Visual training with musical notes changes late but not early electrophysiological responses in the visual cortex

Alan C.-N. Wong
Department of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Terri Y. K. Ng
Department of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Kelvin F. H. Lui
Department of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Ken H. M. Yip
Department of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Yetta Kwailing Wong
Department of Educational Psychology, Faculty of Education, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Visual expertise with musical notation is unique. Fluent music readers show selectively higher activity to musical notes than to other visually similar patterns in both the retinotopic and higher-level visual areas and both very early (e.g., C1) and later (e.g., N170) visual event-related potential (ERP) components. This is different from domains such as face and letter perception, of which the neural expertise marker is typically found in the higher-level ventral visual areas and later (e.g., N170) ERP components. An intriguing question concerns whether the visual skills and neural selectivity observed in music-reading experts are a result of the effects of extensive visual experience with musical notation. The current study aimed to investigate the causal relationship between visual experience and its neural changes with musical notation. Novices with no formal musical training experience were trained to visually discriminate between note patterns in the laboratory for 10–26 hr such that their performance was comparable with fluent music readers. The N170 component became more selective for musical notes after training. Training was not, however, followed by changes in the earlier C1 component. The findings show that visual training is enough for causing changes in the responses of the higher-level visual areas to musical notation while the engagement of the early visual areas may involve additional nonvisual factors.

Introduction

Understanding how the visual cortex represents different object categories remains challenging. It is commonly accepted that visual object identity and category are primarily processed in the ventral visual pathway (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). Specifically, the early visual cortex codes for simple and local visual features (Hubel & Wiesel, 1968), and its activity does not differentiate between noise patterns and objects (Grill-Spector & Malach, 2004; Malach et al., 1995). When the visual information stream reaches the higher visual cortex, category-selective neural responses emerge. For example, the lateral occipital cortex responds selectively to objects compared with noise patterns or scrambled

Citation: Wong, A. C.-N., Ng, T. Y. K., Lui, K. F. H., Yip, K. H. M., & Wong, Y. K. (2019). Visual training with musical notes changes late but not early electrophysiological responses in the visual cortex. *Journal of Vision*, *19*(7):8, 1–16, https://doi.org/10.1167/19.7.8.

https://doi.org/10.1167/19.7.8 Received January 15, 2019; published July 18, 2019 ISSN 1534-7362

objects (Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995). Also, different parts of the fusiform gyrus and the parahippocampal gyrus respond selectively to certain object categories, including faces (Kanwisher, McDermott, & Chun, 1997), body parts (Downing, Jiang, Shuman, & Kanwisher, 2001), buildings and scenes (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998), and letters and words (Cohen et al., 2000; James, James, Jobard, Wong, & Gauthier, 2005).

The functional specialization of higher visual areas for different objects is often associated with visual expertise as suggested in studies comparing real-world experts and novices with an object category (chessboards: Bilalić, Langner, Ulrich, & Grodd, 2011; fingerprints: Busey & Vanderkolk, 2005; birds and cars: Gauthier, Skudlarski, Gore, & Anderson, 2000; radiographs: Harley et al., 2009; words and characters: A. C. Wong, Jobard, James, James, & Gauthier, 2008; musical notes: Y. K. Wong & Gauthier, 2010) as well as in object training studies comparing individuals' neural responses before and after training (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Lochy et al., 2018; Moore, Cohen, & Ranganath, 2006; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002; Scott, Tanaka, Sheinberg, & Curran, 2006, 2008; A. C.-N. Wong, Palmeri, Rogers, Gore, & Gauthier, 2009; Y. K. Wong, Folstein, & Gauthier, 2012). In these studies, real-world expertise or visual training in the laboratory typically resulted in enhanced neural activity for the objects associated with perceptual expertise in the higher visual cortex.

Interestingly, recent evidence using the fMRI technique showed that both early and late visual areas respond selectively for musical notes in expert music readers (Y. K. Wong & Gauthier, 2010). This study compared the activations of music-reading experts and novices when they performed visual judgments on single musical notes. A number of areas were found to respond more to musical notes than Roman letters and mathematical symbols in music-reading experts compared with novices, including V1/V2 and the higher-level visual areas, such as bilateral fusiform gyrus and right inferior temporal sulcus, as well as the auditory, somatosensory, motor, parietal, and frontal regions. An interesting aspect of this study is that the task in the scanner required only visual judgment of single musical notes. So the results suggest that music readers automatically recruit a specialized network of multimodal areas upon seeing musical notes. Using more complex musical-note sequences, the recruitment of V1/V2 was observed less consistently in some studies (Sergent, Zuck, Terriah, & MacDonald, 1992; Y. K. Wong & Gauthier, 2010) but not in others (Mongelli et al., 2016; Nakada, Fujii, Suzuki, & Kwee, 1998; Schön & Besson, 2002; Stewart et al., 2003), which might be related by the use of different types of visual control stimuli and tasks across studies. The selective engagement of both early and late visual areas associated with expertise has not been reported for object categories other than musical notes with the exception of written words in Szwed et al.'s (2011) fMRI study, in which V1/V2, V3v/V4, and the visual word-form area were identified.

A subsequent event-related potential (ERP) study with higher temporal resolution suggested that the selective activation for musical notes is not simply a result of a feedback signal from higher visual cortex, but is at least partly generated locally in the early visual cortex (Y. K. Wong, Peng, Fratus, Woodman, & Gauthier, 2014). During simple visual judgment with single musical notes, selectively higher activation was found in the N170 components and the early C1 (40–60 ms) bilaterally for musical notes in music-reading experts than novices, and this group difference was not found for contrast-matched pseudo-letters. The involvement of the N170 component was expected as it has been shown to be a general expertise marker for many categories, including faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996), cars (Gauthier, Curran, Curby, & Collins, 2003), dogs and birds (Tanaka & Curran, 2001), letters and words (Maurer, Zevin, & McCandliss, 2008; A. C. N. Wong, Gauthier, Woroch, DeBuse, & Curran, 2005), and novel artificial objects (Rossion et al., 2002; Rossion, Kung, & Tarr, 2004). The early C1 component was more surprising as it has been regarded as the first visual ERP component generated in V1 upon visual stimulation (Clark & Hillyard, 1996; Jeffreys & Axford, 1972) and has not been identified for other expertise domains studied before.

Why would the expert recognition of musical notes uniquely engage both early and late visual processes? An intriguing question is to what extent such engagement of both early and late visual processes is caused by experts' extensive visual experience in reading musical notation. One usual caveat of the study of real-world expertise is its correlational nature. Prior studies on expert music reading typically compared visual skills of existing musicians with those of novices. Although it is an intriguing possibility that prolonged experience in note recognition causes changes in the visual system, an alternative is that many people become keenly involved with music reading because they are born with a visual system better equipped for recognizing musical notes. The objective of the study was to investigate whether there is a causal relationship between visual experience and changes in the activity of early and late visual areas elicited by musical notes. To this end, we introduced visual note discrimination training to novices and observed brain activation changes as a result of training. Based on a successful training protocol in improving visual perceptual fluency (Y. K. Wong &

Wong, 2016), we aimed to train music-reading novices to become visually as fluent as real-world experts in discriminating highly similar musical note sequences. No auditory motor labels or meanings were attached to the note sequences. This would allow a test of the effects of pure visual training without the influence of multimodal associations or semantic processing. Before and after training, EEG was used to measure training-induced changes in the neural processes of musical notes.

There are two reasons for focusing on the visual aspect of musical expertise. The primary and theoretical reason is that, understanding the visual aspects of musical expertise would help to build a taxonomy of visual expertise with objects as studied in the literature (faces, cars, dogs, words, characters, fingerprints, medical slides, chess boards, etc.). For some of these categories, evidence has accumulated for the role of training in causing activation changes in the brain. The establishment of the same causal relationship for musical notes would, therefore, allow the comparison with expertise in other domains and, thus, facilitate the understanding of the functional organization of the visual cortex. The secondary and practical reason for focusing on visual expertise is that it is very difficult to create a full-fetched musician in the laboratory with a limited amount of training or follow someone longitudinally for years while he or she is being trained. In contrast, previous studies have shown that hours of visual training with a novel set of objects is enough to cause behavioral and neural changes that are qualitatively similar to real-world expertise, in particular, expertise with face perception (Gauthier et al., 1999; Gauthier, Williams, Tarr, & Tanaka, 1998; A. C. N. Wong, Palmeri, & Gauthier, 2009) and word reading (James & Atwood, 2009; Xue & Poldrack, 2007). Therefore, it is more feasible to target the visual aspect of music training and to observe behavioral and neural changes thereafter.

The ERP components of the C1 and the N170 were the focuses of the study. Based on the selectively larger N170 component observed for musical notes in music readers in the previous study (Y. K. Wong et al., 2014) and the expertise-associated changes of the N170 in previous visual training studies (Cao, Jiang, Li, Xia, & Jackie Floyd, 2015; Rossion et al., 2002; Scott et al., 2006, 2008), we predicted changes in the N170 component for musical notes after visual training. For the C1, selective responses for musical notes in music readers have also been observed before (Y. K. Wong et al., 2014). Previous training studies involving simple feature discrimination in a perceptual learning paradigm have also found changes to the C1 (G.-L. Zhang, Li, Song, & Yu, 2015) although no study has been conducted to test if complex object perception training would lead to changes in the C1. If the V1 recruitment for musical notes in music readers is driven by visual perceptual expertise, then the laboratory training that recreated the visual aspect of musical expertise should be sufficient to cause changes in the C1 component. Alternatively, if nonvisual factors are crucial for the engagement of the early visual cortex for expert music readers, then we should not be able to see changes in the C1 component even if the behavioral visual fluency of the trained participants becomes comparable with that of the real-world experts.

Methods

Participants

Twenty participants (seven females and 13 males; mean age = 20.55, $SD = 1.82$) were recruited from the Chinese University of Hong Kong. All participants were right-handed, reported normal or corrected-to-normal vision, and reported no history of neurological disorders. None of the participants had formal training on any musical instruments except the music curriculum in their primary and secondary schools. They were given informed consent according to the guidelines of the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong and of the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee in adherence to the Declaration of Helsinki. A small monetary compensation was given for their participation. Eight additional participants participated in the pretest but were not allowed to start training: One of them performed during the behavioral pretest at a level approaching the performance level of expert music readers, i.e., the mean threshold for the four- to five-note on a staff sequences before training was <730 ms (Y. K. Wong & Gauthier, 2010), leaving relatively little room for training. The other seven performed poorly (accuracy <85%) in the simple visual judgment tasks in the EEG pretest. Data from three more participants who completed the experiment were discarded as two of them showed low signal-to-noise ratios (SNRs)[1] in the ERP data, and no N170 component can be identified for the other one.

Stimuli and apparatus

The experiment was conducted on Mac Minis (Apple, Cupertino, CA) during training and behavioral testing and a Windows PCs during EEG testing with MATLAB (MathWorks, Natick, MA) and the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). The stimuli were grayscale images presented on a

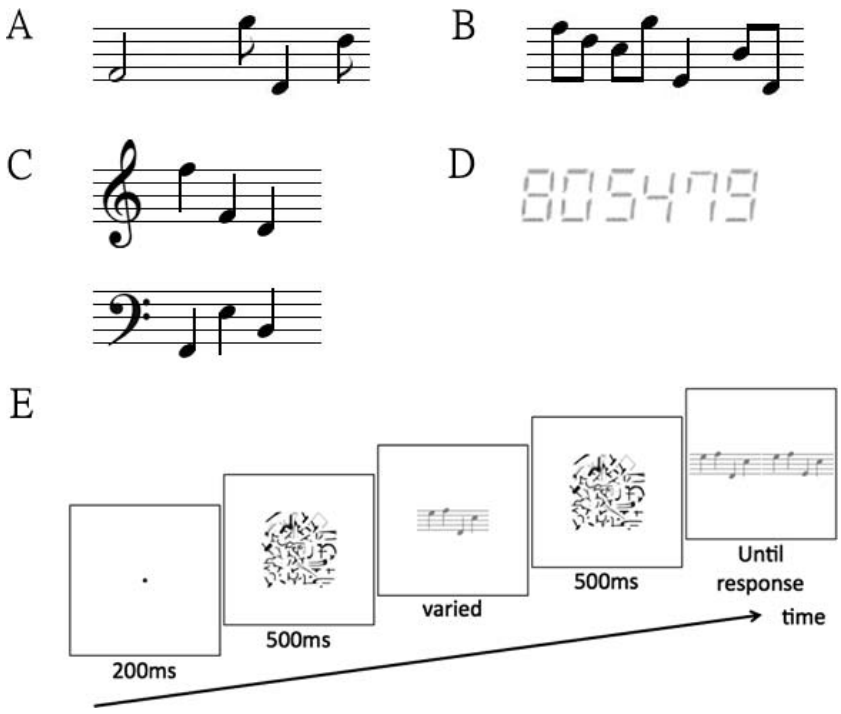

Figure 1. Materials in the behavioral tasks. (A–C) Note sequences (four to five notes on a staff, seven notes on a staff, six notes on two staves) used in the training and behavioral pretests and posttests. (D) Digit strings used in the behavioral pretests/posttests. (E) An example trial in the training and behavioral pretests/posttests.

white background. They were separately generated for the note-discrimination training, behavioral pretests and posttests, and the EEG pretests and posttests.

For the note-discrimination training, three types of random music sequences were generated in MATLAB. First, 12,600 sequences with four to five notes were shown on a five-line staff (Figure 1A). To increase the variability of the visual patterns, the sequences included notes with a range of rhythmic values: a half note (an open circle with a stem), a quarter note (a closed circle with a stem), and an eighth note (a closed circle with a tail at the end of stem or a closed circle with a stem and with a horizontal line connecting it to another eighth note). Second, 10,500 sequences with seven notes were shown on a five-line staff (Figure 1B). To fit in seven notes without increasing the overall width of the sequences, the rhythmic values of the notes included only the quarter notes and eighth notes. Third, 840 sequences with six quarter notes were shown on two five-line staves, each with three quarter notes (Figure 1C). The two staves were shown with a treble clef and a bass clef without any key or time signature. For all sequences, the notes were written ranging from the space below the bottom line of the staff (a "D" note on the treble clef or an "F" note on the bass clef) to the space above the top line of the staff (a "G" note on the treble clef or a "B" note on the bass clef). For each sequence, a distractor sequence was created in which all of the notes were identical except that one of the notes was shifted by one step, and it was chosen randomly. The up- or down-shifting direction of the altered note was counterbalanced across trials. None of the sequences contained notes with repeated pitches. The stimuli were presented on a 17-in. CRT monitor with $1{,}024 \times 768$ pixels resolution and 85 Hz frame rate. The one- and two-staff sequences approximately subtended a visual angle of $6.85° \times 1.84°$ and $5.65° \times 4.86°$ at a distance about 50 cm from the monitor, respectively.

For the behavioral pretests and posttests, similar musical note sequences were generated in MATLAB using a similar method as that used in the training: 450 pairs of note sequences were randomly generated for each type of sequence. Each pair contained a target sequence and a highly similar distractor sequence differing in only one of the notes, and 600 pairs of six-digit strings were used as control stimuli. Each string was composed of six nonrepeating digits from zero to nine (excluding one; Figure 1D). The contrast of all stimuli was lowered to 60% to increase the difficulty level to avoid ceiling effects. Similar to the training, the stimuli were presented on a 17-in. CRT monitor with $1{,}024 \times 768$ pixels resolution and 85 Hz frame rate. Each digit string subtended a visual angle of approximately $4.38° \times 0.90°$ at a distance about 50 cm from the monitor.

For the EEG pretests and posttests, eight musical notes and eight pseudo-letters were used as the basic stimuli (Figure 2A and B). The musical notes consisted

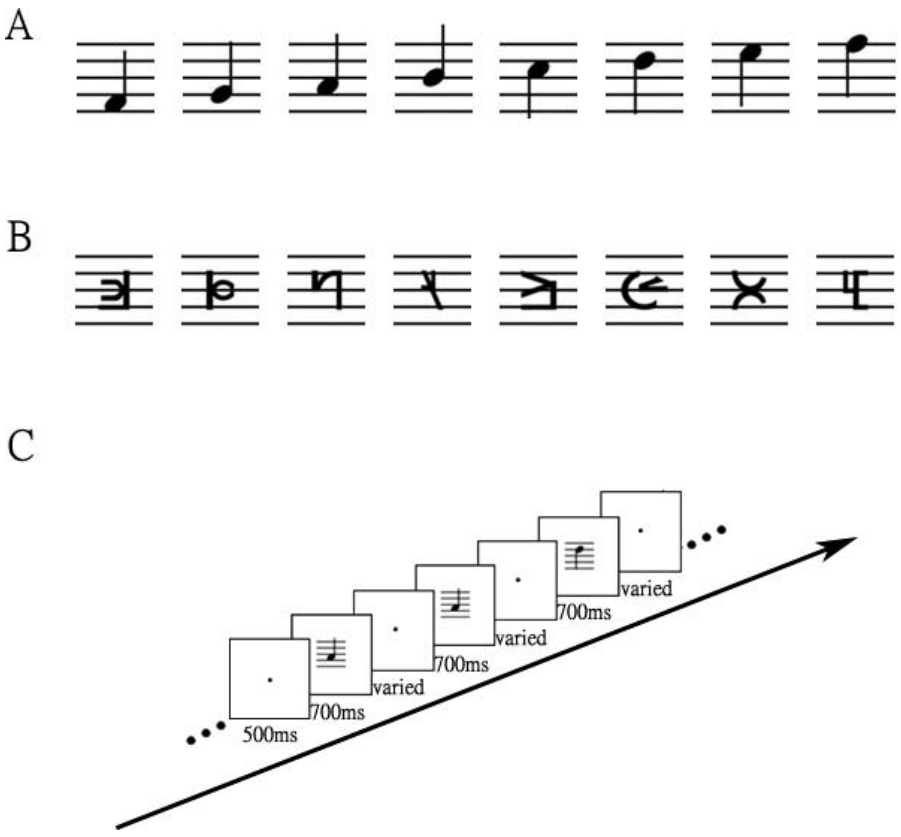

Figure 2. Materials in the EEG pretests/posttests. (A) Musical notations with staff. (B) Pseudo-letters with staff. (C) Example trials in the one-back paradigm.

of eight quarter notes ranging from F3 to F4. The pseudo-letters were generated by combining parts of Roman letters, similar to those used in our previous studies (A. C. N. Wong et al., 2005). Both types of stimuli were shown on the staff. The mean luminance and mean Weber contrast were matched between two object categories. Stimuli were presented on the white background and approximately subtended a visual angle of $2.1° \times 2.4°$ at a distance about 60 cm from the monitor. The testing was conducted on a PC platform using E-Prime 2.0 (Psychology Software Tools, Sharpsburg, PA), and 17-in. CRT monitors were used in all sessions.

Procedure

Each participant went through one behavioral and one EEG pretest session and then for a maximum of 30 hr of note-discrimination training (or until they passed all of the training levels) over 3–6 weeks and, finally, one behavioral and one EEG posttest session.

For the note-discrimination training, participants were required to attend training sessions at least four times per week, 1 hr for each session, and one session per day. The training included 100 levels with graded difficulty. Each level had 20 trials. On each trial (Figure 1E), a black fixation dot was presented at the center of the screen for 200 ms followed by a 500-ms premask. Then a target music sequence was randomly selected and presented for a varied duration from 4,000 ms (level 1) to 80 ms (level 100) followed by a 500-ms postmask. After that, two musical sequences appeared side by side. Participants pressed the "Z" or "M" key on the keyboard to indicate the left or right side of the image was identical to the target, respectively. Participants were required to respond within an 8-s response window (level 1), which was gradually reduced to a 2-s window (level 100). Feedback was given for each trial. Responses later than the response window were considered incorrect. For correct trials, a colored cartoon and a sound effect of "yay" were presented. For incorrect trials, a black and white cartoon with a sad face and a door-banging noise were presented. Participants were required to achieve an accuracy of 90% or above (or 18 out of 20 points) to pass each level. Otherwise, the same level was repeated. To motivate participants, a special trial was randomly given with a 1/200 chance in which a correct response was worth three points. Also, participants were allowed to accumulate one, two, and three tokens with 60%, 75%, and 90% accuracy in a

block, respectively. With 10 tokens, they could initiate the three-point special trial at any time. Participants were allowed to accumulate a maximum of three chances to initiate the special trial at any time point.

In the behavioral pretests and posttests, the perceptual fluency of discriminating music sequences was separately measured for the four- to five-note sequences; seven-note sequences; and six-note, two-staff sequences. For each type of sequence, a three-down, one-up staircase procedure (Levitt, 1971) was used to estimate the presentation duration threshold at which 80% accuracy was attained. On each trial (Figure 1E), a fixation dot in a black color was presented at the center of the screen for 200 ms followed by a 500-ms premask, a target for a varied duration, a 500-ms postmask, and finally two images appeared side by side until response. Participants pressed the "Z" or "M" key on the keyboard to indicate whether the left or right image was identical to the target. The presentation-duration threshold for each type of stimuli was measured three times. The order of the stimulus type was counterbalanced across participants. For each type of stimuli, the participant had five practice trials with feedback, followed by three blocks of test trials without feedback. The presentation duration of the target was 700 ms during the first trial. The presentation duration of the target was then adjusted based on participants' performance, in which the duration would decrease by 30% following three consecutive correct responses and would increase by 35% following each incorrect response. The maximum presentation duration allowed was 5 s to prevent lengthy measurements, and the minimum presentation duration allowed was set to 40 ms to avoid measurement based on largely impossible performance. Each staircase estimation was terminated after 12 reversals, i.e., switching the direction of adjustment from an increase to a decrease or vice versa. The number of reversals was determined based on pilot testing, which showed that participants' performance reached plateau by about 12 reversals. The staircase was also terminated if participants hovering around the maximum presentation duration threshold (i.e., after 20 consecutive trials of presentation duration at 5 s or after 30 trials of presentation duration at 5 s), which showed their difficulty of attaining 80% accuracy within 5 s of presentation duration. Each staircase estimate was averaged across all of the reversals except the first two. The average of the three staircase estimates was taken as the duration threshold for each type of stimuli.

For the EEG pretests and posttests, a one-back paradigm was adopted (Figure 2C). The ERP response for musical notes and pseudo-letters was compared in blocks of six images that were always in the same object category. Each block began with a black fixation dot at the center of the screen for 500 ms, followed by six trials, each with an image presented for 700 ms, and then a black fixation dot presented for 250–450 ms randomly. The fixation dot then turned gray for 2,500 ms for eye blinks. Participants were required to press a "1" key on the number pad of the keyboard as quickly and as accurately as possible if the neighboring trial was repeated. There were 720 trials for each category, including 60 repeated trials (repeated rate = 8.3%). The presentation order of the object categories was counterbalanced across participants. A practice block of trials was administered to ensure that participants fully understood the task prior to the experimental session. The position of the images was spatially jittered for five pixels (0.16°) in random from the center of the screen horizontally and vertically to minimize visual habituation. To decrease the task difficulty for identifying musical notes, the stem of consecutive stimuli always pointed to a different orientation (either upward or downward) unless the images repeated. A repetition suppression paradigm was also introduced after the one-back paradigm in the same session to explore any training effects concerning sensitivity to changes in musical notes. The methods and results are described in the Supplementary Appendix for completion despite the much smaller number of trials used for each condition and the lack of training effects found.

EEG recording and analyses

EEG recording was conducted in a dimly illuminated and air-conditioned room. The EEG was acquired continuously from 27 sites (positions: F3, F4, Fz, C3, C4, Cz, P3, P4, P7, P8, Pz, O1, O2, Oz, T7, T8, PO3, PO4, POz, M1, M2, AFz, and FCz) using Ag/AgCl electrodes mounted in an elastic cap (EASYCAP GmbH, Germany) at a sampling rate of 1,000 Hz with real-time bandpass filter of 0.01–100 Hz. The electrodes were positioned according to the International 10-20 system (Jasper, 1958) and recorded using Neuroscan SynAmps2 amplifiers (Compumedics, El Paso, TX). Horizontal and vertical eye movements were monitored with the bipolar recordings of an electro-oculogram (EOG). Blinks and vertical eye movements were monitored with electrodes affixed above and below the left eye, whereas lateral eye movements were recorded by two electrodes placed on the left and right external canthus. A ground electrode was placed on the forehead (position: AFz). All recordings were initially referenced to FCz and rereferenced off-line to an average of the left and right mastoids. Electrode impedances were kept below 5 kΩ.

Data analysis was performed with MATLAB and the EEGLAB toolbox. The EEG was segmented in

ERP epochs from 100 ms before and 699 ms after stimulus onset. The EEG data were preprocessed with the following steps. Trials with large fluctuation were first discarded by manual examination. All epochs from both tasks were then digitally low-pass filtered at 30 Hz. Trials associated with behavioral responses, i.e., repeated trials, were excluded. Trials with ocular artifacts were discarded by applying a step-like function on the vertical and horizontal EOG channels with a 45-microvolt threshold to 400-ms windows sweeping across the whole epoch in a step size of 10 ms as well as peak-to-peak threshold checking on all channels with a 100-microvolt threshold to 100-ms windows sweeping across the epoch with a step size of 25 ms. On average, 6.84% of the trials were discarded. The ERPs were baseline-corrected with respect to the 100-ms prestimulus interval.

The focus of the ERP analyses was on the amplitudes of two ERP components, the C1 and N170, for notes and pseudo-letters before and after training. Higher activations for notes than pseudo-letters would represent a channel's selectivity for notes, and the main question was how this selectivity increased from pretest to posttest. The C1 time segment was identified at 40–60 ms after stimulus onset at the channels PO3 and PO4, consistent with the setting in our earlier ERP study (Y. K. Wong et al., 2014). The time window corresponds to the earlier phase of C1 measured in previous ERP studies (Bao, Yang, Rios, He, & Engel, 2010; Clark, Fan, & Hillyard, 1995; Clark & Hillyard, 1996; Foxe et al., 2008; Martinez et al., 1999). The N170 time segment was defined for each participant due to the large individual difference in the latency of the component. The grand means of the ERPs across all conditions (musical notes and pseudo-letters in pretest and posttest) of each task for each participant were computed. The prominent negative local trough was searched around 140–200 ms as the peak of the N170 component. The latencies for the component for each participant were then recorded and used as the center of the time window. The channels P3, P4, P7, and P8 were analyzed. P3 and P4 were identified due to the training effect shown in the scalp topography (Figure 5A), and P7 and P8 were the channels showing the largest, i.e., the most negative N170 for all conditions in pretests and posttests collapsed. The average N170 amplitude was, thus, defined as the average amplitude within a 40-ms time window centering at an average of 189 ms ($SD = 23$ ms) across participants. We also examined the N250 component at P7 and P8 where the magnitude was the largest for all conditions in pretests and posttests collapsed. The average N250 amplitude was defined as the average amplitude within a 40-ms time window centering at an average of 299 ms ($SD =$ 20 ms) across participants.

Results

Note-discrimination training

All participants completed 100 levels of the musical-note discrimination task within 30 1-hr sessions. On average, participants completed the training in 16.35 hr (range: 10–26 hr).

Behavioral pretests and posttests

Participants demonstrated a higher perceptual fluency in posttest than pretest for note sequences but not for number strings (Figure 3). A 2 (PrePost: pretest, posttest) × 4 (Stimulus: four to five notes on a staff, seven notes on a staff, six notes on two staves, six-digit string) repeated-measures ANOVA was performed on the average log-transformed duration threshold required to achieve 80% accuracy. The main effect of PrePost was significant, $F(1, 19) = 131.729$, $p \leq 0.0001$, $\eta_p^2 = 0.874$, with a lower threshold duration at posttest than pretest. The main effect of Stimulus was also significant, $F(3, 17) = 84.226$, $p \leq 0.0001$, $\eta_p^2 = 0.937$, with a lower threshold for digits than notes. The interaction between PrePost and stimulus type was also significant, $F(3, 17) = 40.097$, $p \leq 0.0001$, $\eta_p^2 = 0.876$. Planned comparisons showed that the duration threshold was lower in posttest than pretest for all music sequences: four or five notes on a staff, $t(19) = -13.348$, $p \leq 0.0001$, $d = -2.98$; seven notes on a staff, $t(19) = -10.287$, $p \leq 0.0001$, $d = -2.30$; six notes on two staves, $t(19) = -6.712$, $p \leq 0.0001$, $d = -1.50$. This was not the case for the number strings, $t(19) = -0.619$, $p = 0.543$, $d = 0.14$. The threshold durations for digits were lower than that for notes even in pretest, so there was less room to show any improvement with digits as a result of training. Thus, based on the fluency data alone, it remains a possibility that the current note-discrimination training could have improved general visual fluency generalizable to different object categories.

The four- to five-note sequences on one staff have also been used in the previous ERP study (Y. K. Wong et al., 2014). It is, therefore, informative to compare the performance of the current participants before and after training for these sequences with those of the experts and novices in previous studies with the caveat that a staircase method was used to locate the threshold in the current study, and the QUEST procedure (Watson & Pelli, 1983) was used in the previous study. Specifically, the performance of the 20 participants in the current study was compared with the 22 novices and 20 experts in experiments 1 and 2 combined in Y. K. Wong et al. (2014). At pretest, the average log duration threshold of participants in the current study (3.24) was higher than that for novices, 3.03, $t(40) =$

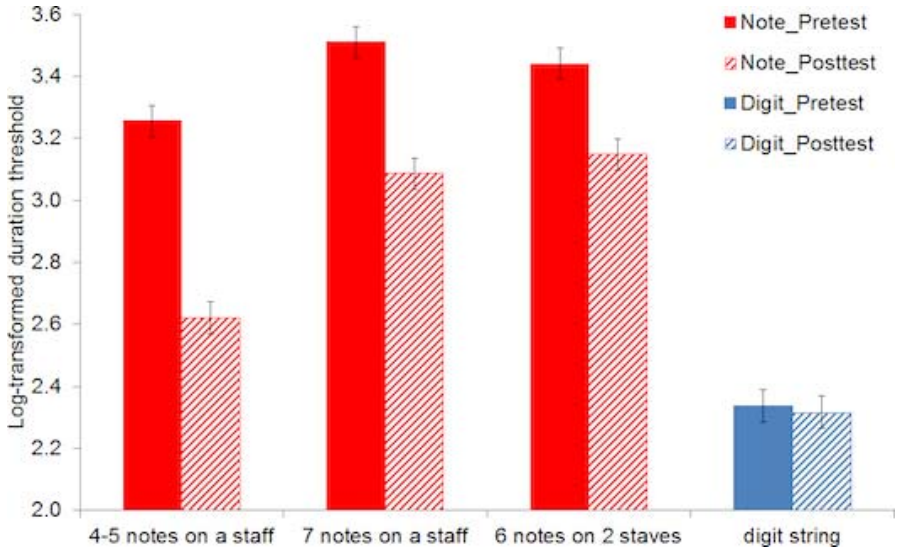

Figure 3. Behavioral pretest and posttest performance. A lower log-transformed duration threshold represents higher perceptual fluency. Error bars represent 95% confidence intervals of the PrePost × Stimulus interaction.

4.05, $p = 0.0002$, and that for experts, 2.51, $t(38) = 11.03$, $p < 0.0001$. At posttest, the average threshold (2.58) was lower than that for novices, $t(40) = 7.20$, $p < 0.0001$, and comparable with that for experts, 2.51, $t(38) = 0.98$, $p = 0.332$.

EEG pretests and posttests

In terms of behavioral performance, performance was high overall (nonparametric sensitivity A = 0.9775; J. Zhang & Mueller, 2005), and 2 × 2 ANOVAs were conducted on the A with PrePost (pretest, posttest) and Stimulus (notes, pseudo-letters) as factors. The main effect of PrePost was significant, $F(1, 19) = 15.001$, $p = 0.001$, $\eta_p^2 = 0.441$, with better performance at posttest than pretest. The main effect of Stimulus was significant, $F(1, 19) = 15.561$, $p = 0.001$, $\eta_p^2 = 450$, with better performance for pseudo-letters than notes. The interaction between PrePost and Stimulus was also significant, $F(1, 19) = 5.775$, $p = 0.027$, $\eta_p^2 = 233$, with a larger improvement from pretest to posttest for notes (from 0.962 to 0.981) than pseudo-letters (from 0.979 to 0.986).

In terms of ERPs, no obvious training effects were observed for the C1 component (Figure 4). A 2 × 2 × 2 repeated-measures ANOVA with PrePost, Stimulus, and Hemisphere as factors was conducted for the average amplitude of the C1 component. The C1 component was more negative in general for pseudo-letters than musical notes as indicated by a main effect of Stimulus, $F(1, 19) = 19.555$, $p < 0.001$, $\eta_p^2 = 0.507$. The interaction between PrePost and Hemisphere was marginally significant, $F(1, 19) = 3.422$, $p = 0.08$, $\eta_p^2 = 0.153$, with more negative C1 at posttest versus pretest in the left hemisphere but less negative C1 at posttest versus pretest in the right hemisphere. No other effects were significant ($ps > 0.088$). The same analyses were also conducted for the C1 component defined with a wider time window (40–100 ms) as in some studies (e.g., G.-L. Zhang et al., 2015), and similarly, no PrePost × Stimulus or PrePost × Stimulus × Hemisphere interaction was significant ($ps > 0.279$).

For the N170, training led to a larger selectivity for musical notes relative to pseudo-letters for the N170 component (Figure 5). We examined the channels P3 and P4 where the N170 showed the most prominent changes on the scalp topography. A similar 2 × 2 × 2 repeated-measures ANOVA with PrePost, Stimulus, and Hemisphere showed a main effect of Stimulus, $F(1, 19) = 9.711$, $p = 0.006$, $\eta_p^2 = 0.338$, with a more negative N170 for notes than pseudo-letters. The main effect of Hemisphere was also significant, $F(1, 19) = 6.201$, $p = 0.022$, $\eta_p^2 = 0.246$, with a more negative N170 for the right than the left hemisphere in general. Most importantly, there was a PrePost × Stimulus interaction, $F(1, 19) = 16.140$, $p < 0.001$, $\eta_p^2 = 0.459$, indicating a larger difference between notes and pseudo-letters at posttest than pretest. No other significant effects were found ($ps > 0.089$).

We also examined the channels P7 and P8, in which the N170 amplitude collapsed across stimuli and pretests/posttests was the largest. It was defined as the average amplitude within a 40-ms time window centering at an average of 190 ms ($SD = 17$ ms) over the posterior channels P7 and P8. A 2 × 2 × 2 repeated-measures ANOVA with PrePost, Stimulus, and Hemisphere as factors was conducted for the average amplitude of the N170 component at P7 and P8. No PrePost × Stimulus or PrePost × Stimulus × Hemisphere interaction was significant ($ps > 0.554$).

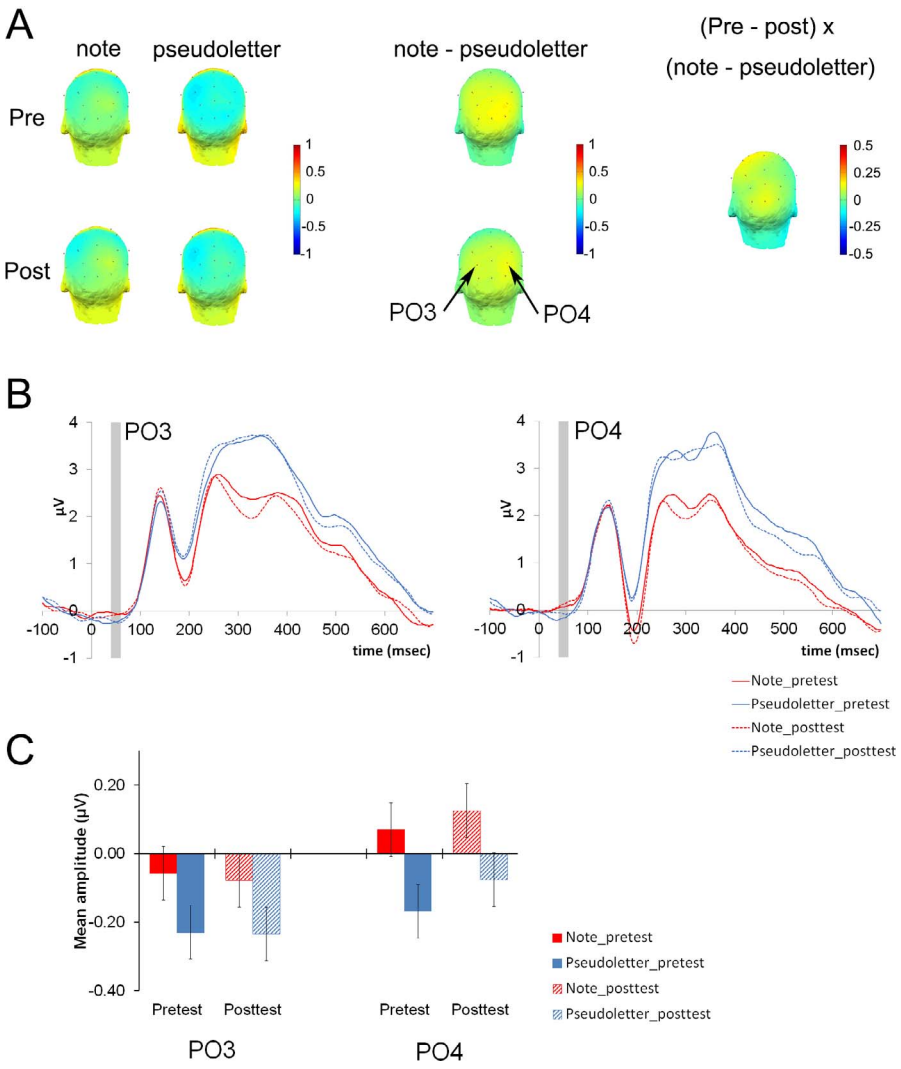

Figure 4. Results for the C1 component in the one-back paradigm at pretests and posttests. (A) Topography maps at 40–60 ms after stimulus onset. (B) Event-related average plots for the channels PO3 and PO4 with the shaded regions highlighting the time window of 40–60 ms. (C) The average amplitude of the C1 component. Error bars represent 95% confidence interval of the PrePost × Stimulus interaction.

In addition, we explored the training effects for the N250 component (Figure 6) given previous reports of changes to this component after fine discrimination training with objects (Scott et al., 2006, 2008). A similar three-way ANOVA was conducted on the average amplitude of the N250 defined at channels P7 and P8 individually (mean peak latency = 299 ms; 40-ms time window). No training effect was found as indicated by the lack of PrePost × Stimulus or PrePost × Stimulus × Hemisphere interactions ($ps > 0.411$).

Discussion

In this study, we asked whether expert-like visual perceptual fluency in recognizing musical notes, created by laboratory training, is sufficient to cause the early electrophysiological responses for musical notes as observed in real-world experts (Y. K. Wong et al., 2014). After 10–26 hr of training, participants successfully acquired visual perceptual fluency comparable with that of real-world experts. The acquisition of expert-like fluency was accompanied by significant changes in the neural responses to musical notes. Whereas the real-world music readers showed neural selectivity to musical notes in both early and late ERP components (C1 and N170; Y. K. Wong et al., 2014), our participants showed changes only to the late ERP component (N170) after training.

Consistent with previous object training studies (Rossion et al., 2002; Scott et al., 2006, 2008), activation changes were found for the musical notes after training. Specifically, training resulted in an

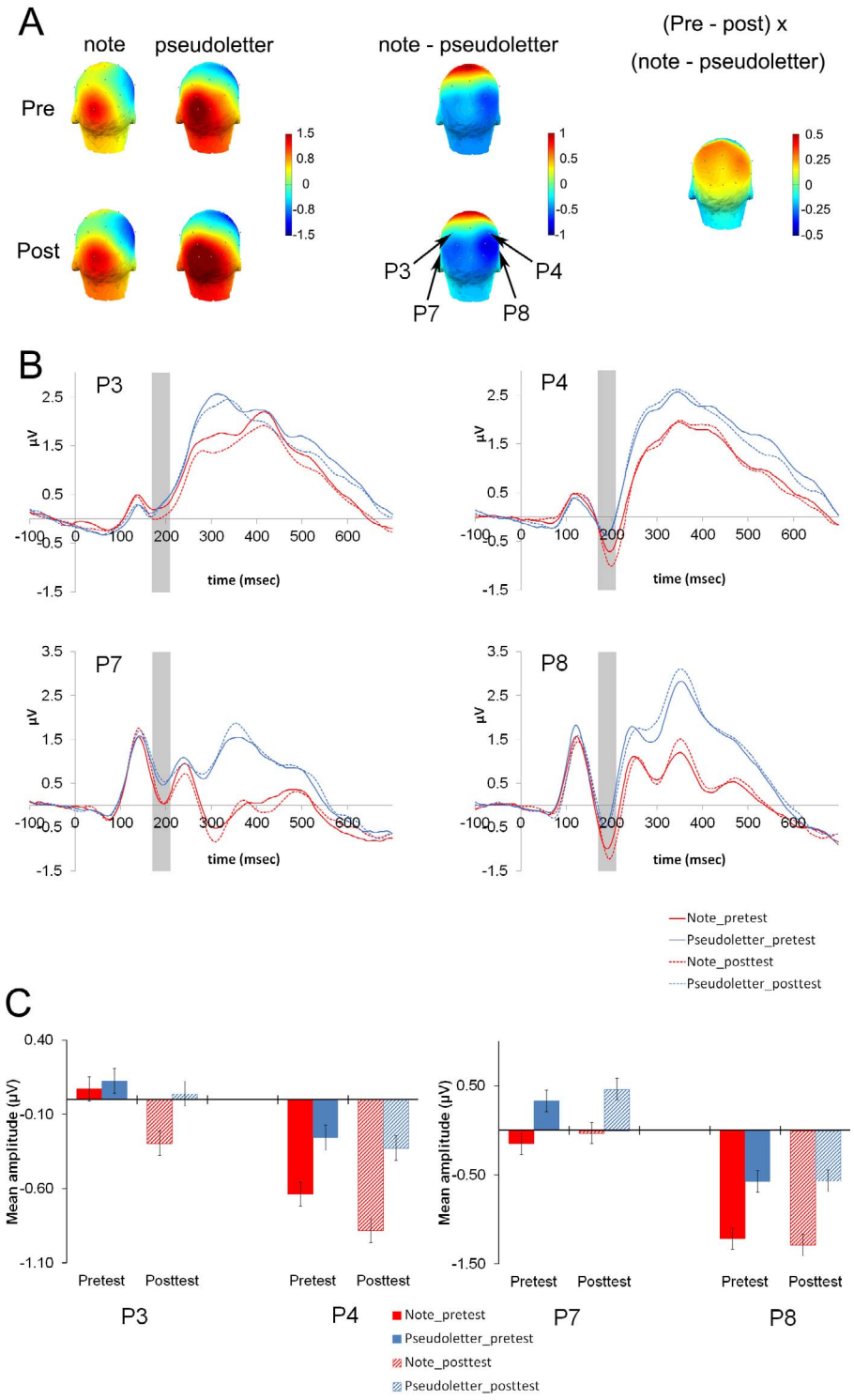

Figure 5. Results for the N170 component in the one-back paradigm at pretests and posttests. (A) Topography maps at 171–211 ms after stimulus onset, where the N170 was most prominent across all posterior channels for all conditions collapsed. (B) Event-related average plots for the channels P3, P4, P7, and P8 with the shaded regions highlighting the time window of 169–209 ms (see Methods). (C) The average amplitude of the N170 component. Note that the same time window was adopted for all participants in the topography and ERP plots in panels A and B while individually defined time windows were identified to compute the average amplitude in panel C. Error bars represent 95% confidence interval of the PrePost × Stimulus interaction.

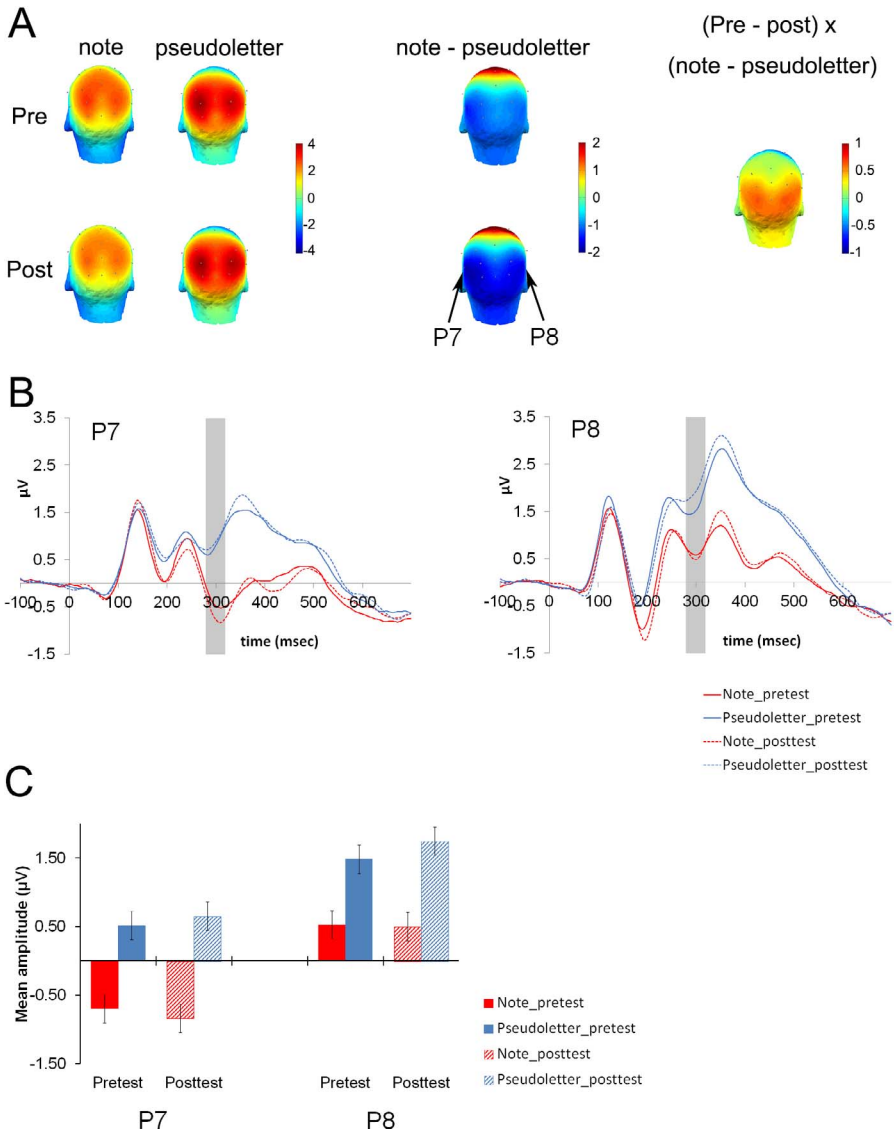

Figure 6. Results for the N250 component in the one-back paradigm at pretests and posttests. (A) Topography maps at 284–324 ms after stimulus onset, where the N250 was most prominent across all posterior channels for all conditions collapsed. (B) Event-related average plots for the channels P7 and P8 with the shaded regions highlighting the time window of 279–319 ms. (C) The average amplitude of the N250 component. Note that the same time window was adopted for all participants in the topography and ERP plots in panels A and B while individually defined time windows were identified to compute the average amplitude in panel C. Error bars represent 95% confidence interval of the PrePost × Stimulus interaction.

increase in the selective response of the N170 component in both hemispheres to musical notes compared with pseudo-letters in the one-back paradigm. This is consistent with the previous findings in which the N170 was larger for musical notes for music reading experts (Y. K. Wong et al., 2014). It is important to note that our training focused solely on the visual aspect of music reading and excluded other crucial aspects of the experience of real-world music readers. For example, classically trained musicians typically learn to associate each musical notes with a pitch name, such as "C" or "doh," depending on the labeling system. In contrast, the current training required participants to discriminate between similar music sequences as visual pattern without labels. Another aspect missing in our current training was the speeded bimanual execution of motor programs corresponding to musical notation when musicians play, sing, or sight-read a piece of music. Still, the training of rapid individuation of multiple notes with high spatial resolution was sufficient to cause changes to the N170 component. This suggests that the selective engagement the higher visual cortex for musical notes in music readers can be explained solely by their visual experience, and response changes in the N170 represent one general neural marker for expert object processing across domains.

As in all training studies, one can question if the training effect found in perceptual fluency and N170 was merely a result of heightened motivation for musical notes compared with control stimuli after training. This is unlikely for several reasons. First, in terms of perceptual fluency, the training effect for notes was large (the average threshold changed from 1,738 ms at pretest to 380 ms at posttest), and it would be difficult to attribute this to heightened motivation alone. Second, in terms of ERP, the training effect was specific to N170 but not to either the earlier component of C1 or the subsequent component of N250. It is difficult to explain this specific training effect with heightened general motivation for notes than control stimuli. Third, according to the motivation explanation, participants could have become more motivated for notes than pseudo-letters especially in the EEG posttest. This would predict a correlation between the training effect in the N170 (PrePost × Stimulus interaction in N170 at P3 and P4 collapsed) and in the one-back task performance. However, no such correlation was found, $r(18) = 0.079$, $p = 0.740$, despite the sufficient variability of the one-back task performance (Spearman–Brown corrected split-half reliability = 0.644). Instead, the training effect in N170 and that in perceptual fluency was marginally significant, $r(18) = 0.381$, $p = 0.098$. This trend was observed even though the two measures were obtained in different testing sessions using different stimuli, suggesting functional significance of the training effect in N170.

Interestingly, changes in the N170 were found only in the one-back paradigm and not in the repetition-suppression paradigm (method and results for the latter in the Supplementary Appendix. Whereas the one-back paradigm tapped onto general activation levels for the category of notes and pseudo-letters, the repetition-suppression paradigm probed sensitivity to differences between different exemplars of notes or between different exemplars of pseudo-letters. The N170 training effect found in the current study may, thus, be more relevant to the better distinction after training between musical notes and pseudo-letters for further visual analyses. Such an interpretation has to be taken with great caution. First, the repetition-suppression paradigm was introduced as an exploratory task, and the primary objective was always the one-back paradigm. The order of the two paradigms was, thus, not counterbalanced across participants, making it more difficult to interpret result differences for the two paradigms. Also, the number of trials analyzed in the repetition-suppression paradigm (200 trials per condition) was much smaller than that in the one-back paradigm (660 trials per condition), so it was more difficult to obtain robust results in the repetition suppression paradigm. Furthermore, the blank duration between the adaptor and the target was set at 100–300 ms. This short interval was meant to increase the repetition-suppression effect, but it also resulted in the carryover of the activations for the adaptor to the target as can be seen in Supplementary Appendix Figures A2 and A3. Last, unfortunately, we presented the target for only 300 ms, making it more difficult to meaningfully compare results for the N250 component with those for other components or for the one-back paradigm. Future studies with more trials devoted to the repetition-suppression paradigm and counterbalanced order of the two paradigms may be worthwhile given that high sensitivity to individual exemplars within an object category is the essence of perceptual expertise (Gauthier & Tarr, 1997; Kovacs et al., 2006; Yue, Tjan, & Biederman, 2006).

The lack of changes to the C1 component as a result of training suggests that the task demands in the training, e.g., visual individuation of multiple visual objects with brief exposure and high spatial resolution, may not be sufficient to cause the selective response for musical notes in the early visual cortex as found in real-world music readers. It is unlikely that the absence of changes in the C1 was caused by insufficiency of visual training or inadequate EEG measurement. First, the training duration (10–26 hr) was two to three times longer than many visual perceptual training studies (Gauthier et al., 1998; Rossion et al., 2002; A. C.-N. Wong et al., 2009). It successfully improved the performance of the participants to a level comparable to that of real-world experts. A similar training of shortened duration (10 hr) was already sufficient to cause reduced crowding that was often associated with early visual processing (Y. K. Wong & Wong, 2016). Second, concerning our EEG measurement, the one-back paradigm used was similar to that used to observe early visual selectivity for musical notes in both fMRI (Y. K. Wong & Gauthier, 2010) and ERP (Y. K. Wong et al., 2014) techniques. Our sample size ($N = 20$) was about two times that used in our prior study with real-world experts with a highly similar EEG task (Y. K. Wong et al., 2014). Moreover, the SNR of our EEG data was satisfactory (Debener et al., 2007; Hu et al., 2010), and the C1 component did show differential responses for musical notes and the pseudo-letter control, demonstrating that the ERP data had sufficient information to differentiate between notes and other images. Furthermore, neural measurements can be more sensitive than behavioral measurements in revealing training effects and condition or participant differences (Hoeft et al., 2011; e.g., Jiang et al., 2007). Given the robust behavioral improvements in the current study, the neural data should have provided sufficient power to observe any potential changes in the C1 component.

There are two potential explanations of the training effects occurring for late but not early ERP compo-

nents. The first possibility is that individuals whose early visual areas are more suited to processing of musical notation would more likely devote themselves to musical training and become a fluent music reader. This explains the observed C1 selectivity for musical notes in real-world experts (Y. K. Wong et al., 2014) but not in trained participants as in the current study. The second, more intriguing possibility concerns the additional nonvisual factors involved in real-life musical training, such as multimodal integration. Musical training is highly demanding in integrating visual, auditory, somatosensory, and motor processes. In Wong & Gauthier (2010) fMRI study, it was found that, even when the task involved only visual judgment of a single musical note, experts automatically showed specialized recruitment of areas involved in auditory, somatosensory, and motor modalities. ERP studies have also shown that C1 responses can be modulated by simultaneously presenting information to two modalities regardless of whether information in the secondary modality is task relevant or not (Fort, Delpeuch, Pernier, & Giard, 2002; Giard & Peronnet, 1999; Karns & Knight, 2008). Therefore, it is possible that extensive experience in reading musical notation, coupled with speeded multimodal processes, causes long-term changes in how the early visual cortex responds to musical notes. Under this hypothesis, it is interesting to consider the case of word-reading expertise, which shares the task demands of fast individuation of multiple letters or words with high spatial resolution and prolonged and highly efficient multimodal integration with pronunciation and writing of the words. In other words, if these task demands are driving the V1 recruitment for musical notes in real-world experts, it should predict that a similar recruitment for words or letters should also be observed in early visual cortex, which is largely feed-forward in nature. Future studies should test this prediction and examine if these task demands are sufficient to explain the V1 selectivity for an object category.

Another intriguing question concerns the correspondence between the C1 and N170 components identified in the current study and in the literature. C1 was typically defined by the activation differences associated with stimuli in the upper versus lower visual field (Clark et al., 1995; Clark & Hillyard, 1996). In the current study, however, all stimuli were presented in the fovea, with small jittering within 0.16° around the fixation. Also, the N170 training effect identified in the current study occurred at the channels P3/P4, different from P7/P8, where the largest N170 amplitude for all conditions collapsed was identified. Therefore, it is reasonable to question whether the current findings apply to the same C1 and N170 components as identified in previous studies. Despite this, it is important to emphasize that, for the purpose of the current study, the conclusion of training effects occurring at later but not earlier ERP would still hold even if it is inconclusive whether the exact C1 and N170 components as identified in other studies were involved.

Conclusions

Our lab training successfully created expert-like performance in visual discrimination of musical notes. The training also resulted in activation changes selective for musical notes in the later (N170) but not earlier (C1) ERP component. A causal relationship is, thus, established between visual experience and high-level visual areas selective for musical notes. The lack of training-induced changes in the early visual areas suggests that additional nonvisual experience may be required to create the full package of musical reading expertise network as observed in real-world fluent music readers.

Keywords: expert object recognition, musical notation, EEG, perceptual learning, perceptual expertise

Acknowledgments

This research was supported by the General Research Fund (14411814) from the Research Grants Council of Hong Kong and the Direct Grant (2021100) from the Chinese University of Hong Kong to A. C.-N. W.

Alan C.-N. Wong, ORCID: 0000-0002-2129-3485; Yetta Kwailing Wong, ORCID: 0000-0002-8243-2047.

Commercial relationships: none.
Corresponding authors: Alan C.-N. Wong; Yetta K. Wong.
Email: alanwong@cuhk.edu.hk; yetta.wong@cuhk.edu.hk.
Address: Department of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; Department of Educational Psychology, Faculty of Education, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

Footnote

[1] We modified the methods used in previous studies (Debener et al., 2007; Hu, Mouraux, Hu, & Iannetti, 2010) and computed the ratio of the standard deviation for the ERP during the 300 ms after stimulus onset and that during the 100 ms before for each participant in

each of the one-back and repetition-suppression paradigms. Twelve SNRs were computed for each participant at each posterior channel of interest (O1/2, PO3/4, P7/8) at pretest and posttest, collapsed across conditions. Data from two participants were discarded because they each had more than 40% of the SNR ratios lower than two. Results were qualitatively the same with data from these two participants included.

References

Bao, M., Yang, L., Rios, C., He, B., & Engel, S. A. (2010). Perceptual learning increases the strength of the earliest signals in visual cortex. *Journal of Neuroscience*, *30*(45), 15080–15084, https://doi.org/10.1523/JNEUROSCI.5703-09.2010.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: Fusiform face area in chess experts and novices. *The Journal of Neuroscience*, *31*(28), 10206–10214.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, *45*, 431–448.

Cao, X., Jiang, B., Li, C., Xia, N., & Jackie Floyd, R. (2015). The commonality between the perceptual adaptation mechanisms involved in processing faces and nonface objects of expertise. *Neuropsychology*, *29*(5), 715–725, https://doi.org/10.1037/neu0000170.

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*, 170–187.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience*, *8*(5), 387–402.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*(Pt 2), 291–307.

Debener, S., Strobel, A., Sorger, B., Peters, J., Kranczioch, C., Engel, A. K., & Goebel, R. (2007). Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: Removal of the ballistocardiogram artefact. *NeuroImage*, *34*(2), 587–597, https://doi.org/10.1016/j.neuroimage.2006.09.031.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001, September 28). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Epstein, R., & Kanwisher, N. (1998, April 9). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Fort, A., Delpeuch, C., Pernier, J., & Giard, M.-H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, *12*, 1031–1039.

Foxe, J. J., Strugstad, E. C., Sehatpour, P., Molholm, S., Pasieka, W., & Schroeder, C. E. (2008). Parvocellular and magnocellular contributions to the initial generators of the visual evoked potential: High-density electrical mapping of the 'C1' component. *Brain Topography*, *21*, 11–21.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, *6*(4), 428–432.

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*(2), 191–197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nat Neurosci*, *2*(6), 568–573, https://doi.org/10.1038/9224.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15/16), 2401–2428.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490.

Goodale, M. A., & Milner, D. A. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, *15*(1), 20–25.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res*, *41*(10–11), 1409–1422.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Reviews Neuroscience*, *27*, 649–677.

Harley, E. M., Pope, W. M., Villablanca, J. P., Mumford, J., Suh, R., Mazziotta, J. C., ... Engel, S. A. (2009). Engagement of fusiform cortex and disengagement of lateral occipital cortex across the acquisition of radiological expertise. *Cerebral Cortex*, *19*(11), 2746–2754.

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., ... Gabrieli, J. D. E. (2011). Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences, USA*, *108*(1), 361–366, https://doi.org/10.1073/pnas.1008950108.

Hu, L., Mouraux, A., Hu, Y., & Iannetti, G. D. (2010). A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. *NeuroImage*, *50*(1), 99–111, https://doi.org/10.1016/j.neuroimage.2009.12.010.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*(1), 215–243.

James, K. H., & Atwood, T. O. (2009). The role of sensorimotor learning in the perception of letter-like forms: Tracking the causes of neural specialization for letters. *Cognitive Neuropsychology*, *26*, 91–110.

James, K. H., James, T. W., Jobard, G., Wong, A. C. N., & Gauthier, I. (2005). Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective & Behavioral Neuroscience*, *5*(4), 452–466, https://doi.org/10.3758/CABN.5.4.452.

Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Clinical Neurophsyiology*, *10*, 371–375.

Jeffreys, D. A., & Axford, J. G. (1972). Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. *Experimental Brain Research*, *16*(1), 1–21.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*(6), 891–903, https://doi.org/10.1016/j.neuron.2007.02.015.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosc.*, *17*, 4302–4311.

Karns, C. M., & Knight, R. T. (2008). Intermodal auditory, visual and tactile attention modeulates early stages of neural processing. *Journal of Cognitive Neuroscience*, *21*(4), 669–683.

Kovacs, G., Zimmer, M., Banko, E., Harza, I., Antal, A., & Vidnyanszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cereb. Cortex*, *16*(5), 742–753.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*, 467–477.

Lochy, A., Zimmermann, F. G. S., Laguesse, R., Willenbockel, V., Rossion, B., & Vuong, Q. C. (2018). Does extensive training at individuating novel objects in adulthood lead to visual expertise? The role of facelikeness. *Journal of Cognitive Neuroscience*, *30*(4), 449–467.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., ... Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A*, *92*(18), 8135–8139.

Martinez, A., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., & Kennedy, W. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*(4), 364–369.

Maurer, U., Zevin, J. D., & McCandliss, B. D. (2008). Left-lateralized N170 effects of visual expertise in reading: Evidence from Japanese syllabic and logographic scripts. *Journal of Cognitive Neuroscience*, *20*(10), 1878–1891.

Mongelli, V., Dehaene, S., Vinckier, F., Peretz, I., Bartolomeo, P., & Cohen, L. (2016). ScienceDirect Music and words in the visual cortex: The impact of musical expertise. *CORTEX*, *86*, 260–274, https://doi.org/10.1016/j.cortex.2016.05.016.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience*, *26*(43), 11187–11196, https://doi.org/10.1523/JNEUROSCI.1873-06.2006.

Nakada, T., Fujii, Y., Suzuki, K., & Kwee, I. L. (1998). "Musical brain" revealed by high-field (3 Tesla) functional MRI. *Neuroreport*, *9*, 3853–3856.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

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*(3), 250–257, https://doi.org/10.1111/1467-9280.00446.

Rossion, B., Kung, C.-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences, USA*, *101*, 14521–14526.

Schön, D., & Besson, M. (2002). Processing pitch and duration in music reading: A RT-ERP study. *Neuropsychologia*, *40*, 868–878.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience*, *18*(9), 1453–1465, https://doi.org/10.1162/jocn.2006.18.9.1453.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, *1210*, 204–215, https://doi.org/10.1016/j.brainres.2008.02.054.

Sergent, J., Zuck, E., Terriah, S., & MacDonald, B. (1992, July 3). Distributed neural network underlying musical sight-reading and keyboard performance. *Science*, *257*, 106–109.

Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage*, *20*, 71–83.

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., & Cohen, L. (2011). Specialization for written words over objects in the visual cortex. *Neuroimage*, *56*(1), 330–344.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43–47.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: The MIT Press.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Wong, A. C.-N., Jobard, G., James, K. H., James, T. W., & Gauthier, I. (2008). Expertise with characters in alphabetic and nonalphabetic writing systems engage overlapping occipito-temporal areas. *Cognitive Neuropsychology*, *26*(1), 111–127, https://doi.org/10.1080/02643290802340972.

Wong, A. C. N., Gauthier, I., Woroch, B., DeBuse, C., & Curran, T. (2005). An early electrophysiological response associated with expertise in letter perception. *Cognitive, Affective & Behavioral Neuroscience*, *5*(3), 306–318, https://doi.org/10.3758/CABN.5.3.306.

Wong, A. C. N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a ziggerin expert - but which type? *Psychological Science*, *20*(9), 1108–1117.

Wong, A. C.-N., Palmeri, T. J., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS One*, *4*(12), e8405.

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2012). The nature of experience determines object representations in the visual system. *Journal of Experimental Psychology: General*, *141*(4), 682–698.

Wong, Y. K., & Gauthier, I. (2010). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, *22*(4), 695–713.

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*(8), 1629–1643, https://doi.org/10.1162/jocn_a_00616.

Wong, Y. K., & Wong, A. C.-N. (2016). Music-reading training alleviates crowding with musical notation. *Journal of Vision*, *16*(8):15, 1–9, https://doi.org/10.1167/16.8.15. [PubMed] [Article].

Xue, G., & Poldrack, R. A. (2007). The neural substrates of visual perceptual learning of words: Implications for the visual word form area hypothesis. *Journal of Cognitive Neuroscience*, *19*(10), 1643–1655.

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *46*, 3802–3811, https://doi.org/10.1016/j.visres.2006.06.017.

Zhang, G.-L., Li, H., Song, Y., & Yu, C. (2015). ERP C1 is top-down modulated by orientation perceptual learning. *Journal of Vision*, *15*(10):8, 1–11, https://doi.org/10.1167/15.10.8. [PubMed] [Article].

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203–212, https://doi.org/10.1007/s11336-003-1119-8.