724/SP.J.1042.2015.00022
- Wong, P., Schwartz, R. G., and Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, mandarin-speaking children. *J. Speech Lang. Hear. Res.* 48, 1065–1079. doi: 10.1044/1092-4388(2005/074)
- Wu, C. L. (2010). Discussion on the promotion of Putonghua since the founding of new China. *Educ. Rev.* 4, 147–149.
- Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y., and Shu, H. (2015). Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Front. Psychol.* 6:436. doi: 10.3389/fpsyg.2015.00436
- Xia, X. Z. (2007). On the influence of dialects on people in dialect areas learning putonghua and the ways to learn putonghua well. *Soc. Sci.* 2, 207–209. doi: 10.3969/j.issn.1002-3240.2007.02.056
- Yao, Y. (2020). *Effects of music training on phonological category perception of 4-5-year-old preschool children*. Changsha, China: Hunan University.
- Yao, Y., and Chen, X. (2020). The effects of music training on categorical perception of mandarin tones in 4- to 5-year-old children. *Acta Psychol. Sin.* 52, 456–468. doi: 10.3724/sp.j.1041.2020.00456
- Yu, Q., and Huang, Y. L. (2019). The influence of dialect experience on categorical perception of mandarin tones. *Appl. Linguist.* 3, 114–123. doi: 10.16499/j.cnki.1003-5397.2019.03.014
- Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science* 342, 585–589. doi: 10.1126/science.1238414