

# A Multimodal Neural Network Recruited by Expertise with Musical Notation

Yetta Kwailing Wong and Isabel Gauthier

## Abstract

■ Prior neuroimaging work on visual perceptual expertise has focused on changes in the visual system, ignoring possible effects of acquiring expert visual skills in nonvisual areas. We investigated expertise for reading musical notation, a skill likely to be associated with multimodal abilities. We compared brain activity in music-reading experts and novices during perception of musical notation, Roman letters, and mathematical symbols and found selectivity for musical notation for experts in a wide-

spread multimodal network of areas. The activity in several of these areas was correlated with a behavioral measure of perceptual fluency with musical notation, suggesting that activity in nonvisual areas can predict individual differences in visual expertise. The visual selectivity for musical notation is distinct from that for faces, single Roman letters, and letter strings. Implications of the current findings to the study of visual perceptual expertise, music reading, and musical expertise are discussed. ■

## INTRODUCTION

Numerous studies suggest an important role for the lateral occipital complex and the fusiform gyrus in visual object recognition. Within these brain areas, different regions are found to respond preferentially to different object categories, including faces (Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996), words (Cohen et al., 2000), letters (James, James, Jobard, Wong, & Gauthier, 2005; Polk et al., 2002), body parts (Peelen & Downing, 2007; Downing, 2001), and buildings (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998). Experience is a viable mechanism to explain such visual selectivity or specialization and may be the only reasonable explanation in some cases such as words and letters. Indeed, expertise in a visual domain, be it acquired in the real world or in the laboratory, can transform how the visual system responds to objects of a category and in some cases lead to visual selectivity for a trained domain of objects in focal regions of the visual cortex (Wong, Jobard, James, Gauthier, in press; Jiang et al., 2007; Moore, Cohen, & Ranganath, 2006; Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006; Reddy & Kanwisher, 2006; Xu, 2005; Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999).

However, visual object recognition depends on more than visual regions of the brain. It also engages a distributed network of areas representing sensory information or conceptual knowledge associated with the objects. For example, pictures of tools engage the motion-selective

middle temporal gyrus and premotor areas (Chao, Haxby, & Martin, 1999; Martin, Wiggs, Ungerleider, & Haxby, 1996). Pictures of food engage gustatory processing areas (Simmons, Martin, & Barsalou, 2005). In addition, brief conceptual associations with novel objects are sufficient to elicit activity in modality-specific areas (for instance in motion-selective or auditory cortex) during a purely visual task with these objects (James & Gauthier, 2003). In sum, studies with familiar objects, presumably associated with rich conceptual information, as well as those with novel objects arbitrarily associated with information from nonvisual modalities, reveal that nonvisual areas are engaged automatically during visual judgments.

Given that simple visual judgments with objects can recruit nonvisual areas, and to the extent that object concepts are grounded in perception and action (Barsalou, 2008), we would expect that, with extensive experience, visual selectivity for objects of expertise can also be found in a multimodal neural network, covering both visual and nonvisual areas. This appears to be the case for the two real-world domains of perceptual expertise that have been studied most extensively with brain imaging, faces, and letters. The presentation of faces automatically engages a distributed network of areas, some of which are responsible for the processing of identity, biological motion, emotional expressions, or eye gaze (Haxby, Hoffman, & Gobbini, 2002). The perception of letters, in turn, also recruits several brain areas more specifically engaged by other tasks such as writing, copying, or visual imagery (James & Gauthier, 2006; Longcamp, Tanskanen, & Hari, 2006; Longcamp, Anton, Roth, & Velay, 2005). It is not clear, however, how activity in these distributed multimodal networks is related to the visual ability of experts.

---

Vanderbilt University, Nashville, TN

What we do know is that activity in visual areas engaged by objects of expertise correlates with behavioral measures of experts' perceptual skills (Xu, 2005; Gauthier, Curran, Curby, & Collins, 2003; Gauthier et al., 2000). In these studies of facelike expertise (expertise at individuating objects in homogeneous categories), the degree of activity in the fusiform face area (FFA) for objects of expertise such as cars or birds, across a variety of tasks in the scanner, predicted performance on a behavioral measure of expertise taken outside the scanner. In these cases, expertise was measured as the ability to match sequentially presented images in the expert domain; for instance, the ability to judge if two cars are from the same make and model regardless of differences in color, pose, or small differences in the year of the model. This is essentially a visual matching task that does not require naming, but performance is likely influenced by nonvisual information accessible to experts. Although past studies have only found activity in the FFA to be correlated with individual differences on this task, it is possible that similar results in nonvisual areas were missed because they were not the focus of the work or because multimodal contributions to car expertise (the most common domain used in this line of work) are limited. Therefore, it remains to be determined whether activity in nonvisual areas during a visual task can also be related to performance in a domain of expertise.

In this study, we focus on musical expertise, a domain for which we can find participants varying greatly in their visual perceptual ability (specifically, in reading musical notation) and for which we also expect a great deal of nonvisual expertise, in particular auditory and motor skills engaged in perceiving music and playing musical instruments (see review in Zatorre, Chen, & Penhune, 2007). Studies have revealed a wide range of neural changes associated with musical training, including a larger cortical representation of fingers in somatosensory cortex (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995), an increased auditory cortical representation (Pantev et al., 1998), a cortical asymmetry in the planum temporale ( Schlaug, Jäncke, Huang, & Steinmetz, 1995), and white matter changes (Bengtsson et al., 2005). These results suggest that musical training results in the recruitment of a multimodal neural network supporting and integrating music processing in various modalities.

Does the visual perception of musical notation recruit this multimodal network associated with musical expertise, similar to the cases for faces and letters? Although few studies have investigated music reading (Peretz & Zatorre, 2003; Zatorre & Peretz, 2001; Deutsch, 1998) or compared the neural response for musical notes with appropriate visual controls, some evidence suggests that parts of the musical neural network may be recruited during music reading (Stewart et al., 2003; Schön, Anton, Roth, & Besson, 2002; Nakada, Fujii, Suzuki, & Kwee, 1998; Sergent, Zuck, Terriah, & MacDonald, 1992). For instance, passive viewing of a music score led to activity

in early visual areas bilaterally and in an occipito-parietal area (Sergent et al., 1992). After training with music reading and keyboard playing, a visual task with musical notation resulted in increased neural responses in parietal and frontal areas (Stewart et al., 2003). Finally, a study contrasting passive viewing of musical scores to Japanese or English texts revealed higher neural activity for musical notes than text in the right transverse occipital sulcus (TOS) in all of eight musicians, but in none of the eight nonmusician controls, suggesting that the right TOS is recruited by expert music reading (Nakada et al., 1998). These studies suggest that part of the multimodal network is engaged with music reading after musical training.

To study brain regions specialized for musical notation as visual objects of expertise, we used single notes and short note sequences and compared the neural activity for these stimuli with visual controls (Roman letters and mathematical symbols) in music-reading experts and novices. Prior work has shown that selectivity for objects of expertise can be obtained in the visual system with a range of tasks, including those that are very different from how experts typically interact with the objects (Xu, 2005; Gauthier, Curran, et al., 2003; Gauthier et al., 2000). In fact, simple tasks that can be performed equally well by novices and experts are often preferred in this literature because expertise effects obtained under those conditions are not confounded by performance differences and more clearly reflect automatic processes engaged by an object category rather than practice with a specific task. Likewise we sought to find visual selectivity for musical notation that would not be task specific. In this study, we combined data sets from two groups of participants for which the tasks were different (for details, see Methods; the stimuli also differed but in relatively superficial ways). Both tasks (a one-back task and a gap detection task; see Methods) involved simple visual judgments that did not refer to the musical meaning of the stimuli, such that both experts and novices could perform the tasks equally well. With the same contrasts of interest (e.g., musical notation vs. Roman letters and mathematical symbols) in both groups and with the identical imaging parameters for scanning, combining these two groups within a single random effects analysis provided better statistical power in our search for brain regions specialized for expert perception of musical notation, to the extent that specialization for musical notation is independent of any specific stimulus sets or tasks performed in the scanner. In addition, we tested the perceptual fluency for musical notation outside the scanner (see Methods) as a measure of individual music-reading ability and studied how it was related to visual selectivity for musical notation in various visual and nonvisual areas.

In the present study, we had several goals. First, we aimed at identifying the loci of visual perceptual expertise for musical notation, a domain that has not been the focus of much prior work. Because the task solved during

music reading is quite different from that solved in cases of expertise that are “facelike” (in which objects that share a common part configuration must be distinguished, as in car or bird identification) and arguably different from reading (in which the vertical position of elements is irrelevant), we predicted that the visual areas showing expertise effects for notes in the visual ventral pathway would differ from face- and letter-selective areas. Also, to test whether the right TOS is recruited with expert music-reading skills, we examined whether the exclusive activation of the right TOS for music-reading experts in Nakada et al. (1998) could be replicated in our study. Second, we were interested in the possibility that single notes or short note sequences, as visual objects of expertise, could automatically engage a distributed multimodal network. Third, we wanted to compare specialization for single notes and five-note sequences to test the counterintuitive prediction that single notes may elicit stronger and more widespread expertise effects in the visual system. This prediction is based on surprising results that selectivity in the visual system was stronger and more widespread but distinct for single letters than for consonant strings (James et al., 2005). Fourth, provided that there is evidence of specialization for notes in a network of visual and nonvisual areas, we wanted to test whether the neural response to notes in these areas would correlate with the degree of visual skill when perceiving musical notation.

## METHODS

### Participants

Participants included 10 music-reading experts (8 women, mean age = 21.4,  $SD = 3.1$ ) and 10 novices (7 women, mean age = 23.8,  $SD = 4.1$ ). Experts had at least 10 years of music-reading experience ( $M = 14.4$  years) and had a high self-rating score on music-reading ability [ $M = 8.8$ ,  $SD = 0.79$  in a 10-point scale, ranging from 1 (*do not read music at all*) to 10 (*expert in music reading*)]. Novices had very limited music-reading experience ( $M = 0.8$  year) and had a low self-rating score ( $M = 1.4$ ,  $SD = 0.97$ ).

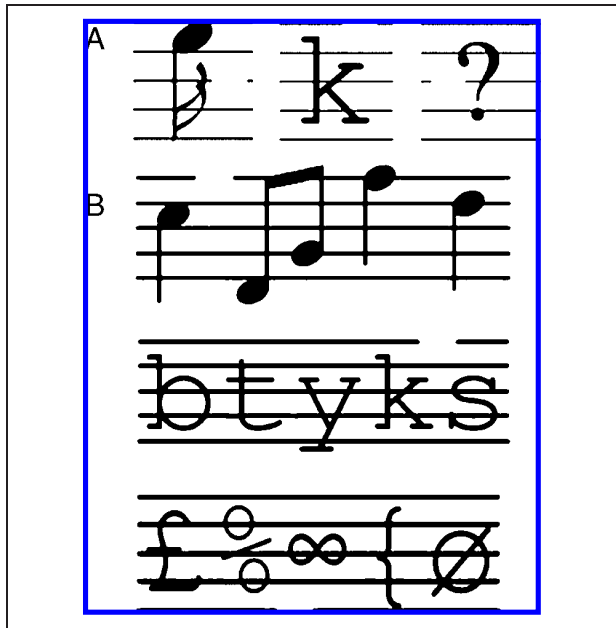
Music-reading ability was further tested in a note-naming task, in which participants were required to write down the letter name of as many musical notes as possible within 2 min on a randomly generated music score. Novices who could not name any of the notes were allowed to skip the test and assigned a score of 0. The mean number of correctly named notes for the expert group was 153.2 ( $SE = 5.54$ ) and that for the novice group was 2.1 ( $SE = 1.79$ ). All participants were right-handed except one left-handed novice, and all reported normal or corrected-to-normal vision and no history of neurological disorders. All gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

### Stimuli

The experiment was conducted on Apple computers using Matlab (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). The stimuli were presented on a liquid-crystal display (LCD) panel and back projected on a screen. Participants viewed the stimuli through a mirror mounted on top of an radio frequency (RF) coil above their head.

Two cohorts of participants were tested with two different tasks (each included five experts and five novices) using slightly different stimuli. For the first cohort of participants, there were 36 pictures in each of eight categories of objects. Faces and common objects were gray-scale images subtending about  $4.4^\circ \times 4.4^\circ$  of visual angle. Single letters, single mathematical symbols, and single musical notes were black-and-white images and subtended approximately  $1.5^\circ \times 1.5^\circ$  of visual angle. The 36 single letters were composed of 12 lowercase letters shown in black in three fonts (Courier, Bernard MT Condensed, and Textile). The 36 mathematical symbols were shown in Times New Roman font. The 36 musical notes included 9 different notes on the five-line staff (ranging from the D below the bottom line to the G above the top line) in four different time values, including half notes (an open circle), quarter notes (a closed circle), eighth notes (a closed circle with one tail), and sixteenth notes (a closed circle with two tails). Finally, five-letter consonant strings, five-mathematical symbol strings, and five-note sequences were also used. The letters in the letter strings were composed of black lowercase letters shown in Courier font. The note sequences always included 3 quarter notes and 2 eighth notes, with the 2 eighth notes appearing equally often in each possible position among the 3 quarter notes. All letter strings and note sequences were randomly generated with Matlab. Only the single notes and the note sequences were presented on the five-line staff.

For the second cohort of participants, stimuli were identical to the first except for the following: First, the stimuli were of different fonts and slightly different sizes. The single letters and mathematical symbols were in the fonts of Courier New, Hansa, and Gulim. The letter strings were in Courier New font and subtended a visual angle of  $2.5^\circ \times 5.1^\circ$ . Second, all the single stimuli and the string stimuli were presented on an identical five-line staff background, respectively. A different version of all stimuli was generated that included a small gap in one of the five staff lines (Figure 1). For single notes, the gap was randomly located anywhere on the staff. For the strings, the gap was always on the top line or the bottom line because a random location appeared too difficult based on pilot work. The position of the whole stimulus was randomly jittered from 0 to 10 pixels horizontally and vertically to avoid visual habituation of the five-line background. Finally, for two novices, 36 numbers were presented in the localizer run instead of mathematical symbols as control



**Figure 1.** Examples of stimuli for the gap detection task in (A) the single run, in which single musical notation (left), single Roman letters (center), or single mathematical symbols (right) were presented on identical five-line staff and the gap (if any) was located in random positions; (B) the string run, in which music sequences (top), letter strings (middle), or symbol strings (bottom) were presented on the staff and the gap (if any) was presented in either the top or the bottom line.

visual stimuli (only in the runs used to localize letter-selective regions). The numbers of “0” to “9” were used in the same fonts and same size as the letters.

### fMRI Task

This experiment used a block design and included two types of experimental scans (identical for both cohorts of participants). The first one was the single run, which included blocks of single notes, single letters, and single mathematical symbols. The second type was the string run, which included blocks of note sequences, letter strings, and mathematical symbol strings. There were three single runs and three string runs, each lasting for 5 min 36 sec and consisting of eighteen 16-sec blocks (six blocks for each stimulus condition), with three 16-sec fixation blocks interleaved at regular interval to establish a baseline of visual activity. In addition, we used two localizer runs showing blocks of faces, objects, single letters, and single mathematical symbols to localize the fusiform face area and letter-selective area(s). Each localizer run lasted for 5 min 20 sec and consisted of sixteen 16-sec blocks of testing stimulus (four blocks for each stimulus condition) and three 16-sec fixation blocks appeared regularly. In all runs, a 10-sec fixation block and a 6-sec fixation block were added at the beginning and

at the end of the run, respectively. During each block, 16 stimuli appeared sequentially, each for 750 msec and followed by a 250-msec blank. The order of conditions was counterbalanced within and across runs.

The two cohorts of participants performed different tasks. The first performed a one-back task, in which they were instructed to press a key with their right index finger as fast as possible if they detected immediate repeats of a stimulus. In each block, the number of repeats ranged from one to three, and the repetition rate for each category of stimulus was around 12%. Our pilot data showed that novices performed worse for the musical notes than other conditions, possibly because it was hard for novices to judge the absolute position of the circle part of the note on the staff under rapid presentation. To make the performance across conditions more comparable, notes were selected so that the participants could perform the task based on other visual features of the note such as the filled or open circle, the number of tails on the stem, whether the stem was pointing up or down, etc. Participants were told that single letters in different fonts were regarded as different. The second cohort of participants was not required to attend to stimulus identity. Instead, they performed a gap detection task where they pressed a key with their right index finger as fast as possible when they detected a gap on any of the five lines. The “gap” trials were about 12% of each type of stimulus.

The task in the localizer run was identical for all participants: They were asked to press the right index finger key as fast as possible if they detected immediate repeats of the stimulus.

### Measure of Perceptual Fluency

Presentation time threshold for matching four-note sequences was measured outside the scanner to quantify perceptual fluency with notes in each individual. Ten experts and nine novices who participated in the fMRI experiment performed this task.

A sequential matching paradigm was used. On each trial, a fixation cross was presented at the center of the screen for 200 msec, followed by a 500-msec premask, and then a target four-note sequence for a varied duration, estimated using QUEST in the Psychtoolbox (Watson & Pelli, 1983) to keep performance at 80% accuracy. After a 500-msec postmask, two four-note sequences would appear side by side, one identical to the target sequence and the other with one of the notes shifted by one step. The task was to select the sequence identical to the target and to respond by keypress. The four notes in the target sequence were randomly generated. The shifted note was randomly chosen out of the four notes, with the number of up/down shifts counterbalanced. Due to a ceiling effect in pilot tests, the contrast of all the stimuli was lowered by about 60% to increase perceptual difficulty. The threshold was measured twice, each with 100 trials, and



the average threshold was used. A smaller perceptual threshold reflects higher perceptual fluency with notes.

The perceptual threshold for matching four letter strings was also measured using an identical procedure as a control measure. The four letter strings were randomly generated with 11 letters: b, d, f, g, h, j, k, p, q, t, and y. These letters were selected because these letters contain parts extending upward or downward, similar to the musical notation used in the four-note sequences. 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.

### **MRI Data Acquisition**

Imaging was performed using a 3-T Philips Intera Achieva scanner at the Institute of Imaging Science at Vanderbilt University. The BOLD-based signals were collected using a T2\*-weighted EPI acquisition (echo time = 35 msec, repetition time = 2000 msec, flip angle = 79°, matrix size = 64 × 64, field of view = 192 mm, 34 slices, slice thickness = 3 mm with no gap). To increase coverage of the brain, the slices were tilted 10° from the horizontal plane so that the ventral temporal cortex and the occipital lobe were always covered, whereas portions of the superior parietal and superior frontal cortex may be left out due to individual differences in brain size. High-resolution T1-weighted anatomical volumes were also acquired using a three-dimensional turbo field echo acquisition (echo time = 4.6 msec, repetition time = 8.9 msec, flip angle = 8°, matrix size = 256 × 256, field of view = 256 mm, 170 slices, slice thickness = 1 mm with no gap).

### **fMRI Data Analysis**

The functional data were analyzed using the Brain Voyager 1.8 (<http://www.brainvoyager.com>) multistudy general linear model procedure. Data preprocessing included three-dimensional motion correction, slice scan time correction, temporal filtering (3 cycles/run high-pass), spatial smoothing (6-mm FWHM Gaussian), and temporal smoothing (2.8-sec FWHM Gaussian). A general linear model analysis computed the correlation of predictor variables or functions with the recorded activation data (criterion variables) across scanning sessions. The predictor functions were based on the blocked stimulus presentation paradigm of each particular run and represented an estimate of the predicted hemodynamic response in that run. The predictors consisted of the stimulus protocol boxcar functions convolved with the gamma function ( $\Delta = 2.5$ ,  $\tau = 1.25$ ) estimate of a typical hemodynamic response (Boynton, Engel, Glover, & Heeger, 1996). Statistical parametric maps were computed for each contrast of interest, treating participants as a random factor. To increase statisti-

cal power, data analysis only included the fourth to eighth volumes of each block when the hemodynamic response should be at peak level. In regions showing significant activation, further data analyses were performed with an ROI 10 × 10 × 10 mm<sup>3</sup> in size, centered on the peak activity, either to produce descriptive statistics or to analyze the data from independent data sets. Contiguous areas of activity were separated as multiple, nonoverlapping, significant 10 × 10 × 10 mm<sup>3</sup> areas if they consisted of multiple local peaks.

## **RESULTS**

### **Behavioral Results**

The data from the two tasks were combined and analyzed for both accuracy and RT. The accuracy was corrected for guessing using the following formula: [(hit rate – false alarm rate) / 1 – false alarm rate]. For RT analysis, only correct responses on trials where a keypress was expected were included (repeated trials for the one-back task; “gap” trials for the gap detection task). Due to technical difficulties, the behavioral data for one expert performing the one-back task were lost. Therefore, behavioral data analyses only included 9 experts and 10 novices (Table 1).

For the single run, performance across groups and stimuli was similar. A 2 × 3 ANOVA (Group × Category) revealed no significant main effects or interaction between Groups (experts and novices) and Category (single notes, single letters, and single symbols) on both accuracy and RT.

For the string run, results indicated that note sequences were more difficult compared with the other conditions only for novices. A 2 × 3 ANOVA (Group × Category) was conducted on accuracy and RT. A main effect was found for Category on both accuracy,  $F(2,34) = 4.06$ ,  $p = .026$ , and RT,  $F(2,34) = 3.36$ ,  $p = .047$ . Scheffé tests ( $p < .05$ ) revealed that the accuracy for note sequences was lower than for letter strings, and the RT for note sequences was slower than symbol strings. Furthermore, a significant interaction between Group and Category was found for accuracy,  $F(2,34) = 6.14$ ,  $p = .005$ . Scheffé tests ( $p < .05$ ) suggest that the accuracy across stimuli conditions was similar for experts. For novices, however, the accuracy for note sequences was lower than letter strings and symbol strings.

### *Measure of Perceptual Fluency with Musical Notation*

This task revealed a large difference in perceptual fluency with musical notation between the two groups. The data from one expert were excluded in this analysis and subsequent correlation analyses because her threshold was 2.7 *SD* from the mean of the rest of the expert group. It could be due to the use of ineffective strategy (such

**Table 1.** Behavioral Data for Musical Notation (N), Letters (L), or Mathematical Symbols (S) for Experts and Novices in (A) the Single Run and (B) the String Run

	One-Back Task			Gap Detection Task			Averaged		
	N	L	S	N	L	S	N	L	S
<i>(A)</i>									
Accuracy									
Expert	0.750	0.842	0.829	0.931	0.901	0.866	0.84	0.871	0.847
Novice	0.909	0.930	0.943	0.798	0.742	0.670	0.853	0.836	0.821
RT (msec)									
Expert	570.6	564.0	557.4	547.7	569.1	565.9	559.1	566.6	561.7
Novice	526.8	490.6	481.5	561.5	575.8	574.8	544.2	533.2	528.1
<i>(B)</i>									
Accuracy									
Expert	0.747	0.735	0.620	0.894	0.923	0.906	0.820	0.829	0.763
Novice	0.837	0.937	0.891	0.748	0.876	0.898	0.792	0.907	0.894
RT (msec)									
Expert	574.2	561.7	581.3	564.3	562.5	547.1	569.2	562.1	564.2
Novice	539.3	516.8	513.9	548.3	540.1	526.1	543.8	528.5	520.0

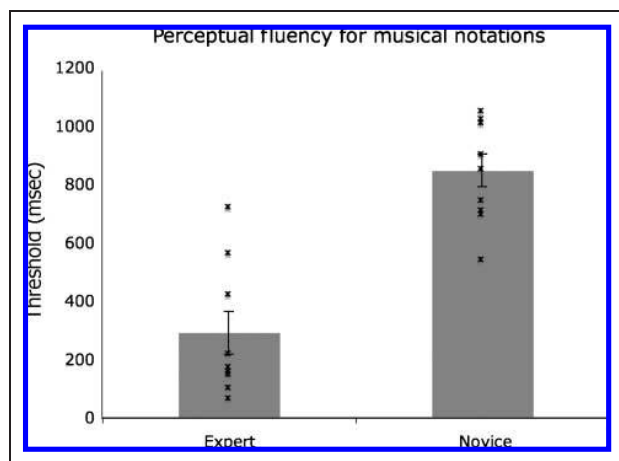
Data were shown separately for the one-back task (left columns), the gap detection task (middle columns), and the average for both tasks (right columns).

as naming all of the notes) with briefly presented stimuli. The mean threshold was 290.8 msec for experts (ranging from 68 to 730 msec) and 846.3 msec for novices (ranging from 547 to 1062 msec; Figure 2). The main effect of Group was highly significant,  $F(1,16) = 32.7, p \leq .0001$ . This measure is very consistent with self-report measures of musical expertise, and the range of perceptual thresholds within each group makes it a useful measure for correlation analyses. In contrast, the perceptual threshold for letters was similar across groups ( $M = 131$  and  $154$  msec for experts and novices, respectively,  $F < 1, ns$ ). Therefore, the difference in perceptual fluency with notes cannot be explained by a generally better perceptual ability in experts.

### Imaging Results

Four types of analyses were conducted. First, statistical parametric maps were generated to look for brain regions showing selectivity for musical notes in the single run. For each of these significant regions, the nature of the activity was further studied by correlating each participant's activity (notes vs. letters and symbols) with our behavioral measure of perceptual fluency for notes. Second, the same analyses were repeated with the string run. Third, ROIs (face-, letter-, and letter-string-selective regions) were identified, and the activity within these regions during the single run and string run was compared.

Finally, visual selectivity for single notes and music sequences was explored in a  $4 \times 4 \times 4$  mm<sup>3</sup> region centered on the Talairach coordinates of the TOS activation reported in Nakada et al. (1998).



**Figure 2.** Perceptual thresholds for music-reading experts and novices for the sequential matching task with four-note music sequences. The gray bars indicate the mean of each group, whereas the crosses represent individual perceptual threshold. It shows that experts can match the music sequences with a much shorter presentation duration and there is a wide range of perceptual threshold within each group. Error bars plot the SEM.

### *Statistical Parametric Maps for Single Run*

To search for brain regions selective for musical notes as a function of expertise, statistical parametric maps were generated for the interaction between Category (single notes vs. single letters and single symbols) and Group (experts vs. novices) for each voxel in the whole brain at the threshold of  $p_{\text{FDR}} < .05$ . Results revealed a widespread multimodal network of cortical areas and subcortical areas (Table 2 and Figure 3). As would be expected for expertise with visual objects, high-level visual areas including bilateral fusiform gyrus and an area along the right inferior temporal sulcus were identified as selective for musical notation. In these areas, musical notes led to higher neural responses in experts than novices, with the activation for letters and symbols similar between both groups (Figure 4A). These areas did not overlap with and were more posterior than the areas specialized for faces or letters as defined in the localizer run. Interestingly, skilled perception of musical notation also engaged early visual areas (V1/V2) bilaterally, which was never reported selective for objects of expertise in previous studies (Figure 4B). In addition, an area in the left occipito-temporal junction showed higher selectivity for musical notation for novices than experts (Figure 4N).

In addition to ventral visual areas, areas in the dorsal pathway also showed expertise effect for notes. This included bilateral occipito-parietal junction, bilateral intraparietal sulcus, left angular gyrus, and left supramarginal gyrus (Table 2). These areas consistently showed higher activity for notes compared with other stimuli in experts, whereas the activity for all conditions was similar for novices (Figure 4C).

In addition, a multimodal network of other areas revealed higher selectivity for musical notation in experts than in novices (Table 2), covering the primary and associative auditory areas along the sylvian fissure bilaterally (Figure 4D and E); the somatosensory areas in the postcentral gyrus bilaterally (Figure 4F); the premotor areas bilaterally (Figure 4G); superior temporal sulcus for audiovisual processing bilaterally (Figure 4H); the frontal areas covering different parts of the inferior frontal gyrus, middle frontal gyrus, and superior frontal sulcus (Figure 4I); the cingulate gyrus (Figure 4J); and the precuneus (Figure 4K). Furthermore, expert music-reading skills also engaged various parts of the cerebellum (Figure 4L) and the corpus callosum (Figure 4M). In contrast, a left motor area showed lower selectivity for musical notes in experts than novices (Figure 4O). This widespread multimodal network showed a higher selectivity for single musical notes in experts in simple visual tasks, demonstrating the strong and automatic association between visual processing of notes and processing in other modalities with the acquisition of musical expertise.

To investigate whether neural activity in these areas can predict performance in a visual task with notes, we examined the correlation between neural activity in these

regions and performance in our behavioral measure of perceptual fluency with notes.

To ensure that the correlations reflected an analysis independent from the ROI definition, we used a four-step method to define ROIs and to extract the response in these ROIs from a separate data set. (1) Areas in the multimodal network defined with all three single runs were used as the reference areas, generated with the interaction contrast between Category (single notes vs. single letters and single symbols) and Group (experts vs. novices) at a threshold of  $p_{\text{FDR}} < .05$ . (2) The same interaction contrast was performed with Run 1 only, and a similar network of areas was obtained, only noisier because it was based on less data. (3) Third, we examined the overlap between areas defined by Run 1 only and the reference areas. If they partially overlapped, the data for each stimulus condition from Runs 2 and 3 were extracted for correlation with behavioral performance, otherwise areas of activity were dropped. This is a conservative process that allowed us simply to reject some spurious areas of activation but has little influence on the specific voxels selected. (4) Finally, Steps 2–3 were repeated with Run 2 and Run 3 separately, and all the data representing the same reference areas were averaged.

Neural selectivity for musical notation was then calculated by subtracting the averaged activity for letters and symbols from that for musical notation. This index was correlated with our measure of perceptual fluency with notes (see Methods). As a lower perceptual threshold indicates better perceptual fluency with musical notation, a negative correlation reveals that the neural activity in a certain region is directly related to perceptual expertise. This negative correlation was significant in several brain areas associated with different modalities, including the right sylvian fissure, the left STS, the right premotor region, the right middle frontal sulcus, the right superior frontal gyrus, and the cingulate gyrus (Table 2 and Figure 5A–E). Interestingly, a significant *positive* correlation was found in the occipito-temporal area; that is, the activity for musical notation was lower with better music-reading skill in this higher visual area (Figure 5F). In contrast, the correlation did not reach significance in the face-, letter-, and letter-string-selective areas (either defined in the group level or individual level, see below) or the bilateral TOS (all  $ps > .15$ ; Figure 5G–L). These results suggest that a network of visual and nonvisual areas predicts individual perceptual fluency with musical notation, and that the neural network involved in expert perception of musical notation is distinct from those associated with face and letter expertise.

### *Statistical Parametric Maps for String Run*

Analyses proceeded for the string run just as with the single run. First, we considered the interaction contrast between Category (note sequences vs. letter strings and symbol strings) and Group (experts vs. novices) at  $p_{\text{FDR}} < .05$ .

**Table 2.** Areas in the Multimodal Network Identified with the Interaction Contrast from the Single Run

<i>Number</i>	<i>Area</i>	<i>Side</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>mm<sup>3</sup></i>	<i>Maximum t</i>	<i>Minimum p</i>	<i>R</i>	<i>p of R</i>	<i>String<sup>a</sup></i>
<i>Occipito-temporal areas</i>											
1	V1/V2	L	-16	-92	6	327	4.51	<.005			
		R	18	-93	9	133	3.94	<.01			
2	V1	L	-3	-82	1	922	5.06	<.001			
3	V1	L	-8	-73	11	559	3.8	<.05			
4	Occipital area	L	-11	-87	25	193	3.91	<.01			
		R	8	-84	36	125	3.64	<.05			
5	Occipito-temporal area	L	-37	-71	2	586	-4.74	<.005	.459	.055	.048
6	Middle temporal gyrus	L	-47	-69	14	128	4.14	<.01			
7	Fusiform gyrus	L	-40	-58	-18	247	4.29	<.005			.048
		R	36	-55	-20	145	3.99	<.01			
8	Lingual gyrus	R	17	-57	6	129	3.43	<.05			
9	Inferior temporal sulcus	R	51	-45	-1	462	4.96	<.005			.04
<i>Parietal areas</i>											
10	Occipito-parietal junction	L	-28	-66	25	129	3.35	<.05			.068
		R	27	-63	31	662	4.61	<.005			
11	Angular gyrus	L	-33	-72	41	320	4.66	<.005			.078
12	Intraparietal sulcus	L	-32	-40	55	370	4.99	<.001			
		L	-30	-38	42	636	4.39	<.005			
		R	33	-43	46	849	5.96	<.001			
13	Supramarginal gyrus	L	-42	-54	34	707	4.44	<.005			
<i>Postcentral gyrus</i>											
14	Postcentral gyrus	L	-33	-27	34	175	3.66	<.05			
		L	-43	-27	53	407	4.46	<.005			.016
		R	44	-19	52	331	4.27	<.005			
<i>Sylvian fissure</i>											
15	Sylvian fissure (BA 41)	L	-37	-22	7	317	4.61	<.005			
		R	50	-21	7	440	4.45	<.005			.056
		R	56	-20	17	717	4.39	<.005			
		R	37	-19	14	264	4.26	<.005			
16	Sylvian fissure (BA 22/44)	R	54	-1	4	996	7.57	<.001	-.633	.005	
		L	-51	11	10	971	6.85	<.001			
<i>Superior temporal sulcus (STS)</i>											
17	STS	L	-42	-49	10	749	4.95	<.005	-.52	.027	
		R	61	-41	5	392	4.29	<.005			
		R	61	-42	25	96	3.43	<.05			



**Table 2.** (continued)

Number	Area	Side	x	y	z	mm <sup>3</sup>	Maximum t	Minimum p	R	p of R	String <sup>a</sup>
18	Anterior STS	L	-57	-26	-2	500	4.32	<.005			
		R	44	-13	-10	572	4.76	<.005			
<i>Motor areas</i>											
19	Primary motor area	L	-23	-18	54	173	-3.73	<.05			
20	Premotor area	L	-30	-2	48	536	4.88	<.005			
		R	36	2	46	896	5.29	<.001	-.563	.015	
<i>Frontal areas</i>											
21	Inferior frontal gyrus	L	-52	8	29	471	4.26	<.005			
22	Inferior frontal gyrus	R	34	23	5	809	6.41	<.001			
23	Middle frontal gyrus	L	-37	26	32	389	4.74	<.005			
		R	40	20	29	340	4.18	<.01			
24	Middle frontal gyrus	R	42	38	22	730	4.12	<.01	-.482	.043	
25	Superior frontal gyrus	R	14	45	26	143	4.13	<.01	-.552	.018	
<i>Precuneus</i>											
26	Precuneus		-6	-68	31	630	4.1	<.01			
27	Precuneus		2	-43	49	710	4.81	<.005			
<i>Cingulate gyrus</i>											
28	Cingulate gyrus		-9	-51	33	635	4.83	<.005			
29	Anterior cingulate gyrus		-9	6	40	796	5.59	<.001	-.477	.045	
<i>Cerebellum</i>											
30	Cerebellum	R	6	-69	-12	660	6.11	<.001			
31	Cerebellum	R	22	-51	-19	387	3.88	<.01			
32	Cerebellum	L	-15	-58	-15	813	5.37	<.001			
33	Cerebellum	L	-6	-51	-21	303	5.01	<.001			.0038
34	Cerebellum		0	-34	-7	796	4.57	<.005			
<i>Corpus callosum</i>											
35	Corpus callosum		-18	-6	28	675	6	<.001	-.45	.06	

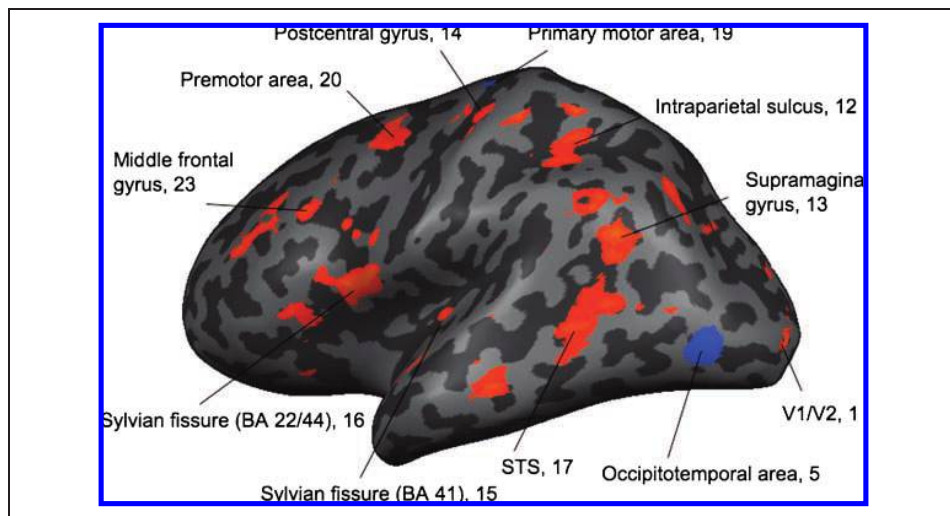
Bilateral areas are grouped under the same area number. [x y z] shows the Talairach coordinates of the peak of the clusters. The minimum p values are p values after false discovery rate (FDR) corrections. R refers to magnitude of the correlation between neural activity of the areas and individual perceptual threshold, and "p of R" shows the p values of the correlation analyses. p Values are shown only when the correlation is significant or marginally significant (p ≤ .06).

<sup>a</sup>Significant or marginally significant (p ≤ .08) interaction contrast using the data from the string run within the areas.

The expertise network specialized for music sequences was less extensive but largely overlapped with that for single musical notes (Table 3). Similar to single notes, higher selectivity for music sequences was found in the early (V1) and the late visual areas (e.g., inferior temporal gyrus) and different parietal regions. Also similar to single notes, a

multimodal network was engaged with expert perception of note sequences, covering the postcentral gyrus (somatosensory processing), the superior temporal gyrus (auditory or audiovisual areas), the premotor areas, the middle frontal gyrus, the cingulate gyrus, and the cerebellum. Interestingly, there were more areas showing a significant negative

**Figure 3.** The widespread multimodal network with the data from the interaction contrast from the single run presented on one of the music-reading expert's inflated brain (left hemisphere) at the threshold of  $p_{FDR} < .05$ . The numbers after each brain region correspond to the area numbers in Table 2. Orange clusters and blue clusters represent higher and lower selectivity for single notes for experts compared with novices, respectively.



interaction, that is, a lower selectivity for musical notation in experts than novices. However, this could be associated with the fact that the task in the string run was more difficult for novices, indicated by a worse performance with notes for novices than for experts (see Table 1).

The analysis exploring the correlation between activity in these regions and perceptual fluency with notes revealed a significant negative correlation in bilateral premotor areas (Table 3), indicating more activity in these areas with a better perceptual skill for musical notation.

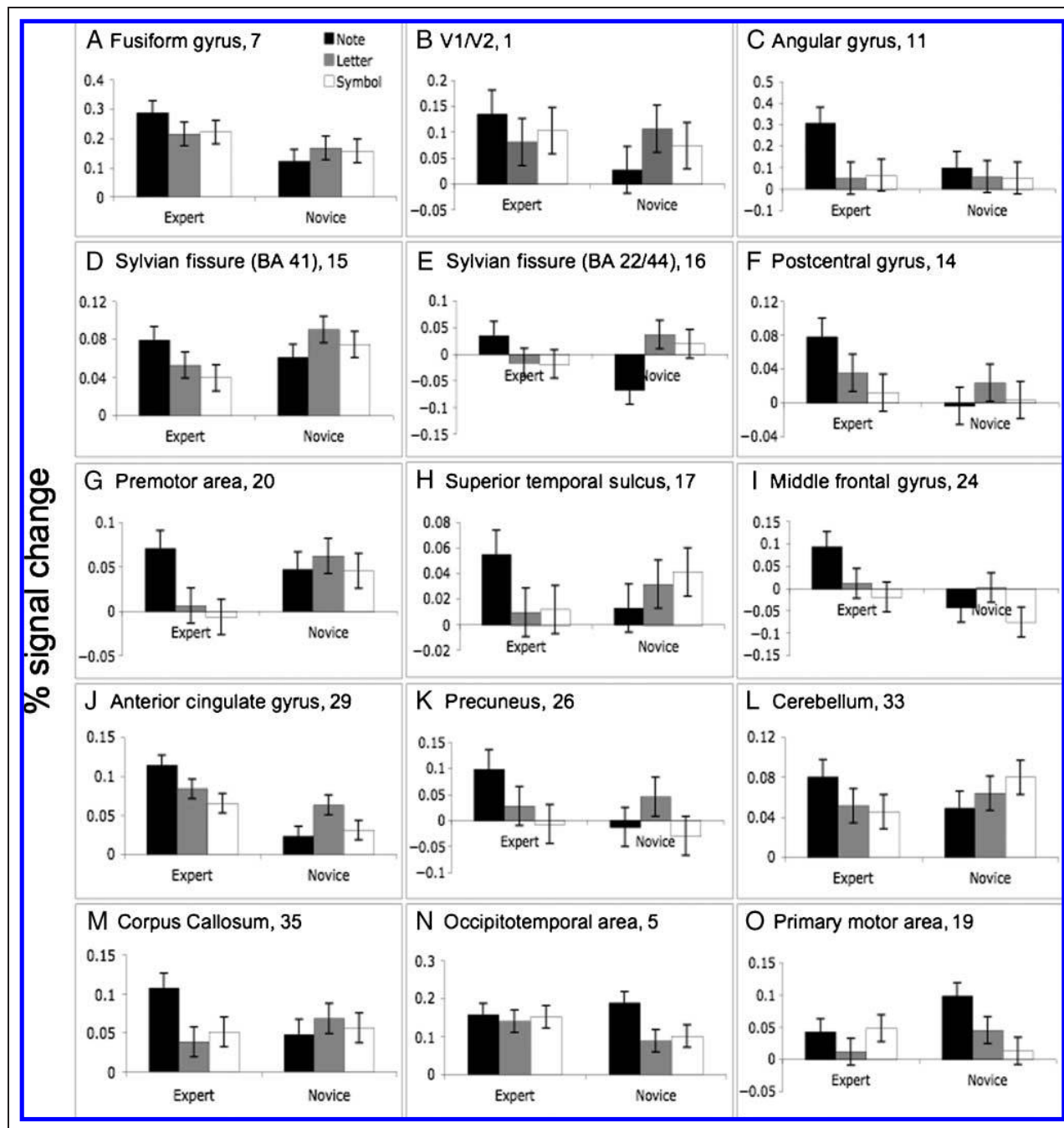
### Comparing Single and String Runs

The multimodal networks revealed in the single run and the string run are qualitatively similar, although more extensive for single notes than note sequences. The number of significant clusters localized in the string run was less than half of that identified in the single run, and the volume of activity across the whole brain showing selectivity for note sequences was 1/10 of that for single notes at a threshold of  $p_{FDR} < .05$  (2439 and 25448  $\text{mm}^3$  for the string run and the single run, respectively). Areas of significant activity were also much smaller in the string run. Half of the regions found in the string run were small clusters of less than 20  $\text{mm}^3$ , whereas in the single run, only one cluster was smaller than 100  $\text{mm}^3$  and half of the clusters were larger than 450  $\text{mm}^3$ . These results are similar to what was previously found for Roman letters (James et al., 2005): More selectivity was obtained for single notes than random sequences of notes just as more selectivity was obtained for single letters than consonant strings.

At the statistical threshold used here ( $p_{FDR} < .05$ ), 58% of the areas selective for note sequences overlapped with those selective for single notes (Table 3). Moreover, most of the areas that did not overlap were very small clusters falling adjacent to significant areas with the single network, suggesting that additional overlap might be obtained with additional power.

In addition, we extracted data from the string runs in areas identified in the single runs and submitted the data to  $2 \times 3$  ANOVAs (Group  $\times$  Category) to see whether the selectivity for single notes replicates in the string run. A higher selectivity for music sequences for experts was found in the right inferior temporal area and the cerebellum and was marginally significant (with  $p$  values around .06–.08) in an occipito-parietal area, a superior parietal area, and an early auditory area along the sylvian fissure (Table 2). The same analysis was performed with data from the single runs in areas identified in the string runs. Only the cingulate gyrus and the cerebellum showed a higher selectivity for single notes for experts (Table 3). Replication of selectivity by looking at activity for single notes in areas found to be engaged for note sequences is limited by the fact that areas of activation tended to be smaller for strings than for single notes. The extensive overlap of the networks identified in the separate analyses for single notes and sequences and the replication of selectivity for music sequences in at least part of this network suggest that the multimodal network recruited by single notes is at least partially shared with that engaged by the perception of music sequences.

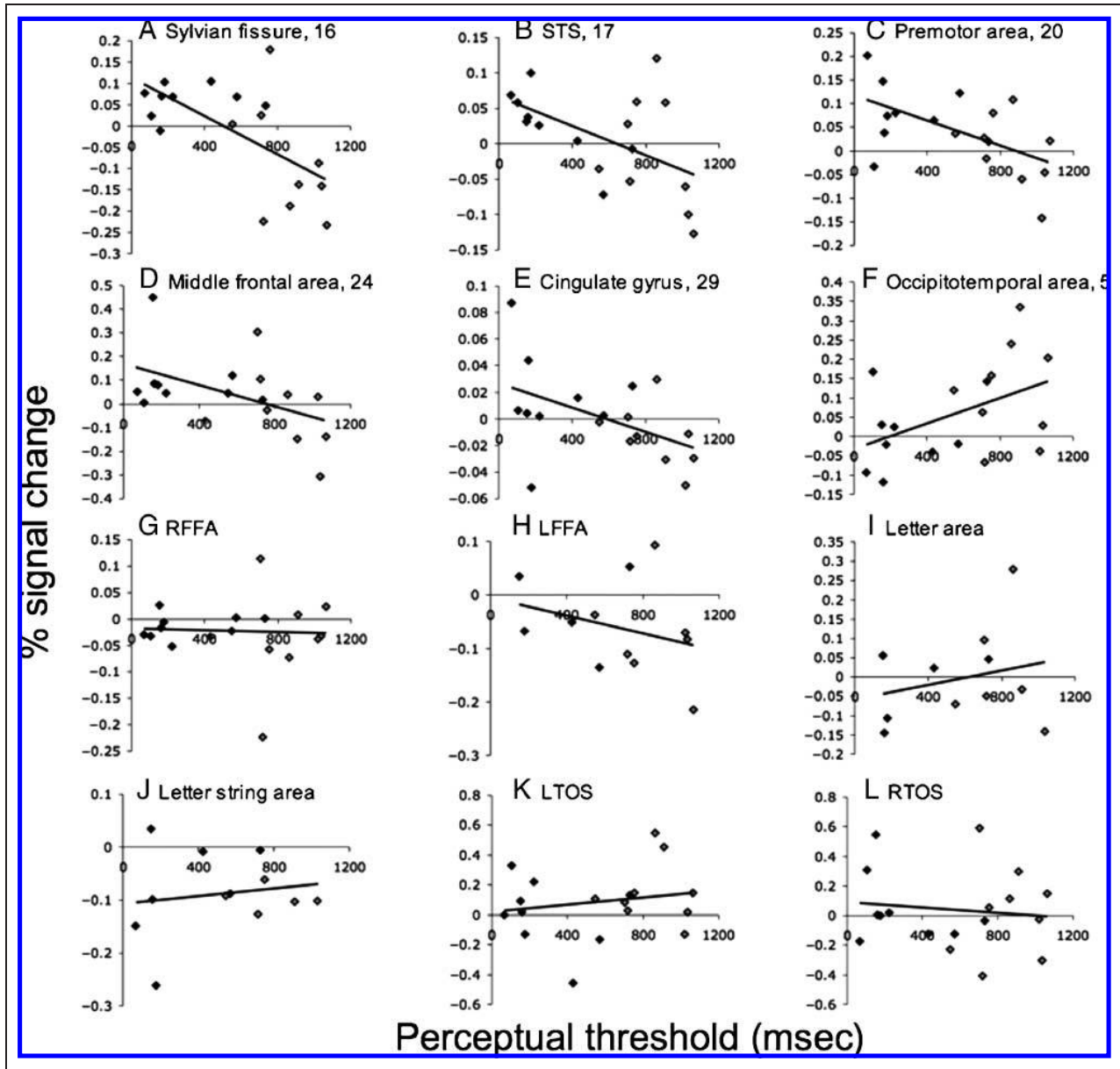
In sum, just as with letters, which showed more extensive selectivity for single letters than consonant strings (James et al., 2005), we observe a more extensive expertise effect for the perception of single notes than for randomly generated sequences of notes. Why do single stimuli elicit stronger category-specific responses than sequences for both domains of expertise? A possible explanation is that the differences in appearance and processes involved with different object categories are obscured by the higher similarity of various types of sequences. When categories of stimuli are compared in the context of sequences, their global visual appearance becomes more similar, as they are all strings of characters and thereby share more energy in the low spatial frequency range. The use of sequences may also trigger similar visual



**Figure 4.** Neural activity in various areas identified in the multimodal network for single run for each stimulus condition (single notes, single letters, and single symbols) for each group. The numbers next to the brain regions correspond to the area numbers in Table 2. (A–M) Higher selectivity for musical notation for experts compared with novices. (N–O) Higher selectivity for musical notation for novices than experts. Error bars plot the *SEM* associated with the Group × Category interaction.

processes, such as sequential processing or comparison of individual components within the sequences. The use of long sequences or of visually similar controls could be one of the reasons why previous studies did not reveal as extensive a network as we obtained here (Stewart et al., 2003; Nakada et al., 1998). Nonetheless, it is worth mentioning that one difference between notes and letters is that for musical notation, the network revealed by single notes and se-

quences of notes appears to be generally shared, if more activated by single notes. In contrast, evidence of double dissociations was obtained for areas selective for single letters and strings, in particular along the occipito-temporal cortex (James et al., 2005). Whether this difference is real or depends on differences in task or spurious variability could be explored in more detailed analyses on individual subjects' data, but this goes beyond the scope of the present work.



**Figure 5.** Correlation analyses between neural activity in various areas in the single run and the individual perceptual threshold. Black and gray dots represent experts and novices, respectively. (A–F) Areas identified from the multimodal network show a significant correlation, whereas none of the correlations reached significance in control areas including the face-selective (G–H), letter-selective (I), and letter-string-selective (J) areas and the bilateral TOS (K–L). RFFA = right fusiform face area; LFFA = left fusiform face area; LTOS = left transverse occipital sulcus; RTOS = right transverse occipital sulcus.

#### Face-, Letter-, and Letter-String-selective Regions

We examined whether musical notation engages areas selective for other domains of expertise: faces, letters, and letter strings. At individual level, the left and the right FFA was defined using the [faces vs. objects, faces vs. fixation] contrast in the localizer run at  $t \geq 2.8$  from all subjects for the right FFA and 14 subjects for the left FFA. The letter area was defined with the contrast [letters vs. faces, objects, and symbols, letters vs. fixation] at  $t \geq 2.0$  from 13 subjects. The letter string area was defined using the [letter strings vs. note se-

quences and symbol strings, letter strings vs. fixation] contrast in the string run at  $t \geq 2.6$  from 14 subjects. Results showed that these regions were not selective for musical notes (Figure 6A–D). The lack of selectivity for musical notation was also found with ROIs defined at the group level.

#### Transverse Occipital Sulcus

In Nakada et al. (1998), eight music-reading experts but none of the eight novices showed selectivity for musical

**Table 3.** Areas in the Multimodal Network Identified with the Interaction Contrast from the String Run

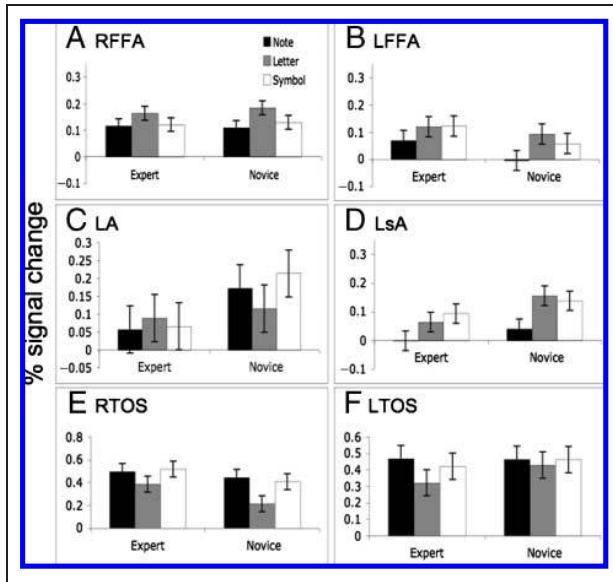
<i>Number</i>	<i>Area</i>	<i>Side</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>mm<sup>3</sup></i>	<i>Maximum t</i>	<i>Minimum p</i>	<i>R</i>	<i>p of R</i>	<i>Overlap<sup>a</sup></i>	<i>Single<sup>b</sup></i>
<i>Occipito-temporal areas</i>												
36	VI		0	-91	-5	3	3.81	<.05				
37	Middle temporal gyrus	L	-39	-72	18	13	4.2	<.05				
38	Middle temporal gyrus	R	59	-55	1	42	4.72	<.05				
39	Inferior temporal gyrus	R	49	-25	-7	5	4.25	<.05				
<i>Parietal areas</i>												
40	Occipito-parietal area	L	-26	-63	22	123	4.73	<.05			y	
41	Intraparietal sulcus	R	31	-69	32	203	4.84	<.05			y	
42	Inferior parietal	R	42	-30	31	3	-4.07	<.05				
<i>Postcentral gyrus</i>												
43	Postcentral gyrus	L	-54	-16	34	5	4.06	<.05				
<i>Sylvian fissure</i>												
44	Sylvian fissure	R	37	-37	28	2	-3.82	<.05				
<i>Superior temporal gyrus</i>												
45	STS	L	-50	-48	3	160	4.59	<.05			y	
<i>Motor areas</i>												
46	Primary motor area	L	-15	-28	56	2	-3.86	<.05				
47	Premotor area	L	-21	-8	38	309	6.61	<.001	-.53	.025	y	
48	Premotor area	L	-46	-1	29	110	4.9	<.05			y	
49	Premotor area	R	40	4	34	327	4.79	<.05	-.45	.059	y	
50	Premotor area	L	-50	4	43	171	4.61	<.05				
<i>Frontal areas</i>												
51	Middle frontal gyrus	R	42	20	31	2	4.08	<.05			y	
<i>Precuneus</i>												
52	Precuneus	L	-11	-58	37	2	-3.92	<.05				
<i>Cingulate gyrus</i>												
53	Cingulate gyrus	L	-10	11	39	105	5.12	<.01			y	.01
54	Cingulate gyrus	L	-2	9	41	39	4.32	<.05			y	.01
<i>Cerebellum</i>												
55	Cerebellum	L	-2	-59	-19	219	5.69	<.005			y	
56	Cerebellum	L	-3	-47	-20	564	6.78				y	
57	Cerebellum	R	28	-52	-23	9	4.06	<.05			y	.07
58	Cerebellum	R	39	-49	-23	3	3.88	<.05			y	.02
59	Cerebellum	R	11	-64	-17	18	3.76	<.05			y	

[x y z] shows the Talairach coordinates of the peak of the clusters. The minimum *p* values are *p* values after false discovery rate (FDR) corrections. *R* refers to magnitude of the correlation between neural activity of the areas and individual perceptual threshold, and "*p of R*" shows the *p* values of the correlation analyses. *p* Values are shown only when significant or marginally significant ( $p \leq .06$ ).

<sup>a</sup>y indicates areas overlapped with the areas identified in the single run.

<sup>b</sup>Significant (or marginally significant,  $p \leq .07$ ) interaction contrast using the data from the single run in these areas.





**Figure 6.** Neural activity for musical notation, letters, and mathematical symbols in the single run in the individually defined face-selective (A–B), letter-selective (C), and letter-string-selective (D) areas and bilateral TOS (E–F). Error bars plot the *SEM* associated with the Group  $\times$  Category interaction. Unlike areas covered in the multimodal network, none of these areas show a higher selectivity for musical notation for experts. RFFA = right fusiform face area; LFFA = left fusiform face area; LA = letter area; LsA = letter string area; LTOS = left transverse occipital sulcus; RTOS = right transverse occipital sulcus.

notation around the right TOS. This result was not replicated here, as the interaction of Group  $\times$  Category did not reveal any areas around the TOS in either experimental run. To further examine whether we could replicate this finding, a  $4 \times 4 \times 4 \text{ mm}^3$  ROI centered at  $[\pm 36, -87, 4]$  was defined bilaterally, the center of which was the Talairach Coordinates approximated from Figure 3 in Nakada et al. (1998).

We found no selectivity for musical notation in experts in this ROI (Figure 6E and F). A  $2 \times 3$  ANOVA (Group  $\times$  Category) was conducted in bilateral TOS for each experimental run. In the single run, the main effect of Category was significant in the left TOS,  $F(2,36) = 7.60, p = .0018$ . Scheffé tests ( $p < .05$ ) showed a smaller activation for single letters than both single notes and single symbols. In the string run, the main effect of category was significant in the left TOS,  $F(2,36) = 5.64, p < .0074$ , and the right TOS,  $F(2,36) = 16.93, p < .0001$ . Scheffé tests ( $p < .05$ ) indicated a larger activation for note sequences than letter strings in bilateral TOS for both groups and for symbol strings than letter strings only in RTOS. For both single run and string run, the Group  $\times$  Category interaction was not significant ( $p > .4$ ) for bilateral TOS.

Therefore, similar to Nakada et al. (1998), we found a higher activity for notes than Roman letters (English and Japanese words in their study) in the TOS. The difference is that we found this effect in both experts and novices;

that is, the selectivity for notes was not exclusive for experts and that this area was also selective for symbols compared with letters.

## DISCUSSION

The goal of this study was to investigate expertise effects during visual judgments with musical notation in both visual and nonvisual areas of the brain. By studying musical expertise, an ability of a particularly multimodal nature, we tested whether activity in both visual and nonvisual areas predicts fluency in a visual task with objects of expertise. A widespread multimodal network of brain regions was identified with higher selectivity for musical notation in music-reading experts than novices. This network included visual, auditory, audiovisual, somatosensory, motor, parietal, and frontal areas. Although we assume that specialization in this multimodal network depends on experience with reading, playing, and hearing music, it is striking that visual stimuli as simple as single notes or five-note music sequences are enough to activate this distributed network. This and the fact that one of the tasks we used (the gap detection task) did not even require participants to attend to the musical notes suggest that the recruitment of these areas is relatively automatic.

### Musical Notation: A Different Kind of Visual Expertise?

We observed expertise effects in various visual areas selective for musical notation for experts but not novices, including bilateral early visual areas, bilateral fusiform gyrus, and inferior temporal areas. However, we failed to replicate the expertise effect for the reading of musical scores reported by a previous study in the TOS (Nakada et al., 1998). This could be explained by our use of a more engaging task for both groups. Nakada et al. (1998) used a passive viewing task, which likely led to experts paying much higher attention reading the musical scores than novices (to whom musical scores were not meaningful). In contrast, our study used either one-back or gap detection task that involved simple visual judgments and in which both experts and novices performed well. Therefore, our tasks possibly engaged both groups in a more comparable way. Furthermore, with an additional stimulus condition, mathematical symbols, we found that the TOS responded similarly to notes and symbols (both higher than letters) in both groups. It suggests that this area is not specially recruited for musical notes per se or for expert perception of musical notation.

Although previous studies suggest that face selectivity is stronger in the right hemisphere (Kanwisher et al., 1997) and letter selectivity is left dominant (Cohen et al., 2000; see review in Wong et al., in press), we did not observe a hemispheric asymmetry in the processing of notes. Expert perception of musical notation recruited ventral temporal areas in both hemispheres (with peaks

at  $[-40, -58, -18]$  and  $[36, -55, -20]$  for left and right areas, respectively) posterior to the face- and the letter-selective areas. Both the location of these foci and their level of selectivity for notes are similar across hemispheres.

Neural areas engaged by expertise for musical notation do not overlap with areas showing selectivity for objects in other domains of expertise such as faces and letters. Although this difference could be attributed to the differences in shape between these domains, there are reasons to believe that this is not the driving factor. Indeed, objects that vary greatly in geometry, such as birds, cars, and novel objects called Greebles (Xu, 2005; Gauthier et al., 1999, 2000), all engage the same visual area as faces, provided that participants are expert at discriminating between exemplars in these object domains. In addition, simple Chinese characters (in Chinese readers) as well as digits engage the same area as single Roman letters (Wong et al., in press; James et al., 2005). The recruitment of the letter- and the face-selective areas may be driven more by processing than by geometry. The processes recruited by face-like expertise are likely to be those that facilitate the individuation of visually similar objects (Gauthier, 2000). The processes recruited by letter-like expertise, in contrast, may be those involved in rapid categorization of shapes, in the context of regularity in the font, size, and orientation (for further discussion, see Gauthier, Wong, Hayward, & Cheung, 2006; Wong & Gauthier, 2006). In the case of expertise with musical notation, we lack behavioral studies that would characterize the factors that distinguish their processing from other kinds of object recognition. It is certainly possible to speculate; for instance, visual-motor transcoding could be much more important for musical notation than other objects, as musical notation is often used as a cue for motor execution in musicians. Some argued that practice in this task creates a general advantage for speeded motor responses to any visual stimuli (Brochard, Dufour, & Després, 2004). In addition, musical notation is generally presented on a five-line staff, in which spatial positions and spaces between clustering of notes are useful cues in music reading (Sloboda, 1981). Therefore, the use of spatial information in the perception of musical notation could be more important than in most other object domains, perhaps with the exception of chess expertise, in which the spatial configuration of chess pieces is thought to be important (Chase & Simon, 1973). Previous neuroimaging studies using musical notation revealed dorsal pathway activity, consistent with the involvement of spatial processes in music reading (Stewart et al., 2003; Schön et al., 2002; Sergent et al., 1992).

The recruitment of early, retinotopic visual areas has not been reported in previous studies with objects of expertise. It suggests that processing in these retinotopic areas is important for music reading. The selectivity for musical notation is not simply a higher response to the five-line staff, as this selectivity was also common in the condition where all stimuli had an identical five-line staff background. One possibility is that music reading involves

simultaneous recognition of multiple musical notes located on different positions of the staves and different parts of the visual field. With extensive experience, this task demand may eventually lead to the establishment of multiple representations of musical notation in the retinotopic cortex. Indeed, perceptual learning in a visual search task revealed performance-relevant increases in retinotopic cortex (Sigman et al., 2005).

In at least a very general sense, the specialization in different areas for letter perception, for face recognition and individuation in other domains, and for the perception of musical notation is consistent with the process map hypothesis (Gauthier, 2000). The process map hypothesis suggests that category selectivity reflects the effect of the differences in tasks associated with various categories during the development of perceptual expertise. With extensive experience, visual presentation of objects of expertise may automatically recruit brain regions that were once necessary to perform the required task, even if the task is no longer currently required. If the perception of faces, letters, and notes typically occurs in tasks that differ in their demands, then different visual areas should be specialized.

### Different Patterns for Expertise Effects

The most intuitive expertise effect that we could obtain is one where notes elicit more activity than control stimuli, only in experts (e.g., Gauthier et al., 2000). Indeed, most areas showed a response of this type, including early and high-level visual areas, all the parietal areas, the postcentral gyrus, part of the sylvian fissure, the superior temporal gyrus, the premotor areas, the frontal areas, the precuneus, the cingulate gyrus, the cerebellum, and the corpus callosum. However, we also found other patterns of expertise effect in this study. For example, novices responded less to notes compared with letters and symbols in various areas, including early visual areas (Figure 4B), part of the sylvian fissure (Figure 4E), an inferior frontal area (area 22, Table 2), and part of the cerebellum (area 34, Table 2). This pattern suggests a familiarity effect: Novices are less familiar with musical notation compared with other visual conditions.

A third pattern was obtained in the left occipito-temporal junction (Figure 4N) and a primary motor area (Figure 4O), where no difference between notes and control stimuli was observed for experts, whereas notes elicited a stronger response than letters and symbols in novices. The occipito-temporal area even reflected perceptual ability of reading music, with higher activity reflecting a lower perceptual fluency for musical notation. A decrease in neural activity for objects of expertise has been found in several visual studies. For example, a visual training with Korean characters (with naïve subjects) found decreased activity in the left inferior temporal gyrus (Xue & Poldrack, 2007). In a study that trained participants to discriminate gratings with one training session, a decrease in activity was found in both early and late visual areas (Mukai et al., 2007), which

was correlated with behavioral improvement (similar to our finding in the occipito-temporal cortex). Reduced activity in the motor cortex for musicians compared with nonmusicians has been reported in a simple motor tapping task (Hund-Georgiadis & von Cramon, 1999). Different authors offer different interpretations of what decreased neural activity with practice implies. Some proposed the possibility that it reflects sharpened neuronal tuning curves for the trained stimuli (e.g., Schiltz et al., 1999) or the computations in the area have become more efficient with expertise (Zatorre et al., 2007), whereas others assumed the processing of the trained stimuli has shifted to other areas (Sigman et al., 2005). Further work, for example, using a repetition suppression design, could test some of these alternatives.

### **Characterizing the Multimodal Network for Musical Expertise**

Similar to other domains of perceptual expertise such as faces and letters (James & Gauthier, 2006; Haxby et al., 2002), musical notation automatically engages an extensive multimodal network of areas. Various areas outside the visual cortex, including the primary and associative auditory areas, the somatosensory areas, the audiovisual areas, the parietal areas, the premotor areas, other frontal areas, the precuneus, the cingulate gyrus, and the cerebellum, all showed selectivity for musical notation compared with control visual stimuli. One characteristic of this multimodal network is that most of the areas were found bilaterally (Table 2).

Our design cannot resolve the function of each of these areas in the processing of musical notation, but prior neuroimaging work in the musical domain suggests important roles of some of these regions in different aspects of musical expertise. Parietal areas may be recruited because of the visuospatial nature of musical notation (Stewart et al., 2003; Schön et al., 2002; Sergent et al., 1992). The supramarginal gyrus could be involved in visuomotor processing, as it was recruited after training with reading music and playing the keyboard (Stewart et al., 2003) and was also engaged in writing (Sugihara, Kaminaga, & Sugishita, 2006; Katanoda, Yoshikawa, & Sugishita, 2001). Primary and associative auditory areas are important for pitch and rhythmic processing (Zatorre et al., 2007) and have an increased representation after musical training (Pantev et al., 1998) in a task that did not require attention to the auditory meaning of musical notation. Activity for reading musical notes in these areas suggests that, with expertise, notes become truly multimodal objects. The superior temporal gyrus (STS) is often characterized as a key multisensory area. Single-cell studies find that the whole STS generally responds to audiovisual stimuli, with the posterior part more engaged by visual stimuli and the anterior part by auditory stimuli (see review in Beauchamp, 2005). We observed activity in both anterior and posterior portions of the STS, again consistent with the idea that the auditory

meaning of notes is easily accessed from visual presentation alone.

In the somatosensory cortex, increased cortical representation for the fingers of the left hand has been reported in string players (Elbert et al., 1995). The music-reading experts in this study were trained on various instruments, including piano, flute, percussion, guitar, bassoon, etc., and most reported playing multiple instruments. Nevertheless, most of the musical instruments require skilled movements of fingers and arms, which could underlie the automatic recruitment of the somatosensory areas with coordinates roughly matching finger or arm representations. We also observed expertise effects in several areas involved in motor execution. Premotor regions are thought to be important for internal planning of learned sequences (Roland & Zilles, 1996) and auditory-motor interactions (Zatorre et al., 2007). The cerebellum is involved in the coordination of movement, including synchronization of movement in time (Ramnani, Toni, Passingham, & Haggard, 2001) and rhythm and motor skill acquisition (Sakai, Hikosaka, & Nakamura, 2004). The anterior cingulate gyrus is important for monitoring consequences of actions and error detection (Ito, Stuphorn, Brown, & Shall, 2003; Carter et al., 1998). These motor-related regions could be recruited because of long-term experience with motor execution in response to visual perception of musical notation (Levine, Morsella, & Barge, 2007). The extent of activity in motor areas in our simple visual tasks is surprising. In fact, a recent study suggests that reading music sequences produce a larger electromyography response at the throat than text reading or a Math task, suggesting that reading musical melodies elicits subvocal activity (Brodsky, Kessler, Rubinstein, Ginsborg, & Henik, 2008). Whether the perception of single musical notation may automatically elicit similar muscle activity, outside of a motor task, is an interesting question for future research. Likewise, the expertise effects in inferior and superior frontal regions in our study suggest that, despite the simple nature of the stimuli and tasks, more complex processes important in musical expertise were engaged. These areas have been associated with higher level functions such as the processing of temporal coherence (Levitin & Menon, 2003), tonality structure (Janata et al., 2002), and syntax (Patel, 2003; Bookheimer, 2002).

Given the recruitment of areas in distant parts of the brain, it is interesting to note an expertise effect in the corpus callosum of experts viewing musical notation. As hemodynamic activity is thought to underlie fMRI BOLD signal (e.g., Buxton, Uludag, Dubowitz, & Liu, 2004), white matter activity detected with fMRI is often ignored. However, the expertise effect in the corpus callosum was one of the strongest effects we observed (maximum  $t$  value = 6.0 compared with 4.69, the averaged maximum  $t$  value of all significant areas), and it even showed a marginally significant correlation with individual perceptual threshold ( $r = -.451$ ,  $p = .06$ ). Examination of individual subjects' data indicated that activity in this

area did not overlap with gray matter in any of the subjects. In fact, fMRI activity in the corpus callosum has been reported and replicated (Aramaki, Honda, Okada, & Sadato, 2006; Weber et al., 2005; Tettamanti et al., 2002), and several possible mechanisms have been proposed for BOLD signal in white matter (Tettamanti et al., 2002). Furthermore, anatomical evidence supports the relationship between the corpus callosum and the musical training. For example, musicians have a larger anterior corpus callosum compared with nonmusicians ( Schlaug et al., 1995), and the white matter density in the corpus callosum is correlated with the amount of time practicing piano (Bengtsson et al., 2005). Our findings, therefore, suggest that interhemispheric transfer of information through the corpus callosum is a feature of the automatic response of the musical brain to the presentation of musical notation.

### **Nonvisual Areas Correlate with Fluency in a Visual Task**

Our correlation analyses with perceptual fluency for notes revealed that those subjects showing most selectivity for notes in auditory, audiovisual, motor, and frontal regions (and the least selectivity in occipito-temporal cortex) were also those who could match note sequences at the shortest presentation durations. There are several possible interpretations to this finding. One is that with increasing musical expertise, notes become multimodal objects that are represented in nonvisual areas across the brain such that the nonvisual areas participate in or even facilitate the decisional processes during visual judgments. In that sense, music-reading experts could automatically and rapidly recode a visual object as a multimodal concept, regardless of whether it is a simple visual task that even novices can perform well (e.g., the one-back task) or a difficult task that experts tend to outperform novices (e.g., the threshold task). Indeed, previous studies have provided evidence that established multimodal associations can facilitate visual perception. For example, associating nonvisual information with visual shapes can facilitate visual judgments (Gauthier, James, Curby, & Tarr, 2003). Practice with audiovisual stimuli can facilitate subsequent performance when only visual information is available (Seitz, Kim, & Shams, 2006). Therefore, a more effective recruitment of this multimodal musical network may actually facilitate the visual perception of notes and in turn lead to a higher perceptual fluency in experts.

A different interpretation is that the activity represents a somewhat epiphenomenal reactivation of a network of areas that generally functions together during musical experience (Rogers, Morgan, Newton, & Gore, 2007) or a full representation of a multimodal concept that involves both sensory and motor information (Mahon & Caramazza, 2008), with little causal influence on visual fluency. It is even possible that activity in some other visual areas was more directly and causally associated with visual perfor-

mance, but we failed to detect it because of individual variability in its location. Thus, we cannot reject the possibility that visual performance in matching notes depends on the activity of visual areas, which in turn spreads to the rest of the network associated with musical expertise. Further work with ERPs or TMS may be able to test the functional role of nonvisual areas in the visual perception of musical notation.

### **Comparing the Multimodal Networks for Music and Speech**

Apart from musical expertise, speech perception is another domain that involves different multimodal processes with which most people have extensive experience (Rosenblum, 2008). Similar to music processing, listening to speech elicits a widespread network of areas, including primary and associative auditory areas, superior temporal gyrus, premotor areas, and inferior frontal gyrus (e.g., Callan et al., 2006; Hickok, Buchsbaum, Humphries, & Muftuler, 2003). This network overlaps greatly with that for music processing, at least qualitatively (see the previous discussion and Table 2), although it has been argued that speech and music are relatively specialized in the left and the right auditory cortex, respectively (Zatorre, Belin, & Penhune, 2002). Interestingly, perceiving the visual inputs related to speech also activates part of this multimodal network related to speech. For example, silent lip reading elicits response in the primary and associative auditory areas that are also active for heard speech (Calvert et al., 1997). Judgments on visually presented strings activate the inferior frontal gyrus, which is also engaged when listening to aurally presented strings in similar tasks (Booth et al., 2004). Presenting a single letter visually activates the STS and the superior temporal gyrus, which are related to the processing of speech sound (Van Atteveldt, Formisano, Goebel, & Blomert, 2004). In sum, it is common for both music and speech that simple visual stimuli (a single note or a single letter) can recruit extensive multimodal networks that are not directly relevant to the task at hand.

### **Acknowledgments**

This work was supported by grants from the James S. McDonnell Foundation to the Perceptual Expertise Network; by the Temporal Dynamics of Learning Center (NSF Science of Learning Center SBE-0542013); and by a Discovery award from Vanderbilt University.

Reprint requests should be sent to Yetta K. Wong, Department of Psychology, Vanderbilt University, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240-7817, or via e-mail: yetta.wong@vanderbilt.edu.

### **REFERENCES**

- Aramaki, Y., Honda, M., Okada, T., & Sadato, N. (2006). Neural correlates of the spontaneous phase transition during bimanual coordination. *Cerebral Cortex*, *16*, 1338–1348.



- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.
- Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, *15*, 145–153.
- Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullen, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, *8*, 1148–1150.
- Bookheimer, S. Y. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188.
- Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2004). Development of brain mechanisms for processing orthographic and phonologic representations. *Journal of Cognitive Neuroscience*, *16*, 1234–1249.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, *16*, 4207–4221.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical experience on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, *54*, 103–109.
- Brodsky, W., Kessler, Y., Rubinstein, B.-S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 427–445.
- Buxton, R. B., Uludag, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, *23*(Suppl. 1), S220–S233.
- Callan, D. E., Tsytasarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, *31*, 1327–1342.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*, 593–596.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*, 913–919.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.
- 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.
- Deutsch, D. (1998). *The psychology of music* (2nd ed.). London: Academic Press.
- Downing, P. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.
- Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*, 305–307.
- 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.
- Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex? *Trends in Cognitive Sciences*, *4*, 1–2.
- 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., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, *20*, 507–523.
- 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., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–573.
- Gauthier, I., Wong, A. C.-N., Hayward, W. G., & Cheung, O. S. (2006). Font tuning associated with expertise in letter perception. *Perception*, *35*, 541–559.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*, 59–67.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*, 673–682.
- Hund-Georgiadis, M., & von Cramon, D. Y. (1999). Motor-learning-related changes in piano players and non-musicians revealed by functional magnetic-resonance signals. *Experimental Brain Research*, *125*, 417–425.
- Ito, S., Stuphorn, V., Brown, J. W., & Shall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, *302*, 120–122.
- 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.
- James, T. W., & Gauthier, I. (2003). Auditory and action semantic features activate sensory-specific perceptual brain regions. *Current Biology*, *13*, 1792–1796.
- Janata, P., Birk, J. K., van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying western music. *Science*, *298*, 2167–2170.
- 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*, 891–903.
- 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.
- Katanoda, K., Yoshikawa, K., & Sugishita, M. (2001). A functional MRI study on the neural substrates for writing. *Human Brain Mapping*, *13*, 34–42.
- Levine, L. R., Morsella, E., & Barge, J. A. (2007). The perversity of inanimate objects: Stimulus control by incidental musical notation. *Social Cognition*, *25*, 267–283.
- Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *Neuroimage*, *20*, 2142–2152.
- Longcamp, M., Anton, J. L., Roth, M., & Velay, J. L. (2005). Premotor activations in response to visually presented



- single letters depend on the hand used to write: A study on left-handers. *Neuropsychologia*, *43*, 1801–1809.
- Longcamp, M., Tanskanen, T., & Hari, R. (2006). The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *Neuroimage*, *33*, 681–688.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology (Paris)*, *102*, 59–70.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*, 649–652.
- Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience*, *26*, 11187–11196.
- Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *Journal of Neuroscience*, *27*, 11401–11411.
- Nakada, T., Fujii, Y., Suzuki, K., & Kwee, I. L. (1998). “Musical brain” revealed by high-field (3 Tesla) functional MRI. *NeuroReport*, *9*, 3853–3856.
- 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.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*, 811–814.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681.
- 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.
- Peretz, I., & Zatorre, R. J. (2003). *The cognitive neuroscience of music*. New York: Oxford University Press.
- Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D’Esposito, M., Detre, J. A., et al. (2002). Neural specialization for letter recognition. *Journal of Cognitive Neuroscience*, *14*, 145–159.
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*, 5205–5215.
- Ramnani, N., Toni, I., Passingham, R. E., & Haggard, P. (2001). The cerebellum and parietal cortex play a specific role in coordination: A PET study. *Neuroimage*, *14*, 899–911.
- Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology*, *16*, 408–414.
- Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, *25*, 1347–1357.
- Roland, P. E., & Zilles, K. (1996). Functions and structures of the motor cortices in humans. *Current Opinion in Neurobiology*, *6*, 773–781.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409.
- Sakai, K., Hikosaka, O., & Nakamura, K. (2004). Emergence of rhythm during motor learning. *Trends in Cognitive Sciences*, *8*, 547–553.
- Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., et al. (1999). Neuronal mechanisms of perceptual learning: Changes in human brain activity with training in orientation discrimination. *Neuroimage*, *9*, 46–62.
- Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*, 699–701.
- Schön, D., Anton, J. L., Roth, M., & Besson, M. (2002). An fMRI study of music sight-reading. *NeuroReport*, *13*, 2285–2289.
- Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, *16*, 1422–1427.
- Sergent, J., Zuck, E., Terriah, S., & MacDonald, B. (1992). Distributed neural network underlying musical sight-reading and keyboard performance. *Science*, *257*, 106–109.
- 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.
- Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*, 1602–1608.
- Sloboda, J. (1981). The uses of space in music notation. *Visible Language*, *25*, 86–110.
- 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.
- Sugihara, G., Kaminaga, T., & Sugishita, M. (2006). Interindividual uniformity and variety of the “writing center” a functional MRI study. *Neuroimage*, *32*, 1837–1849.
- Tettamanti, M., Paulesu, E., Scifo, P., Maravita, A., Fazio, F., Perani, D., et al. (2002). Interhemispheric transmission of visuomotor information in humans: fMRI evidence. *Journal of Neurophysiology*, *88*, 1051–1058.
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*, 271–282.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.
- Weber, B., Treyer, V., Oberholzer, N., Jaermann, T., Boesiger, P., Brugger, P., et al. (2005). Attention and interhemispheric transfer: A behavioral and fMRI study. *Journal of Cognitive Neuroscience*, *17*, 113–123.
- Wong, A. C.-N., Jobard, G., James, K. H., James, T. W., & Gauthier, I. (in press). Expertise with characters in alphabetic and nonalphabetic writing systems engage overlapping occipito-temporal areas. *Cognitive Neuropsychology*.
- Wong, C.-N., & Gauthier, I. (2006). An analysis of letter expertise in a levels-of-categorization framework. *Visual Cognition*, *15*, 854–879.
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*, 1234–1242.
- 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*, 1643–1655.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547–558.
- Zatorre, R. J., & Peretz, I. (2001). *The biological foundations of music* (Vol. 930). New York: New York Academy of Sciences.