

Perceptual Expertise and Top-Down Expectation of Musical Notation Engages the Primary Visual Cortex

Yetta Kwailing Wong¹, Cynthia Peng², Kristyn N. Fratus²,
Geoffrey F. Woodman², and Isabel Gauthier²

Abstract

■ Most theories of visual processing propose that object recognition is achieved in higher visual cortex. However, we show that category selectivity for musical notation can be observed in the first ERP component called the C1 (measured 40–60 msec after stimulus onset) with music-reading expertise. Moreover, the C1 note selectivity was observed only when the stimulus category was blocked but not when the stimulus category was randomized. Under blocking, the C1 activity for notes

predicted individual music-reading ability, and behavioral judgments of musical stimuli reflected music-reading skill. Our results challenge current theories of object recognition, indicating that the primary visual cortex can be selective for musical notation within the initial feedforward sweep of activity with perceptual expertise and with a testing context that is consistent with the expertise training, such as blocking the stimulus category for music reading. ■

INTRODUCTION

It is widely accepted that object recognition critically depends on activity reaching later stages of the visual hierarchy in anterior inferotemporal cortex (DiCarlo, Zoccolan, & Rust, 2012; Riesenhuber & Poggio, 1999). In V1, cells are typically considered local detectors that do not differentiate between objects, noise patterns, and textures (Grill-Spector & Malach, 2004; Malach et al., 1995), and V1 activity remains high even when object recognition performance drops to chance level (Grill-Spector, Kushnir, Hendler, & Malach, 2000). In contrast, higher visual areas, such as the lateral occipital cortex, respond selectively to objects compared with noise patterns or scrambled objects (Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995), and such activity correlates with behavioral object recognition (Grill-Spector et al., 2000). In the fusiform gyrus and the parahippocampal gyrus, various small and focal regions are functionally specialized for different object categories, including faces (Kanwisher, McDermott, & Chun, 1997), body parts (Peelen & Downing, 2007; Downing, 2001), buildings and scenes (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998), and letters and words (James, James, Jobard, Wong, & Gauthier, 2005; Cohen et al., 2000). Regions in the ventral temporal cortex are often recruited for objects of expertise, with neural activity predicting behavioral performance for objects of expertise (McGugin, Gatenby, Gore, & Gauthier, 2012; Wong, Folstein, & Gauthier, 2012; Wong, Palmeri, Rogers, Gore,

& Gauthier, 2009; Gauthier, Curby, Skudlarski, & Epstein, 2005; Xu, 2005; Gauthier & Tarr, 2002; Gauthier, Skudlarski, Gore, & Anderson, 2000). Therefore, object recognition is thought to be achieved in later processing stages along the visual hierarchy, after combining local and featural information from early visual cells (DiCarlo et al., 2012; DiCarlo & Cox, 2007; Kourtzi & DiCarlo, 2006; Grill-Spector & Malach, 2004; Riesenhuber & Poggio, 1999).

Contrary to these theoretical predictions, bilateral V1 activity differentiates musical notation from other object categories (e.g., letters and symbols) among music-reading experts but not in novices (Wong & Gauthier, 2010a). This is unlikely because of the general attention-grabbing nature of objects of expertise, because similar results are not found in other domains of expertise, such as faces (Ishai, 2008; Kanwisher et al., 1997), Roman letters (James & Gauthier, 2006), and other objects (van der Linden, van Turenout, & Indefrey, 2010; Wong, Palmeri, Rogers, et al., 2009; van der Linden, Murre, & van Turenout, 2008; Moore, Cohen, & Ranganath, 2006; Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006; Yue, Tjan, & Biederman, 2006; Cohen et al., 2000; Gauthier et al., 2000; Gauthier, Williams, Tarr, & Tanaka, 1998).¹ Although such expertise-dependent object selectivity in V1 is largely unexpected, it is consistent with findings that V1 activity is highly alterable with short visual perceptual training (Pourtois, Rauss, Vuilleumier, & Schwartz, 2008; Kourtzi, Betts, Sarkheil, & Welchman, 2005; Sigman et al., 2005; Furmanski, Schluppeck, & Engel, 2004; Schwartz, Maquet, & Frith, 2002; Crist, Li, & Gilbert, 2001; Schoups, Vogels, Qian, & Orban, 2001) and that V1 receives detailed information about object categories through feedback

¹City University of Hong Kong, ²Vanderbilt University

mechanisms (Williams et al., 2008) and evidence that the cortical thickness is significantly greater for musicians than nonmusicians in V1 (Bermudez, Lerch, Evans, & Zatorre, 2009).

In this study, we investigate the time course of the V1 object selectivity for musical notation to address its underlying mechanisms. Experiment 1 asked whether the V1 object selectivity depends on feedback from higher visual areas that includes information about object categories (Williams et al., 2008; Gilbert & Sigman, 2007; Sigman et al., 2005; Tong, 2003; Lee, 2002; Lee, Yang, Romero, & Mumford, 2002) or whether this selectivity can be observed within the initial sweep of feedforward visual activity among V1 cells (Bao, Yang, Rios, He, & Engel, 2010; Pourtois et al., 2008; Schoups et al., 2001). To tease apart these two alternatives, music-reading experts and novices performed a 1-back task with blocks of single musical notes, Roman letters, or pseudoletters. The stimuli and tasks were similar to those of a prior fMRI study (Wong & Gauthier, 2010a). Using ERPs, we tested whether the expertise-dependent selectivity for musical notation can be observed in the early part of the C1 component (40–60 msec; Rauss, Schwartz, & Pourtois, 2011; Bao et al., 2010; Proverbio & Adorni, 2009; Foxe et al., 2008; Kelly, Gomez-Ramirez, & Foxe, 2008; Pourtois et al., 2008; Stolarova, Keil, & Moratti, 2006; Pourtois, Grandjean, Sander, & Vuilleumier, 2004; Martinez et al., 1999; Clark & Hillyard, 1996; Clark, Fan, & Hillyard, 1995; Jeffreys & Axford, 1972). The C1 component is the first visual ERP component that typically onsets 40–60 msec poststimulus, peaks 80–100 msec poststimulus, and is largest at posterior electrode sites (i.e., PO3 and PO4; see Luck, 2005; Clark et al., 1995). The polarity of the C1 is not fixed but changes as a function of stimulus position in the visual field. Specifically, the C1 is positive for stimulus presented in the lower visual field and is negative for stimulus presented in the upper visual field (Clark et al., 1995), consistent with the anatomy of the calcarine fissure in which the lower visual field is represented on its upper bank whereas the upper visual field is represented on its lower bank. Although the C1 is typically measured with stimuli presented in the upper or lower visual fields (Rauss et al., 2011; Proverbio & Adorni, 2009; Kelly et al., 2008; Pourtois et al., 2004; Clark et al., 1995; Jeffreys & Axford, 1972), it can also be measured with foveally presented stimuli as in the current design (Foxe et al., 2008; Giard & Peronnet, 1999). We chose to present stimuli at the fovea because parafoveal presentation would both depart from the fMRI study in which we observed V1 activation in music-reading experts (Wong & Gauthier, 2010a) and could produce large behavioral differences in novice versus expert performance (Wong & Gauthier, 2012) that could complicate the interpretation of the results (e.g., more small eye movements toward the stimuli in one group). We focused on the early part of the C1 because, in this time window, the activity is too early to be explained by extrastriate activity in the

next P1 component (onset around 60–90 msec) or in higher-level visual cortex (Bao et al., 2010; Foxe et al., 2008; Luck, 2005; Martinez et al., 1999; Clark & Hillyard, 1996; Clark et al., 1995), and most of the activated cells are in the LGN and V1 during this latency (Schmlesky et al., 1998). Note that the waveform of the C1 may not always be prominent, especially for stimuli presented at fovea, because its amplitude is small and overlaps with the subsequent P1 component (Luck, 2005). However, our goal was to test for expertise effects within the typical temporal window of the C1 (onset around 40–60 msec). Although we predicted a group difference in C1 amplitude if the V1 object selectivity for musical notation indeed occurs during the feedforward sweep of neural activity, we did not have specific predictions about the direction of the C1 changes because of perceptual experience.

Our results indicate that, in the early C1, category selectivity for musical notation was higher in experts compared with novices. Experiment 2 replicated this effect and found that the category selectivity for notes emerges in V1 only when the stimulus category was blocked (as in Experiment 1 and in our prior fMRI work; Wong & Gauthier, 2010a) but not when the stimulus category was randomized. The C1 category selectivity for notes in the blocked condition predicted individual music-reading ability. Experiment 3 extended the novel finding of blocking dependent expertise effect from the neural level to the behavioral level. Specifically, experts showed a performance advantage judging note-like stimuli when the stimulus orientation was blocked or regularly alternating, whereas the performance advantage was abolished when the stimulus orientation was randomized. Altogether, our results suggest that V1 is selective for an object category of expertise, a finding not predicted by existing theories of object recognition.

METHODS

Experiment 1

Participants

Twenty-two participants (including the author YW) were recruited from Vanderbilt University and the Nashville community for cash payment. All participants reported their experience in music reading and rated their music-reading ability (1 = *do not read music at all*; 10 = *expert in music reading*), and their handedness was assessed with the Edinburgh Handedness Inventory (Oldfield, 1971). Eleven participants (including the author) who have at least 10 years of music-reading experience and/or consider themselves music-reading experts were recruited in the expert group (seven women and four men; mean age = 21.7 years, *SD* = 3.0 years; nine right-handed and one left-handed), with 14.5 years of music-reading experience and a self-rating score of 9.45 on average. The main instruments practiced by these 11 experts were piano (6), violin (2), double bass (1), flute (1),

and guitar (1). They had been playing it for 14.5 years on average (range = 9–23 years, $SD = 3.9$ years). Eleven participants who reported being unable to read music were recruited in the novice group (six women and five men; mean age = 25.0 years, $SD = 6.1$ years; nine right-handed and one left-handed), with 0.45 years of music-reading experience and a self-rating score of 1.54 on average. All novices were nonmusicians except for a drummer and a guitarist, who played without reading musical notes. All participants had normal or corrected-to-normal vision, including far and near acuity (20/20 or 20/30) and functional contrast sensitivity. All gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

Stimuli and Design

The experiment was conducted on a Mac Pro using Matlab (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Stimuli were presented on the screen of a calibrated Sony GDM-FW900 monitor (1024H \times 768V resolution; 100 Hz frame rate; 34.63 cd/m^2 mean luminance) in a dimly illuminated room. Spectral power distributions of the red (R), green (G), and blue (B) phosphors were measured using a spectroradiometer (Ocean Optics, Dunedin, FL, Model USB4000). The relative light level of each gun at every digital value (256; 2^8 levels) was measured with a Minolta colorimeter (Osaka, Japan, model CA-100). Stimuli were presented with a visual angle of approximately $1.28^\circ \times 1.28^\circ$ at a viewing distance of about 114 cm from the monitor.

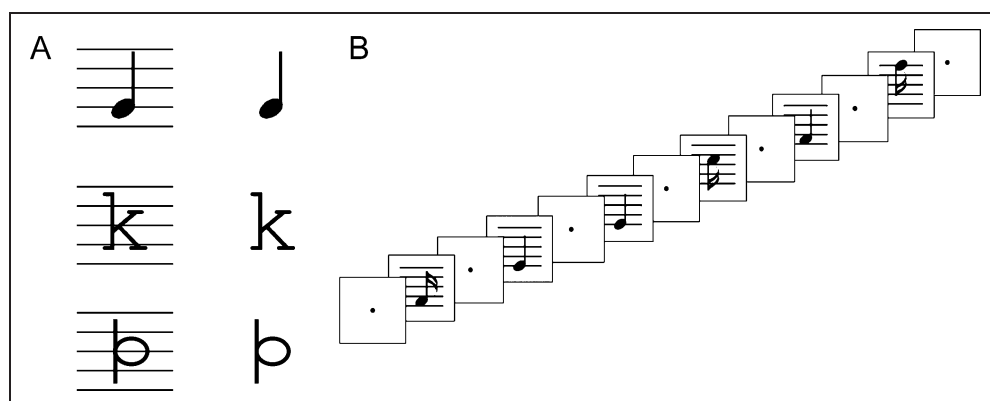
There were 18 black-and-white images in each of three object categories (musical notes, Roman letters, and pseudoletters; Figure 1). The 18 musical notes were generated in Matlab and were nine different notes (ranging from the “E” on the bottom line to the “F” on the top line) in two different time values, including quarter notes (a closed circle) and sixteenth notes (a closed circle with two tails). The Roman letters included 18 uppercase letters (excluding A, E, I, J, O, T, X, and Z) in the Courier font. The 18 pseudoletters were created by various combina-

tions of the parts from the Roman letters with comparable complexity (Wong, Gauthier, Woroch, Debuse, & Curran, 2005). The stimuli in all categories were shown either on a five-line staff or not. For no-staff stimuli, six musical notes were used, including a quarter note (a closed circle), an eighth note (a closed circle with one tail), and a sixteenth note (a closed circle with two tails), either pointing upward or downward. Six Roman letters and six pseudoletters were drawn from the set to keep the stimulus variability similar across stimulus conditions, and the chosen letters and pseudoletters were counterbalanced across participants. The mean luminance and mean contrast (Weber contrast) were matched across the three object categories. One-way ANOVAs on stimulus category revealed that the luminance and the contrast between different object categories for the with-staff conditions or for the no-staff conditions were similar, with all $F_s(1, 51) < 1$.

Following our prior fMRI study (Wong & Gauthier, 2010a), a 1-back task was used, in which each of the six object categories (notes, letters, and pseudoletters, with or without staff) was presented in blocks of six trials. Each block began with a black fixation dot at the center of the screen for 500 msec, followed by six trials, each with a stimulus presented for 700 msec, and then the black fixation dot presented for a randomized period of 250–450 msec. The fixation dot then turned gray for 2 sec, then black for 200 msec, cuing the start of the next block. Participants were required to press a key on a gamepad (with the right thumb) as fast as possible when they detected a repeat of the stimulus. Participants were instructed to maintain fixation throughout the whole block and were encouraged to blink only during the gray dot period.

For the note condition, the stimulus order was constrained to reduce task difficulty for novices (Wong & Gauthier, 2010a). That is, for notes on the staff, consecutive stimuli always pointed to different orientations or had different number of tails unless they were repeated. For notes with no staff, consecutive stimuli always pointed to different orientations unless they were repeated. Participants were explicitly informed about these constraints.

Figure 1. Sample stimuli and example of trial sequence in Experiment 1. (A) Examples of the single notes (top), Roman letters (middle), and pseudoletters (bottom) in the ERP study either on an identical five-line staff (left column) or not (right column). (B) The 1-back task used in the ERP study, in which participants were required to press a key as fast as possible when they detected an immediate repeat of the stimulus.



The stimuli were spatially jittered for five pixels from the center of the screen in all four directions randomly to reduce visual adaptation (Wong & Gauthier, 2010a). There were 720 trials for each of the six object categories, including 60 repeated trials (repeat rate = 8.3%). The order of the blocks for different object categories was counterbalanced. The trials were divided into eight runs, and participants were encouraged to take frequent rests (provided about every 4.5 min of testing). Participants received feedback on accuracy and RT on the screen every 30 blocks of trials and were given constant verbal feedback on eye movements. They were given 72 trials for practice before the test. The whole experiment took around 3 hr, including the setup of the electrode cap for EEG recording.

Recording and Analysis

The EEG was recorded from tin electrodes held on the scalp by an elastic cap (Electrocap International, Eaton, OH). A subset of the International 10/20 System sites was used (F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, PO3, PO4, O1, and O2) in addition to nonstandard sites OL (halfway between O1 and T5) and OR (halfway between O2 and T6). The right mastoid electrode served as the reference site. The signals were re-referenced offline to the average of the left and the right mastoids (Nunez, 1981). The EOG was recorded using electrodes positioned 1 cm lateral to the external canthi to measure horizontal eye movements and with an electrode beneath the left eye, referenced to the right mastoid, to measure vertical eye movements and blinks. The EEG and EOG were amplified by an SA Instrumentation amplifier with a gain of 20,000 and a band-pass filter of 0.01–100 Hz. The amplified signals were digitized at 250 Hz by a PC-compatible computer and averaged offline. Trials associated with behavioral responses (all repeated trials) or those accompanied with artifacts (eye movement or blocking) were excluded from the averages.

Data analysis was performed with ERPSS (The Event-Related Potential Software System; 1993), Matlab, and the EEGLAB toolbox (Version 8.0). For ocular artifact rejection, a two-step procedure was used that has been described previously (Woodman & Luck, 2003). Briefly, the cross covariance between the single-trial EOG waveform and a 100-msec step function was computed, and trials with maximum covariance exceeding a certain threshold were rejected. The threshold was adjusted for each participant according to visual inspection of the EOG waveforms. The averaged horizontal EOG (HEOG) waveforms were used to reject any participants with systematic unrejected eye movements. Trials with blocking (saturated activity) were rejected if the signal was at the maximum or minimum value for 20 msec or consistently hovering around the maximum or minimum value (>40 data points in a 1-sec time window). One expert and one novice with more than 25% of the trials rejected were

excluded from all analyses. On average, 9.9% and 10.7% of the trials were rejected for the expert group and novice group, respectively. The ERPs were baseline-corrected with respect to 200 msec prestimulus interval (except the analyses for the contingent negative variation [CNV]). For each ERP component, the average scalp voltage was computed for each stimulus condition within the corresponding time window. Grand-averaged waveforms were low-pass filtered (cutoff at 35 Hz with an *SD* of 6 msec) for presentation purposes only. All analyses were performed on the unfiltered data.

Perceptual Fluency

All participants except Y. W. performed the perceptual fluency task to quantify individual music-reading ability (Wong & Gauthier, 2010a, 2010b, 2012). The task used a sequential matching paradigm with music sequences containing four notes each. On each trial, a fixation cross was presented at the center of the screen for 200 msec, followed by a 500-msec premask, a target four-note sequence for a varied duration and, after a 500-msec postmask, 2 four-note sequences appeared side-by-side, one identical to the first sequence and the other with one of the notes shifted by one step (randomly chosen out of the four notes, with the up/down shifts counterbalanced). Participants had to select the matching sequence by key press. The perceptual threshold was estimated using QUEST (Psychtoolbox; Watson & Pelli, 1983), as the duration of the target sequence required to keep performance at 80% accuracy. Sequences were randomly generated using notes ranging from the note below the bottom line (a “D” note) to the note above the top line (a “G” note). Contrast for all the stimuli was lowered by about 60% to avoid a ceiling effect. The threshold was measured four times, each with 40 trials, and the thresholds were averaged.

To control for individual differences not specifically tied to expertise with notes, perceptual fluency for four-letter strings was also measured in an identical procedure. The four-letter strings were randomly generated with 11 letters: b, d, f, g, h, j, k, p, q, t, y. These letters were selected because they contain parts extending upward or downward, similar to musical notation. To create the distractor string, one of the four letters was chosen (counterbalanced across stimuli) and replaced by a different letter randomly drawn from the set. The string stimuli were also shown with the same lowered contrast as the note sequences.

Experiment 2

Participants

Twenty-two participants were recruited from Vanderbilt University and the Nashville community for cash payment. The participants were assigned into the two groups based on similar criteria as Experiment 1, except an additional

criterion that their performance in the perceptual fluency task was in the same range as that for Experiment 1 (with a duration threshold of <500 msec for experts and <1650 msec for novices). Ten participants (seven women and three men; mean age = 20.6 years, $SD = 2.07$ years) were included in the expert group with 12.8 years of music-reading experience and a self-rating score of 8.90 on average. There were six pianists, three violinists, and one singer among the experts, and they had been playing their main instrument for 13.4 years on average (range = 9–21 years, $SD = 4.0$ years). Twelve participants (four women and eight men; mean age = 25.3 years, $SD = 4.68$ years) were included in the novice group with 1.67 years of experience and a self-rating score of 1.75 on average. All novices were nonmusicians, except for two guitarists and a cello player who did not read musical notation efficiently. All participants had normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

Stimuli and Design

The stimuli and design were identical to Experiment 1 except for the following. First, the stimuli were always shown on a five-line staff. Second, the stimulus category was kept the same (“blocked”) or pseudorandomized (“randomized”) within each block of trials. The “blocked” condition was identical to that of Experiment 1. For the “randomized” condition, the stimulus category was pseudorandomized, such that the stimulus category of consecutive images was always different except for the repeated trials. The order of the two blocking conditions was counterbalanced across participants, such that half of the participants performed in the “blocked” condition before the “randomized” condition and half performed in the reversed order. Participants were explicitly told about the blocking manipulations and were instructed to press a key on a gamepad (with the right thumb) as fast as possible when they detected a repeat of the stimulus regardless of the blocking conditions. The number of “same” trials was the same for the blocked and the randomized conditions.

The setup, recording, and analysis procedures of the EEG was identical to that in Experiment 1. On average, 13.5% and 18.8% of the trials were rejected for the expert group and novice group, respectively, and the rejection rate was similar across groups ($p > .1$).

Experiment 3

Participants

Forty-three participants were recruited from Vanderbilt University for course credits or cash payment. One participant did not complete the experiment, and another was excluded from the analyses because his note threshold

was 12 SD larger than that for the other participants. This resulted in a sample of 22 women and 19 men, average age of 21.8 years old ($SD = 2.6$ years), all right-handed except for two left-handed and one ambidextrous participants. Thirteen of these participants with 10 or more years of music-reading experience were included in the “high-expertise” group, whereas the other 28 participants with 5 or less years of music-reading experience were included in the “low-expertise” group. All gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

Stimuli and Design

The experiment was conducted on Mac Mini using Matlab (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). On the stimuli, a black circle was positioned on a set of staff lines, the edges of which were gradually blended to the white background using Photoshop (see sample stimuli in Figure 4A). Ten staff lines were used instead of five (typical for musical notation), such that experts could not easily name the musical notes. The black circle was either on a line (bisected one of the lines) or off a line (in the space between two of the lines). The position of the black circle was centered at the middle or at either of the four quadrants of the image to create 10 images for the “on” or “off” condition such as to increase the position variability of the circles. Vertical versions of all stimuli were created by rotating the stimuli by 90°. The contrast of all images was reduced by 50% to avoid ceiling performance. A visual mask was created with overlapping lines of horizontal, vertical, and oblique orientations packed with randomly positioned black circles.

On each trial, a central fixation cross was presented for 500 msec, followed by the stimulus for 66 msec and a mask for 500 msec. Participants were asked to judge if the black circle was on or off a line by key press, emphasizing accuracy without any time limit. These stimuli were presented in two blocks (with 80 trials per block) in three different conditions: In the Blocked condition (one block for each orientation), all trials were in the same orientation; in the Random condition (two blocks), horizontal and vertical trials were presented in a randomized order; in the Predictable condition (two blocks), trials always alternated between vertical and horizontal lines.

RESULTS

Eleven music-reading experts and 11 novices participated in Experiment 1. A separate perceptual fluency test confirmed that experts perceived music sequences faster than novices (for experts: mean = 341.6 msec, $SD = 111.4$ msec; for novices: mean = 1098.0 msec, $SD = 330.4$ msec), but their perceptual ability for letter strings were similar (for experts: mean = 194.4 msec, $SD = 97.6$ msec; for novices: mean = 232.9 msec, $SD =$

131.6 msec); the interaction of Experts/Novice Group \times Music/Letter Stimuli was significant, $F(1, 18) = 50.3, p \leq .0001, \eta_p^2 = 0.74$.

In a 1-back task, participants were presented with blocks of single musical notes, Roman letters, or pseudo-letters, either on a five-line staff or not (Figure 1A). The average luminance and contrast of the stimuli were matched across categories (see Methods). Participants were required to press a key as fast as possible when they detected an immediate repeat of the stimulus (Figure 1B). One expert and one novice with more than 25% of the trials rejected for artifacts were excluded from all analyses. Behavioral performance on the 1-back task was similar across the two groups ($F_s < 1$ for the Group \times Stimulus interaction effects), replicating the findings in the prior fMRI study (Wong & Gauthier, 2010a) and indicating that this task engaged the two groups similarly.

Category-selective C1 Effect With Expertise

The early C1 (40–60 msec) was examined to test whether the category selectivity for notes was generated in V1 within the initial feedforward sweep of activity. As shown in Figure 2A and B, the C1 category selectivity for notes with staff showed maximal group differences along the posterior parietal midline recording sites, consistent with the typical topographic distribution of the C1 (Luck, 2005; Clark et al., 1995; Jeffreys & Axford, 1972).

At PO3/PO4 where the effect was maximal, the C1 was more positive for notes than for pseudoletters in experts but not in novices (Figure 2C). In the Expert/Novice Group \times Note/Pseudoletter Stimulus \times Left/Right Hemisphere ANOVA, this observation was confirmed by a significant Group \times Stimulus interaction, $F(1, 18) = 7.05, p = .016, \eta_p^2 = 0.28$ (after Greenhouse–Geisser correction), and subsequent post hoc pairwise comparisons with the Scheffé tests at $p < .05$. This interaction effect was independent of hemisphere ($p > .6$) and was observed within the left hemisphere at PO3 ($p = .0084$) and the right hemisphere at PO4 ($p = .040$).²

Testing Alternatives to the Category Selectivity for the C1

First, prestimulus activity was examined to test if the expertise difference was already observed before the stimulus onset. The same Group \times Stimulus \times Hemisphere ANOVA was performed on the voltage at the same channels (PO3/PO4) in each of the 20-msec time bins between –80 msec and 0 msec. The Group \times Stimulus interaction effect did not reach significance for any of the time bins (all $p_s > .19$), suggesting that any Group \times Stimulus interaction on the voltage was negligible before stimulus onset.

Second, can the category selectivity for notes on the C1 be explained by group differences in terms of eye movement artifacts? The Group \times Stimulus ANOVA be-

tween the 40 and 60 msec window was performed on the voltage at the vertical (VEOG) and the horizontal (HEOG) eye channels. The Group \times Stimulus interaction was not significant for the VEOG ($p = .13$) or the HEOG ($p = .28$), suggesting that eye movements cannot account for the C1 effect.

Third, if the C1 effect is driven by a stronger attentional bias to notes in experts, a general attentional boost should not be limited to the early C1 (40–60 msec), but also be observable in the subsequent P1 component (60–120 msec). However, the same Group \times Stimulus \times Hemisphere ANOVA on the P1 on the same channels did not reveal any Group \times Stimulus interaction ($F < 1$), suggesting that the C1 category selectivity for notes is not driven by an attentional difference.

Fourth, although the five-line staff was shared by all stimulus categories and unlikely explains the C1 effect, we tested whether the C1 category selectivity for notes is dependent on the five-line staff. As shown in Figure 2C and D, the C1 effect was highly similar regardless of the presence of the staff. The Group \times Stimulus \times Hemisphere \times Staff presence ANOVA confirmed that no effects involving the factor of staff reached significance (all $F_s < 1$), whereas a Group \times Stimulus interaction was observed, $F(1, 18) = 6.31, p = .022, \eta_p^2 = 0.26$, with a more positive C1 for notes than for pseudoletters in experts, but not in novices (Scheffé tests, $p < .05$).

Lastly, given a blocked design, participants could anticipate the stimulus category after seeing the first image of each block, leading to a slow negative potential in the frontocentral region called the CNV (Luck, 2005; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). Did the C1 category selectivity merely reflect expertise effects on the CNV that might have already existed before stimulus onset?

To address this question, we first confirmed that expertise effects on the CNV (–200 to 0 msec) were observed. The CNV followed a typical topographic distribution around the Cz (Rose, Verleger, & Wascher, 2001; Travis, Tecce, & Guttman, 2000; McEvoy, Smith, & Gevins, 1998; Coles & Rugg, 1995). The CNV for notes was more negative than that for pseudoletters only in the experts, supported by a significant Group \times Stimulus interaction, $F(1, 18) = 4.65, p = .045, \eta_p^2 = 0.21$, and subsequent post hoc pairwise comparisons with the Scheffé tests ($p < .05$).

Although the CNV and the C1 had different topographic and temporal characteristics, making the CNV unlikely to explain the C1 effects, we tested whether the C1 effect could still be observed after filtering away the expertise effects on the CNV. A high-pass filter of 2 Hz was applied to the data such that the category selectivity in the CNV was no longer found at Cz (Group \times Stimulus interaction, $F < 1$) or at any other channels (all $p_s > .2$). However, the Group \times Stimulus interaction on the C1 was still significant at PO3/PO4, $F(1, 18) = 8.13, p = .011, \eta_p^2 = 0.31$, demonstrating that the category-selective expertise

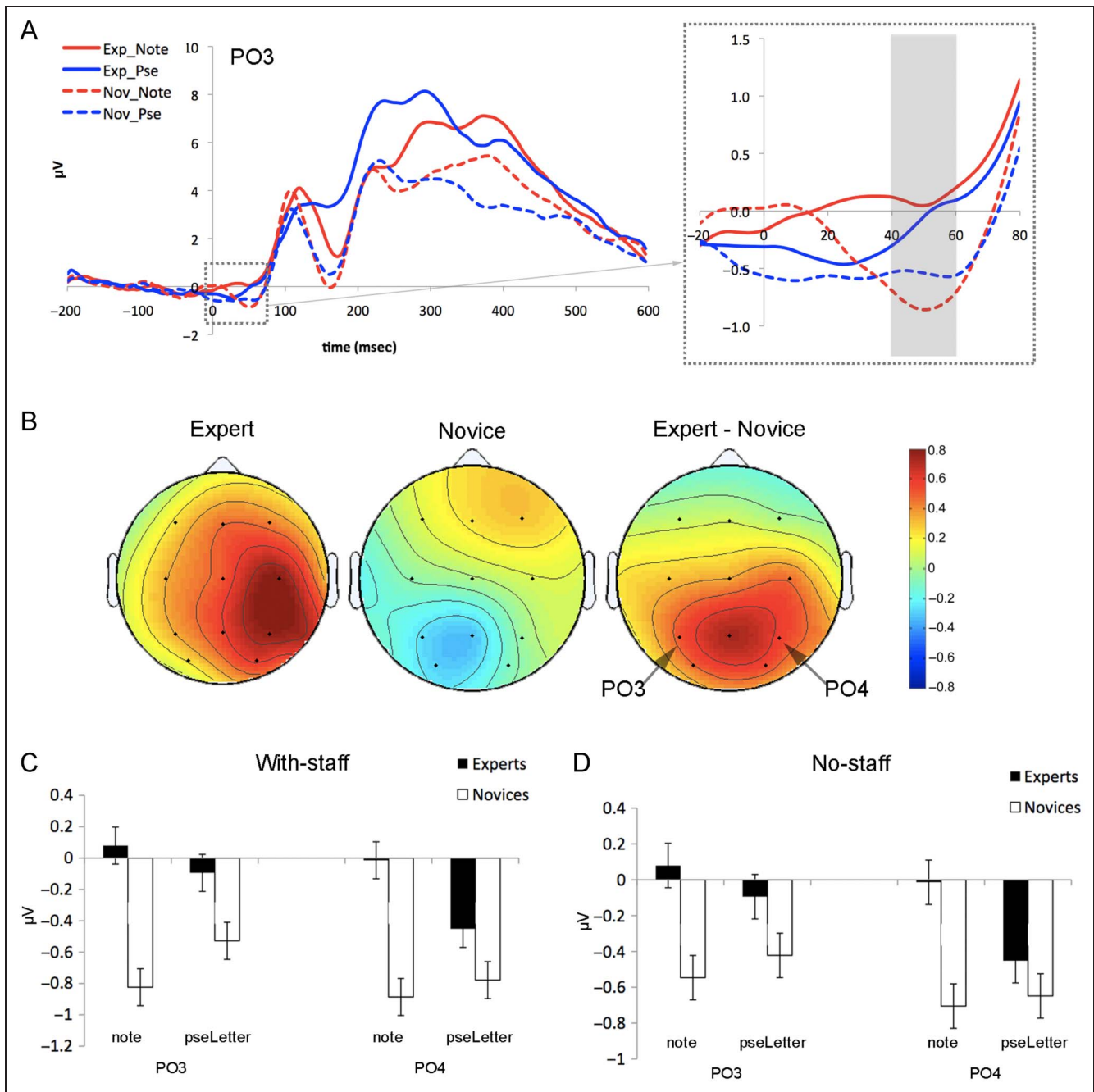


Figure 2. Grand-averaged waveform, scalp topography, and average voltage for the C1 in Experiment 1. (A) ERP waveform for the C1 at PO3 for stimuli with staff. Solid lines refer to the experts, and dotted lines refer to the novices. Red and blue lines show the ERP for notes and pseudoletters, respectively. The gray bar highlights the time window for the early C1 (40–60 msec). (B) The topographic distribution for the early C1 for the Expert/Novice Group \times Note/Pseudoletter Category interaction within experts (left), novices (middle), and the difference plot between the two groups (right). (C, D) The averaged early C1 at PO3/PO4 for the with-staff conditions (C) and the no-staff conditions (D). The black and white bars indicate data for experts and novices, respectively. The error bars plot 95% confidence interval for the Group \times Stimulus \times Hemisphere interaction. “pseLetter” refers to the pseudoletter condition.

effects on the C1 and the CNV effect are dissociable and independent.

Experiment 2: The Category-selective C1 Effect Is Dependent on Blocking

In Experiment 2, we examined whether the category-selective C1 effect is dependent on the blocked design

(as in Experiment 1 and in the prior fMRI experiment; Wong & Gauthier, 2010a) or is stimulus-driven (regardless of the blocking condition). Ten experts and 12 novices performed the 1-back task when the stimulus category was either blocked (as in Experiment 1) or randomized. The stimuli were always shown on a five-line staff. Similar to Experiment 1, experts demonstrated a higher perceptual fluency than novices for music sequences (for experts:

mean = 377.0 msec; $SD = 165.1$; for novices: mean = 1148 msec; $SD = 396.9$), but not for letter strings in a separate perceptual fluency measure (for experts: mean = 182 msec; $SD = 89.3$; for novices: mean = 231.1 msec; $SD = 137.8$); the Expert/Novice Group \times Note/Letter Stimulus interaction was significant, $F(1, 20) = 24.0, p \leq .0001, \eta_p^2 = 0.55$). Behavioral performance on the 1-back task was similar across groups (no effects involving group reached significance, all $ps > .14$), again indicating that this task engaged the two groups similarly.

We observed that the category-selective C1 effect with expertise is dependent on the blocking condition. As illustrated in Figure 3, the C1 was more positive for notes than for pseudoletters in experts only in the blocked condition. In contrast, for novices, the C1 stayed similar between notes and pseudoletters regardless of blocking. In the Group \times Stimulus \times Hemisphere \times Blocking ANOVA, our observations were supported by a Significant Group \times Stimulus \times Blocking interaction, $F(1, 20) = 4.79, p = .041, \eta_p^2 = 0.19$, (after Greenhouse–Geisser correction), and subsequent post hoc pairwise comparisons with the Scheffé tests at $p < .05$. This three-way interaction was not modulated by hemisphere ($F < 1$).³

Although both Experiments 1 and 2 included the same blocked conditions, it appears that the C1 for pseudoletters was more positive in experts than in novices only in Experiment 1 (Figure 2C) but not in Experiment 2 (Figure 3E). To test whether such difference across experiments was reliable, an Experiment \times Group \times Stimulus \times Hemisphere ANOVA on the C1 was performed. Importantly, the Group \times Stimulus interaction was robust, $F(1, 38) = 11.3, p = .0017$, showing no evidence of being different across experiments ($p > .9$ for the three-way interaction). These suggest that the apparent group difference for pseudoletters observed in the graphs was not reliable.

The Category Selectivity Effect for the N170 and the C1

To test whether the expertise- and blocking-dependent category selectivity is unique for the C1 or is a shared property for any expertise effects associated with music reading, we investigated the properties of the N170, which has been found to be a general expertise marker for many other object categories, including faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996), cars (Gauthier, Curran, Curby, & Collins, 2003), dogs and birds (Tanaka & Curran, 2001), letters (Wong et al., 2005) and Greebles (Rossion, Kung, & Tarr, 2004; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002).

We did not observe any evidence that the N170 is dependent on stimulus blocking. First, the category selectivity for musical notation on the N170 reached significance only in Experiment 1 but not in Experiment 2, suggesting that the N170 selectivity for notes was not as robust as that for the C1. The N170 followed the typical distribution where the effects were maximal at T5/6 (Wong et al.,

2005; Gauthier et al., 2003; Tanaka & Curran, 2001). In Experiment 1, the N170 for notes on staff was more negative than that for pseudoletters, in experts but not in novices, supported by a significant Group \times Stimulus interaction in the Group \times Stimulus \times Hemisphere ANOVA, $F(1, 18) = 5.99, p = .025, \eta_p^2 = 0.25$, and subsequent post hoc pairwise comparisons with the Scheffé tests ($p < .05$). This interaction was independent of hemisphere ($p = .14, \eta_p^2 = 0.12$). However, in Experiment 2, the Group \times Stimulus interaction in the blocked condition was not observed ($p > .5$), although the numerical trend for a more negative N170 for notes than pseudoletters was still observed among the experts in the left hemisphere. Second, in Experiment 2, the interaction between group, stimulus, and blocking on the N170 did not reach significance ($F < 1$) and the interaction effect was not modulated by hemisphere ($F < 1$). Importantly, the Group \times Stimulus \times Blocking \times C1/N170 component effect was significant, $F(1, 20) = 4.45, p = .048, \eta_p^2 = 0.18$, indicating that the C1 was more dependent on stimulus blocking than the N170.

Behavioral Relevance of the Category Selectivity on the C1

Does the category selectivity on the C1 for musical notation reflect processes that are behaviorally relevant during music reading? We addressed this question by testing whether individual music-reading ability (measured by the perceptual fluency task) predicts the amplitude of the C1 for notes across the two experiments. In this analysis, the C1 in the blocked condition was averaged across the two hemispheres because the C1 effects were similar across hemispheres. Both experts and novices were included in the analyses.

The C1 for notes was predicted by note fluency in the zero-order correlation, $r(39) = -.32, p = .043$ (Table 1). The correlation holds after controlling for factors including the perceptual fluency for letters, the amplitude of the C1 for pseudoletters, and that for letters, $r(36) = -.32, p = .049$, confirming that this effect cannot be explained by other general, non-category-specific effects. In contrast, the N170 for notes was not predicted by note fluency, either before, $r(39) = .11, p = .51$ (Table 1), or after partialing out the contributions of the other three variables, $r(36) = .25, p = .13$. This suggests that the category selectivity for the C1 reflects neural computations important for music-reading ability.

Experiment 3: The Behavioral Advantage of Blocking

On the basis of the blocking-dependent C1 effects, we expected that behavioral advantages afforded by V1 within the feedforward sweep of visual activity should depend on blocking too. Such blocking-dependent perceptual training effects have not been discussed in prior

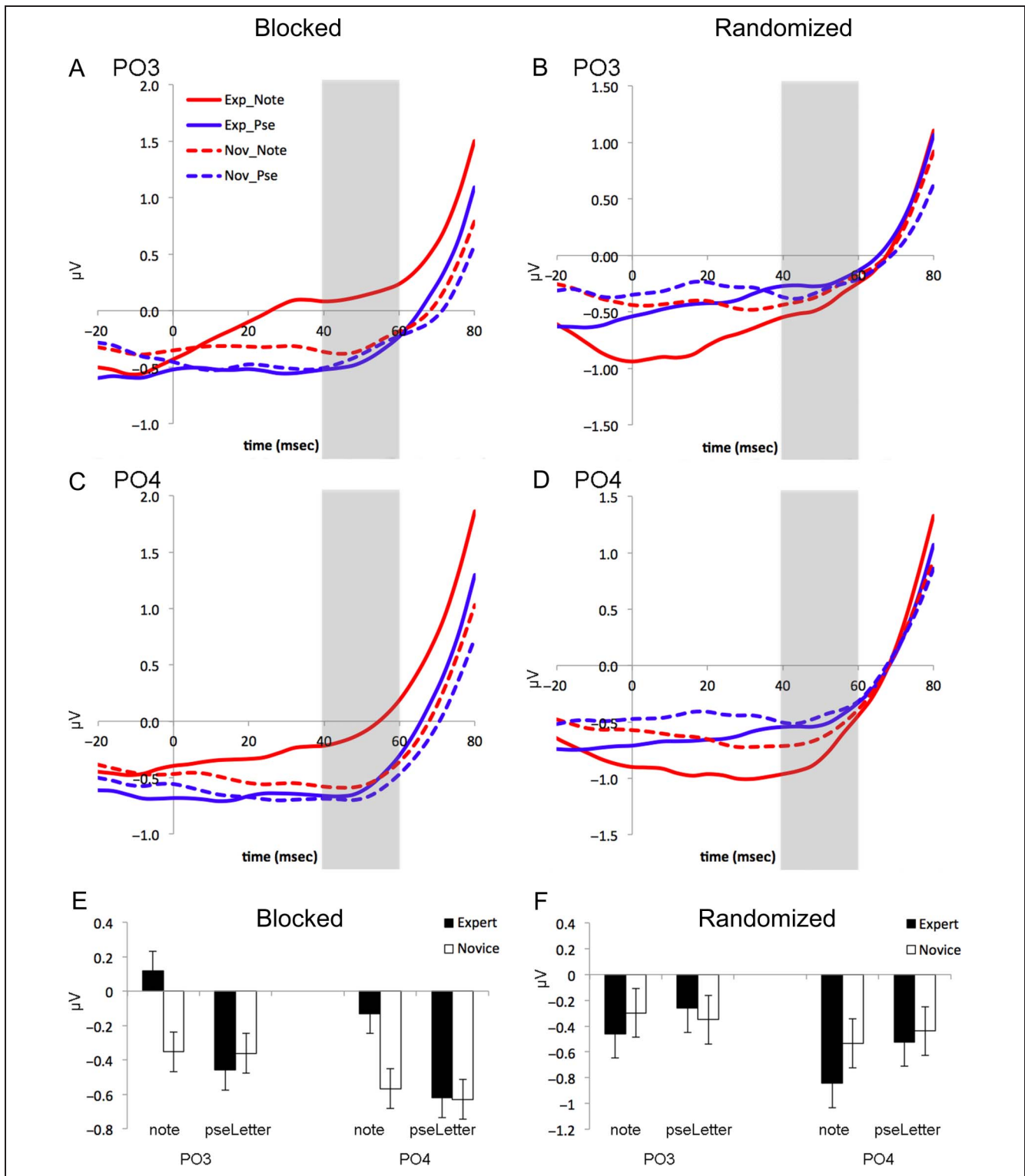


Figure 3. ERP waveform and average voltage for the C1 under different blocking conditions in Experiment 2. At PO3, the early C1 showed a selectively higher response for notes in experts only in the blocked condition (A) but not in the randomized condition (B). Similar patterns were observed for PO4 (C, D). Solid lines refer to the experts, and dotted lines refer to the novices. Red and blue lines show the ERP for notes and pseudoletters, respectively. The gray bar highlights the time window for the early C1 (40–60 msec). The bar graphs show the averaged C1 for the blocked condition (E) and the randomized condition (F). The black and white bars indicate data for experts and novices, respectively. The error bars plot 95% confidence interval for the Group \times Stimulus \times Hemisphere \times Blocking interaction. “pseLetter” refers to the pseudoletter condition.

Table 1. Correlation Coefficients of the Regression Analyses for the Behavioral Fluency for Notes and Letters, the C1, and the N170

	<i>Behavioral Fluency</i>		<i>C1</i>			<i>N170</i>	
	<i>Notes</i>	<i>Letters</i>	<i>Notes</i>	<i>PseLetters</i>	<i>Letters</i>	<i>Notes</i>	<i>PseLetters</i>
Behavioral letter fluency	.210						
C1 notes	-.318*	-.082					
C1 PseLetters	.007	.036	.652*				
C1 letters	-.192	.071	.716*	.652*			
N170 notes	.106	.05	.071	-.124	.176		
N170 PseLetters	.011	.028	.178	.008	.293	.888*	
N170 letters	.074	.074	.174	.003	.295	.807*	.945*

The asterisks indicate significant correlations ($p < .05$). “PseLetters” refers to the pseudoletter condition.

perceptual training studies (to our best knowledge), and this novel behavioral prediction about the perceptual abilities of music notation experts was entirely derived from the ERP experiments. This prediction was tested in Experiment 3.

Forty-three participants were recruited, and two were excluded from data analysis (one did not complete the experiment and the other had a note threshold of 12 *SD* larger than that for the rest of the group). They were presented with note-like stimuli (Figure 4A) in horizontal or

vertical orientations in three different blocking conditions, in which the stimulus orientation was kept the same within each block (“blocked”), randomized (“randomized”), or always alternating between the two orientations (“predictable”). Individual skill in music reading was separately measured with the perceptual fluency test (see Methods).

We first categorized the participants into high-expertise ($n = 13$; average note fluency = 444.5 msec, $SD = 149$ msec) and low-expertise groups ($n = 28$; average note fluency = 1115.9 msec, $SD = 493$ msec) based on their music training background (see Methods). A mixed ANOVA with Group \times Orientation \times Blocking revealed a three-way interaction, $F(2, 39) = 3.06$, $p = .052$, $\eta_p^2 = .07$ (Figure 4). To unpack this interaction, we treated music-reading expertise as a continuous variable because it is statistically more powerful and that there is no absolute cutoff for these group definitions. Music-reading expertise was defined as [the note threshold minus the letter threshold] measured in the perceptual fluency task. Multiple regression analyses were performed to test if music-reading expertise can be predicted with individual performance for horizontal and vertical stimuli under different blocking conditions.

For the blocked and predictable conditions, with performance for horizontal and vertical stimuli entered as simultaneous predictors, performance with the horizontal stimuli predicted music-reading ability but the vertical stimuli did not (see Table 2). In contrast, under randomized condition, performance with neither orientation predicted music-reading ability. Therefore, behavioral performance with musical stimuli reflects music-reading skill only when the stimulus orientation is predictable, indicating the importance of blocking on trained perceptual skill, at least in the case of music-reading expertise.

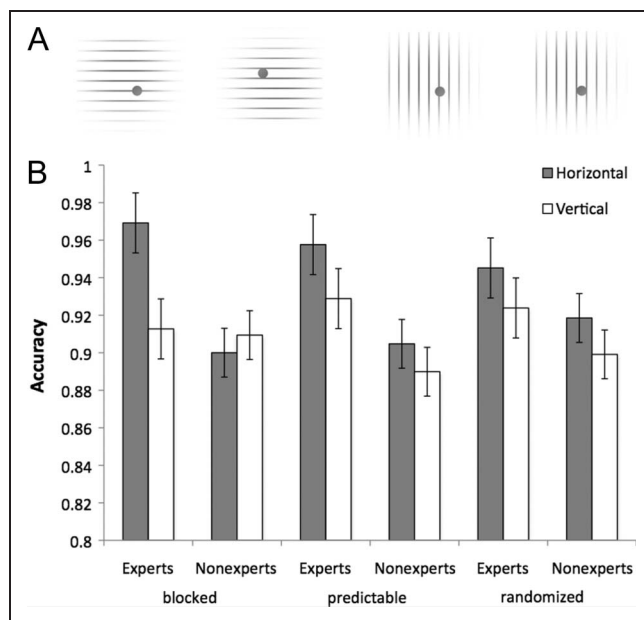


Figure 4. Examples of the note-like stimuli and results for Experiment 3. (A) Examples of the horizontal and vertical note-like stimuli, in which the black dot was either positioned on a line or in the space between two lines. Participants were asked to judge if the black dot was on or off a line by key press with accuracy emphasized and without time limit. (B) Accuracy for the experts and nonexperts with horizontal and vertical stimuli when the stimulus category was blocked, predictable (constantly alternating between horizontal and vertical orientations), or randomized. Error bars plot the *SE* of each condition.

DISCUSSION

In two independent experiments, we replicated category-selective activity for musical notation in bilateral C1

Table 2. Results of Three Separate Multiple Regression Analyses on Music-reading Ability (Perceptual Fluency for Notes Minus that for Letters)

<i>Model</i>	<i>Predictors</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>0-order r</i>	<i>p</i>
1. Blocked condition (R^2 adj. = 11.1%)	Intercept	3361.57	1073	3.13	.003		
	Horizontal	-1990.84	967.2	-2.06	.046	-.374	.016
	Vertical	-914.74	1072	-0.853	.399	-.249	.117
2. Randomized condition (R^2 adj. = 0.1%)	Intercept	2358.91	1223	1.93	.061		
	Horizontal	-1385.46	2446	-0.566	.574	-.224	.16
	Vertical	-420.45	1906	-0.221	.827	-.208	.192
3. Predictable condition (R^2 adj. = 15.1%)	Intercept	2971.61	841.1	3.53	.001		
	Horizontal	-4407.66	1822	-2.42	.021	-.404	.009
	Vertical	1976.66	1653	1.2	.239	-.265	.096

All predictors were entered simultaneously in each model. The last two columns provide the zero-order correlations of each predictor with music-reading ability.

(40–60 msec poststimulus onset) with music-reading expertise (Experiments 1 and 2). The C1 category selectivity for notes is observed only when the stimulus category is blocked, and the degree of selectivity reflects individual level of music-reading skill (Experiment 2). This is novel, as a blocking-dependent expertise effect has not been reported in the visual perceptual training literature (although lack of blocking, under the term “roving,” has been shown to impair some kinds of perceptual learning; Herzog, Aberg, Fremaux, Gerstner, & Sprekeler, 2012). We demonstrated that blocking the stimulus category does not only lead to category-selective activity on the neural level (Experiment 2) but also result in a behavioral advantage in judging musical stimuli among experts (Experiment 3).

This category selectivity on the C1 is likely generated within the initial feedforward sweep of V1 activity, as neural activity in this early time window is not attributed to feedback from extrastriate or from higher visual cortex in which the onset of activity occurs after 60 msec (Foxy et al., 2008; Luck, 2005; Schmolesky et al., 1998; Clark & Hillyard, 1996; Clark et al., 1995; Jeffreys & Axford, 1972). Our results cannot be explained by prestimulus noise, eye movement artifacts, or by the presence of the five-line staff. The early effects cannot be explained by the prestimulus CNV because the category selectivity on the C1 was still observed after filtering out the CNV effects. Also, the C1 selectivity for notes cannot be explained by voluntary attention or top-down engagement (Harel, Gilaie-Dotan, Malach, & Bentin, 2010) because an attentional effect would be expected to last through the next P1 component (60–120 msec), for which similar category selectivity was not observed.

The C1 category selectivity for notes in this study converges with our prior fMRI findings that bilateral V1 is selective for musical notation with expertise in music reading (Wong & Gauthier, 2010a). These findings challenge

widely accepted theories suggesting that object recognition is only achieved in higher visual cortex (DiCarlo et al., 2012; DiCarlo & Cox, 2007; Kourtzi & DiCarlo, 2006; Grill-Spector & Malach, 2004; Riesenhuber & Poggio, 1999). Instead, with perceptual expertise and under an appropriate testing context (e.g., blocking the stimulus category), cells in V1 can respond selectively to the object category of musical notes within the initial sweep of feedforward activity, and the level of V1 activity predicts behavioral performance in recognizing music sequences.

Why Is V1 Recruited for Music-reading Expertise?

There are at least two possible factors that may drive the recruitment of V1 by musical notation in experts. The first factor is that V1 may best fulfill the task demands of music-reading expertise (Wong et al., 2012; Swzed et al., 2011; Sigman et al., 2005; Ahissar & Hochstein, 2004; Sigman & Gilbert, 2000). Our prior work shows that, with identical stimuli and amount of training, differences in training tasks can lead to recruitment of different visual areas (Wong et al., 2012). It suggests that the demands of the training task can in some cases be sufficient to explain how different kinds of perceptual expertise recruit different visual areas. Efficient music reading requires speeded identification of multiple notes within a glance that are on the five-line staff and are often spatially spread out. The need for a high spatial resolution representation and simultaneous recognition of multiple stimuli may underlie the recruitment of V1 for music reading (Wong et al., 2012; Sigman et al., 2005; Lee et al., 2002; Mumford, 1991).

A different account involves the multimodal integration of music-reading expertise with other auditory, somatosensory, and motor processes. Previous work has shown that simultaneously processing information in an additional nonvisual modality results in changes in the C1

response, regardless of whether the second modality is task relevant or not (Karns & Knight, 2008; Fort, Delpeuch, Pernier, & Giard, 2002; Giard & Peronnet, 1999). The mere presentation of a single musical note on staff line automatically engages a widespread multimodal network, including auditory, somatosensory, motor, and other frontal regions (Wong & Gauthier, 2010a). It is possible that extensive training in integrating multimodal information with musical notes have induced long-term changes in V1 cells, such as by increasing the neural responses of V1 cells toward musical notes or by recruiting more V1 cells for musical notes that are not normally activated by visual stimuli (Giard & Peronnet, 1999).

Is the recruitment of V1 specific for the category of musical notes? If not, why is it never reported in other domains of perceptual expertise? Prior studies may have missed the early visual selectivity simply because it is theoretically not expected or because they did not have enough statistical power to reveal the early effects (there were typically 100–200 trials per condition for prior ERP studies but 660 trials per condition in this study). Including a novice group, as in the present work, may have provided a more powerful contrast to reveal any expertise effects in this early time window. In contrast, there was no novice group for letter recognition in our design, which might explain why letter selectivity was not observed with the C1 component in the current study. Indeed, reading expertise (e.g., English or Chinese) may be another candidate domain for recruiting V1, because it shares similar task demands of speeded recognition of multiple objects with high spatial resolution and also the requirement of multimodal integration. This possibility is consistent with recent fMRI findings that reading expertise leads to various changes in the early visual cortex (Swzed et al., 2011; Dehaene et al., 2010).

The Role of Blocking of Stimulus Categories

Why does the C1 selectivity for notes only emerge when the stimulus category is blocked instead of randomized? Although the effect of blocking has rarely been discussed, recent evidence is consistent with the idea that visual selectivity is modulated by top–down activity, tasks, and contexts. For example, deactivating the top–down activity from visuo-parietal cortex can affect the direction selectivity in early visual areas (Galuske, Schmidt, Goebel, Lomber, & Payne, 2002). For trained monkeys, the shape selectivity of the V1 neurons can be modulated by whether the to-be-matched shape was lines, circles, or sinusoids (McManus, Li, & Gilbert, 2011). After perceptual training of human participants, the category selectivity for object silhouettes in both early and late visual cortex depends on whether the participants were performing a visual search task or a shape matching task (Wong et al., 2012). These findings support theories stating that changes induced during perceptual learning are expressed under top–down influences defined by the tasks (Gilbert

& Sigman, 2007), that task demands can reprogram an entire network of sensory neurons that does not only gate the magnitude but also influence the function of neural responses (McManus et al., 2011) or that category selectivity may emerge as a network property between the interaction of feedforward and feedback influences rather than from local cells tuned selectively to a category (Price & Devlin, 2011). Note that such top–down processing does not necessarily imply conscious influences such as effort level or voluntary attention (Price & Devlin, 2011). Instead, the task-dependent tuning of the neural network may be largely implicit, nonverbalizable, and highly associated with different parts of the neural network and may occur automatically when the participants know that they are about to perform a well-learned task with a set of highly familiar stimuli. Such flexibility in tuning up the neural network may be important especially when early retinotopic cortex is involved in perceptual learning, because changing the low-level representation in a permanent, hard-wired way may be detrimental to performing other visual tasks (Schafer, Vasilaki, & Senn, 2007; Fahle & Poggio, 2002; Gilbert, Sigman, & Crist, 2001).

In music reading, blocking of the stimulus category is the rule—notes hang together in long sequences of musical notation on music scores. Experts learn to process these sequences very quickly and be prepared that more notes are coming. Similarly, we process print under blocked conditions, which is thought to allow the use of perceptual regularities, such as font information, for efficient recognition (Gauthier, Wong, Hayward, & Cheung, 2006; Sanocki, 1988). In other words, for some kinds of expertise, stimulus blocking may provide an optimal testing context for experts to set up the trained neural networks and express their acquired visual skill (McManus et al., 2011; Price & Devlin, 2011; Gilbert & Sigman, 2007).

Experiment 3 further confirmed that blocking serves as a predictable testing context that does not only determine the early category selectivity on the neural level but also affect behavioral performance with note-like stimuli. Performance reflected individual levels of music-reading skill only when the staff lines of the stimuli were horizontal (consistent with the trained music reading orientation) and when the staff line orientation was predictable. These indicate the importance of predictability in learned perceptual skill: Experts can express their music-reading skill only when the participants can predict whether the upcoming stimulus is in the familiar trained orientation or not.

It remains unclear whether the blocking effect is specific to the domain of music reading or it is a common characteristic associated with expertise effects observed in V1. At least some of the perceptual expertise effects can be observed when stimulus category is intermixed with control categories, both on the neural level (e.g., category selectivity on the N170 with faces, dogs, birds, and fingerprints (Busey & Vanderkolk, 2005; Xu, 2005; Tanaka & Curran, 2001; Rossion et al., 2000; Bentin et al., 1996)

and on the behavioral level (e.g., performance advantage for novel objects intermixed with objects transformed in part configuration; Wong, Palmeri, & Gauthier, 2009). However, none of the expertise effects observed in randomized conditions are associated with V1. Further studies need to clarify whether expertise effects associated with V1 can be found in other expertise domains and whether those effects are dependent on blocking.

In conclusion, our findings indicate that perceptual expertise can penetrate and influence neural activity as early as 40–60 msec poststimulus onset, and the C1 is thus the earliest perceptual expertise marker ever reported. In the context of existing studies in other domains of expertise, perceptual expertise for musical notation appears to stand out in two important ways: its feedforward recruitment of V1 and its dependence on blocking.

Acknowledgments

We would like to thank Magen Speegle for her help with data collection. This work was supported by the JSMF and NSF (SBE-0542013), the Vanderbilt Vision Research Center (P30-EY008126), the National Eye Institute (R01 EY013441-06A2), the National Institutes of Health (R01-EY019882, P30-EY08126, and P30-HD015052), and National Science Foundation (BCS-0957072).

Reprint requests should be sent to Yetta Kwailing Wong, Department of Applied Social Studies, Y7419, Academic Building I, City University of Hong Kong, Tat Chee Road, Kowloon Tong, Hong Kong, or via e-mail: yetta.wong@gmail.com.

Notes

1. Although one study reported an effect of expertise for cars in early visual cortex (Harel et al., 2010), it may be caused by the larger image size of the expertise-related stimuli than the control category, resulting in more early visual activity even in novices.
2. We did not observe any category selectivity for the C1 for letters in Experiment 1. In the Expert/Novice Group \times Letter/Pseudoletter Stimulus \times Hemisphere ANOVA on the C1, the Group \times Stimulus interaction was not significant regardless of whether the letters were on the five-line staff or not ($F_s < 1$), and this effect was not different across hemispheres ($F_s < 1$ for the three-way interactions). In addition, we became aware after data collection that letters may not serve as a good baseline for musical note perception because musical training can influence reading ability (Moreno et al., 2009). Therefore, the letter conditions were excluded in the rest of the analyses for simplicity's sake.
3. Similar to findings in Experiment 1, we did not observe any category selectivity for the C1 for letters in Experiment 2. In the Group \times Letter/Pseudoletter Stimulus \times Blocking \times Hemisphere ANOVA on the C1, the Group \times Stimulus interaction was not significant ($F < 1$), which was not modulated by blocking or hemisphere ($p_s > .11$ for the three-way or four-way interactions). Therefore, the letter conditions were excluded in the rest of the analyses for simplicity.

REFERENCES

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8, 457–464.

- 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, 15080–15084.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551–565.
- Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, 19, 1583–1596.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Busey, T., & Vanderkolk, J. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, 45, 431–448.
- 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, 387–402.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., et al. (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, 291–307.
- Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 1–26). New York: Oxford University Press.
- Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. *Nature Neuroscience*, 4, 519–525.
- Dehaene, S., Pegado, F., Brago, L. W., Ventura, P., Filho, G. N., Jobert, A., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330, 1359.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11, 333–341.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73, 415–434.
- Downing, P. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470–2473.
- Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, 23, 115–125.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598–601.
- Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.
- 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., Pasiaka, W., Schroeder, C. E., et al. (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.
- Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology*, 14, 573–578.
- Galuske, R. A., Schmidt, K. E., Goebel, R., Lomber, S. G., & Payne, B. R. (2002). The role of feedback in shaping neural representations in cat visual cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, 99, 17083–17088.

- Gauthier, I., Curby, K. M., Skudlarski, P., & Epstein, R. A. (2005). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 222–234.
- Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, *6*, 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*, 191–197.
- Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 431–446.
- Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training “Greeble” experts: A framework for studying expert object recognition processes. *Vision Research*, *38*, 2401–2428.
- Gauthier, I., Wong, A. C.-N., Hayward, W. G., & Cheung, O. S.-C. (2006). Font-tuning associated with expertise in letter perception. *Perception*, *35*, 541–559.
- 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*, 473–490.
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696.
- Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*, 681–697.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409–1422.
- Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*, 837–843.
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677.
- Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cerebral Cortex*, *20*, 2304–2318.
- Herzog, M. H., Aberg, K. C., Fremaux, N., Gerstner, W., & Sprekeler, H. (2012). Perceptual learning, roving and the unsupervised bias. *Vision Research*, *61*, 92–99.
- Ishai, A. (2008). Let’s face it: It’s a cortical network. *Neuroimage*, *40*, 415–419.
- James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory-motor brain network. *Neuropsychologia*, *44*, 2937–2949.
- James, K. H., James, T. W., Jobard, G., Wong, A. C., & Gauthier, I. (2005). Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective & Behavioral Neuroscience*, *5*, 452–466.
- 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*, 1–21.
- 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.
- Kams, C. M., & Knight, R. T. (2008). Intermodal auditory, visual and tactile attention modulates early stages of neural processing. *Journal of Cognitive Neuroscience*, *21*, 669–683.
- Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cerebral Cortex*, *18*, 2629–2636.
- Kourtzi, Z., Betts, L., Sarkheil, P., & Welchman, A. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLOS Biology*, *3*, e204.
- Kourtzi, Z., & DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Current Opinion in Neurobiology*, *16*, 152.
- Lee, T. (2002). Top-down influence in early visual processing: A Bayesian perspective. *Physiology & Behavior*, *77*, 645–650.
- Lee, T., Yang, C., Romero, R., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience*, *5*, 589–597.
- Luck, S. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *92*, 8135–8139.
- Martinez, A., Anillo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*, 364–369.
- McEvoy, L. E., Smith, M. E., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice. *Cerebral Cortex*, *8*, 563–574.
- McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of Sciences*, *109*, 17063–17068.
- McManus, J. N., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences*, *108*, 9739–9746.
- Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience*, *26*, 11187–11196.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*, 712–723.
- Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics*, *65*, 135–145.
- Nunez, P. L. (1981). *Electric fields of the brain*. New York: Oxford University Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience*, *26*, 13025–13036.
- Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–648.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex*, *14*, 619–633.
- Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Research*, *48*, 55–62.

- Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences, 15*, 246–253.
- Proverbio, A. M., & Adorni, R. (2009). C1 and P1 visual responses to words are enhanced by attention to orthographic vs. lexical properties. *Neuroscience Letters, 463*, 228–233.
- Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: A predictive coding framework. *Neuroscience and Biobehavioral Reviews, 35*, 1237–1253.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.
- Rose, M., Verleger, R., & Wascher, E. (2001). ERP correlates of associative learning. *Psychophysiology, 38*, 440–450.
- Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left lateralized face-like electrophysiological responses. *Psychological Science, 13*, 250–257.
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *NeuroReport, 11*, 69–74.
- 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, U.S.A., 101*, 14521–14526.
- Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 472–480.
- Schafer, R., Vasilaki, E., & Senn, W. (2007). Perceptual learning via modification of cortical top-down signals. *PLOS Computational Biology, 3*, e165.
- Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79*, 3272–3278.
- Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature, 412*, 549–553.
- Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences, U.S.A., 99*, 17137–17142.
- Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nature Neuroscience, 3*, 264–269.
- Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron, 46*, 823–835.
- Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the C1 visual event-related component by conditioned stimuli: Evidence for sensory plasticity in early affective perception. *Cerebral Cortex, 16*, 876–887.
- Swzed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabregue, R., Amadon, A., et al. (2011). Specialization for written words over objects in the visual cortex. *Neuroimage, 56*, 330–344.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12*, 43–47.
- Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience, 4*, 219–229.
- Travis, F., Tecce, J. J., & Guttman, J. (2000). Cortical plasticity, contingent negative variation, and transcendent experiences during practice of the transcendental meditation technique. *Biological Psychology, 55*, 41–55.
- van der Linden, M., Murre, J. M. J., & van Turennout, M. (2008). Birds of a feather flock together: Experience-driven formation of visual object categories in human ventral temporal cortex. *PLOS One, 3*, e3995.
- van der Linden, M., van Turennout, M., & Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience, 22*, 1270–1282.
- Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature, 203*, 308–384.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*, 113–120.
- Williams, M., Baker, C., Op De Beeck, H., Mok Shim, W., Dang, S., Triantafyllou, C., et al. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience, 11*, 1439–1445.
- 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, and Behavioral Neuroscience, 5*, 306–318.
- Wong, A. C.-N., Palmeri, T. P., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a ziggerin expert —But which type? *Psychological Science, 20*, 1108–1117.
- Wong, A. C.-N., Palmeri, T. P., 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*, e8405.
- Wong, Y. K., Folstein, J. R., & Gauthier, I. (2012). The nature of experience determines object representation in the visual system. *Journal of Experimental Psychology: General, 141*, 682–698.
- Wong, Y. K., & Gauthier, I. (2010a). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience, 22*, 695–713.
- Wong, Y. K., & Gauthier, I. (2010b). Holistic processing of musical notation: Dissociating failures of selective attention in experts and novices. *Cognitive, Affective & Behavioral Neuroscience, 10*, 541–551.
- Wong, Y. K., & Gauthier, I. (2012). Music-reading expertise alters visual spatial resolution for music notation. *Psychonomic Bulletin & Review, 19*, 594–600.
- Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 121–138.
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*, 1234–1242.
- Yue, X., Tjan, B., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*, 3802–3811.