Original Article

QJEP

Music-reading expertise associates with face but not Chinese character processing ability

Quarterly Journal of Experimental Psychology
2022, Vol. 75(5) 854–868

DOI: 10.1177/17470218211053144
qjep.sagepub.com
SAGE

Jie Chen[1], Panpan Yuan[1], Hong Li[1,2], Changming Chen[3], Yi Jiang[4] and Kang Lee[5]

Abstract

A growing number of behavioural and neuroimaging studies have investigated the cognitive mechanisms and neural substrates underlying various forms of visual expertise, such as face and word processing. However, it remains poorly understood whether and to what extent the acquisition of one form of expertise would be associated with that of another. The current study examined the relationship between music-reading expertise and face and Chinese character processing abilities. In a series of experiments, music experts and novices performed discrimination and recognition tasks of musical notations, faces, and words. Results consistently showed that musical experts responded more accurately to musical notations and faces, but not to words, than did musical novices. More intriguingly, the music expert's age of training onset could well predict their face but not word processing performance: the earlier musical experts began musical notation reading, the better their face processing performance. Taken together, our findings provide preliminary and converging evidence that music-reading expertise links with face, but not word, processing, and lend support to the notion that the development of different types of visual expertise may not be independent, but rather interact with each other during their acquisition.

Keywords

Visual expertise; face processing; Chinese character processing; musical notation processing; music-reading

Received: 24 August 2020; revised: 28 July 2021; accepted: 30 July 2021

Introduction

People acquire various forms of visual object expertise during the course of development. Certain forms of expertise, such as face, are the natural category, and the ability to process faces is typically acquired implicitly through everyday experience (Anzures et al., 2013). In contrast, some other forms of expertise, such as word and musical notation, are cultural artefacts, and the ability to read words and musical notations is often taught explicitly and acquired through formal instruction; individuals tend to acquire the expertise to process them in a highly deliberate manner that requires hours and hours of practice (Dehaene et al., 2010; Sergent et al., 1992). For over a century, extensive behavioural and neuroscience research has been devoted specifically to understanding how one acquires visual expertise in a particular category and its underlying cognitive mechanisms and neural substrates (Dehaene & Cohen, 2007; Gauthier et al., 2000; Kanwisher et al., 1997; Rossion et al., 2002; Tanaka & Curran, 2001; A. C.-N. Wong et al., 2005, 2009).

However, most of the existing research on visual expertise had been done in isolation. When several forms of

[1]School of Educational Science, Cognition and Human Behavior Key Laboratory of Hunan Province, Hunan Normal University, Changsha, China
[2]School of Psychology, South China Normal University, Guangzhou, China
[3]Department of Psychology, Xinyang Normal University, Xinyang, China
[4]State Key Laboratory of Brain and Cognitive Science, Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[5]Dr. Eric Jackman Institute of Child Study, University of Toronto, Toronto, Ontario, Canada

Corresponding authors:
Hong Li, School of Psychology, South China Normal University, Guangzhou 510631, China; School of Educational Science, Cognition and Human Behavior Key Laboratory of Hunan Province, Hunan Normal University, Changsha 410081, China.
Email: lihong_psych@m.scnu.edu.cn

Yi Jiang, State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China.
Email: yijiang@psych.ac.cn

Kang Lee, Dr. Eric Jackman of Child Study, University of Toronto, 45 Walmer Road, Toronto, Ontario M5R 2X2, Canada.
Email: kang.lee@utoronto.ca

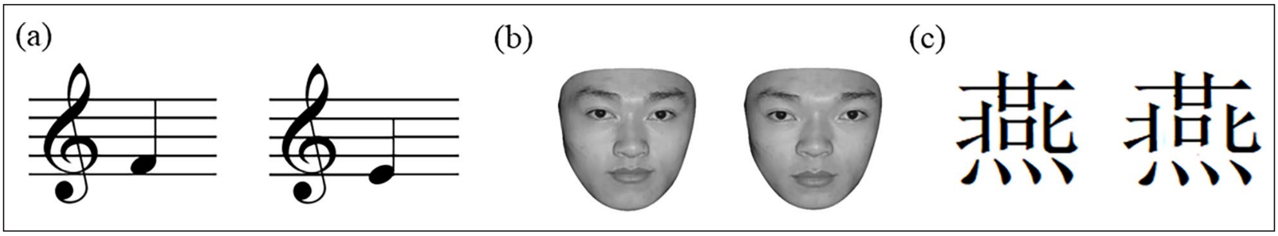

Figure 1. (a) These two musical notes are identical, but their spatial locations on the five-line staff change, so their identities (pitch) change. (b) Although the two face images have same internal features, they are still recognised as different people due to the distances between the eyes and between the nose and mouth are changed. (c) The character in left means "swallow." However, the character in right still means "swallow" despite the displacement of its features from their original location.

visual expertise were considered together (e.g. face, word, car, bird, or greeble expertise), the focus had been mainly on how similar or dissimilar different forms of expertise are in behaviour or in the brain to elucidate whether similar neurocognitive mechanisms are underlying different forms of visual expertise (Boggan et al., 2012; Gauthier et al., 2003; McGugin et al., 2012). These studies have advanced greatly our understanding of the similarities and differences in terms of the cognitive and neural mechanisms underlying different forms of visual expertise.

In contrast to the vast research literature about various forms of individual visual object expertise, there is relatively limited research on the interaction between these forms of visual expertise and whether the acquisition of one type of expertise associates with another (Behrmann & Plaut, 2012; Dundas et al., 2013, 2014, 2015; S. Li et al., 2013; S. T. K. Li & Hsiao, 2018). Recently, researchers have begun to investigate the relationship between word and face processing and to determine whether the development of word and face processing is independent or interact with each other during their acquisition. For example, Dundas et al. (2013) found that the magnitude of face lateralisation (left visual field [LVF] vs right visual field [RVF] accuracy) has a positive correlation with reading skill and thus suggested that word-reading experience drives the right lateralisation for face processing. Dehaene et al. (2010) found that with increased literacy, neural responses to faces decrease slightly in the visual word form area (VWFA) but increases strongly in the right fusiform face area (Dehaene et al., 2010). S. Li et al. (2013) found that among preschoolers (5–6 years of age), earlier and better readers tended to have a delayed right lateralisation of the face N170. Ventura et al. (2013) further found that learning to read reduced the automatic engagement of holistic processing for faces. These recent findings, albeit limited, taken together revealed a close link between learning to read and the changes of face processing in both behavioural and neural levels.

However, it is as yet unknown whether the acquisition of other forms of visual expertise associates with the acquisition of face processing or that of word processing or both. To bridge this significant gap, we examine whether learning to read musical notations would be associated with changes in the face and Chinese character processing.

We focused on music-reading experience because there are both similarities and differences among musical notation, word, and face. However, face is the natural category, whereas musical notation and Chinese character are both culture artefacts, and music is also governed by culture-dependent combinatorial rules as speech. Different from face processing expertise that is typically acquired implicitly through everyday experience since birth (Anzures et al., 2013), the acquisition of the abilities to read words and musical notations requires explicit and intensive training and practice usually starting from childhood (Dehaene et al., 2010; Sergent et al., 1992).

However, it is well known that musical notation reading requires the processing of spatial information, as the spatial location of musical notes is directly related to pitch, and that their relative height separation is related to pitch intervals. In other words, the spatial information of the musical notation determines its identity (see Figure 1a). Thus, skillful and accurate analysis of spatial relations is highly crucial during musical notation reading, and it is thus not surprising that visuospatial abilities are enhanced in musicians compared with non-musicians (Brochard et al., 2004). Similarly, when the spatial relationships between the major facial features (i.e., eyes, nose, and mouth) are changed but the facial features themselves remain the same, the face takes on a new identity (Freire et al., 2000; Ge et al., 2006) (see Figure 1b). By contrast, although Chinese words contain configural information (A. C.-N. Wong et al., 2011, 2012, 2019), unlike music notes and faces, changing the configural relation of the features of Chinese words does not alter the identity of the words (see Figure 1c). In Chinese orthography, a word retains its meaning even when the spacing between its components is altered, as long as its components and its first-order relationship (i.e., relative positions among character parts) remain unchanged. Tso et al. (2020) suggested that featural information (e.g., individual strokes) played a more critical role in Chinese character recognition than did configural information (e.g., exact distances between strokes). Thus, the spatial information is crucial for processing individual musical notations and faces but not for Chinese characters (Ge et al., 2006).

Therefore, music-reading expertise provides a good opportunity to examine the interactions among the acquisition of various forms of visual expertise. In a series of experiments, we recruited Chinese participants who either had formal music-reading training (musical experts) or had no music-reading experience (musical novices). We aimed to examine whether music-reading expertise would be associated with performances for face and word identification. We used a same/different identity-matching task in Experiments 1, 2, 3, and 4 to measure participants' abilities (response speed and accuracy) at processing musical notations, faces, and words. It has been widely demonstrated that music literacy acquisition has positive transfer effects on musical and even nonmusical areas of cognition (Chan et al., 1998; Cheek & Smith, 1999; Kraus & Chandrasekaran, 2010; Schellenberg, 2001, 2004; Schellenberg & Peretz, 2008). In addition, recent findings regarding the transfer effect in the motor domain suggested that the visual perception anticipation skill acquired in one domain can be transferred to another closely related or similar domain (Moore & Müller, 2014; Müller et al., 2015; Rosalie & Müller, 2012). Given both similarities and differences among musical notation, word and face as mentioned above, we expect that music-reading expertise would be associated with enhanced face or word processing abilities or both.

Experiment 1

Methods

Participants. A total of 50 music undergraduate or postgraduate students (38 females, $M=21.4$ years, $SD=2.77$) were recruited from the College of Music of the Southwest University in China. All of the musicians had received formal musical training and had many years of music-reading experience ($M=9.2$ years, $SD=5.38$). The mean age when they began music-reading was $M=10.96$ years ($SD=4.62$, range: 3–20 years of age). All of them must pass a professional music examination before entering the college, including the examination of music theory, vocal music, musical instrument, and solfeggio.

In addition, 50 non-music students (39 females, $M=20.58$ years, $SD=1.75$) were recruited from other colleges of the Southwest University. They reported that they could neither play nor read music and had not received any musical training. All participants were healthy, right-handed, with normal or corrected to normal vision, and reported no history of any neurological or psychological disorders. The study was approved by the local human research ethics committee, and it adhered to the tenets of the Declaration of Helsinki. All participants signed an informed consent form before participating.

Stimuli and procedure. In total, 80 Chinese faces and 80 Chinese characters were used for the perception matching task. Faces were young Chinese adult faces (40 women and 40 men). Faces or words were divided into two groups; half (40 faces or words) were used as targets and the other half as distractors. Musical notations with eight different time values (the half note, quarter note, eighth note, sixteenth note, dotted half note, dotted quarter note, dotted eighth note, and dotted sixteenth note) were set on the different position of the five-line staff. The paired distractor notations shared the same time values with the targets, with only the position shifted by one step on the five-line staff (half trials with up shifts and the other with down shifts). Some notes repeated once. Thus, each of the three experimental conditions contained 40 trials. The size of each face, word, and musical notation was about 2.6 cm wide and 3.6 cm high and subtended a visual angle of $1.86° \times 2.58°$ at a viewing distance of 80 cm.

Participants took part in the experiment individually. Each trial was initiated by a 200 ms presentation of a small white cross on the grey computer screen. Afterwards, a target stimulus (the face, word, or musical note) was presented for a duration of around 300 ms, followed by a 500 ms blank screen. Then, two probe stimuli, the target stimulus and a distractor stimulus, were presented side by side for 1000 ms (see Figure 2). Participants' task was to detect whether the target in the probe stimuli was on the left or the right side. Participants were instructed to press the F key with their left forefingers as soon and accurately as the target was on the left side and press the J key with their right forefingers as soon and accurately as the target was on the right side. The left–right position of the target was randomised across trials. Each stimulus was followed by a 1,000 ms blank screen. All trials were presented randomly.

Results and discussion

Response times (RTs) analyses included only trials with correct responses. Analyses of variance (ANOVAs) of mean RTs with Stimulus type (musical notations, faces, and words) as a within-subjects variable and Participant group (musical experts and musical novices) as a between-subjects variable showed a significant main effect of Stimulus type, $F(2, 196)=65.62$, $p<.001$, $\eta_p^2=.4$, and a Stimulus type × Participant group interaction effect, $F(2, 196)=34.13$, $p<.001$, $\eta_p^2=.26$, $1-\beta>.99$; post hoc power tested by G-power, Faul et al., 2009. RTs in the word condition were shorter than those in the face condition ($p<.001$, Bonferroni corrected), in which RTs were also shorter than those in the musical notation condition ($p=.008$, Bonferroni corrected). The simple effect analysis showed that musical experts responded significantly faster to musical notations and faces than musical novices did, Note: $t(98)=-5.84$, $p<.001$, $d=1.17$; Face: $t(98)=-2.28$,

Figure 2. Illustration of the temporal sequence of an individual trail in Experiment 1.

$p = .025$, $d = 0.45$, but there was no significant group difference for word processing, Word: $t(98) = -1.07$, $p = .29$ (see Figure 3).

Furthermore, ANOVAs of mean accuracies also showed a significant main effect of Stimulus type, $F(2, 196) = 61.1$, $p < .001$, $\eta_p^2 = .38$, and a Stimulus type × Participant group interaction effect, $F(2, 196) = 59.93$, $p < .001$, $\eta_p^2 = .38$, $1-\beta > .99$; post hoc power tested by G-power, Faul et al., 2009. Accuracies in the word condition were higher than those in the face condition ($p < .001$, Bonferroni corrected), in which accuracies were also higher than those in the musical note condition ($p < .001$, Bonferroni corrected). The simple effect analysis showed that the accuracy of musical notation processing was higher for musical experts than for novices, Note: $t(98) = 8.69$, $p < .001$, $d = 1.72$, and there were no group differences for faces and words, Face: $t(98) = 0.31$, $p = .76$; Word: $t(98) = 0.17$, $p = .87$ (Figure 3).

To control for the speed–accuracy trade-off, we also computed an inverse efficiency score (RT divided by accuracy). The same pattern of findings held when we used the index of RT/accuracy, with smaller RT/accuracy in musical notation and face processing for music-reading experts than for novices (all ps $< .035$), and no difference in word processing between the two groups ($p = .27$).

Results shown above suggested that music-reading experts showed superior abilities for face and musical notation, but not word perception as compared to music-reading novices. To examine specifically the relationship between music training experience and the processing of musical notations and faces, we performed a partial correlation analysis and correlated RT/accuracy of face and musical notation processing with the age at which music-reading began and the number of years of music-reading for musicians after controlling for RT/accuracy in word processing. Results showed that RT/accuracy in musical notation and face processing were all significantly related to age at which music-reading began ($r_{face} = .34$, $p = .018$; $r_{note} = .45$, $p = .001$; Figure 4a). The RT/accuracy of musical notation processing was significantly related to the number of years of music-reading ($r_{note} = -.47$, $p = .001$), but the RT/accuracy of face processing was not ($r_{face} = -.13$, $p = .39$; Figure 4b).

In sum, we found that musical experts unsurprisingly responded faster and more accurately to musical notation than did musical novices. More importantly, musical experts also responded faster to faces than did musical novices. The correlational analyses showed that the age of music-reading training onset and the length of music training predicted different outcomes: (a) the longer music-reading experience, the superior their musical notation processing ability, which was not true for face processing ability; (b) the earlier participants began musical notation reading training, the superior their face and musical notation processing ability. Thus, the present results suggested that learning to read musical notation was associated with better face processing but not word processing, likely due to the fact that the identity recognition of musical notation and faces both depends on the analysis of spatial information. Furthermore, the age of onset, but not length, of musical notation training, predicted face processing performance, suggesting an important role of early music learning in the acquisition of face processing expertise.

However, the present experiment used a delayed two-alternative, forced-choice matching-to-sample task, in which the target stimulus and the probe stimulus were presented sequentially. Participants should keep a mental image of the target stimulus until they accomplished the perception matching task. Thus, the observed differences

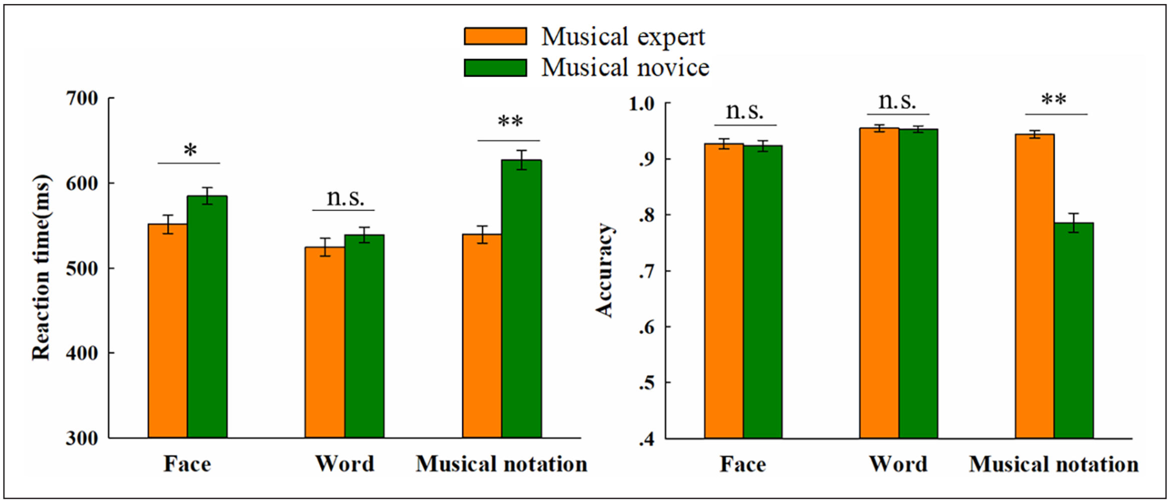

Figure 3. The results of reaction time and accuracy for face, word, and musical notation processing in musical experts and novices. Error bars plot s.e. $^*p < .05$, $^{**}p < .01$, and n.s. means no significant difference.

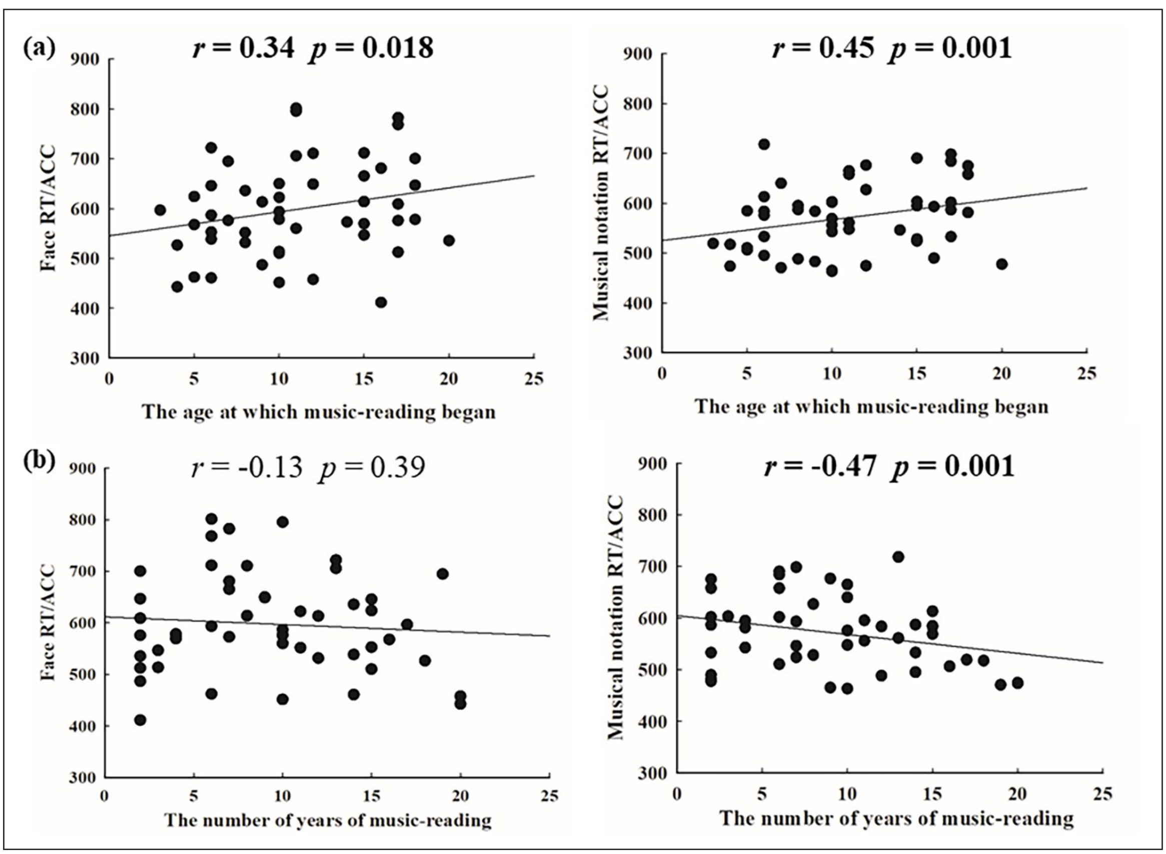

Figure 4. (a) Correlations of the age at which music-reading began with RT/ACC for face and musical notation processing after controlling for RT/ACC in word processing. (b) Correlations of the number of years of music-reading with RT/ACC for face and musical notation processing after controlling for RT/ACC in word processing.

in the perception matching task might be due to the different mental visual imagery or visual working memory ability between musical experts and novices. To explore the role of the lower-level perception processing, which was not contaminated by visual working memory, we conducted Experiment 2, in which the target stimulus and the probe stimulus were presented simultaneously. We tested whether the positive association between the acquisition of

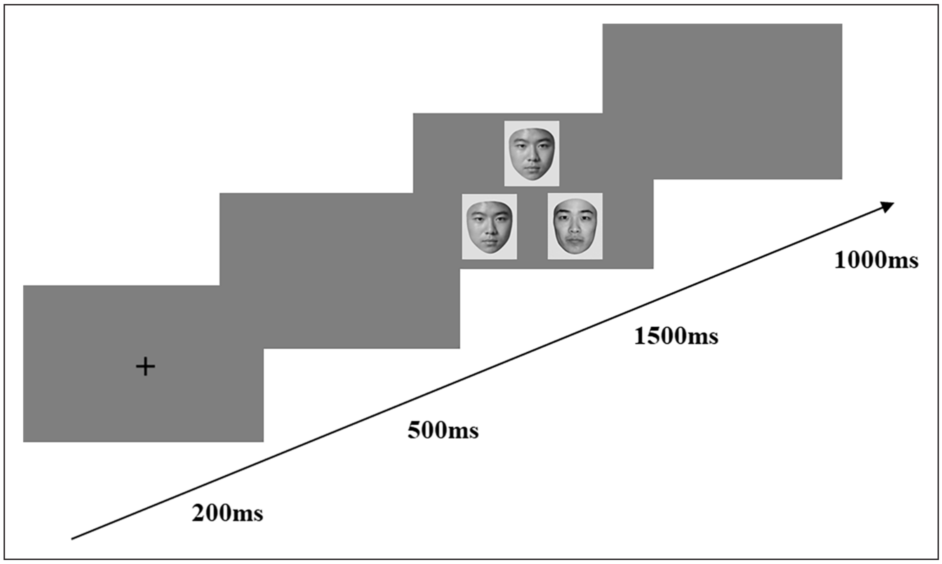

Figure 5. Illustration of the temporal sequence of an individual trail in Experiment 2.

music-reading and face processing still existed after controlling for the influence of visual working memory.

Experiment 2

Method

Participants. A total of 45 new music undergraduate students (29 females, $M=20.44$ years, $SD=1.08$) were recruited from the Department of Music of the Huainan Normal College in China. All of the musicians had received formal musical training and had long years of music-reading experience ($M=5.64$ years, $SD=2.6$). The mean age of beginning their regular music-reading was 13.69 years ($SD=3.84$, range: 5–20 years of age). All of them must pass through the professional examination of music before entering the college, including the examination of music theory, vocal music, musical instrument, and solfeggio.

In addition, 45 new non-music students (27 females, $M=20.02$ years, $SD=1.14$) were recruited from other departments of this College. They reported that they could not read music or play musical instruments, and none of them had received any other musical training except for those taught in public school. All subjects were healthy, right-handed, with normal or corrected to normal vision, and reported no history of any affective disorder. This investigation was approved by the local human research ethics committee, and it adhered to the tenets of the Declaration of Helsinki. All participants signed an informed consent form before participating.

Stimuli and procedure. The experimental stimuli in Experiment 2 were identical to those in Experiment 1 except for the presentation of the stimuli. Each trial was initiated by a 200 ms presentation of a small white cross on the grey computer screen. Afterwards, the target stimulus (face, word, or musical notation) and the probe stimuli (target and distractor stimuli) were presented simultaneously for 1,500 ms, with the probe stimuli presented below the target stimulus (Figure 5). The task was to determine whether the target in the probe stimuli was on the left or right sides, and subjects pressed the F key with their left forefingers if the target was on the left side and pressed the J key with their right forefingers if the target was on the right side. Each stimulus was followed by a 1,000 ms blank screen. Each of the three experimental conditions contained 40 trials, and all trials were presented randomly.

Results and discussion

RT analyses included only trials with correct responses. ANOVAs of mean RTs with Stimulus type as a within-subjects variable and Participant group (musical experts and musical novices) as a between-subjects variable showed a significant main effect of Stimulus type, $F(2, 176)=246.5$, $p<.001$, $\eta_p^2=.74$, and a Stimulus type × Participant group interaction effect, $F(2, 176)=34.03$, $p<.001$, $\eta_p^2=.28$, $1-\beta>.99$; post hoc power tested by G-power, Faul et al., 2009. RTs in the word condition were shorter than those in the face condition ($p<.001$, Bonferroni corrected), in which RTs were also shorter than those in the musical note condition ($p<.001$, Bonferroni corrected). The simple effect analysis showed that musical experts responded significantly faster to musical notations and faces than musical novices, Note: $t(88)=-7.17$, $p<.001$, $d=1.51$; Face: $t(88)=-2.4$, $p=.02$, $d=0.51$, but not to words, Word: $t(88)=-1.12$, $p=.27$. Furthermore, ANOVAs of mean accuracies also showed a significant main effect of Stimulus type, $F(2, 176)=99.66$, $p<.001$, $\eta_p^2=.53$, and a Stimulus type × Participant group interaction effect,

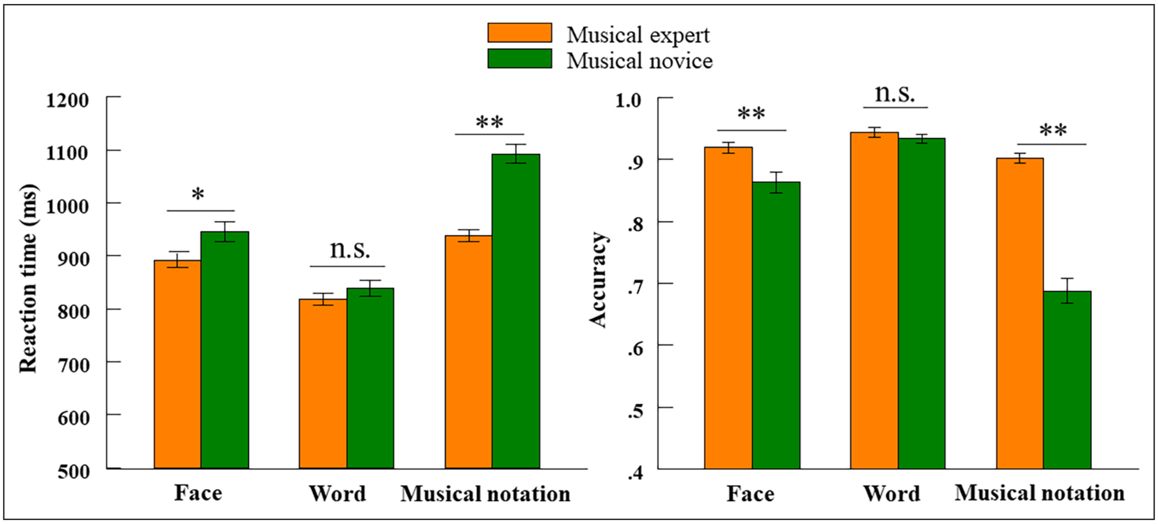

Figure 6. The results of reaction time and accuracy for face, word, and musical notation processing in musical experts and novices. Error bars plot s.e. $*p<.05$, $**p<.01$ and n.s. means no significant difference.

$F(2, 176)=53.21$, $p<.001$, $\eta_p^2=.38$, $1-\beta>.99$; post hoc power tested by G-power, Faul et al., 2009. Accuracies in the word condition were higher than those in the face condition ($p<.001$, Bonferroni corrected), in which accuracies were also higher than those in the musical note condition ($p<.001$, Bonferroni corrected). The simple effect analysis showed that response accuracies of musical notation and face processing were significantly higher for musical experts than for musical novices, Note: $t(88)=9.61$, $p<.001$, $d=2.03$; Face: $t(88)=2.92$, $p=.004$, $d=0.61$, and there was no group difference for word processing, Word: $t(88)=0.93$, $p=.36$ (see Figure 6).

The same pattern of findings held when we used the index of RT/accuracy, with smaller RT/accuracy in musical notation and face processing for music-reading experts than for novices (all $ps<.002$), and no difference in word processing between the two groups ($p=.14$).

Results shown above suggested that music-reading experts showed superior abilities for face and musical notation, but not word perception as compared to music-reading novices. To examine specifically the relationship between music training experience and the processing of musical notations and faces, we correlated RT/accuracy of music notation and face processing with the age at which music-reading began and the number of years of music-reading for music experts after controlling for RT/accuracy of word processing. Results showed that RT/accuracy in musical notation and face processing were all significantly related to age at which music-reading began ($r_{face}=.33$, $p=.029$; $r_{note}=.58$, $p<.001$; see Figure 7a). The RT/accuracy of musical notation processing was significantly related to the number of years of music-reading ($r_{note}=-.55$, $p<.001$), but the RT/accuracy of face processing was not ($r_{face}=-.22$, $p=.15$) (see Figure 7b).

Thus, the correlational analyses of Experiment 2 also replicated the results in Experiment 1 by showing that the earlier the participants began music-reading, the superior their face and musical notation processing ability, and that the longer music-reading experience, the superior their musical note processing ability.

In sum, these results consistently showed that, after controlling for the influence of visual working memory, musical experts still responded faster and more accurately to musical notations and faces than did musical novices, suggesting that music-reading expertise was associated with the perceptual processing of faces and musical notations but not that of words. However, it should be noted that the accuracy was relatively higher and the RT was relatively lower for word processing than for music notation and face processing in both Experiments 1 and 2, which may lead to the lack of group difference in word processing. Thus, to rule out this possibility, we conducted Experiment 3, in which we increased the task difficulty of word processing through increasing the similarity between the target and the probe word stimuli.

Experiment 3

Method

Participants. A total of 38 new music undergraduate students (13 females, $M=19.29$ years, $SD=0.84$) were recruited from the College of Music of the Liaoning Normal University in China. All of the musicians had received formal musical training and had long years of music-reading experience ($M=6.61$ years, $SD=4.43$). The mean age of beginning their regular music-reading was 10.76 years ($SD=4.4$, range: 4–18 years of age). All of them must pass through the professional examination of music before

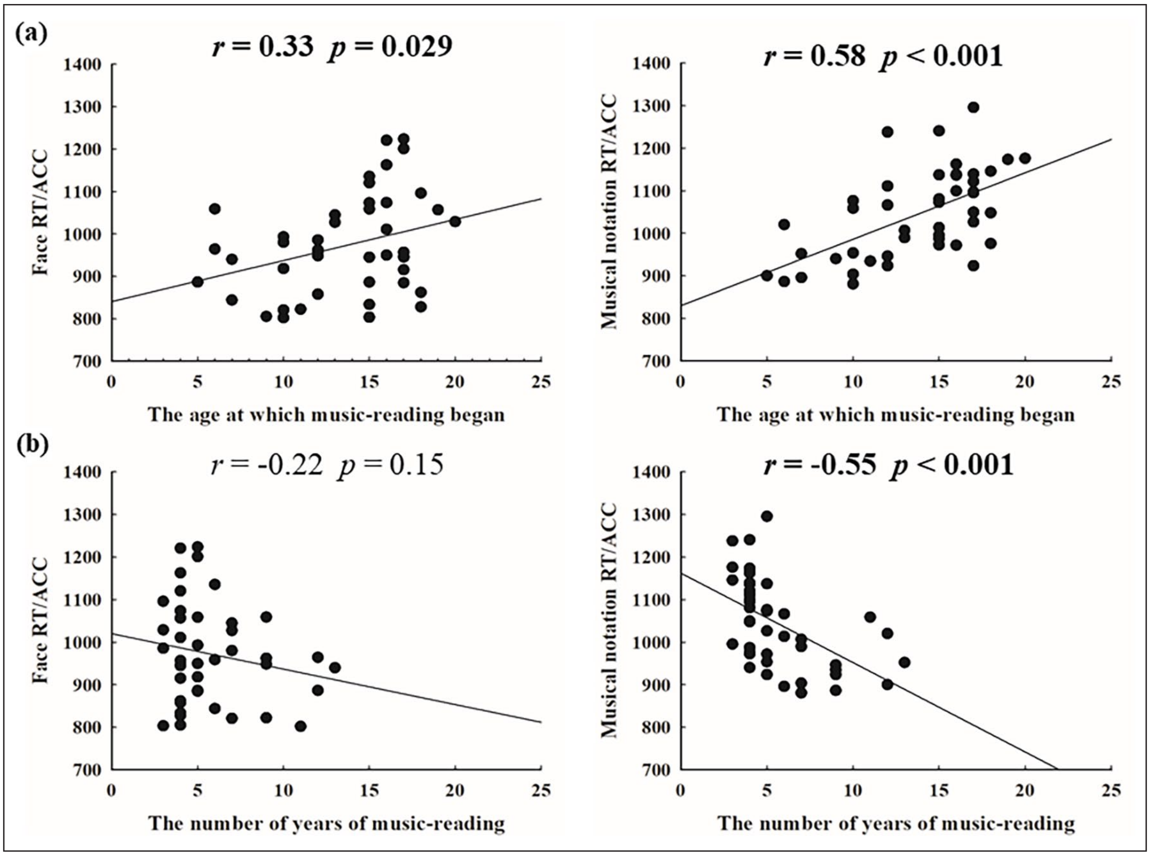

Figure 7. (a) Correlations of the age at which music-reading began with RT/ACC for face and musical notation processing after controlling for RT/ACC in word processing. (b) Correlations of the number of years of music-reading with RT/ACC for face and musical notation processing after controlling for RT/ACC in word processing.

entering the college, including the examination of music theory, vocal music, musical instrument, and solfeggio.

In addition, 40 new non-music students (14 females, M=19.75 years, SD=1.28) were recruited from other colleges of the Liaoning Normal University. They reported that they could not read music or play musical instruments, and none of them had received any other musical training except for those taught in public school. All subjects were healthy, right-handed, with normal or corrected to normal vision, and reported no history of any affective disorder. This investigation was approved by the local human research ethics committee, and it adhered to the tenets of the Declaration of Helsinki. All participants signed an informed consent form before participating.

Stimuli and procedure. The experimental stimuli and procedure in the Experiment 3 were identical to those in the Experiment 2 except for the selection of word stimuli. We changed the similarity between the target and the probe words to increase the task difficulty of word processing.

Results and discussion

ANOVAs of mean RTs and accuracies both showed significant main effect of Stimulus type, RT: $F(2, 152)=49.4$, $p<.001$, $\eta_p^2=.39$; ACC: $F(2, 152)=39.8$, $p<.001$, $\eta_p^2=.34$. The RTs and accuracies for word processing (1,007.41 ms, 88.01%) were in the middle of those for face (960.23 ms, 93.46%) and musical notation processing (1,041.14 ms, 83.56%).

Consistent with the results of Experiment 2, ANOVAs of mean RTs and accuracies both showed a Stimulus type × Participant group interaction effect, RT: $F(2, 152)=27.1$, $p<.001$, $\eta_p^2=.26$, $1-\beta>.99$; ACC: $F(2, 152)=26.62$, $p<.001$, $\eta_p^2=.26$, $1-\beta>.99$; post hoc power tested by G-power, Faul et al., 2009. The simple effect analysis showed that musical experts responded significantly faster to musical notations and faces than musical novices, Note: $t(76)=-6.04$, $p<.001$, $d=1.37$; Face: $t(76)=-3.1$, $p=.003$, $d=0.7$, but not to words, Word: $t(76)=-1.32$, $p=.19$. Response accuracy for musical notation and face processing was higher for musical experts than for musical novices, Note: $t(76)=5.94$, $p<.001$, $d=1.36$; Face: $t(76)=2.15$, $p=.035$, $d=0.49$, and there was no group difference for word processing, Word: $t(76)=-1.43$, $p=.16$ (see Figure 8).

The same pattern of findings held when using the index of RT/accuracy, with smaller RT/accuracy in musical notation and face processing for music-reading experts than for novices (all $ps<.01$), and no difference in word processing between the two groups ($p=.8$).

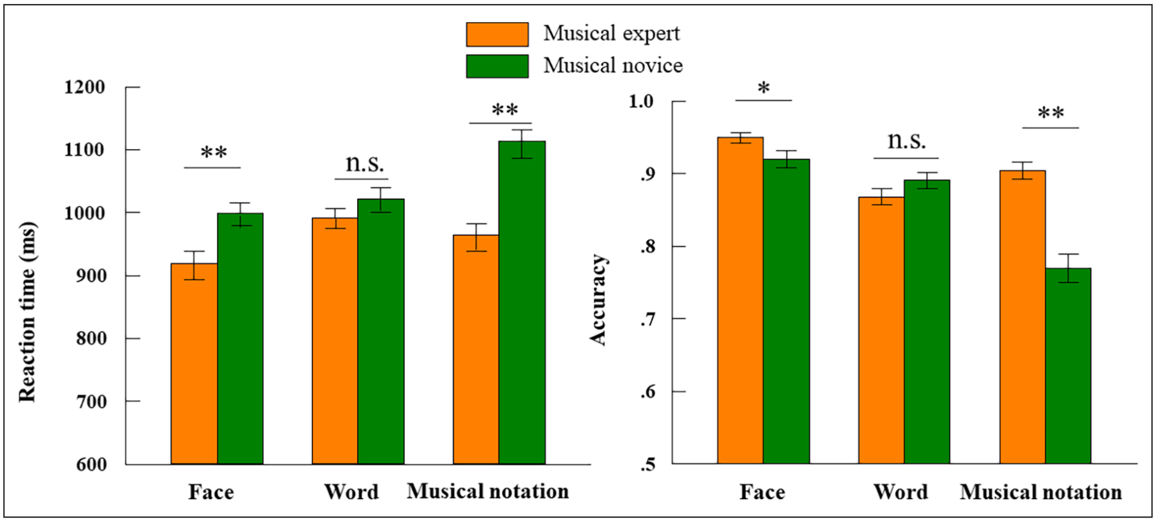

Figure 8. The results of reaction time and accuracy for face, word, and musical notation processing in musical experts and novices. Error bars plot s.e. * $p<.05$, ** $p<.01$ and n.s. means no significant difference.

Results showed above suggested that music-reading experts showed superior abilities for face and musical notation, but not word perception as compared to music-reading novices. The partial correlation analysis showed that the age at which music-reading began significantly correlated with the RT/accuracy in face processing ($r_{face}=.41$, $p=.011$), and a significant correlation was detected between the age at which music-reading began and RT/accuracy in musical notation processing after controlling for RT/accuracy in word processing ($r_{note}=.33$, $p=.043$; see Figure 9a). In addition, the RT/accuracy of musical notation processing was significantly related to the number of years of music-reading ($r_{note}=-.43$, $p=.009$), but the RT/accuracy of face processing was not ($r_{face}=-.31$, $p=.064$) after controlling for RT/accuracy of word processing (see Figure 9b).

In sum, the ANOVAs and correlational analyses of Experiment 3 confirmed the findings of Experiment 2 by showing that musical experts responded faster and more accurately to musical notations and faces than did musical novices, and the earlier the participants began music-reading, the superior their face and musical note processing ability. However, the stimuli were presented in a framed area in light grey background during Experiments 1, 2, and 3. This might have unnecessarily crowded the visual perception of the images, and hence the reported effects might also be related to group differences in crowding. To rule out this possibility, we carried out Experiment 4, in which the framed area has been removed (see Figure 10). Moreover, previous studies showed a left visual field superiority for face and Chinese character processing and a right visual field superiority for English word processing (Dundas et al., 2013; Hsiao & Lam, 2013), and stimuli were presented in the central visual field in the present Experiments 1, 2, and 3. Thus, Experiment 4 further examined group differences when stimuli were presented in the left and right visual fields to test the association between music-reading expertise and the hemispheric lateralisation for face and Chinese character recognition.

Experiment 4

Method

Participants. A total of 30 new music undergraduate students (27 females, $M=19.30$ years, $SD=0.60$) were recruited from the College of Music of the Hunan Normal University in China. All of the musical experts had received formal musical training and had long years of music-reading experience ($M=8.43$ years, $SD=4.31$). The mean age of beginning their regular music-reading was 10.33 years ($SD=4.67$, range: 4–18 years of age). All of them must pass through the professional examination of music before entering the college, including the examination of music theory, vocal music, musical instrument, and solfeggio.

In addition, 30 new non-music students (27 females, $M=19.03$ years, $SD=0.89$) were recruited from other colleges of this University. They reported that they could not read music or play musical instruments, and none of them had received any other musical training except for those taught in public school. In addition, both groups were matched in age, $F(1, 58)=1.86$, $p=.18$, parents' education level, $F(1, 58)=2.53$, $p=.12$, and the score of Raven's Advanced Reasoning Test, $F(1, 58)=1.09$, $p=.3$. All subjects were healthy, right-handed, with normal or corrected to normal vision, and reported no history of any affective disorder. This investigation was approved by the local human research ethics committee, and it adhered to the tenets of the Declaration of Helsinki. All participants signed an informed consent form before participating.

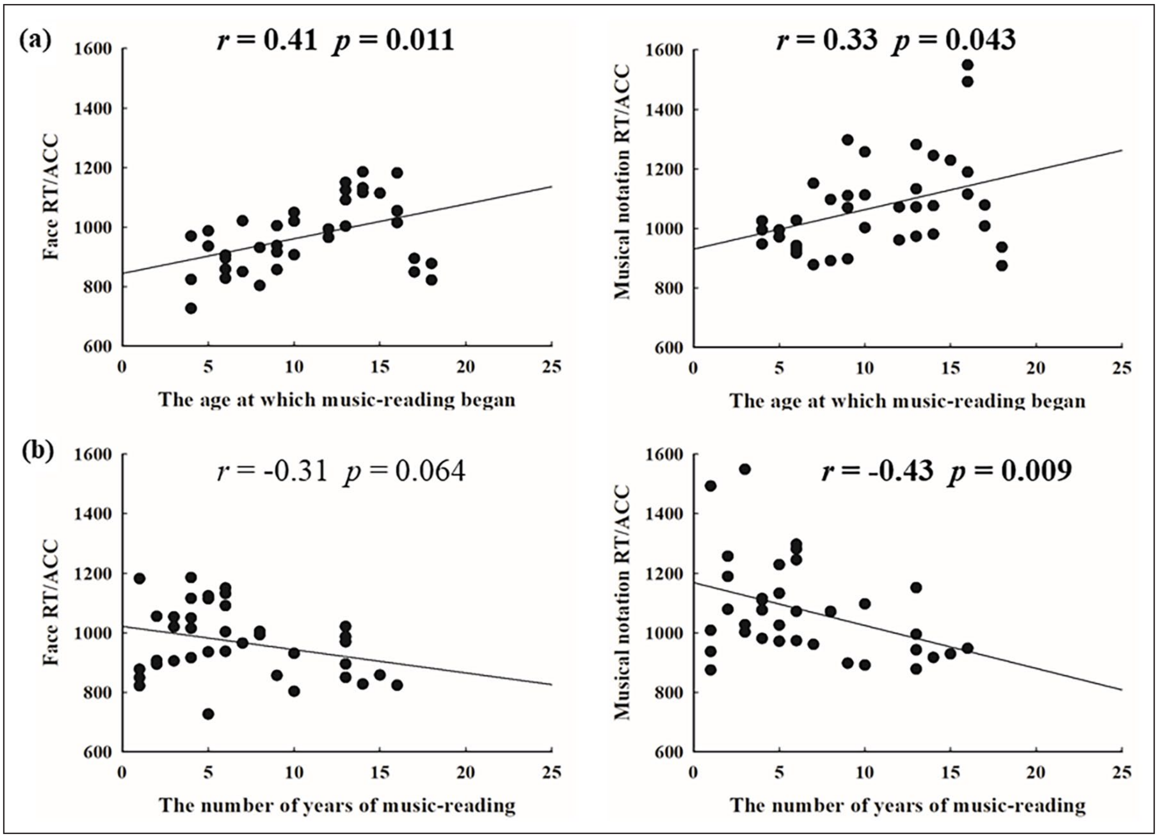

Figure 9. (a) Correlations of the age at which music-reading began with RT/ACC for face and musical notation processing after controlling for RT/ACC in word processing. (b) Correlations of the number of years of music-reading with RT/ACC for face and musical notation processing after controlling for RT/ACC in word processing.

Stimuli and procedure. In total, 60 Chinese faces (30 women and 30 men), 60 Chinese characters, and 60 musical notes were used for the perception matching task. The size of each face, word, and musical notation was about 2.6 cm wide and 3.6 cm high, and subtended a visual angle of 1.86° × 2.58° at a viewing distance of 80 cm. Unlike Experiments 1, 2, and 3, the experimental stimuli in Experiment 4 did not contain a grey framed area (see Figure 10). In addition, because the accuracies in processing musical notations were relatively low in musical novices in Experiments 1, 2, and 3, to reduce the task difficulty in processing musical notations, the paired target and distractor notations differed both in time value and position on the five-line staff.

Participants took part in the experiment individually. Each trial was initiated by 800–1200 ms presentation of a small black cross on the white computer screen. Afterwards, a target stimulus (the face, word, or musical notation) was presented in the centre of the screen for a duration of around 750 ms, followed by a 150 ms blank screen. Then, a probe stimulus was presented on the left or right side of the cross for 150 ms (see Figure 10). The centre of the lateralised stimulus was 4.15° from fixation. The left–right position of the probe stimulus was randomised across trials. Participants were required to keep their gaze fixated centrally throughout the experiment and detect whether the probe stimulus was the same with the target stimulus as soon and accurately as possible. Each probe stimulus was followed by a 1500 ms blank screen. Each of the six experimental conditions contained 30 trials, and all trials were presented randomly.

After the experiment, participants were also instructed to fill out a questionnaire about their musical experiences, parents' education level, and the abbreviated version of Raven's Advanced Reasoning Test (Arthur & Day, 1994).

Results and discussion

The design of the experiment entailed a between-subjects variable of Participant group (musical experts and musical novices) with two within-subject factors: Stimulus type (musical notation, face, and Chinese character) and Field (left and right). ANOVAs of mean RTs showed no Participant group × Stimulus type, $F(2, 116) = 2.38$, $p = .1$, Participant group × Field, $F(1, 58) = 0.11$, $p = .74$, and Participant group × Stimulus type × Field effects, $F(2, 116) = 0.29$, $p = .74$. However, the main effect of Stimulus type, $F(2, 116) = 27.94$, $p < .001$, $\eta_p^2 = .33$, and

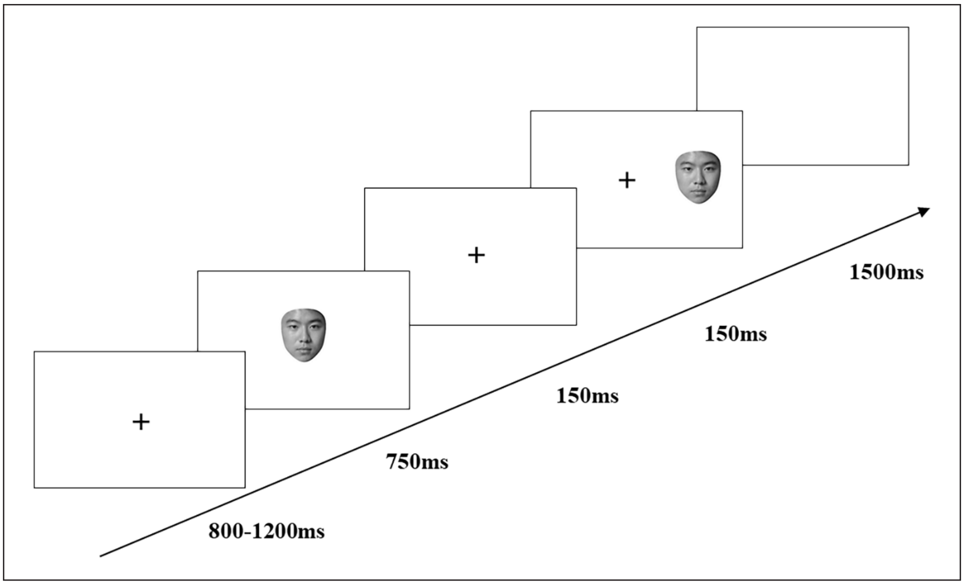

Figure 10. Illustration of the temporal sequence of an individual trail in Experiment 4.

Stimulus type × Field interaction effect was significant, $F(2, 116)=10.32$, $p<.001$, $\eta_p^2=.15$. RTs in the musical notation condition were shorter than those in the word condition ($p=.033$, Bonferroni corrected), in which RTs were also shorter than those in the face condition ($p<.001$, Bonferroni corrected). The simple effect analysis showed that RTs were shorter in the left field than in right field conditions, irrespective of the type of stimulus, and the degree of visual lateralisation effect (RTs in the right field minus RTs in the left field) was stronger in processing Chinese character than in processing musical notation and face (all ps $<.01$, Bonferroni corrected).

More importantly, ANOVAs of mean accuracies showed significant main effect of Stimulus type, $F(2, 116)=47.33$, $p<.001$, $\eta_p^2=.45$, and a Stimulus type × Participant group interaction effect, $F(2, 116)=3.65$, $p=.035$, $\eta_p^2=.06$, $1-\beta>.99$; post hoc power tested by G-power, Faul et al., 2009. Accuracies in the word condition were higher than those in the musical notation condition ($p<.001$, Bonferroni corrected), in which accuracies were also higher than those in the face condition ($p<.001$, Bonferroni corrected). The simple effect analysis showed that the accuracies of faces and musical notations processing were higher for musical experts than for novices, Face: $t(58)=2.96$, $p=.004$, $d=0.76$; Note: $t(58)=4.45$, $p<.001$, $d=1.15$, and there were no group differences for words, Word: $t(58)=1.48$, $p=.15$ (Figure 11). In addition, no significant main effect of Field, Stimulus type × Field, and Participant group × Stimulus type × Field interaction effects were observed (all ps $>.1$). Thus, extending the previous three experiments, Experiment 4 further suggested that musical experts showed superior face recognition abilities than musical novices when face stimuli were presented in the left and right visual fields. This finding, to some extent, reflects that the association between music-reading expertise and face recognition is stable.

However, there were no group differences between faces, words, and musical notations processing when we used the index of RT/accuracy, and also no any significant correlations between musical experiences (the number of years of music-reading and the age at which music-reading began) and face or musical notation processing (RT, accuracy or RT/accuracy; all ps $>.05$). These results may reflect that processing stimuli in central and lateral visual fields involve different cognitive mechanisms.

General discussion

This study for the first time examined the association among music-reading expertise and music notation, face, and Chinese character processing. Experiments 1, 2, 3, and 4 provided convergent evidences showing that music-reading expertise is associated with enhanced musical notation, face, but not Chinese character processing abilities. Moreover, Experiments 1, 2, and 3 revealed that the earlier the music experts began musical notation reading, the superior their musical notation and face processing abilities.

The present findings are generally in line with the existing findings that musical training is positively associated with a wide range of cognitive abilities, such as verbal memory ability (Chan et al., 1998), mathematical ability (Cheek & Smith, 1999), reading ability (Douglas & Willatts, 1994), and even IQ (Schellenberg, 2004).

Musical training also has significant effects on brain function and structure (Gaser & Schlaug, 2003; Schlaug

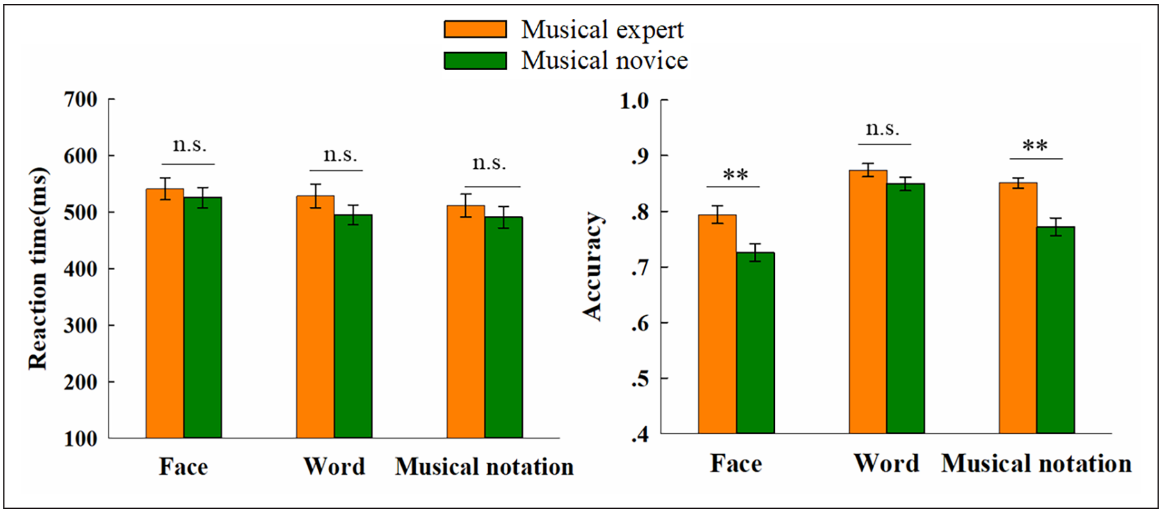

Figure 11. The results of reaction time and accuracy for face, word, and musical notation processing in musical experts and novices.
Error bars plot s.e. $^{**}p < .01$ and n.s. means no significant difference.

et al., 1995; Y. K. Wong et al., 2014; Y. K. Wong & Gauthier, 2010). However, the present findings extend these previous findings by demonstrating that the visual expertise for processing musical notation does not interact with all other forms of visual expertise. Rather, acquisition of music-reading skills is associated with enhanced face, but not Chinese character processing abilities, and this association is specific and limited to the visual expertise with which the musical notation processing expertise shares similar cognitive processes. Li and Hsiao's study showed music-reading expertise modulated hemispheric lateralisation in English word processing but not in Chinese character processing. Similarly, S. T. K. Li et al. (2019) found music-reading expertise modulates the visual span for English letters rather than Chinese characters. These modulation effects were suggested to be due to the similarities between the processes involved in reading English words and musical notations but the dissimilarities between the processes involved in reading Chinese characters and musical notations (S. T. K. Li et al., 2019; S. T. K. Li & Hsiao, 2018). In other words, whether different expertise domains influence each other depends on their similarities in the cognitive processes involved (S. T. K. Li & Hsiao, 2018). Our findings were also consistent with previous studies regarding the transfer effect in the motor domain (Moore & Müller, 2014; Müller et al., 2015; Rosalie & Müller, 2012). For example, Moore and Müller (2014) found that the visual anticipatory skill in baseball batting can be transferred to cricket batting because these two sports share similar time constraints and are both classified as striking sports, but not to rugby due to the different kinematic patterns between them (Moore & Müller, 2014; Müller et al., 2015). Thus, the association between music-reading expertise and face processing ability may be due to the fact that both musical notation and face identity recognition depend heavily on spatial information processing (Freire et al., 2000). However, it has been suggested that both configural and featural processing make important contributions to face recognition (Cabeza & Kato, 2000; Chuk et al., 2017). Zhou et al. (2012) found that face drawing experts demonstrated less holistic face processing as compared with people with no face drawing experience. This is because that face drawing involves quick and accurate grasp of the main facial features of each individual, and thus can reduce the holistic face processing. It would be an interesting question for future studies to investigate the association between music-reading expertise and configural and featural processing of faces and to further determine whether music-reading expertise has a stronger association with configural than with featural processing of faces.

Our correlational results shed further light on the relationship between the acquisition of musical notation reading expertise and that of face processing. We found that the age of onset of musical notation reading correlated with the music experts' face processing performance. Thus, the finding of the importance of the age of music training onset on the music experts' face processing suggests that the positive association between musical notation reading expertise and face processing is a developmental phenomenon: Whether musical notation reading associates with face processing depends on at which age, not how long, the music experts began their musical notation reading training, and the early commencement of music-reading training plays a more important role than the total accumulative music-reading experience. This finding is consistent with previous studies showing that some cognitive abilities (e.g., executive functions, rhythm synchronisation, and sensorimotor synchronisation skills) are strongly associated with early musical training (Bailey & Penhune, 2010, 2013; Chen et al., 2021). Neuroimage studies found that early musical training could lead to more extensive changes in brain structure than later training (Schlaug et al., 1995). Moderate amount of musical training early in life could contribute to faster neural timing in response to speech later in life, long after training stopped (>40 years) (White-Schwoch et al., 2013). In other words, early experience of musical training, not the length of training, can carry meaningful and long-lasting benefits well into adulthood.

As shown in Figures 4, 7, and 9, 11–12 years of age appear to be a transitional age: the music experts who began music-reading training before 11–12 years of age had markedly better face processing performances than those beginning music-reading training after 11–12 years—Experiment 1: RT/ACC_{face}, 566.92 versus 631.81, $t(48) = -2.55$, $p = .014$; Experiment 2: RT/ACC_{face}, 903.22 versus 998.71, $t(43) = -2.63$, $p = .012$; Experiment 3: RT/ACC_{face}, 911.44 versus 1034.7, $t(36) = -3.81$, $p = .001$. One possible reason for this age-related shift is that before 11–12 years of age, children's face processing ability is far from mature (Mondloch et al., 2002, 2003, 2006), whereas after 11–12 years of age, their face processing expertise becomes similar to that of adults. The association between music-reading expertise and face processing ability may be easier to realise when children's face processing system is immature than when it has reached maturity. Given the less mature face processing system's plasticity, it is more susceptible to the influence of non-face-specific experiences. As a result, music training during the earlier developmental period of face processing exerted stronger association with face recognition ability than that during the later developmental period.

Although music-reading experts outperformed novices on face processing, it is worth noting that the present design did not allow for establishing a causal relationship between music-reading experience and the face processing advantage (Schellenberg, 2001). It has been suggested that pre-existing characteristics among musicians, such as demographics, cognitive abilities, and personality traits, can predict their later involvement in musical training (Corrigall et al., 2013). Thus, it may be possible that it is a pre-existing face processing advantage among musicians that bias them towards learning to read music early in life. In addition, the long-term music-reading is always accompanied by other forms of musical training (e.g., music listening and musical instrument practice). Thus, the musicians' face processing advantage may due to the general musical training from music-reading, music listening, and musical instrument practice. Thus, building on the present findings, future studies need to use the true experimental design by randomly assigning participants to either music notation reading or control conditions with pre- and post-tests on musical notation, face, and word processing to better determine the causal relationship between music-reading experience and the face processing advantage.

In summary, the results of Experiments 1, 2, 3, and 4 provide preliminary and converging evidence to support the association between music-reading expertise and face identity recognition ability, and this association appears to be greater at the early developmental stage of face processing expertise acquisition. The present findings add to the growing body of recent research demonstrating the acquisition of one form of visual expertise interacting with that of another (Dundas et al., 2013; S. T. K. Li & Hsiao, 2018; Moore & Müller, 2014; Müller et al., 2015), and the short- and long-term consequences of such interactions in behaviour and in the brain (Dehaene et al., 2010; S. Li et al., 2013). The existing, albeit very limited, evidence taken together suggests that the development of different types of visual expertise may not be independent, but rather interact with each other during their acquisition.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the National Natural Science Foundation of China (NSFC31771240, NSFC31300866) and Postgraduate Scientific Research Innovation Project of Hunan Province (CX20210466).

ORCID iDs

Jie Chen https://orcid.org/0000-0002-6264-9812
Yi Jiang https://orcid.org/0000-0002-5746-7301

References

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., & Lee, K. (2013). Development of own-race biases. *Visual Cognition*, *21*(9), 1165–1182. https://doi.org/10.1080/13506285.2013.821428

Arthur, W., & Day, D. V. (1994). Development of a short form for the raven advanced progressive matrices test. *Educational & Psychological Measurement*, *54*(2), 394–403. https://doi.org/10.1177/0013164494054002013

Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, *204*(1), 91–101. https://doi.org/10.1007/s00221-010-2299-y

Bailey, J. A., & Penhune, V. B. (2013). The relationship between the age of onset of musical training and rhythm synchronization performance: Validation of sensitive period effects. *Frontiers in Neuroscience*, *7*, Article 9. https://doi.org/10.3389/fnins.2013.00227

Behrmann, M., & Plaut, D. C. (2012). Bilateral hemispheric processing of words and faces: Evidence from word impairments in prosopagnosia and face impairments in pure alexia. *Cerebral Cortex*, *24*, 1102–1118. https://doi.org/10.1093/cercor/bhs390

Boggan, A. L., Bartlett, J. C., & Krawczyk, D. C. (2012). Chess masters show a hallmark of face processing with chess. *Journal of Experimental Psychology: General*, *141*(1), 37–42. https://doi.org/10.1037/a0024236

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain & Cognition*, *54*(2), 103–109. https://doi.org/10.1016/S0278-2626(03)00264-1

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, *11*(5), 429–433. https://doi.org/10.1111/1467-9280.00283

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, *396*, Article 128. https://doi.org/10.1038/24075

Cheek, J. M., & Smith, L. R. (1999). Music training and mathematics achievement. *Adolescence*, *34*(136), 759–761. https://doi.org/10.1177/105971239900700311

Chen, J. J., Scheller, M., Wu, C. Y., Hu, B. Y., Peng, R., Liu, C. H., Zhu, L. W., & Chen, J. (2021). The relationship between early musical training and executive functions: Validation of effects of the sensitive period. *Psychology of Music*, *50*(1), 86–99. https://doi.org/10.1177/0305735620978690

Chuk, T., Crookes, K., Hayward, W. G., Chan, A. B., & Hsiao, J. H. (2017). Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, *169*, 102–117. https://doi.org/10.1016/j.cognition.2017.08.003

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*, Article 222. https://doi.org/10.3389/fpsyg.2013.00222

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*, 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., NunesFilho, G., Jobert, A., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*, 1359–1364. https://doi.org/0.1126/science.1194140

Douglas, S., & Willatts, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading*, *17*, 99–107. https://doi.org/10.1111/j.1467-9817.1994.tb00057.x

Dundas, E. M., Plaut, D. C., & Behrmann, M. (2013). The joint development of hemispheric lateralization for words and faces. *Journal of Experimental Psychology: General*, *142*(2), 348–358. https://doi.org/10.1037/a0029503

Dundas, E. M., Plaut, D. C., & Behrmann, M. (2014). An ERP investigation of the co-development of hemispheric lateralization of face and word recognition. *Neuropsychologia*, *61*, 315–323. https://doi.org/10.1016/j.neuropsychologia.2014.05.006

Dundas, E. M., Plaut, D. C., & Behrmann, M. (2015). Variable left-hemisphere language and orthographic lateralization reduces right-hemisphere face lateralization. *Journal of Cognitive Neuroscience*, *27*(5), 913–925. https://doi.org/10.1162/jocn_a_00757

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*, 159–170. https://doi.org/10.1068/p3012

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*(27), 9240–9245. https://doi.org/10.1523/JNEUROSCI.23-27-09240.2003

Gauthier, I., Curran, T., Curby, K., & Collins, D. (2003). Perceptual interference support a non-modular account of face processing. *Nature Neuroscience*, *6*, 428–432. https://doi.org/10.1038/nn1029

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*, 191–197. https://doi.org/10.1038/72140

Ge, L. Z., Wang, Z., McCleery, J. P., & Lee, K. (2006). Activation of Face Expertise and the Inversion Effect. *Psychological Science*, *17*, 12–16. https://doi.org/10.1111/j.1467-9280.2005.01658.x

Hsiao, J. H., & Lam, S. M. (2013). The modulation of visual and task characteristics of a writing system on hemispheric lateralization in visual word recognition—A computational exploration. *Cognitive Science*, *37*, 861–890. https://doi.org/10.1111/cogs.12033

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Li, S., Lee, K., Zhao, J., Yang, Z., He, S., & Weng, X. (2013). Neural competition as a develop-mental process: Early hemispheric specialization for word processing delays specialization for face processing. *Neuropsychologia*, *51*, 950–959. https://doi.org/10.1016/j.neuropsychologia.2013.02.006

Li, S. T. K., Chung, S. T., & Hsiao, J. H. (2019). Music-reading expertise modulates the visual span for English letters but not Chinese characters. *Journal of Vision*, *19*(4), Article 10. https://doi.org/10.1167/19.4.10

Li, S. T. K., & Hsiao, H. W. (2018). Music reading expertise modulates hemispheric lateralization in English word processing but not in Chinese character processing. *Cognition*, *176*, 159–173. https://doi.org/10.1016/j.cognition.2018.03.010

McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the FFA related to perceptual performance. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(42), 17063–17068. https://doi.org/10.1073/pnas.1116333109

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology*, *86*, 67–84. https://doi.org/10.1016/S0022-0965(03)00102-4

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*, 553–566. https://doi.org/10.1068/p3339

Mondloch, C. J., Maurer, D., & Ahola, S. (2006). Becoming a face expert. *Psychological Science*, *17*, 930–934. https://doi.org/10.1098/rstb.1992.0012

Moore, C., & Müller, S. (2014). Transfer of expert visual anticipation to a similar domain. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *67*, 186–196. https://doi.org/10.1080/17470218.2013.798003

Müller, S., Mclaren, M., Appleby, B., & Rosalie, S. M. (2015). Does expert perceptual anticipation transfer to a dissimilar domain? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 631–638. https://doi.org/10.1037/xhp0000021

Rosalie, S. M., & Müller, S. (2012). A model for the transfer of perceptual-motor skill learning in human behaviors. *Research Quarterly for Exercise and Sport*, *83*(3), 413–421. https://doi.org/10.1080/02701367.2012.10599876

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, *13*, 250–257. https://doi.org/10.1111/1467-9280.00446

Schellenberg, E. G. (2001). Music and nonmusical abilities. *Annals of the New York Academy of Sciences*, *930*, 355–371. https://doi.org/10.1093/acprof:oso/9780198744443.003.0008

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G., & Peretz, I. (2008). Music, language and cognition: Unresolved issues. *Trends in Cognitive Science*, *12*(2), 45–46. https://doi.org/10.1016/j.tics.2007.11.005

Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*(8), 1047–1055. https://doi.org/10.1016/0028-3932(95)00045-5

Sergent, J., Zuck, E., Terriah, S., & MacDonald, B. (1992). Distributed neural network underlying musical sight-reading and keyboard performance. *Science*, *257*(5066), 106–109. https://doi.org/10.1126/science.1621084

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*, 43–47. https://doi.org/10.1111/1467-9280.00308

Tso, R. V. Y., Chan, R. T. C., & Hsiao, J. H. (2020). Holistic but with reduced right-hemisphere involvement: The case of dyslexia in Chinese character recognition. *Psychonomic Bulletin & Review*, *27*(3), 553–562. https://doi.org/10.3758/s13423-020-01721-y

Ventura, P., Fernandes, T., Cohen, L., Morais, J., Kolinsky, R., & Dehaene, S. (2013). Literacy acquisition reduces the influence of automatic holistic processing of faces and houses. *Neuroscience Letters*, *554*, 105–109. https://doi.org/10.1016/j.neulet.2013.08.068

White-Schwoch, T., Woodruff Carr, K., Anderson, S., Strait, D. L., & Kraus, N. (2013). Older adults benefit from music training early in life: Biological evidence for long-term training-driven plasticity. *The Journal of Neuroscience*, *33*, 17667–17674. https://doi.org/10.1523/JNEUROSCI.2560-13.2013

Wong, A. C.-N., Bukach, C. M., Hsiao, J., Greenspon, E., Ahern, E., Duan, Y., & Lui, K. F. H. (2012). Holistic processing as a hallmark of perceptual expertise for nonface categories including Chinese characters. *Journal of Vision*, *12*(13), Article 7. https://doi.org/10.1167/12.13.7

Wong, A. C.-N., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011). Holistic processing of words modulated by reading experience. *PLOS ONE*, *6*(6), Article e20753. https://doi.org/10.1371/journal.pone.0020753

Wong, A. C.-N., Gauthier, I., Woroch, B., Debuse, C., & Curran, T. (2005). An early electrophysiological potential associated with expertise in letter perception. *Cognitive, Affective, and Behavioral Neuroscience*, *5*(4), 306–318. https://doi.org/10.3758/CABN.5.3.306

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face—like expertise with objects: Becoming a Ziggerin expert—But which type? *Psychological Science*, *20*(9), 1108–1117. https://doi.org/10.1111/j.1467-9280.2009.02430.x

Wong, A. C.-N., Wong, Y. K., Lui, K. F. H., Ng, T. Y. K., & Ngan, V. S. H. (2019). Sensitivity to configural information and expertise in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(1), 82–99. https://doi.org/10.1037/xhp0000590

Wong, Y. K., & Gauthier, I. (2010). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, *22*(4), 695–713. https://doi.org/10.1162/jocn.2009.21229

Wong, Y. K., Peng, C., Fratus, K. N., Woodman, G. F., & Gauthier, I. (2014). Perceptual expertise and top–down expectation of musical notation engages the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*, 1629–1643. https://doi.org/10.1162/jocn_a_00616

Zhou, G., Cheng, Z., Zhang, X., & Wong, A. C.-N. (2012). Smaller holistic processing of faces associated with face drawing experience. *Psychonomic Bulletin & Review*, *19*, 157–162. https://doi.org/10.3758/s13423-011-0174-x