Music and words in the visual cortex: the impact of musical expertise

Valeria Mongelli, Stanislas Dehaene, Fabien Vinckier, Isabelle Peretz, Paolo Bartolomeo, Laurent Cohen

Music and words in the visual cortex:

the impact of musical expertise

Valeria Mongelli [1,2,4,7], Stanislas Dehaene [6], Fabien Vinckier [1,2,4,7], Isabelle Peretz [5], Paolo Bartolomeo [1,2,4,7,8], Laurent Cohen [1,2,3,4,7]

[1] *Inserm, U 1127, F-75013, Paris, France*

[2] *Sorbonne Universités, UPMC Univ Paris 06, UMR S 1127, F-75013, Paris, France*

[3] *AP-HP, Hôpital de la Pitié Salpêtrière, Department of Neurology, F-75013, Paris, France*

[4] *CNRS, UMR 7225, F-75013, Paris, France*

[5] *International Laboratory for Brain, Music and Sound Research (BRAMS), Department of Psychology, University of Montreal, Montreal, Canada*

[6] *Cognitive Neuroimaging Unit, CEA DSV/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France*

[7] *Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France*

[8] *Department of Psychology, Catholic University, Milan, Italy*

Corresponding author: laurentcohen2@gmail.com

Abstract

How does the human visual system accommodate expertise for two simultaneously acquired symbolic systems? We used fMRI to compare activations induced in the visual cortex by musical notation, written words and other classes of objects, in professional musicians and in musically naïve controls. First, irrespective of expertise, selective activations for music were posterior and lateral to activations for words in the left occipitotemporal cortex. This indicates that symbols characterized by different visual features engage distinct cortical areas. Second, musical expertise increased the volume of activations for music and led to an anterolateral displacement of word-related activations. In musicians, there was also a dramatic increase of the brain-scale networks connected to the music-selective visual areas. Those findings reveal that acquiring a double visual expertise involves an expansion of category-selective areas, the development of novel long-distance functional connectivity, and possibly some competition between categories for the colonization of cortical space.

Keywords

1 Introduction

A substantial part of individual differences in cognitive abilities results from the individual practice of highly trained skills. The brain correlates of such acquired expertise have been accessible to anatomical and functional brain imaging for over two decades, in a broad variety of domains ranging from navigation in space (Maguire et al., 2000), juggling (Gerber et al., 2014), to olfactory (Delon-Martin et al., 2013) and visual (Gauthier et al., 2000; Harel et al., 2013) expertise. Word reading may be the culturally most important and widely shared expertise, entailing both anatomical and functional differences between the brain of literate and illiterate individuals (Carreiras et al., 2009; Dehaene et al., 2010; Thiebaut de Schotten et al., 2012). Learning to read yields functional changes in the visual cortex, including the development of a word-selective area in the left occipitotemporal region (Cohen et al., 2000), and the displacement of the neighboring face-selective area toward the right hemisphere (Dehaene et al., 2010). The acquisition of expertise thus entails both the emergence of novel local specialization and a competition between classes of stimuli. How then does the visual system accommodate expertise for several independent symbolic systems, as occurs for instance in high-level musicians who simultaneously acquire alphabetic and musical notations? Category-selectivity is a dominant organizing feature of the ventral visual cortex (Grill-Spector and Weiner, 2014), and it could be expected that distinct occipitotemporal regions may be devoted to different symbolic systems. The preferences of a given cortical patch for a given type of visual objects may result from two causes (Hannagan et al., 2015). First, *a priori* perceptual biases, e.g. preference for foveal vs. peripheral stimuli (Hasson et al., 2002), make each cortical site more or less suitable to represent different types of stimuli (Srihasam et al., 2014). Second, preferences may also arise from privileged connections with distant regions involved in domain-specific functions (Mahon and Caramazza, 2009; Plaut and Behrmann, 2011; Saygin et al., 2012).

Both factors may contribute to the reproducible placement of the reading-selective Visual Word Form Area (VWFA, Cohen et al., 2000). First, it falls in a cortical region with a preference for foveal over peripheral stimuli (Hasson et al., 2002), for analytical over configural processing (Ventura et al., 2013), and for sensitivity to line junctions (Szwed et al., 2011). Second, the VWFA also shows stronger anatomical connectivity with perisylvian language areas compared to the neighboring Fusiform Face Area (FFA) (Bouhali et al., 2014).

Alphabetic and musical notations make use of substantially different visual features, and the two sets of symbols are used to trigger different cognitive abilities. Therefore, both perceptual biases and long-distance connectivity would predict segregated cortical preferences for the two domains.

Paradoxically, strong empirical support for the cortical segregation of different symbolic systems comes from a recent study in monkeys. Macaques were trained at identifying alphabetic symbols, Tetris-like shapes, and sketchy cartoon faces. Following training, monkeys developed distinct regions selective for each

of the three sets, at reproducible locations within the occipitotemporal cortex (Srihasam et al., 2012; Srihasam et al., 2014). In humans, there is also evidence for segregated activations for letters and numerals in the fusiform gyrus (Polk et al., 2002; Shum et al., 2013; Abboud et al., 2015), and for Chinese vs alphabetic stimuli in the early visual cortex (Szwed et al., 2014). There are also some indications of segregated activations for printed words and musical notation in the occipitotemporal cortex (Wong and Gauthier, 2010a), but this evidence is controversial (Muayqil et al., 2015) and is not supported by statistical comparisons of activation topography between words and music.

Here, in order to study the cortical segregation of symbolic systems, we used fMRI to assess activations induced by musical notation in the left occipitotemporal cortex, in professional musicians and naive controls, comparing them to activations by written words and other classes of objects (Figure 1). From the above considerations we derived four core predictions, which we assessed using novel individual analyses of activation topography and volume. First, music notation differs from words both in terms of visual features and of associated representations, semantic and other. Therefore we expected to find music-selective activations topographically distinct from other category-selective regions in the ventral stream. Second, visual encoding of music scores differs between musicians and naïve subjects (Stewart, 2005). Hence we predicted that music-related activations should differ depending on musical expertise. Such differences might affect the location, the amplitude, or the spatial expansion of music-related activations. Third, we predicted that musical expertise could interact with other categories, particularly words, affecting the location, the amplitude, or the volume of nearby word-related activations. Fourth, if indeed we identify distinct visual regions selective for words and music, the functional connections of the music-selective area to the rest of the brain should differ depending on musical expertise, while the connections of the word-selective area should not differ across groups.

2 Materials and Methods

2.1 Subjects

Twenty-four musicians and 24 control subjects gave written informed consent to participate in the study. Three musicians were excluded because of technical defects or claustrophobia during fMRI acquisition; one control was excluded because after scanning he admitted to some musical competence. The two groups were matched for age (mean age: controls 29.8 years, SD=10; musicians 32.6 years, SD=10), gender (controls: 13 men; musicians: 12 men), and educational level (schooling: controls 14.7 years, SD=1.8; musicians 15.7 years, SD=1.8). All were right-handed according to the Edinburgh inventory (Oldfield, 1971) and had normal or corrected-to-normal vision. Three musicians, although perfectly fluent French speakers, were not native French speakers. Musicians were either Master students at the CNSM (Conservatoire National Supérieur de Musique et de Danse de Paris), or professional musicians. Their main instrument was the cello (n=3), the violin (n=6), the trumpet (n=1), the oboe (n=3), the recorder (n=1), the

viola (n=1), the piano (n=6), the harpsichord (n=1), and the clarinet (n=1). Two participants were equally proficient with two instruments. Professional musicians were concert performers, music teachers, orchestra and choir conductors, composers, music arrangers, sound engineers. In musicians, the mean age of onset of instrumental training was 5.7 years (SD=2.1); the mean age of the onset of reading acquisition was 4.9 years (SD=1.3) for words and 5.8 years (SD=2.8) for music (t(20)=1.5, NS). Controls were unable to read any musical notation. All subjects were paid for their participation and were naïve about the aims of the study. The experiment was approved by the local ethical committee.

2.2 Stimuli

We used five categories of black and white pictures: faces, tools, houses, pairs of words, music scores (Figure 1A). Each category contained 38 pictures. All stimuli were black line drawings on a white background. Faces, houses and tools were derived from highly contrasted gray-level photographs matched for size and overall luminance. Faces (17 females, 21 men) were front or slightly lateral views of non-famous people. Houses comprised outside pictures of houses and buildings. Tools were common hand-held household objects (e.g. knife, hair-dryer) presented in a normal orientation. The faces, tools, and houses images used here were used in previous studies in order to map category selectivity in the occipitotemporal cortex (Gaillard et al., 2006; Thirion et al., 2007; Dehaene et al., 2010; Pegado et al., 2014; Pinel et al., 2015). Music notation corresponded to one bar of classical piano music (G and F clefs), containing no alphabetic symbol. They were extracted from the "Mutopia project" database (http://www.mutopiaproject.org/), and selected from Mozart, Bach and Beethoven pieces. Pairs of words were semantically congruent adjective plus noun pairs (e.g. *amitié sacrée*), written in lower-case, both 6-letters long and of high lexical frequency (http://lexique.org). Music scores were 100 pixel wide x 71 pixel high, word pairs were 108 pixel wide x 87 pixel high, faces were 86 pixel wide x 108 pixel high, houses were 108 pixel wide x 96 pixel high, tools were 107 pixel wide x 102 pixel high. All images were padded with random visual noise to reach a size of 300 x 300 pixels (7° x 7° of visual angle).

2.3 Experimental design

The experiment was programmed using E-Prime software. Subjects were presented with an alternation of blocks of pictures (8000 ms per block) and blocks of rest (7800 ms per block). Each stimulation block included eight pictures from one of the five categories of stimuli. Each picture was displayed for 600 ms and followed by a 400 ms blank screen. During rest and inter-trials intervals, a central fixation cross remained present in order to minimize eye-movements. The experiment included 10 s of initial rest, followed by 30 blocks of pictures (six for each category) and 30 blocks of rest. Blocks were presented in pseudorandom order so as to maximize the variety of transitions between conditions while avoiding repetition of the same condition in successive blocks. Subjects were asked to press a button with their right thumb whenever a

picture was identical to the previous one, which was the case for 20% of stimuli (one to three repetitions per block).

2.4 fMRI acquisition and analysis

We used a 3-Tesla MRI (Siemens Trio) with a 32 channel head coil, and a multiband echo-planar imaging sequence sensitive to brain oxygen-level-dependent (BOLD) contrast (54 contiguous axial slices, 2.5 mm isotropic voxels, in-plane matrix = 80 x 80; TR = 1160 ms; angle = 62°, TE = 25 ms). 420 volumes were acquired. The first 4 volumes were discarded to reach equilibrium. Five additional BOLD volumes with reverse phase encoding direction were also acquired. Functional images were realigned, treated with the FSL "Topup" toolbox in order to correct EPI distortions due to B0 field inhomogeneity (Andersson et al., 2003), normalized to the standard MNI brain space and spatially smoothed with an isotropic Gaussian filter (3 mm FWHM). We chose a low-smoothing filter because we wanted to take advantage of the high-resolution MRI. A two-level analysis was then implemented in SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/).

For each participant, data were high-pass filtered and modelled by regressors obtained by convolution of the 5 experimental conditions plus the button presses with the canonical SPM haemodynamic response function (a combination of 2 gamma functions, with a rise peaking around 6 s followed by a longer undershoot). Individual contrast images for the 5 types of stimuli minus rest were smoothed with an isotropic Gaussian filter (2 mm FWHM) to take into account between-subject differences in functional anatomy, and entered into a second-level whole-brain ANOVA with subjects as random factor, stimulus category as within-subject factor, and musical status as between-subject factor.

For the analysis of activation asymmetry, individual normalized anatomical images were flipped, and then normalized back to the original normalized anatomy; the corresponding normalization matrices were applied to the flipped contrast images, allowing for an accurate match of the left and right hemispheres; flipped contrast images were then subtracted from the original contrast images and smoothed with an isotropic Gaussian filter (2 mm FWHM). Difference images were entered in the same group-level ANOVA as before, allowing us to test the interaction of any given contrast with the left/right hemisphere factor.

For individual methods of analysis (activation topography and volume, and psychophysiological interaction), we needed to identify individual activation peaks to words and to music. To this end, we first identified an unbiased group-level activation peak by contrasting words plus music minus the other 3 categories, pooling controls and musicians (MNI -54 -55 -15, Z>8). We then defined a left occipitotemporal region of interest (ROI), by taking an anatomical mask including the inferior occipital, inferior temporal, and fusiform gyri (Tzourio-Mazoyer et al., 2002), and restricting this mask from 15 mm anterior to 15 mm posterior to the unbiased activation peak (ie MNI y=-40 to y=-70) (Figure 4). In each subject, we searched this ROI for the voxel most strongly activated (i.e. with the highest t value) by music minus all other stimuli (except words), and by words minus all other stimuli (except music). We also measured the volume of the

activation clusters surrounding those peaks, within the same ROI, at the P<0.01 voxelwise threshold. In one of the analyses, we wished to compare the intensity of activations by words and music at the individual word-selective and music-selective peaks. In order to avoid double dipping biases in this analysis, data were again modelled as before, but using distinct regressors for the even and the odd blocks of each type of stimuli (Kriegeskorte et al., 2009). We used the odd blocks to identify the coordinates of the peaks, and the even blocks to compute activation intensity.

We performed psychophysiologic interaction (PPI) analyses, using SPM8 functionalities (Friston et al., 1997). In essence, PPI models the response across the brain as the influence of a seed region and its interaction with experimental conditions. We identified individual peaks as explained before. We then defined two individual ROIs as 4 mm radius spheres centered on those peaks, then extracted BOLD signal from those ROIs, and for each ROI performed a GLM analysis whose regressors were the ROI's extracted signal (or "physiological variable"), an experimental regressor (or "psychological variable", in the present case, music vs. words), and the element-by-element product between the first two variables (or "psychophysiological interaction variable"). Contrasts defined over the psychophysiological interaction allowed us to identify regions whose correlation with the ROI changed according to experimental conditions, i.e. was higher or lower during word reading than during music viewing. Finally, contrast images were smoothed (FWHM 2 mm), and we computed second-level t-tests for controls and musicians, as well as comparisons between groups, masking by the adequate contrast relative to rest (voxelwise P<0.01). For instance, when looking, in controls, for brain regions more correlated to some ROI during music viewing than during word reading, we masked the analysis by the contrast of music minus rest in controls. For the comparisons between groups, we masked by the first term of the contrasts. For instance, when looking for brain regions more correlated to music ROI during music viewing than during word reading, in musicians more than in controls, we masked the analysis by the contrast of music minus rest in musicians.

Unless stated otherwise, we used a clusterwise threshold of P<0.05, corrected for multiple comparisons across the whole brain, and an uncorrected voxelwise threshold of P<0.001 for standard functional analyses and P<0.01 for PPI analyses. Whenever analyses were masked by other contrasts, masking contrasts were thresholded at voxelwise P<0.01.

2.5 Behavioral measures

2.5.1 Elementary music reading test

This brief test, checking for the lack of any musical reading abilities in controls, contained four different questions. First, participants had to name 6 printed notes on a G clef stave, and 6 on a F clef stave. Second, six notes with different durations were presented and subjects had to rank them from the shortest to the longest.

2.5.2 Profile of Music Perception Skills (PROMS)

The PROMS (Profile of Music Perception Skills, Law and Zentner, 2012) assesses perceptual musical skills across multiple dimensions (melody, timbre, rhythm, etc.) (http://www.uibk.ac.at/psychologie/forschung/tests_and_diagnostics/). In its brief version, it comprises 4 subtests: melody, tuning, speed and beat. Each subtest comprises eighteen trials, preceded by a practice trial. In all subtests, the task is a same/different judgment: participants hear twice the same stimulus, followed by a short interval and a comparison target. They are asked to indicate whether the target matches the initial stimulus or not using a 5-level scale ("definitely same/different", "probably same/different", "I don't know"; correct "definitely" answers score 1, correct "probably" answers score 0.5, all other answers score 0).

3   Results

3.1   Behavioral results

3.1.1   Elementary music reading test

None of the control subjects was able to name even a single note on the G key, nor to order correctly 6 notes according to their duration. A minority of controls were able to name up to 3 out of 6 notes on the F key (some of them had some very basic musical training during secondary school).

3.1.2   Profile of Music Perception Skills (PROMS)

On the PROMS (Profile of Music Perception Skills, Law and Zentner, 2012), which assesses auditory perceptual musical skills across multiple dimensions, musicians outperformed controls on each of the subtests (all Ps<0.001; see Table 1).

3.1.3   Behavioral responses during fMRI

In the scanner, subjects were presented with musical notations, words, faces, houses, and tools, and had to perform a repetition detection task (Figure 1B). Controls and musicians correctly detected 95% and 97% of repeated stimuli (hits), respectively, while erroneously responding to 0.5% and 0.2% of non-repeated stimuli (false alarms). Those overall rates did not differ across groups (Mann-Whitney test, P=0.0816 for hits, and P=0.111 for false alarms). Hit rates differed across the 5 categories of stimuli in controls (Friedman test, P<0.001), but not in musicians (P=0.47). The only difference between groups was the better performance for music in musicians than in controls (96% vs 85%; Mann-Whitney test, P=0.025), while groups did not differ for the other 4 categories. There were no significant differences involving false alarm rates.

3.2   fMRI results

We will briefly report the overall correlates of musical expertise at the whole-brain level, then focus on activations in the ventral visual cortex.

3.2.1 Whole-brain correlates of musical expertise

We contrasted music minus non-musical stimuli (i.e. words, faces, houses and tools), separately in musicians and controls. In both groups, this showed an extensive bihemispheric network with a noticeable enhancement in musicians. In order to identify the commonalities between controls and musicians, we computed the conjunction of activations in the two groups (voxelwise P<0.001 each, clusterwise P<0.05 corrected), masking by music minus rest in controls and musicians, respectively. This showed bilateral activations in the lateral and mesial prefrontal cortex, anterior insula, intraparietal sulcus (IPS), and inferior temporal gyrus (left: -54 -58 -6, Z=4.70; right: MNI 54 -55 -9, Z=6.20) (Figure 2A, red activations). We then looked for specific correlates of musical expertise. To do so, we compared the contrast of music minus non-musical stimuli, in musicians relative to controls. In order to restrict the analysis to regions actually involved in music processing, this comparison between groups was restricted to regions activated by music minus rest in musicians (see Methods section). This showed bilateral left-predominant cortical activations in the lateral and mesial prefrontal and rolandic cortex, IPS, temporal and occipital lobe, and right-predominant cerebellum (Figure 2A, green activations). This set included bilateral inferior temporal activations (left: MNI -54 -55 -18, Z=5.87; right: MNI 54 -46 -21, Z=6.08). Asymmetry analysis confirmed that this expertise-related network predominated in the left hemisphere (Figure 2B, blue activations), including the inferior frontal gyrus (MNI -54 8 12, Z=6.69), frontal eye field (MNI -21 -13 60, Z=5.08), intraparietal (MNI -21 -52 48, Z=3.56) and supramarginal areas (MNI -63 -19 21, Z=5.58). The contrast opposite to the one assessing musical expertise, i.e. the subtraction of music minus other categories in controls more than in musicians, masking by music minus rest in controls, showed no activations.

In summary, as compared to all the other types of stimuli, musical notations yielded extensive bilateral brain activations. Some of those activations were common to musicians and controls, while activations reflecting musical expertise included left-lateralized perisylvian regions, left intraparietal cortex, and bilateral ventral temporal cortex. Ventral occipito-temporal activations, which are the focus of this study, will now be examined in more detail.

3.2.2 Activations in category-selective regions

As a first approach, we identified category-selective occipitotemporal regions, and studied their activation by musical stimuli, and how those activations were modulated by musical expertise (Figure 3). We apply the term of selectivity to situations in which a cortical region is activated significantly more by one category of items than by other visually comparable stimuli. Selective regions were identified by contrasting each of the four non-musical categories minus the three other ones, pooling controls and musicians. Note that we took care to exclude music from the procedure of selecting regions of interest, so as to avoid any biases in the subsequent study of music-related activations. This showed a typical mosaic of ventral occipitotemporal areas (Figure 3, top panel): the VWFA for written words (MNI -54 -55 -15 Z= 7.10), the right-hemispheric FFA for faces (MNI 42 -43 -21, Z>8), the bilateral PPA for houses (left: MNI -27 -49 -6, Z>8; right: MNI 30

-43 -9, Z>8), and the bilateral LOC for tools (left: MNI -48 -70 -3, Z>8; right: MNI 48 -64 -3, Z>8). Activations for the five types of stimuli minus rest were computed in the six peak voxels for each subject, and were entered in ANOVAs with subjects as random factor, type of stimuli and location as within-subject factors, and musical expertise as between-subject factor (Figure 3, bottom panel). We focused our analyses not on overall differences in activation level between musicians and controls, but on the differences between groups in the relative activation for the various categories of stimuli.

In each category-selective region, we compared music-related activation with the activation induced by the locally preferred type of stimulus (e.g. words for the VWFA, or faces for the FFA). In all regions but one, the activation evoked by written music was weaker than the activation induced by the locally preferred category (all Ps<0.0012). The only exception was the VWFA, where music activation did not differ from activation by words (F(1,42)=1.5, P=0.23) (interaction of the "music vs preferred" factor x region : F(5,210)=22.2, P<0.001). Zooming in on the VWFA, we observed a clear difference between groups: in control subjects, music activated the VWFA less than words (F(1,22)=7.48, P=0.012). Conversely, in musicians, activation was stronger for music than for words (F(1,20)=19.46, P<0.001; interaction group x type of stimulus F(1,42)=25.6, P<0.001). In all other regions, the group x stimulus interaction failed to reach significance.

In summary, musical expertise changed the profile of activation across types of stimuli in the VWFA, while the properties of the other category-selective regions were not affected. The activation induced by music in the VWFA was pushed higher than the activation by the usually preferred stimuli, namely words.

3.2.3 Music- and word-related activations in the ventral occipitotemporal cortex

In the above analyses, we pooled activations by controls and musicians in order to identify the peak with the highest overall selectivity for words. At this location, we showed that activation to musical notation in musicians actually exceeded activation to words. This demonstrates that there is at least a close proximity between the regions that subtend the visual processing of music and words. However, we cannot yet determine whether those regions are actually identical, or whether they differ according to category (i.e. music vs words) and to musical expertise.

3.2.3.1 Group analyses

As a first step to clarify this issue, we identified inferior temporal activations, contrasting words or music stimuli to the average of the other three categories (faces, houses, tools), separately in controls and musicians. Activations to words peaked at MNI -54 -55 -15 (Z=6.83) in controls, and at MNI -57 -58 -15 (Z=4.61) in musicians. Activations to music were bilateral, peaking in controls at MNI -54 -58 -6 (left, Z=4.79) and MNI 54 -55 -12 (right, Z=5.91), just anterior to the LOC as identified in the previous section. In musicians, there were similar bilateral peaks at MNI -51 -58 -9 (left, Z>8) and MNI 54 -49 -18 (right, Z>8).

Those music-related activations were significantly stronger in musicians than in controls (Left: MNI -54 -55 -18, Z=5.47; Right: MNI 54 -46 -21, Z=5.87). Thus, in the left hemisphere, the peak coordinates of activations to words and music in controls and musicians were packed within a few millimeters. However, such group-level analyses are blurred by individual variability and may miss small but systematic within-subject differences in topography (Cohen et al., 2004).

3.2.3.2 Individual analyses of left occipitotemporal activations

To circumvent the blurring of group-level analyses, and to compare the topography of activations statistically, we localized individual peak activations by music minus all other stimuli (except words), and by words minus all other stimuli (except music), within a left occipitotemporal ROI (see Methods section. Figure 4A).

Before moving to topographical analyses, note that individual analyses provided an improved understanding of regional selectivity for words and music. The activation profile at the group-level peak of the VWFA showed an interaction of stimulus type and expertise (Figure 3, bottom panel), with activations stronger for words than music in naïve subjects, and stronger for music than for words in musicians. However, the average of activation profiles at individual peaks (Figure 4B) reveals a simpler pattern: word peaks show more activation for words than music (F(1,42)=38.6, P<0.001), and music peaks more activation for music than words (F(1,42)=35.4, P<0.001), with no significant interaction with the level of expertise (interaction at the word peak : F<1; interaction at the music peak : F(1,42)=2.4, P=0.14). Note that for this analysis, the location of individual peaks and activation strength were identified from independent datasets, avoiding double dipping biases (see Methods; Kriegeskorte et al., 2009).

The extraction of individual activation peaks allowed us to compare the coordinates for words and music between musicians and controls. Rather than using the raw MNI coordinates, we basically rotated the coordinate axes using Principal Component Analysis (PCA). Indeed, activations fit in an almost flat cortical space, in which coordinates are not independent, due to the overall 3D slant of the temporal cortex, and the arbitrary MNI axes may not be the most informative representation of the topography of individual peaks. This procedure simply provides the most synthetic way to describe the topography of activations. PCA resulted in 3 components accounting for 60%, 27%, and 13% of the variance. The first component corresponded to an axis with an anterior, mesial, and slightly ventral orientation, to which the second and third axes were orthogonal (Figure 4C). The localization of activation peaks was quite consistent across individual : Extreme individual coordinates were less than 2.2 SD away from the mean, along the 3 axes, in both groups, for words and music, except for a single control subject whose peak for music was 3.5 SD above the mean along the second axis (ie more postero-mesial). Coordinates along the 3 axes of the new reference frame were entered in ANOVAs with subjects as random factor, musical expertise as between-subject factor and type of stimuli as within-subject factor (Table 2 and Figure 4C). Along the first axis, activations were more anteromesial for words than for music (F(1,42)=23.3, P<0.001), without any

difference as a function of musical expertise. Along the second axis, activations were more anterolateral for musicians than for controls (F(1,42)=6.9, P=0.012). This effect of expertise along the second axis was strong for words (F(1,42)=7.85, P=0.0076), and non-significant for music (F<1), although the interaction was not significant (F=1). There were no other significant differences or interactions.

Thus, musical expertise changed the peak location of word-selective regions in the ventral stream, but the peak location of music-related activations did not change. Indeed, ventral activations to music differed neither in location nor in peak intensity (F<1) between musicians and naïve subjects. We reasoned that expertise could be reflected in a third parameter: the spatial extent of activations. For each subject, we computed the volume of the clusters comprising the individual peaks for words and music described in section 2.4. This volume was entered in an ANOVA with subjects as random factor, musical expertise as between-subject factor and type of stimuli as within-subject factor (Figure 5A). There was an interaction between expertise and type of stimuli (F(1,42)=15.7, P<0.001). The volume activated by music was more than twice as large in musicians than in controls (mean: 1251 and 585 mm^3, respectively; F(1,42)=18.4, P<0.001), while the volume activated by words did not differ between musicians and controls (mean: 594 and 702 mm^3, respectively; F<1).

In summary, (1) activations for music were posterior and lateral to activations for words, a difference which could not be established using group-level analyses; (2) in musicians, activations for words were more anterolateral than in controls; (3) the volume of activation to music was larger in musicians than in controls.

3.2.4 Psychophysiological interaction

The previous results show that the organization of ventral activations varies with perceptual expertise: despite identical word reading competence, the location of word-selective activations differed between controls and musicians, while music-related activations peaked at the same location despite major difference in musical expertise. We reasoned that, although subjects without musical training could partially process the musical stimuli at a visual level, using an occipito-temporal area whose peak location is identical to that of musicians, they should differ in the ability to use these stimuli to activate additional areas involved in auditory, semantic or motor knowledge. Thus, expertise may depend on the brain-scale networks to which occipitotemporal visual areas give access, irrespective of their exact location. If so, then ventral music peaks should be connected to very different functional networks in musicians and controls, while word peaks should show similar connections. To test this hypothesis, we studied the functional networks associated with those regions using PPI analyses. As explained in the Methods section, we identified individual left occipitotemporal word-selective and music-selective ROIs, and looked for brain regions which would be more (or less) correlated to each of those ROIs during music viewing than during word reading. We did this separately for musicians and controls, and also compared the two groups (Figure 6).

This analysis revealed, first, a set of left-predominant language-related regions more connected to the word ROI during word reading than during music viewing. This network was found in both controls and

musicians, with no significant differences between groups. Second, we found a set of bilateral fronto-temporo-parietal regions more connected to the music ROI during music viewing than during word reading. This music network was stronger and more extensive in musicians than in controls, with a left-hemispheric predominance. The opposite contrasts (stronger correlations to the word ROI during music perception, or to the music ROI during word perception) showed no significant activations.

Thus, the PPI analysis showed that although word ROIs had different locations in musicians and controls, they were connected to the same functional language networks. Conversely, music ROIs did not differ in location, but were larger and connected to more developed networks in expert musicians than in naive controls.

4 Discussion
4.1 Summary

We used fMRI to study activations induced by musical notation in the occipitotemporal cortex, in professional musicians and naive controls, comparing them to activations by written words and other classes of objects. Whole-brain analyses showed strong activations to music in the bilateral occipitotemporal cortex of both musicians and controls, with an additional increase in musicians. Among the usual occipitotemporal category-selective regions, the VWFA was the only one to demonstrate a significant impact of musical expertise on its activation profile: in group analyses, activations for words and music were overlapping, with activations stronger for words than for music in controls, and stronger for music than for words in musicians. Because group-level analyses blurred potential topographical distinctions between music- and word-related activations, we moved to individual-level analyses. Those allowed us to reach a more accurate description of the topography of activations and of the correlates of expertise, and to address the four predictions which were put forward at the outset. First, activations for music were significantly posterior and lateral to activations for words. Second, the volume of activation to music was larger in musicians than in controls. Third, activations for words were more anterolateral in musicians than in controls. Fourth, PPI analyses showed that the functional connectivity of the ventral word-selective regions did not differ according to musical expertise, whereas the connectivity of the music-selective regions was strongly enhanced in musicians.

We will first discuss the topography of music-related activations, irrespective of musical expertise, and then turn to the mechanisms of expertise.

4.2 Music-related activations

4.2.1 Whole-brain study

Whole-brain analyses showed extensive bilateral activations for music, as compared to all other stimuli averaged together (Figure 2), including regions presumably involved in attentional processing and sensorimotor interactions (see Supplemental material).

4.2.2 Dissociation between words and music reading

Music-related activations in the left occipitotemporal cortex peaked close but posterior and lateral to word-related activations (Figure 4), but, as shown in group-level analyses, activations for those two categories were largely overlapping. In this context, should we consider that music-selective activations exist in the left occipitotemporal cortex? Discussing the case of word reading, Cohen and Dehaene (2004) proposed to break down the ambiguous notion of selectivity into three logically independent features: functional specialization, reproducible location, and regional selectivity. Functional specialization refers to the fact that the visual system as a whole develops properties only relevant to one class of stimuli, such as for instance invariance for changes between upper- and lower-case letters (Dehaene et al., 2001; Qiao et al., 2010). We assume that such specialization occurs here in expert musicians, since they exhibit enhanced perceptual performance with music stimuli and increased intensity and volume of the activations induced by music stimuli.

The distinct notion of reproducible location applies whenever the neurons subtending functional specialization tend to be grouped together in some fixed region of visual cortex. It does not imply that this region should not also harbor other neurons with different preferences. Here, group-level analyses showed that music activates the occipitotemporal cortex at a definite location, different from the category-selective activations for faces, houses or tools, suggesting reproducible location. Still, group-level analyses failed to show topographical differences between activations for music and activations for words. This leads us to the third concept, regional selectivity, which applies whenever some region is devoted solely to the processing of one category of objects. On the basis of imaging data, extreme regional selectivity would imply activation by the relevant category of items only, with no activation at all by other types of stimuli. This criterion is met in some studies using high-resolution techniques (Nobre et al., 1994; Tsao et al., 2006). In the present case, individual analyses showed that the average peak location of ventral activations differs between music and words. However this difference is small, and activations largely overlap, showing no evidence for regional selectivity for music as compared to words, but not excluding it either.

Neuropsychological observations may play a critical role, as dissociated impairments affecting only one category of objects are direct arguments for regional selectivity. Typically, in the few published cases of

pure alexic musicians, the reading impairment affects both words and music, while sparing other categories of objects (Dejerine, 1892; Fasanaro et al., 1990; Beversdorf and Heilman, 1998). However, Cappelletti et al. (2000) reported a case of selective music reading deficit. The patient was a professional musician, a piano and guitar performer, a prolific song writer, an experienced music proof-reader, and she could sight-sing and sight-read at the piano. Following encephalitis, she suffered from a left ventral posterior temporal lesion which, based on the published brain images, may be located at MNI -55 -45 -20, plus a more posterior right occipital lesion. She suffered from alexia affecting selectively musical notes, while sparing words, numbers and other symbols. This carefully studied case corresponded to a "classical" dissociation according to Shallice's typology (1988), as word reading was spared entirely, while the patient "was no longer able to read even a single musical note of melodies, including those she had herself composed". This report suggests that, despite the substantial overlap we observed with printed words, regional selectivity for music notation may prevail, at least in some expert musicians. This conclusion should however remain cautious, as this finding has not been replicated in another patient as yet.

4.2.3 The topography of music-related activations

The origins of the distinct neural locations of the music and word reading mechanisms are presently unclear. Srihasam et al. (2014) trained monkeys at identifying alphabetic symbols, Tetris-like geometrical shapes, and sketchy cartoon faces. Following training, monkeys developed regions selective for each of the three sets, at reproducible locations within the occipitotemporal cortex. Identical reinforcement rules were used for the 3 sets of stimuli, showing that regional segregation was not a result of task-related function. Srihasam et al. (2014) also showed that activations were identically located irrespective of the order in which the sets had been trained. Similarly to Srihasam et al. (2014), we would argue that the observed pattern of selectivity for music versus words - is likely related to innate constraints underlying the mosaic of occipitotemporal category-selectivity. Srihasam et al. (2014) put forward such candidate constraints by showing that category-selective areas differ in their sensitivity to central vs peripheral location, and to curvy vs straight shapes. At present, the logic underlying the functional organization of the occipitotemporal cortex is not understood with sufficient detail to allow us to understand the precise location of activations to words and music, one may point to both visual commonalities and differences between the two systems. On the one hand, words and musical notation both resort to foveal, high-contrast, monochromatic linear shapes, with a high number of line junctions and important high-resolution details. Such commonalities may explain the proximity, within the visual cortex, of word-related and music-related activations. On the other hand, there are visual differences potentially relevant to cortical selectivity. Compared to intrinsically linear letter strings, musical notes occupy a 2-dimensional portion of space: while letter identity is invariant for position, note identity varies according to its vertical position and also according to external cues such as clef symbols and time signature, a type of long-distance spatial dependency which has no equivalent with letters.

4.3 Correlates of musical expertise

At the whole-brain level, musical expertise was reflected by left-predominant activations in a set of frontal and parietal regions. This overall result is congruent with Muaykil et al. (2015), who recently found left frontal and parietal activations in subjects with elementary musical competence, and with Stewart et al. (2003), who found training effects of music reading in the left inferior parietal lobule. Wong & Gauthier (2010a), report bilateral activations in musical experts, and also mention left inferior parietal activations. The present study focusses on occipitotemporal processing, and it is out of our scope to speculate on the interpretation of whole-brain activations, beyond checking that our results fit with the literature.

We have shown that musicians have larger music-related activations in the left occipitotemporal cortex than controls, with no increase in activation amplitude. Activation enhancement is one of the possible correlates of extensive practice, while it is often unclear whether it corresponds to increases in the extent or the strength of activations, or both (Kelly and Garavan, 2005; Chang, 2014). For instance, the acquisition of reading expertise leads to enhanced responses in the VWFA, both in terms of peak amplitude and of distance to the peak (Dehaene et al., 2010). Likewise, activation at or around the location of the Fusiform Face Area has been reported to increase with the acquisition of expertise for cars, birds, or a class of novel "greeble" shapes (Gauthier et al., 1999; Gauthier et al., 2000). More recently, Harley et al. (2009) showed that visual expertise in analyzing radiological images is correlated with enhanced activation in the right fusiform region. In music reading, Wong & Gauthier (2010a) found stronger activation in musicians than in controls in the fusiform gyrus, slightly more mesial but only 2 mm off along the anteroposterior axis (MNI -40 -58 -18), as compared to the present results. These activations contrasting single notes versus letters and other symbols are described as posterior to letter-selective areas, but no coordinates are reported for the latter. Curiously, Wong & Gauthier (2010a) found no significant effect of expertise when using strings of notes, letters and symbols. In an EEG study, Proverbio et al. (2013) found an increase in the amplitude of the occipitotemporal N170 wave in response to music notation, with left-hemispheric predominance, in musicians relative to controls. The coordinates of the left temporal generators explaining the surface voltage for music stimuli (Talairach coordinates -48.5 -55 -17.6) are in good agreement with the present observations.

We can only speculate on the behavioral correlates of the increase in activation volume in musicians. Musicians exhibit enhanced perceptual abilities for musical stimuli, as reflected for instance by a reduction of perceptual crowding, selectively for musical stimuli, in proportion to music reading fluency (Wong and Gauthier, 2012). Another facet of expertise may be the encoding of frequent combinations of notes, such as rhythmic, melodic, or chord patterns, as a single object. Indeed, musicians develop automatic holistic perception for printed music (Wong and Gauthier, 2010b). Following brain lesions, such encoding may be lost, resulting in "note-by-note" reading (Stanzione et al., 1990). Expertise in word reading involves analogous perceptual tuning, namely reading-selective reduction of crowding (Huckauf and Nazir, 2007;

Pelli and Tillman, 2008), and synthetic perception of complex objects such as frequent groups of letters (Vinckier et al., 2007; Grainger and Ziegler, 2011).

In order to enrich the analysis of individual differences beyond the main contrast between musicians and naïve subjects, we tried to identify correlates of variants of musical expertise. We compared musicians with a more extensive training to read "harmonic" music (e.g. piano) vs musicians playing monodic instruments (e.g. flute). Furthermore, despite limited variance in our sample, we tried to study the impact of the order of acquisition of music relative to words. Those do not reveal significant effects relevant to ventral occipitotemporal activations (see Supplemental Material).

Changes in the ventral cortex probably account for only a part of musical perceptual expertise. Indeed, we have shown fingerprints of expertise in the left-predominant dorsal visual stream. Considering the importance of spatial information in musical notation, it is likely that music reading expertise involves a close interaction of skilled dorsal and ventral processing, possibly subtending spatial and featural properties of musical notation, respectively (Stewart, 2005; Vinckier et al., 2006; Vinckier et al., 2007). Note also that the role of the music-selective region under study concerns the identification of notes, but possibly not of other musical symbols (crescendo, sharp, etc). Indeed, the recognition of those symbols can be spared in patients with left occipitotemporal lesions disrupting music reading (Fasanaro et al., 1990; Cappelletti et al., 2000).

4.4 Impact of musical expertise on the processing of words

A consequence of musical expertise was the anterolateral displacement of activations for words. Note that in musicians and controls, who had the same expertise in word reading, word-related activations differed neither in volume nor in amplitude. One may hypothesize that, during the simultaneous acquisition of written words and musical notation, the two systems compete for cortical space, the development of musical skills "pushing" words to a more anterior location. Proverbio et al. (2013) found that left fusiform activations to words were about one centimeter more anterior in musicians than in controls, without assessing the significance of this difference and using the spatially imprecise method of event-related potentials. Another example of interaction between categories of visual objects in defining the layout of cortical specialization was reported by Dehaene et al. (2010), who showed that literacy was correlated with a shrinking of face-selective areas contiguous to the VWFA, and an increase in face-induced activation in the right FFA (see also Plaut and Behrmann, 2011; Pegado et al., 2014). Srihasam et al. (2014) showed that when monkeys learned a second symbol set, activations induced by a first-learned set moved away slightly, again suggestive of competition between categories for cortical space.

4.5 Expertise and long-distance connectivity

The PPI analysis provided independent confirmation that the ventral music-selective area is involved in expert music processing, while the neighboring word-selective area is not. Previous studies have shown correlates of music processing in the form of increased functional connectivity across distributed cerebral networks, during beat perception (Grahn and Rowe, 2009), musical improvisation (Pinho et al., 2014), and motor synchronization to auditory stimuli (Krause et al., 2010). We extend these results by putting to light an increase of long-distance connectivity of the music-selective visual region, during the perception of musical notation in musicians. When studying the connectivity of the word-selective area, no difference was found between controls and musicians during music viewing, because this area is not the gateway to the music network, nor during word viewing, because the two groups shared the same level of expertise for word reading.

To conclude, expert object recognition involves perceptual aspects, which are subtended by changes in the visual cortex, but also attentional processes and domain-specific knowledge, which rely on distributed cerebral regions (for a review see Harel et al., 2013). In the present study, which focused on the visual component of expertise, we observed an expansion of the area which is activated by music notation more than by other categories of stimuli. Increases in activation have been observed in other fields of visual expertise (Kelly and Garavan, 2005; Chang, 2014), including word reading (Dehaene et al., 2010) and mirror-reading (Poldrack et al., 1998; Poldrack and Gabrieli, 2001). Moreover, we observed a displacement of word-related activations in musicians, suggesting the existence of some form of competition between categories for the colonization of visual cortex (Op de Beeck and Baker, 2010). Finally, establishing a bridge between the perceptual and the distributed components of music expertise, we showed that musicians develop novel functional connectivity between the ventral music-selective area and distant non-perceptual regions.

Acknowledgements

References

Abboud S, Maidenbaum S, Dehaene S, Amedi A (2015) A number-form area in the blind. Nat Commun 6:6026.

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870-888.

Beversdorf DQ, Heilman KM (1998) Progressive ventral posterior cortical degeneration presenting as alexia for music and words. Neurology 50:657-659.

Cabeza R, Nyberg L (2000) Imaging cognition II: An empirical review of 275 PET and fMRI studies. J Cogn Neurosci 12:1-47.

Cappelletti M, Waley-Cohen H, Butterworth B, Kopelman M (2000) A selective loss of the ability to read and to write music. Neurocase 6:321-332. .

Carreiras M, Seghier ML, Baquero S, Estevez A, Lozano A, Devlin JT, Price CJ (2009) An anatomical signature for literacy. Nature 461:983-986.

Chang Y (2014) Reorganization and plastic changes of the human brain associated with skill learning and expertise. Frontiers in human neuroscience 8:35.

Cohen L, Jobert A, Le Bihan D, Dehaene S (2004) Distinct unimodal and multimodal regions for word processing in the left temporal cortex. Neuroimage 23:1256-1270.

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

Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. Science 330:1359-1364.

Dejerine J (1892) Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. Mem Soc Biol 4:61-90.

Delon-Martin C, Plailly J, Fonlupt P, Veyrac A, Royet JP (2013) Perfumers' expertise induces structural reorganization in olfactory brain regions. Neuroimage 68:55-62.

Fasanaro AM, Spitaleri DLA, Valiani R, Grossi D (1990) Dissociation in Musical Reading: A Musician Affected by Alexia without Agraphia. Music Perception: An Interdisciplinary Journal 7:259-272.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218-229.

Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3:191-197.

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci 2:568-573.

Gerber P, Schlaffke L, Heba S, Greenlee MW, Schultz T, Schmidt-Wilcke T (2014) Juggling revisited - a voxel-based morphometry study with expert jugglers. Neuroimage 95:320-325.

Grahn JA, Rowe JB (2009) Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:7540-7548.

Grainger J, Ziegler JC (2011) A dual-route approach to orthographic processing. Front Psychol 2:54.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nature reviews Neuroscience 15:536-548.

Hannagan T, Amedi A, Cohen L, Dehaene-Lambertz G, Dehaene S (2015) Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. Trends in cognitive sciences 19:374-382.

Harel A, Kravitz D, Baker CI (2013) Beyond perceptual expertise: revisiting the neural substrates of expert object recognition. Frontiers in human neuroscience 7:885.

Harley EM, Pope WB, Villablanca JP, Mumford J, Suh R, Mazziotta JC, Enzmann D, Engel SA (2009) Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. Cereb Cortex 19:2746-2754.

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34:479-490.

Huckauf A, Nazir TA (2007) How odgcrnwi becomes crowding: stimulus-specific learning reduces crowding. Journal of Vision 7:18 11-12.

Kelly AM, Garavan H (2005) Human functional neuroimaging of brain changes associated with practice. Cereb Cortex 15:1089-1102.

Krause V, Schnitzler A, Pollok B (2010) Functional network interactions during sensorimotor synchronization in musicians and non-musicians. Neuroimage 52:245-251.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535-540.

Law LN, Zentner M (2012) Assessing musical abilities objectively: construction and validation of the profile of music perception skills. ploS One 7:e52508.

Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RS, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A 97:4398-4403.

Mahon BZ, Caramazza A (2009) Concepts and categories: a cognitive neuropsychological perspective. Annu Rev Psychol 60:27-51.

Muayqil T, Davies-Thompson J, Barton JJS (2015) Representation of visual symbols in the visual word processing network. Neuropsychologia 69:232-241.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97-113.

Pegado F, Comerlato E, Ventura F, Jobert A, Nakamura K, Buiatti M, Ventura P, Dehaene-Lambertz G, Kolinsky R, Morais J, Braga LW, Cohen L, Dehaene S (2014) Timing the impact of literacy on visual processing. Proceedings of the National Academy of Sciences of the United States of America 111:E5233-5242.

Pelli DG, Tillman KA (2008) The uncrowded window of object recognition. Nat Neurosci 11:1129-1135.

Pinho AL, de Manzano O, Fransson P, Eriksson H, Ullen F (2014) Connecting to create: expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:6156-6163.

Plaut DC, Behrmann M (2011) Complementary neural representations for faces and words: a computational exploration. Cogn Neuropsychol 28:251-275.

Polk TA, Stallcup M, Aguirre GK, Alsop DC, D'Esposito M, Detre JA, Farah MJ (2002) Neural specialization for letter recognition. J Cogn Neurosci 14:145-159.

Proverbio AM, Manfredi M, Zani A, Adorni R (2013) Musical expertise affects neural bases of letter recognition. Neuropsychologia 51:538-549.

Saygin ZM, Osher DE, Koldewyn K, Reynolds G, Gabrieli JD, Saxe RR (2012) Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. Nat Neurosci 15:321-327.

Shum J, Hermes D, Foster BL, Dastjerdi M, Rangarajan V, Winawer J, Miller KJ, Parvizi J (2013) A brain area for visual numerals. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:6709-6715.

Srihasam K, Vincent JL, Livingstone MS (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. Nat Neurosci 17:1776-1783.

Srihasam K, Mandeville JB, Morocz IA, Sullivan KJ, Livingstone MS (2012) Behavioral and anatomical consequences of early versus late symbol training in macaques. Neuron 73:608-619.

Stanzione M, Grossi D, Roberto L (1990) Note-by-Note Music Reading: A Musician with Letter-by-Letter Reading. Music Perception: An Interdisciplinary Journal 7:273-283.

Stewart L (2005) A neurocognitive approach to music reading. Annals of the New York Academy of Sciences 1060:377-386.

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

Szwed M, Qiao E, Jobert A, Dehaene S, Cohen L (2014) Effects of literacy in early visual and occipitotemporal areas of Chinese and French readers. J Cogn Neurosci 26:459-475.

Szwed M, Dehaene S, Kleinschmidt A, Eger E, Valabregue R, Amadon A, Cohen L (2011) Specialization for written words over objects in the visual cortex. Neuroimage 56:330-344.

Thiebaut de Schotten M, Cohen L, Amemiya E, Braga LW, Dehaene S (2012) Learning to Read Improves the Structure of the Arcuate Fasciculus. Cereb Cortex.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273-289.

Ventura P, Fernandes T, Cohen L, Morais J, Kolinsky R, Dehaene S (2013) Literacy acquisition reduces the influence of automatic holistic processing of faces and houses. Neurosci Lett.

Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L (2007) Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55:143-156.

Vinckier F, Naccache L, Papeix C, Forget J, Hahn-Barma V, Dehaene S, Cohen L (2006) "What" and "where" in word reading: ventral coding of written words revealed by parietal atrophy. J Cogn Neurosci 18:1998-2012.

Wong YK, Gauthier I (2010a) A multimodal neural network recruited by expertise with musical notation. J Cogn Neurosci 22:695-713.

Wong YK, 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 YK, Gauthier I (2012) Music-reading expertise alters visual spatial resolution for musical notation. Psychonomic Bulletin & Review 19:594-600.

Tables

	Musicians	Controls
Melody score	14.6 (1.5)	8.8 (3.0)
Tuning score	13.8 (3.4)	9.4 (3.2)
Speed score	14.8 (1.8)	11.6 (3.1)
Beat score	13.2 (1.4)	9.5 (2.5)
Total score	56.4 (4.2)	39.4 (10.4)

Table 1: Mean (SD) behavioral scores on the PROMS test. Musicians performed significantly better than controls in all 4 subtests (t-test, all P<0.001)

		Mean MNI coordinates		
Stimuli	Group	x	y	z
Music	Controls	-52	-61	-9
	Musicians	-54	-60	-9
Words	Controls	-47	-56	-13
	Musicians	-51	-52	-12

Table 2: Mean MNI coordinates of individual peak activations by words and music, in controls and musicians.

Captions to figures

Figure 1. (A) Examples of stimuli used in the fMRI experiment. (B) Hit rate in detecting repetitions of words, faces, houses, tools and music scores, in musicians (yellow bars) and controls (green bars). Musicians were more proficient than controls with music scores (P=0.025).

Figure 2. Whole-brain music-selective responses, assessed by the contrast of music minus non-music stimuli. (A) Red activations: conjunction of music-selective responses in controls and in musicians, masking by music minus rest in controls and musicians, respectively. Green activations: Correlates of musical expertise, assessed by comparing music-selective responses in musicians minus in controls, masking by music minus rest in musicians. (B) Asymmetry of the correlates of musical expertise, showing a set of left-hemispheric activations.

Figure 3. Top panel: category-selective activations assessed by contrasting each non-musical category minus the other three, pooling controls and musicians (voxelwise P<0.001, clusterwise 0.05 corrected). Bottom panel: Barcharts show the change in BOLD signal (arbitrary units) at the peak of the category-selective regions relative to baseline, in controls (green bars) and in musicians (yellow bars), induced by music (M), words (W), faces (F), houses (H) and tools (T). Error bars represent 1 SEM across subjects after subtraction of each subject's overall mean. Musical expertise changed the profile of activation of the VWFA, while the properties of the other category-selective regions were not affected. In controls, music activated the VWFA less than words, while in musicians, activation was stronger for music than for words.

Figure 4. Analyses of individual left occipitotemporal activations to music and words. (A) Individual peaks were identified within a left occipitotemporal region of interest (blue). (B) Mean amplitude of activations at individual peaks, showing larger activation at word peaks for words (W) than music (M), and larger activation at music peaks for music than words, both in musicians and controls. (C) Mean MNI x and y coordinates (+/- 1 SEM) are shown for music (triangles) and words (circles), in controls (green) and musicians (yellow). The first two axes of the PCA are shown in red. Activations for music were posterior and lateral to activations for words, and activations for words were more anterolateral in musicians than in controls.

Figure 5. Mean volume of the activation clusters surrounding individual word peaks during the perception of words (W), and music peaks during the perception of music (M), respectively. The volume of activation to music was larger in musicians than in controls (P<0.001).

Figure 6. Functional connectivity of ventral visual cortex during the processing of written words and music. Left occipitotemporal word-selective and music-selective ROIs were identified in each subject (see Figure 4). Psycho-physiological interaction (PPI) analysis assessed which brain regions were correlated with the word ROI (left column) and the music ROI (right column), during word more than music perception (green), and during music more than word perception (red), separately for musicians and controls. A set of left-predominant language-related regions were more connected to the word ROI during word reading than during music viewing in controls (A) and musicians (B), with no difference across groups (C). A bilateral network was more connected to the music ROI during music than during words viewing in controls (D) and musicians (E), with a strong predominance in the latter group (F).

Figure 1

(A) Sample stimuli

(B) Hit rate in detecting picture repetitions
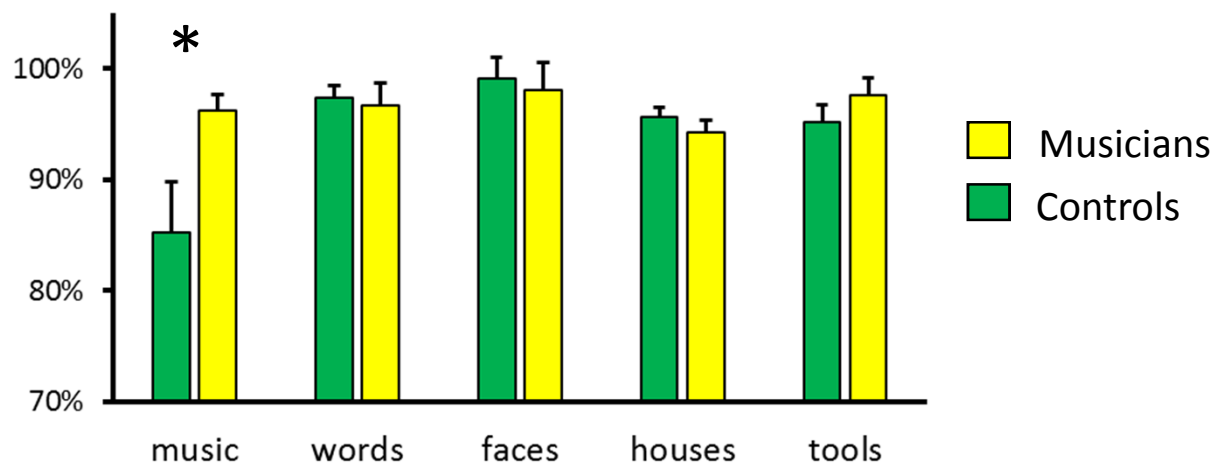
Figure 2

Whole-brain correlates of musical expertise

(A) Music > other stimuli

■ in controls and musicians   ■ in musicians > controls (musical expertise contrast)

z=-12   z=-5   z=19   z=39   z=62

L   R
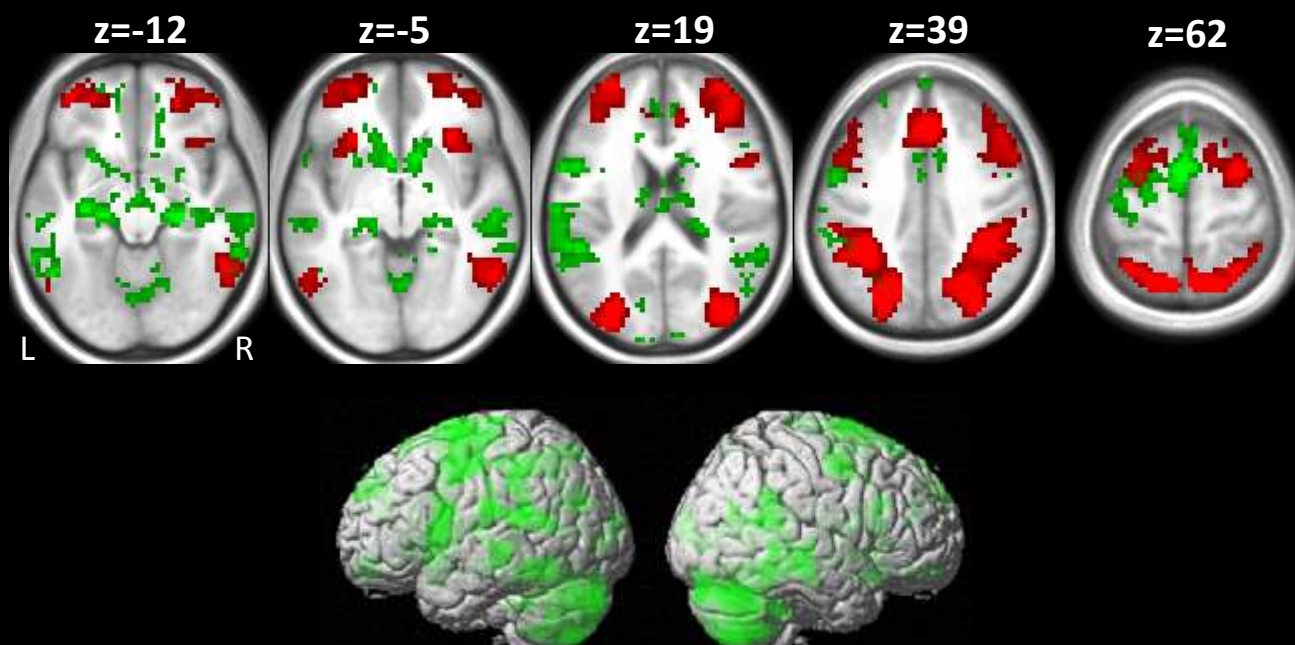
(B) Asymmetry of the musical expertise contrast

z=-12   z=-5   z=20   z=37   z=59

L   R
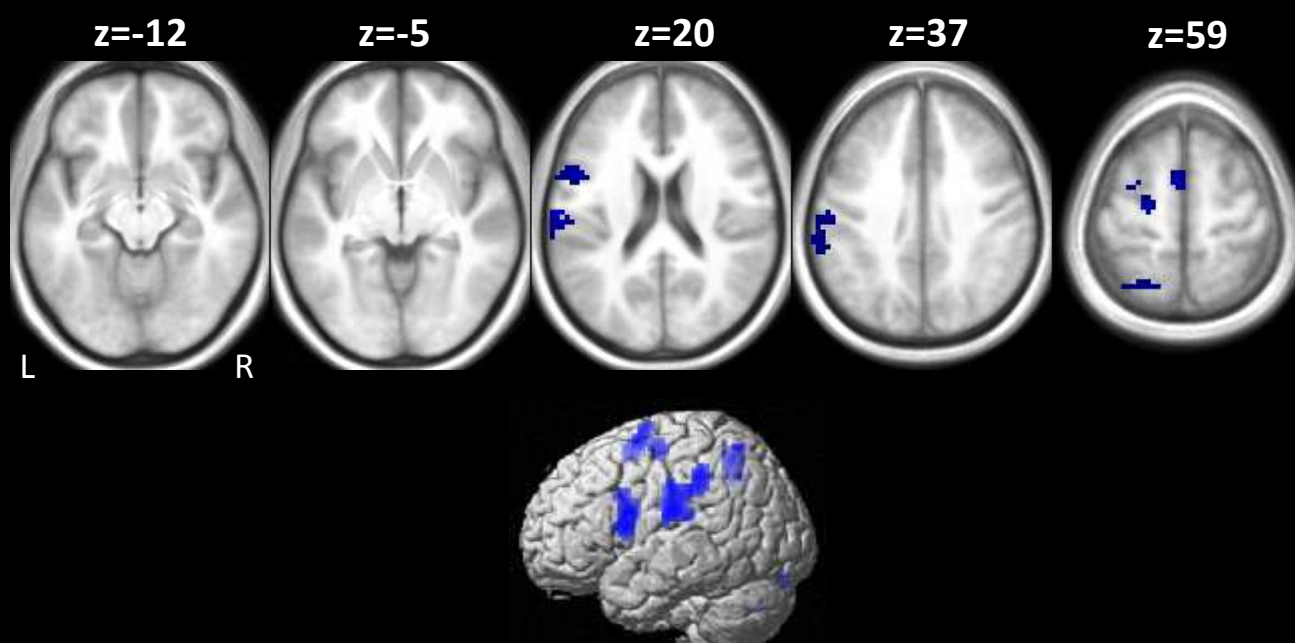
Figure 3
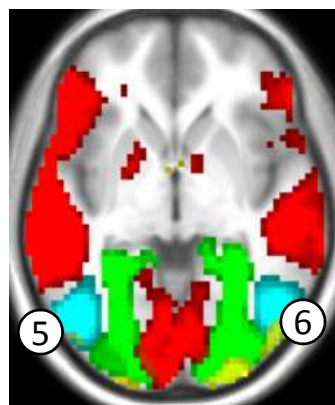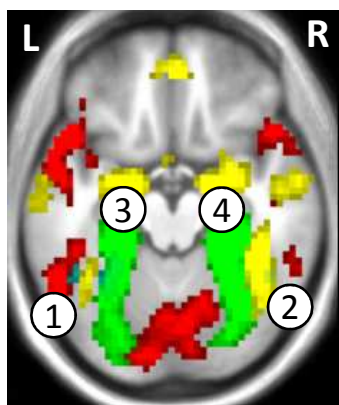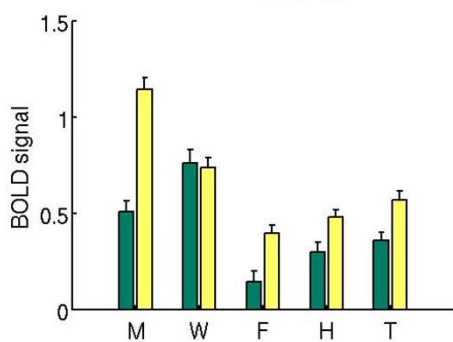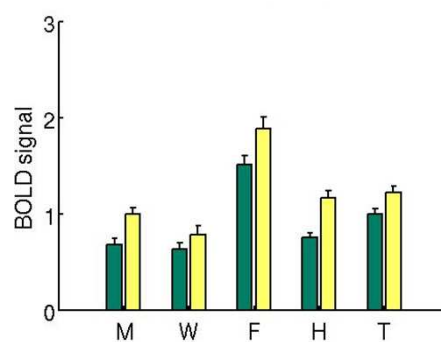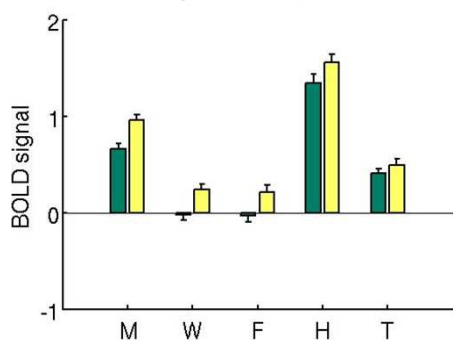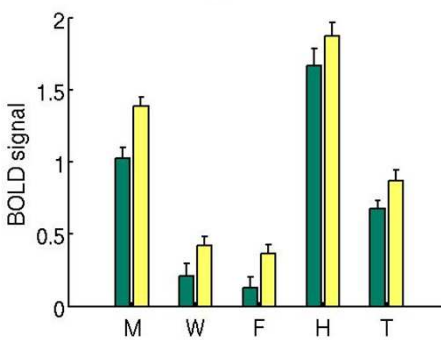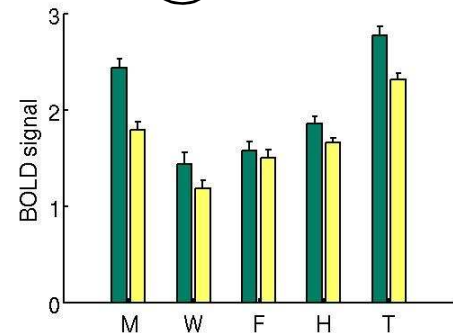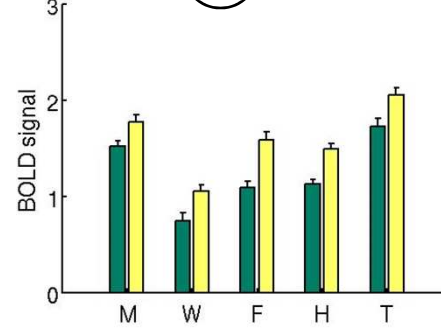
Activations in category-selective regions

Figure 4

Individual left occipitotemporal activations to music and words

(A)

(B) Activation profile
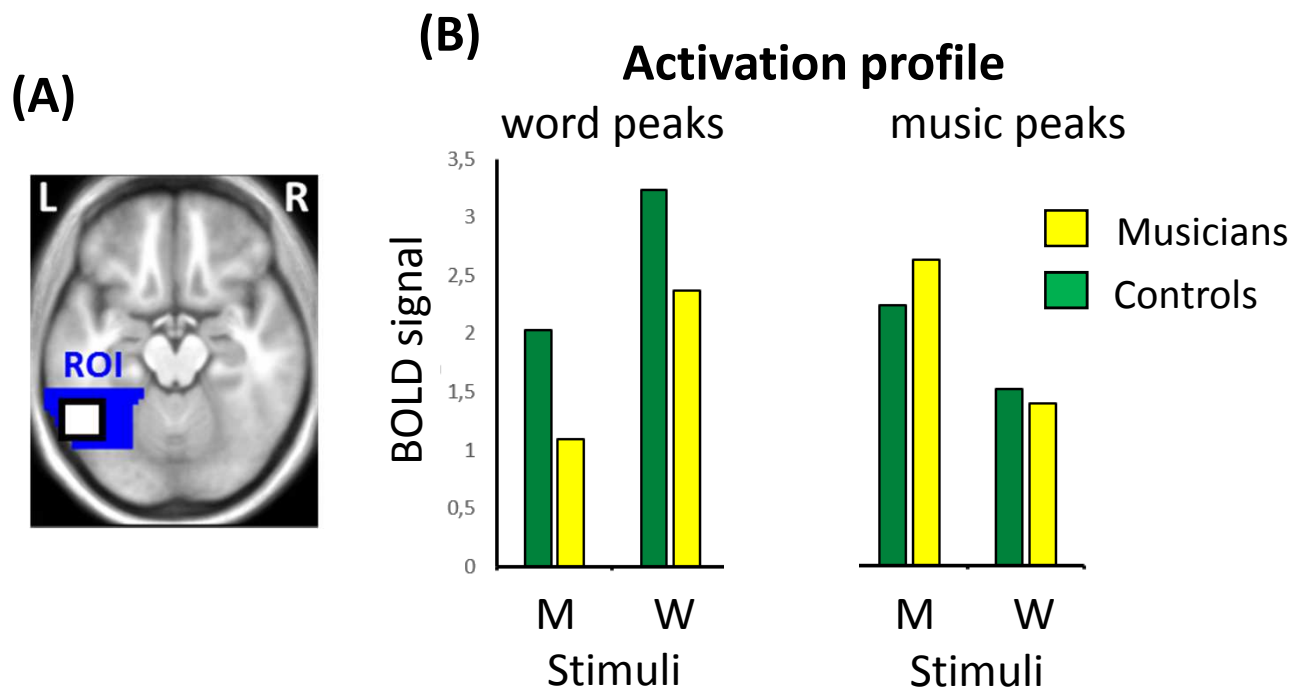
(C) Topography of peak activations

○ Words – (faces+houses+tools)

△ Music – (faces+houses+tools)

Figure 5

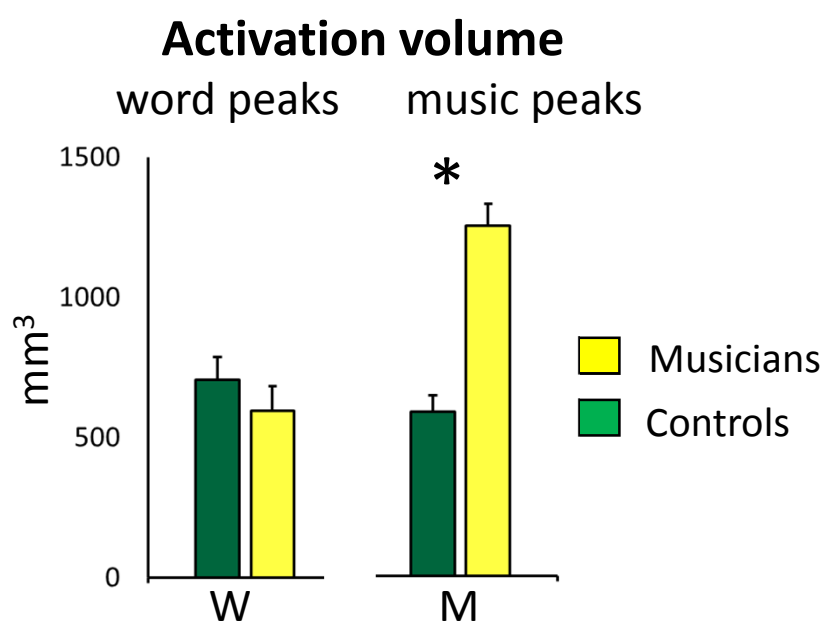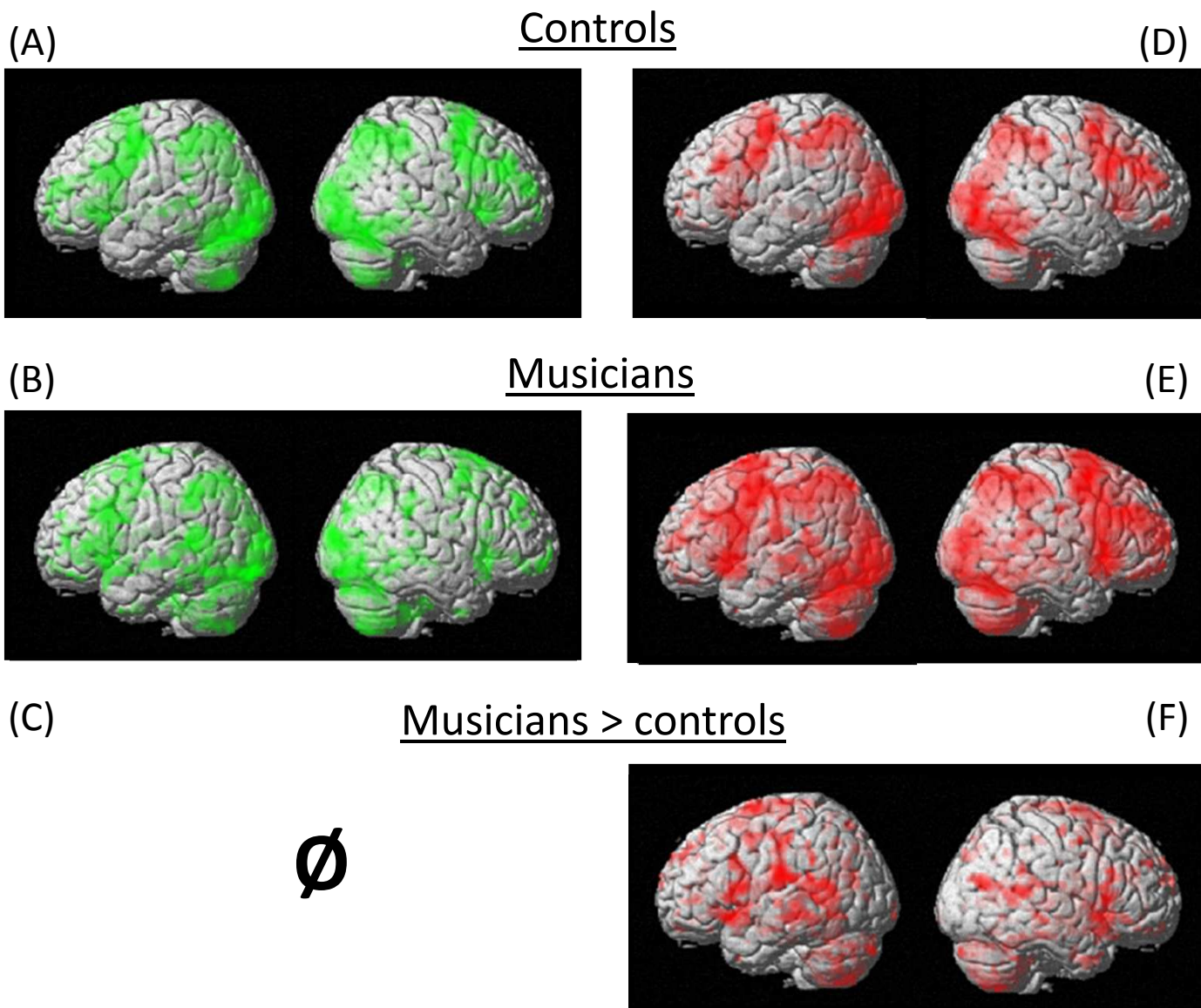**Activation volume**

word peaks   music peaks
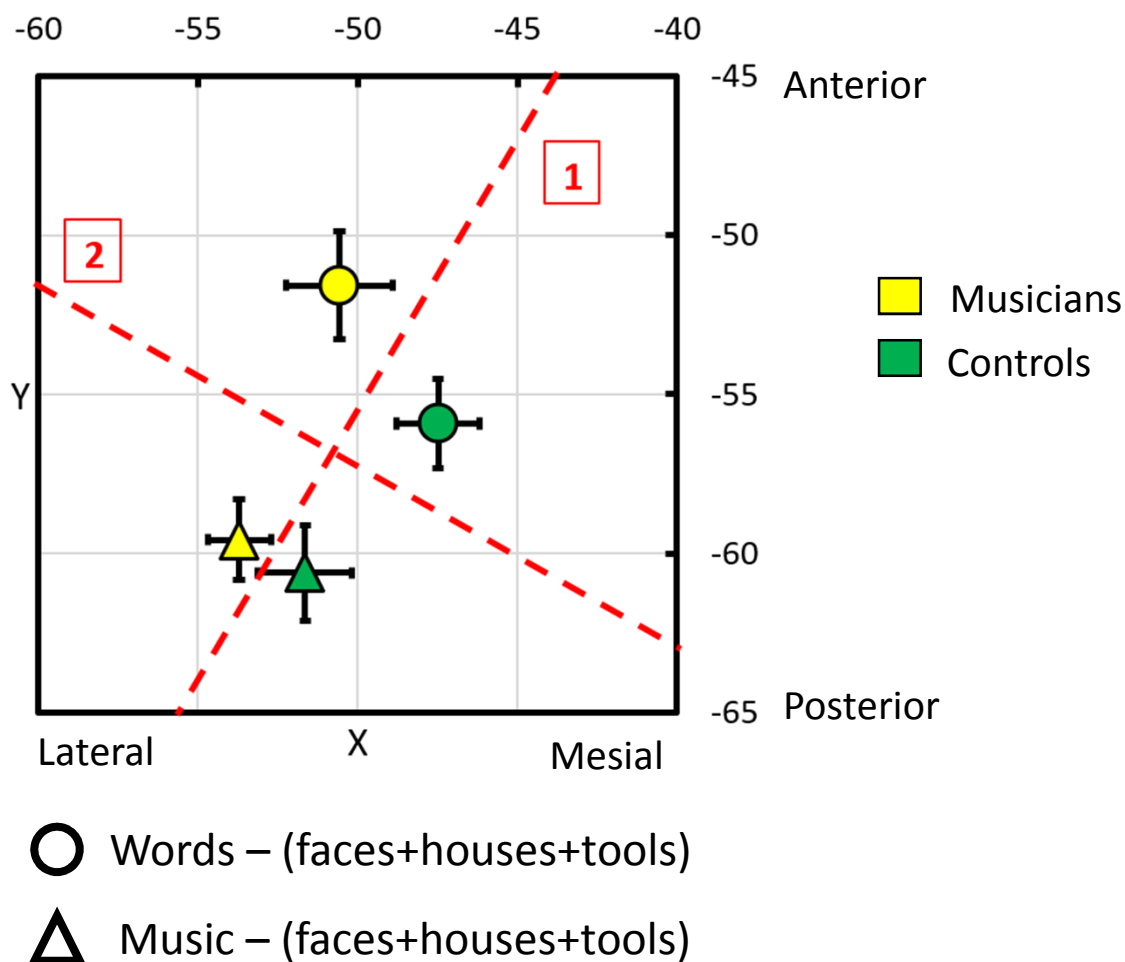
Figure 6

Psychophysiological interaction

Seed region: Individual peaks for words	Seed region: Individual peaks for music

Controls

(A)

(D)

Musicians

(B)

(E)

Musicians > controls

(C)

(F)

Ø

Regions more correlated with the seed:
🟩 during words than music perception
🟥 during music than words perception