Music Lexical Networks

The Cortical Organization of Music Recognition

Isabelle Peretz, Nathalie Gosselin, Pascal Belin, Robert J. Zatorre, Jane Plailly, and Barbara Tillmann

International Laboratory for Brain, Music, and Sound Research (BRAMS), Montreal, Canada

University of Montreal, Montreal, Canada

Department of Psychology and Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, Scotland, United Kingdom

Montreal Neurological Institute, McGill University, Montreal, Canada

Neurosciences Sensorielles, Comportement, Cognition, UMR CNRS 5020, Université Lyon 1, Lyon, France

Successful recognition of a familiar tune depends on a selection procedure that takes place in a memory system that contains all the representations of the specific musical phrases to which one has been exposed during one’s lifetime. We refer to this memory system as the musical lexicon. The goal of the study was to identify its neural correlates.

The brains of students with little musical training were scanned with functional magnetic resonance imaging (fMRI) while the student listened to familiar musical themes, unfamiliar music, and random tones. The familiar themes were selected from instrumental pieces well-known to the participants; the unfamiliar music was the retrograde version of the familiar themes, which the participants did not recognize; and the random sequences contained the same musical tones but in a random order. All stimuli were synthesized and played with the sound of a piano, thereby keeping low-level acoustical factors similar across conditions. The comparison of cerebral responses to familiar versus unfamiliar music reveals focal activation in the right superior temporal sulcus (STS). Re-analysis of the data obtained in a previous study by Plailly et al. also points to the STS as the critical region involved in musical memories. The neuroimaging data further suggest that these auditory memories are tightly coupled with action (singing), by showing left activation in the planum temporale, in the supplementary motor area (SMA), and in the inferior frontal gyrus. Such a cortical organization of music recognition is analogous to the dorsal stream model of speech processing proposed by Hickok and Poeppel.

Key words: music lexicon; memory; familiar music; superior temporal sulcus (STS); functional magnetic resonance imaging (fMRI)

Introduction

Humans from infancy to old age are able to recognize a familiar tune. Many moments of life are associated with a song or a particular piece of music. Yet, we still know very little about this basic and widespread musical ability. This may be due in part to the apparent ease and speed with which music recognition is normally carried out, that is, automatically. The goal of the present study was to identify the neural correlates of the major processing components involved in the recognition of a familiar tune.
Figure 1. Schematic representation of the main processing components involved in the recognition of a tune. What distinguishes the perceptual analysis of the random tone sequences from the analysis of the familiar and unfamiliar musical stimuli used here is the musical structure that arises from the presence of tonal and metric regularities (not shown here). (Adapted from Peretz and Coltheart.)

Recognition of a familiar tune depends on a set of computations that transform the acoustic signals into a perceptual (auditory) representation that makes contact with the musical lexicon (Fig. 1). The musical lexicon is presently best understood as a perceptual representational system for isolated tunes, much in the same way as the mental word lexicon represents isolated words. Access to the musical lexicon is conceived as automatic and incremental by involving three stages: access, selection, and integration. In the access stage, the beginning of the music activates a series of potential tune candidates. This set of candidates is determined solely on the basis of bottom-up information deriving from computations performed along the melodic and temporal perceptual analysis of the input. In the selection stage, the cohort of tune candidates is progressively reduced with the increase of available musical information about the specific tune. During this discrimination phase, the level of activation of the candidates is raised or lowered depending on their compatibility with the musical input, until the best-fitting match is selected and hence recognition is achieved. Singing from memory or inner singing can be initiated at the selection stage and proceed onward through the vocal plan formation for overt singing. This cohort model has been originally developed for word recognition by Marslen-Wilson and has been shown to account for tune recognition as well. Finally, the integration stage implies that the position and function of the selected tune or musical phrase is integrated into a larger musical context, that is, within a higher-level representation of the piece. This integration phase may involve syntactic rules and extra-musical associations. Here, we focus on the access and selection stages in the musical lexicon.

The neural correlates of the musical lexicon point to the superior temporal gyrus as
a key brain structure. Long ago, Penfield and Perot\(^9\) reported that electrical stimulation of the exposed surface of the auditory cortex could elicit music-specific memories in the form of a musical hallucination. Stimulation on both sides of the temporal lobes, with a slight predominance for the right side, could result in the patients’ reporting illusory musical memories. Similarly, bilateral brain damage to the auditory cortex can result in persistent loss of recognition abilities for music despite normal or near-normal perceptual processing of musical input.\(^{10-12}\) Such memory loss can be limited to music. For instance, an amusic patient with bilateral damage to the auditory cortex was normal at recognizing and memorizing spoken lyrics, whereas she performed at chance when required to recognize or to relearn the corresponding melody (played without lyrics). The deficit was selective because the patient had no difficulties with other nonmusical auditory materials, such as voices and animal cries, and had no memory impairment for visual stimuli.\(^{12}\) Hence, the memory failures can be not only modality-specific, but also music-specific.

While difficulties in recognizing familiar melodies can occur after damage to either left or right superior temporal structures,\(^{10}\) the participation of the bilateral or left superior temporal structure and of the left frontal areas have been emphasized in several neuroimaging studies.\(^{1,13-15}\) These findings may be attributed in part to the fact that familiar melodies are associated with extra-musical and extra-experimental events that may contribute to recognition. For instance, song melodies (played without lyrics) automatically trigger the lyrics with which they are typically paired.\(^{16}\) The presence of these associated memories may even confer an advantage in the case of lesion. As uncovered by Steinke et al.,\(^{17}\) brain damage can impair recognition of instrumental music, but spare recognition of song melodies. However, in two neuroimaging studies,\(^{1,14}\) both instrumental music and perceptual tasks were used, thereby diminishing the contribution of associated lyrics. Yet in both studies, the superior temporal sulcus (STS) located in the left hemisphere appeared critical. Such left-lateralization may relate to acoustical factors. For instance, in these two studies,\(^{1,14}\) different excerpts served as familiar and unfamiliar stimuli. Given the intricacy of the perceptual requirements and the variety of representations relevant to music recognition, it is essential to control the acoustical and the perceptual attributes. The present study represents such an attempt, as explained later.

Another way to tap the musical lexicon is to study musical imagery. Imagery refers to the subjective experience of being able to imagine music in the absence of real sound input. By applying behavioral methods developed by Halpern\(^{18}\) to patients with focal auditory cortex lesions, it was possible to demonstrate that imagery deficits are found after damage to right auditory cortical areas.\(^{19}\) This general conclusion has been supported in subsequent work using functional neuroimaging, which consistently indicates that secondary auditory cortices are recruited during a variety of tasks that involve imagery or rehearsal of melodies.\(^{13,20,21}\) Further evidence comes from electrophysiological measures showing that the scalp topography elicited by imagining the continuation of a melody is similar to the N100 component elicited by a real tone.\(^{22}\) These demonstrations of auditory cortex activity in the absence of an acoustical stimulus, or at least not driven solely by external input, support the contention that access to the memory representations in the musical lexicon is determined by perceptual processes. In addition, selection of a music candidate, when generating a familiar tune, tends to engage inferior frontal regions\(^{20,21}\) and the supplementary motor area (SMA\(^{20,23}\)), which may relate to subvocalization or inner singing.

In sum, recognition of a familiar tune involves a number of procedures that makes the isolation and brain localization of the musical lexicon relatively difficult. In this endeavor, control over the perceptual input appears essential,
especially when functional magnetic resonance imaging (fMRI) is the selected method of investigation. To this aim, we use here musical stimuli that are as similar as possible in terms of acoustical features, such as pitch, timbre, intensity, and tone density. Furthermore, for the unfamiliar music, we use retrograde adaptations of the familiar themes so as to keep both acoustical structure and musical structure as similar as possible. Because access to the musical lexicon is mainly a bottom-up process, as pointed out by results obtained with the gating paradigm,⁷,⁸ and because music recognition is mostly driven by pitch structure,²⁴,²⁵ we expect the familiar–unfamiliar contrast to reflect activation of the musical lexical representations in the STS not only on the left side, but also on the right side of the brain. This activation of the STS and more generally of the superior temporal gyrus should also be seen in the contrast between unfamiliar music and random sequences, as the result of perceptual analysis of musical structure and lexical search. Indeed, it has been shown that pseudowords may produce greater activation of the neural networks involved in word recognition than real words.²⁶ In order to tap this bottom-up access to the lexicon, participants listened to the auditory stimuli without task instructions. The idea was that mere exposure would be more likely to reflect automatic access to the lexicon than a familiarity decision task that might activate associate memories.

Materials

Twenty-eight melodies were selected so as to be highly familiar to the subjects.²⁷ They were taken from instrumental pieces that were not originally sung with lyrics. The 28 unfamiliar melodies were retrograde adaptations of the familiar melodies (as done by Hébert et al.²⁸). The adaptations consisted of slight corrections to the temporal structure in order to eliminate some awkward metrical results and avoid recognizability. From the 1375 tones that made up the 56 melodies, 708 were randomly selected and concatenated to create 28 random sequences. Overall energy (RMS) was matched across melodies (Cool Edit; Syntrillium Software Corporation, Phoenix, AZ), which were computer-generated, played with the sound of a piano for 8.5 s, and faded out over the last 0.5 s. An example is presented in musical notation in Figure 2 and can be heard at http://www.brams.umontreal.ca/plab/publications/article/112.

Procedure

Scanning was performed on a 1.5-T Siemens Sonata imager (Erlangen, Germany). A high-resolution T1-weighted anatomic scan was obtained for each subject (voxel size: 1 × 1 × 1 mm³; matrix size: 256 × 256). Then one run of 115 T2*-weighted gradient-echo planar images of blood oxygen level–dependent (BOLD) signal, an indirect index of neural activity, was acquired. A head coil was used to obtain 20 interleaved slices (TE: 50 ms, TR: 11.5 s, flip angle: 90 deg, matrix size: 64 × 64, voxel size: 5 × 5 × 7 mm³) with an acquisition time of 1.9 s. The long inter-acquisition time minimized the effects of scanner noise on participants’ ability to hear the auditory stimuli and avoided contaminating the BOLD signal response to the stimuli in the auditory cortices.²⁹

The inter-acquisition interval could be filled with either one of the 28 familiar melodies, one of the 28 unfamiliar melodies, one of
the 28 random sequences, or left empty in 28 trials. Order of presentation was pseudo-randomized so that a stimulus condition was never presented twice in sequence. Participants were required to close their eyes and to listen to the auditory stimuli, which were presented binaurally at 80-dB sound-pressure level A via custom MR-compatible pneumatic sound-transmission headphones. After scanning, participants performed a familiarity judgment task on a 10-point scale (with 1 = very unfamiliar–10 = very familiar) on the 28 familiar melodies that were mixed with the 28 unfamiliar melodies in a random order.

Data Analysis

BOLD signal images were smoothed using a 12-mm gaussian kernel, corrected for motion, and transformed into standard stereotaxic space using in-house software applying the MNI 305 target. The statistical analysis of the fMRI data was based on a linear model with correlated errors and implemented in a suite of MATLAB programs (Natick, MA). First, the stimulus conditions were set up in a design matrix corresponding to each acquisition. Second, the linear model was solved for, yielding the effects, standard deviations, and t-statistics for each run and for each contrast. In a final step, these results were combined, yielding the group statistical maps (across all participants) for each contrast.

Group-average statistical images were obtained for each condition by computing an omnibus test on individual t-maps using a pooled estimate of standard deviation and a corrected threshold established at $t > 4.32$ ($P < 0.05$), based on the number of resolution elements in the acquisition volume (2880 resels). Furthermore, we used a volume of interest (VOI) approach for specific structures identified as important in retrieving melodies from memory (i.e., STS, SMA) and for brain areas that were significantly activated in the t-maps corrected for multiple comparisons (i.e., the planum temporale and the inferior frontal gyrus). The maximum number of voxels obtained in the individual t-map for a condition contrast was identified within a search radius of 7 mm from these maxima. BOLD signal values for each condition were then converted to percent signal increases by reference to the mean value for the silence condition and subjected to analyses of variance with condition (familiar, unfamiliar and random) as within-subjects factor.

Results

As can be seen in Figure 3, participants correctly classified the melodies as familiar and unfamiliar with the exception of one familiar melody, which was judged to be unfamiliar by more than half of the participants. This stimulus was discarded from subsequent analyses.

Familiar–Unfamiliar Contrast

The BOLD responses associated with listening to the familiar melody relative to the unfamiliar, time-reversed melody revealed significant activation in the right STS [$x, y, z$ spatial coordinates: 48, −24, −10; $t(16) = 4.66$;...
Figure 3. Mean ratings of familiarity judgments are presented for the 28 familiar themes (solid circles) and for their 28 retrograde versions (open circles). The gray circle (#23) indicates the familiar tune that was discarded from the analyses. Vertical bars indicate between-participants standard errors.

\[ P < 0.05; \text{see Fig. 4} \] and less in the left STS \([-48, -24, -10; t(16) = 2.38, \text{n.s.}]\). The subsequent VOI analysis in this brain structure confirms that the right STS is more activated by the familiar melodies as compared to both the unfamiliar melodies and random sequences \(F(2,16) = 6.51, P < 0.01; \text{see Fig. 4}\). On the left side, the VOI analysis does not reach significance \(F(2,16) = 3.61, \text{n.s.}\). It is worth noting that activation in the right STS for the unfamiliar melodies relative to silence is actually a deactivation (Fig. 4). Of interest, ventral striatum and precuneus subthreshold deactivation can be seen (Table 1).

Familiar–Random Contrast

The BOLD responses associated with listening to the familiar melody relative to random sequences activates three brain regions: the left SMA \((-4, -4, 66)\), the planum temporale \((-60, -40, 24)\), and the inferior frontal gyrus \((-56, 14, -2)\). Listening to a familiar melody is associated with more activation in the SMA as compared to random sequences \(t(16) = 4.00; P < 0.07, \text{by the whole-brain analysis, and } F(2,16) = 6.44, P < 0.01, \text{by the VOI analysis; see Fig. 5C}\). More activation is also seen at the planum temporale \([-60, -40, 24]; t(16) = 5.94; P < 0.05, \text{by the whole-brain analysis; see Fig. 5A}\) and in the inferior frontal gyrus \([-56, 14, -2; t(16) = 4.36; P < 0.05, \text{by the whole brain analysis; see Fig. 5B}\]. As previously seen in the comparison between familiar and unfamiliar music, familiar music versus random sequence contrast reveals subthreshold activation in both the ventral striatum and the precuneus (Table 1 and Fig. 5).

Unfamiliar–Random Contrast

This contrast between the unfamiliar melodies and the random sequences does not reveal any significant activation in the whole-brain analysis. However, the VOI analysis reveals that listening to an unfamiliar music activates more significantly the planum temporale \((-60, -40, 24)\) and the inferior frontal gyrus \((-56, 14, -2)\) than random tones \(F(2,16) = 13.00, P < 0.001, \text{and } F(2,16) = 4.15, P < 0.05, \text{respectively; see Fig. 5}\).

\(^{f}\)Another contrast that might be interesting to examine is the feeling of novelty that might be elicited by the melodic reversal. However, this contrast was not associated with activation in a particular area of the brain in the whole-brain analysis.
Figure 4. The familiar–unfamiliar contrast reveals significant activation in the right superior temporal sulcus (STS), which is superimposed upon horizontal (top), coronal (middle), and sagittal (bottom) sections taken from the MNI stereotaxic space. The right hemisphere is on the right. Graphs show the percent signal change relative to the silence condition for the three conditions (familiar, unfamiliar, and random music) from 7-mm radius arrays centered on voxels corresponding to location of the peak height maxima of the left and right STS. Asterisks indicate significant t-test comparison. Vertical bar indicates standard error. (In color in Annals online.)

Discussion

Music recognition refers to a set of operations that transform an auditory signal into a mental representation that can make contact with memories. There are multiple transformations to consider in the mapping from auditory signal to memory retrieval. At the very least, the acoustical input must be transformed in an abstract musical representation that can be mapped on memory representations. Comparing brain responses to musically structured stimuli (i.e., unfamiliar melodies) relative to random tones was expected to tap perceptual processing of Western musical structure. However, this contrast did not reveal any significant cortical activation. The data were most revealing about the contact with music memories in the musical lexicon.

Recognition of a familiar instrumental melody involves a network including the right, and to some extent the left STS, the left planum temporale, the left SMA, and the left inferior frontal gyrus. The STS appears central in the recognition process. The STS activation is clearly modulated by familiarity, with familiar melodies eliciting increases and unfamiliar melodies decreases of the same brain structure. This pattern of BOLD responses squares nicely with the selection stage in the cohort model. At the selection stage, the familiar melody has found a perceptual match in memory, and the unfamiliar melody has been eliminated as a possible candidate in the cohort. This elimination may be associated with a deactivation, as seen here at the level of the STS. Indeed, this deactivation most likely reflects termination of lexical search and not perceptual processes because perceptual processing was similar in the case of the unfamiliar music.

In principle, this pattern should also be observed in prior neuroimaging studies. The study that is closer in design to the present one is the study conducted by Plailly et al. In that study, odors and instrumental excerpts of commercial recordings of music were presented in a familiarity decision task. The results are reported in terms of bimodal brain activations associated with feeling of familiarity (in comparison to feeling of unfamiliarity) elicited by both odors and instrumental music. Music-specific and odor-specific activations are also reported, but these modality-specific activation patterns excluded brain activation common to both modalities. For the purpose of comparison, Jane Plailly and Barbara Tillmann re-analyzed their data with a small volume correction centered on the STS activation coordinates found in the present study (48, −24, −10) for the contrast of interest (familiar music–unfamiliar music). Interestingly, this new contrast yields a significant right
TABLE 1. Stereotaxic Coordinates (x, y, z; in MNI Space) and Significance Levels (t-Values) of Activation Foci in the Familiar–Unfamiliar and Familiar–Random Contrasts

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Familiar–unfamiliar</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right superior temporal sulcus</td>
<td>48</td>
<td>−24</td>
<td>−10</td>
<td>4.67a</td>
</tr>
<tr>
<td>Left superior temporal sulcus</td>
<td>−48</td>
<td>−18</td>
<td>−16</td>
<td>2.43</td>
</tr>
<tr>
<td>Left ventral striatum</td>
<td>−24</td>
<td>2</td>
<td>−8</td>
<td>3.75</td>
</tr>
<tr>
<td>Right ventral striatum</td>
<td>16</td>
<td>4</td>
<td>−8</td>
<td>3.44</td>
</tr>
<tr>
<td>Left precuneus</td>
<td>−8</td>
<td>−72</td>
<td>46</td>
<td>−4.16</td>
</tr>
<tr>
<td>Right precuneus</td>
<td>4</td>
<td>−70</td>
<td>46</td>
<td>−4.25</td>
</tr>
<tr>
<td><strong>Familiar–random</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left planum temporale</td>
<td>−60</td>
<td>−40</td>
<td>24</td>
<td>5.94a</td>
</tr>
<tr>
<td>Right planum temporale</td>
<td>60</td>
<td>−30</td>
<td>10</td>
<td>3.54</td>
</tr>
<tr>
<td>Left inferior frontal gyrus</td>
<td>−56</td>
<td>14</td>
<td>−2</td>
<td>4.37a</td>
</tr>
<tr>
<td>Supplementary motor area</td>
<td>−4</td>
<td>−4</td>
<td>66</td>
<td>4.00</td>
</tr>
<tr>
<td>Left ventral striatum</td>
<td>−24</td>
<td>2</td>
<td>−5</td>
<td>3.64</td>
</tr>
<tr>
<td>Right ventral striatum</td>
<td>20</td>
<td>6</td>
<td>0</td>
<td>3.54</td>
</tr>
<tr>
<td>Left precuneus</td>
<td>−8</td>
<td>−72</td>
<td>46</td>
<td>−3.12</td>
</tr>
<tr>
<td>Right precuneus</td>
<td>8</td>
<td>−72</td>
<td>36</td>
<td>−4.05</td>
</tr>
</tbody>
</table>

*a Significant at \( P < 0.05 \) by whole-brain analysis.

*b Based on probabilistic map.30

Figure 5. Activation in the planum temporale (A), in the inferior frontal gyrus (B), and in the SMA (C) in the familiar–random contrast. Graphs show the percent signal change relative to the silence condition for the three conditions (familiar, unfamiliar, and random music) from 7-mm radius arrays centered on voxels corresponding to location of the peak height maxima of each region. Asterisks indicate significant t-test comparison. Vertical bar indicates standard error. A (\( z = 6 \) shows the bilateral activation of the planum temporale. In B (\( z = 2 \)), one can see both the activation in the left inferior frontal gyrus and the bilateral activation of the ventral striatum. (In color in Annals online.)
STS activation $(45, -24, -11; Z = 3.10)^4$ and a subthreshold activation in the left STS $(−54, −33, 14; Z = 2.64)$. However, the BOLD signal in the STS is diminished, not reversed, in response to unfamiliar music. Thus, the dynamic of the BOLD response of the STS to unfamiliar music remains to be determined. What is remarkably consistent across the two studies is the involvement of the right STS in listening to familiar music.

It is also clear that additional regions participated in the process of music recognition in both Plailly et al.’s study and the present situation. Here, listening to familiar music is associated with activity in the planum temporale, the SMA, and the inferior frontal gyrus. One possible role of these associated areas is that they reflect (inner) singing. For example, listening and covert singing of familiar songs involve similar structures, namely, the right planum temporale. Similarly, overt singing involves bilaterally the STS and the inferior frontal gyrus.

In line with previous studies, we also observed some activation in brain regions involved in reward (e.g., ventral striatum), as recently revealed in singing. Thus, the implication of a dorsal pathway from the STS to the inferior frontal gyrus, which appears involved in covert and overt singing, may have played a similar role here when listening to familiar music.

Covert singing while listening to music is a common experience. Both the familiar and unfamiliar selections used here are “singable” despite the fact that they are coming from an instrumental repertoire. Moreover, we have found in a prior study that singing is a much better index of music recognition than is naming (title retrieval) even though the musical cues were played on a piano. Thus, we propose that participants have spontaneously sung along with the music, albeit covertly and mostly with the familiar stimuli, less so with the unfamiliar music and not at all with the random sequences. This gradient of “singability” corresponds to the differential degree of activation seen here along the dorsal pathway (Fig. 5). Because we did not record overt singing here, this account remains a hypothesis to be tested in future studies.

In sum, comparing familiar music to unfamiliar music, we found, in two independent studies, focal and bilateral activation in the STS with a right-hemisphere bias. This region probably contains musical lexical networks. What the present neuroimaging data further suggest is that these auditory memories are tightly coupled with action (singing). Such a cortical organization of music recognition is analogous to the dual-stream model of speech processing proposed by Hickok and Poeppel. In this model, the speech signal is initially analyzed in the STS, also referred to as the phonological lexicon. From this early phonological analysis two processing streams are distinguished: a ventral stream for comprehension, and a dorsal stream for articulation, including the angular gyrus as well as the premotor cortex and the inferior frontal gyrus. The present data suggest that the process of music recognition might be similarly organized in the brain.

Acknowledgments

Preparation of this paper was supported by grants from Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and from a Canada Research Chair.

Conflicts of Interest

The authors declare no conflicts of interest.

References


