Music Experience Enhances Categorical Perception of English Vowels in Mandarin Speakers

Jiayu LIANG, Bowen JIA, Jianing LIU, Xinyi LI, and Hao ZHANG
Shandong University

Categorical perception (CP) refers to the inclination to perceive speech sounds as discrete categories rather than continuous variations. While the CP of consonants is well-established, the perception of vowels remains a topic of ongoing debate. Recent studies have focused on identifying factors that may influence CP of vowels, and most of them have yielded relatively consistent results. However, research investigating the impact of music experience on CP of vowels has produced conflicting findings. Furthermore, these studies have overlooked the inclusion of non-native vowels as experimental stimuli and the involvement of Mandarin speakers as participants. Consequently, the primary objective of this study is to explore the influence of music experience on the categorical perception of English vowels among Mandarin speakers.

To achieve this objective, our study recruited a total of 12 Mandarin-speaking participants who were divided into two distinct groups: the music group and the non-music group. The individuals in the music group had undergone a minimum of six years of formal musical training (M=12.17, SD=4.74), whereas the participants in the non-music group had not received any musical training. In order to control for potential effects of English proficiency, we evaluated the CET4 scores of the participants (mandatory test for non-English major university students in China), which were all below 500 (M=425.16, SD=66.05). The experimental materials utilized in this study consisted of resynthesized waveform (wav) files of the English words "bet," "bat," and "bait." TANDEM-STRAIGHT, a method renowned for generating stimuli that closely resemble natural speech, was employed for this purpose. Notably, only the formants of the speech sounds were altered. Throughout the experimental procedure, participants were instructed to complete an online music perception test, an identification test, and a discrimination test. We assessed various measures of the identification function for each participant, including boundary position, width, and slope. Furthermore, discrimination function scores, encompassing between-category and within-category score, and peakedness score, were also evaluated to provide a comprehensive analysis.

The obtained data were subjected to analysis of variance (ANOVA) and Pearson correlation analysis. The results revealed statistically significant differences in boundary position [$F(2, 48) = 6.08$, $p = 0.02$], width [$F(2, 48) = 5.83$, $p = 0.02$], and slope [$F(2, 48) = 6.03$, $p = 0.02$], between the music group and the non-music group. Additionally, a robust negative correlation was observed between music scores (derived from the music perception test) and boundary width for bet-bat continuum ($r = -0.78$, $p < 0.001$). Specifically, the music group exhibited more anterior positions, narrower width, sharper boundaries, higher discrimination accuracy, a stronger alignment between discrimination peaks and categorical boundaries, and more consistent peakedness scores in comparison to the non-music group, regarding both stimuli. These findings suggest that musical training or exposure has the potential to enhance perceptual abilities associated with the categorization of English vowels.

In summary, the present study contributes to the expansion of the OPERA-e theory by presenting empirical evidence derived from the CP of vowels. These findings underscore the existence of shared sensory and cognitive processing mechanisms between speech and music perception, highlighting the potential of musical training to enhance speech processing abilities.