The structural components of music perception
A functional anatomical study

Hervé Platel,1,2 Cathy Price,3 Jean-Claude Baron,1 Richard Wise,3 Jany Lambert,1,2
Richard S. J. Frackowiak,3 Bernard Lechevalier1,2 and Francis Eustache1,2

1INSERM U.320, 2Service de Neurologie Dejerine,
CHU de Caen, University of Caen, France and 3Wellcome
Department of Cognitive Neurology, Institute of Neurology,
London, UK

Summary
This work explores the cerebral structures involved in the appreciation of music. We studied six young healthy subjects
(right handed, French, without musical talent), using a high resolution PET device (CTI 953B) and 15O-labelled water.
In three tasks, we studied the effects of selective attention to pitch, timbre and rhythm; a final task studied semantic
familiarity with tunes (considered as divided attention for pitch and rhythm). These four tasks were performed on the
same material (a tape consisting of 30 randomly arranged sequences of notes). We selected a paradigm, without a
reference task, to compare the activations produced by attention to different parameters of the same stimulus. We
expected that the activations recorded during each task would differ according to the differences in cognitive
operations. We found activations preferentially in the left hemisphere for familiarity, pitch tasks and rhythm, and in
the right hemisphere for the timbre task. The familiarity task activated the left inferior frontal gyrus, Brodmann area (BA)
47, and superior temporal gyrus (in its anterior part, BA 22). These activations presumably represent lexico-semantic
access to melodic representations. In the pitch task, activations were observed in the left cuneus/precuneus (BA
22). These results were unexpected and we interpret them as reflecting a visual mental imagery strategy employed to
carry out this task. The rhythm task activated left inferior Broca’s area (BA 44/6), with extention into the neighbouring
insula, suggesting a role for this cerebral region in the processing of sequential sounds.

Keywords: PET; music; activation paradigm; cognitive strategy; selective attention; functional imaging

Abbreviations: ANCOVA = analysis of covariance; BA = Brodmann area; rCBF = regional cerebral blood

Introduction
We need only a few seconds to identify ‘Le Boloero’ by Ravel. One of the most notable features of this famous piece of
music is that it induces the listener into focusing his/her attention successively on the basic components of the piece:
first the characteristic rhythm given by the snare drum, secondly the enchanting melody played several times and
finally the timbre of the various wind instruments on which the melody is played. We are endowed with a unique
capability to direct attention to different sound characteristics of a musical tune and to virtually extinguish perception of
one feature to highlight another. This behaviour shows that perceiving music is a complex cognitive process that is based
on several ways of categorizing auditory stimuli. This in turn suggests that listening to music is based on distinct neural
processes corresponding to the basic components of music. However, little is known of the neural structures involved in
such processing, and most of our knowledge comes from the neuropsychological literature.

Thus, clinico-pathological studies, particularly of auditory agnosia, have shown that the perception of language and
music are differentiated at the neuroanatomical level and that within music perception, specific abilities may be selectively
impaired. For instance, rare cases of aphasia without amusia following damage to the left hemisphere (Luria et al., 1965;
Signoret et al., 1987), and the reverse pattern following right hemisphere lesions (Judd et al., 1979; McFarland and Fortin,
1982) have been reported. Although the belief that perception of music is a capacity specific to the right hemisphere is
common, studies in brain-damaged subjects suggest that musical perception involves both hemispheres (Lechevalier
et al., 1985). As emphasized by Peretz (1994), ‘the bias seen in favour of the right hemisphere is based on an accumulation of data in a very specific sector of musical perception, the perception of different pitches, presented alone or in sequence’ (p. 213).

In agreement with this view, reports suggest that, although identification and recognition of a musical piece may involve both hemispheres, the integrity of the left hemisphere is critical (Lechevalier et al., 1995), as illustrated by a left brain-damaged patient who exhibited impaired melody identification despite intact melody discrimination (Eustache et al., 1990).

Likewise, both dichotic listening experiments in normal subjects and studies in brain-damaged subjects support the idea that the right hemisphere is involved in discrimination of melodies (Kimura, 1964; Zatorre, 1979; Peretz, 1985). However, other studies have shown impaired melodic outline utilization followed right hemisphere damage, and impaired pitch–interval discrimination followed lesions to either hemisphere (Peretz, 1990), suggesting at least two distinct neural systems underlying this melodic processing.

There are data suggesting that timbre perception is preferentially processed by the right hemisphere. For instance, Milner (1962) and Chase (1967) reported that right temporal lobectomy induced significant impairment in timbre perception tasks compared with left temporal lobectomy. Similarly, Samson and Zatorre (1988, 1991, 1994) have shown that patients with right anterior temporal lobectomy performed significantly worse in discriminating spectral information than both healthy volunteers and patients with left anterior temporal lobectomy. Mazzuchi et al. (1982) reported a single case with relatively pure disturbance in timbre perception following right temporal brain damage.

The left hemisphere appears to be devoted to the processing of the rhythmic, temporal and sequential components of music. Thus, Papcun et al. (1974) reported a left hemispheric advantage in both naïve subjects and Morse code operators performing a dichotic letter perception task in Morse code, a task equated with rhythm discrimination by the authors. Right hemisphere advantage was found in naïve subjects if sequences exceeded seven elements, consistent with a working memory component to the left hemisphere and perhaps a more holistic contribution of the right hemisphere. A similar left hemisphere advantage was reported by Robinson and Solomon (1974) and Gordon (1978) in auditory dichotic tasks of rhythmic patterns without phonetic support. Consistent with these findings from normal subjects, Peretz (1985) reported disruption in rhythm perception after left hemisphere damage, while Mavlov (1980) described a right-handed music teacher with left temporoparietal damage and impaired recognition and production of rhythm patterns, independent of sensory modality.

To summarize the above literature, the left hemisphere appears to be specialized for both rhythm and access to musical semantic representations (i.e. identification and recognition of melodies), while the right hemisphere is particularly engaged in melodic perception (pitch contour) and timbre. However, some specific abilities, within these broad functions, such as the processing of pitch intervals in melodic perception, appear to have reverse hemispheric dominance (Peretz and Morais, 1980; Peretz et al., 1994) indicating the need to address individual aspects of music perception. Furthermore, the studies reported above have all lacked precision at the neuroanatomical level.

PET is one method for mapping the neuronal networks active during perceptual-cognitive processing. However, in contrast to the sophistication of current psychological models (McAdams and Bigand, 1993) little work has been devoted to the perception of different components of music. In a pioneering study Mazziotta et al. (1982) reported one of the rare experiments on musical perception (Seashore test). In timbre discrimination tasks, increased cerebral glucose utilization was observed bilaterally in posterior temporal regions, particularly in right superior/posterior temporal region and the right/left asymmetry was most pronounced in frontal regions. The same results were observed for chords and musical sequences. The PET study of Mazziotta et al. (1982) is the only one, to our knowledge, in which different components of music perception have been studied, but the methodology used was suboptimal by today’s standards. In a study of right monaural listening to pure sounds (500 Hz and 4 kHz), Lauter et al. (1985) confirmed a tonotopic distribution in primary auditory cortex (left temporal superior gyrus of Heschl) with anterior and lateral activation for low sounds and more medial and posterior activation for high-pitched sounds. In a pitch discrimination task using verbal material (phonemes), Zatorre et al. (1992) reported activation in the right inferior frontal region in contrast to passive listening. Zatorre et al. (1994) also studied melodic listening and memorizing tonal information. The tasks involved the perception of pitch change, either between the first two notes of eight-note melodic sequences, or between the first and last notes of the sequence. The first task (pitch judgement between the first two notes) induced right frontal lobe activations. The second task induced a much more complex activation pattern, with the engagement of right frontal and temporal lobes, insula, and parietal cortex.

No PET activation study has, as yet, assessed either the specificity of musical semantic processing or explored the different sub-components of musical expression, such as timbre, pitch and rhythm. We therefore set up a PET study which consisted of four different activation tasks on the same auditory material: (i) an identification/familiarity task; (ii) an attention to pitch task; (iii) an attention to timbre task; (iv) an attention to rhythm task. Our purpose is not to exhibit all of the cerebral regions activated during these processes, but on the contrary, to pick out of the common processing areas those activation foci which are specific for each of these tasks. The first objective of this research is therefore to confirm with a PET study in healthy subjects the existence of functional independance (suggested by the neuropsychological literature) between selected musical components. In
order to demonstrate this independance, we conceived an experimental paradigm that capitalized on the enhancement of activation with ‘mental focusing’ (Corbetta et al., 1990; Pardo et al., 1991; Zatorre et al., 1992). In other words, focusing attention on changes in pitch, for example, should enhance cerebral activity associated with the processing of this musical parameter. The validity of this principle has been demonstrated in other sensory areas (Roland and Frieberg, 1985; Corbetta et al., 1990; Meyer et al., 1991; Sergent et al., 1992b; Demonet, 1994). The underlying hypothesis of our study is that the psycho-acoustic parameters pitch, rhythm and timbre, and the semantic representations of melody, are processed by neural networks for the most part common and in part specific. Functional independence should manifest itself by a marked hemispheric dominance for certain processes and by different activation foci for processes involving the same hemisphere. From the literature on music perception, we predicted specific activations principally in the right hemisphere for the rhythm and pitch tasks, and in the left hemisphere for the rhythm and familiarity tasks.

Material and methods
Four distinct processes were studied: (i) identification/familiarity, related to the recognition of ‘semantic clues’ taken from well-known tunes (see below); (ii) changes in timbre; (iii) changes in pitch; (iv) rhythm regularity.

These four parameters were the material of four distinct tasks carried out on the same sound material. Following an instruction, which was always identical, the subject produced a motor response by pressing on the buttons of a computer mouse (index or middle finger of the right hand). This paradigm allows us to investigate whether there are separate cerebral activation foci for the processing of four musical components when the musical material is kept constant.

Description of the sound material
The sound tape was made of 30 sequences of notes. Each sequence (or pattern) contained 5–10 notes, and lasted <2.5 s. Between each sequence there was a silent interval of ≥1.5 s allowing for the subject’s response. The total duration of the tape was 2 min.

The temporal division of the recordings was carried out by computer, using the CUBASE sequence program. The sounds were provided by a Kawai Klm synthesizer. They resembled the digital recording of real instruments. The tests were recorded on digital tapes, using a DAT Sony 57 ES system.

The 30 sequences of notes covered four classes, in part overlapping, as follows.

Fifteen sequences containing a change in timbre
Two synthesized timbres in the sound spectrum of the oboe were used. One of the timbres was more brilliant (contained more harmonics) than the other. Patterns with tonal change corresponded to sequences of notes in which the two tones of the oboe could be heard alternately. In other patterns, one instrument alone played all the notes of a sequence.

Fifteen sequences containing changes in pitch
These corresponded to sequences in which one or more notes were produced at a different fundamental pitch, from the first note The pitch intervals between the notes were variable, but the notes were always contained within the same octave (centred around A, 440 Hz). This task is not the appreciation of a melodic contour or of intervals between notes, but simply a judgement of whether the sequence contains a pitch change or not. Of the 15 sequences containing pitch changes, 10 fitted melodies that could be recognized. The remaining five ‘non-recognizable’ sequences were created with musical melody to reduce differences between ‘plausible’ and ‘artificial’ melodies.

Fifteen sequences containing a rhythmic irregularity
In these sequences the length of the notes, or the intervals between them, was not always identical. There were 15 sequences out of the 30 in which the lengths of the notes, and/or the intervals between them, were equal, and 15 where they were irregular.

Among the 30 sequences, certain had multiple changing characteristics (e.g. only changes of pitch for one, but changes of rhythm, pitch and ‘semantic anchorage’ for another). To avoid any learning effect, four different

Fifteen sequences with ‘semantic anchorage’ (or clue)
Patterns with ‘semantic anchorage’ had either a melodic contour and/or a rhythmic arrangement that brought to mind a familiar piece of music (e.g. Mozart’s ‘Eine Kleine Nachtmusik’, ‘The Blue Danube’ of Strauss, military marches, etc.). We use the term ‘semantic anchorage’ or ‘clues’ because none of the sequences was produced with the orchestration, tempo or original key of the scores. Despite this, the melodic outline and particular rhythm of each was so preserved that each ‘clue’ was identifiable. The extracts were chosen such that the subjects, who were French, would not recall words that could be sung, and in turn that the sequences would not be easily verbalized. For some excerpts, only the title of the musical piece could possibly be remembered. We balanced 10 such sequences with clues by creating five sequences, each solely recognizable by rhythm (no change of pitch), by melody (no change of rhythm) or by both melody and rhythm. Examples of sequences solely recognizable by rhythm were ‘Bossa nova’, ‘Boogie’ beats, Ravel’s ‘Boléro’, and tapped sequences such as those from pop songs or famous advertisements.
arrangements of the 30 sequences were made. The order of presentation of the sequences was different for each of the tasks.

In addition to the four different arrangements of the ‘definitive tape’, a ‘test tape’ was also made to train and familiarize the subjects. It used the same matrix sequence as the ‘definitive tape’ but differed in the changes made in certain of the parameters (particularly the semantic clues). It also existed in four different arrangements.

The sounds were reproduced binaurally by two small hi-fi earphones. Pressing on the right or left button of a computer mouse held in the subject’s right hand made it possible to record responses and, by means of a chronometric computer program, calculate the reaction time for responses. The start of the sound tape and the chronometric program were synchronized by subjects, when they pressed one of the buttons of the computer mouse.

**Instructions**

The six selected subjects were trained with the type of stimuli and tasks to be performed. Nevertheless the sound material and the instructions were again explained at the time of the scan. The four instructions given to the subjects for each sequence were as follows. ‘You are going to hear 30 sequences of notes’, then one of the following.

(i) For semantic anchorage: ‘Press the left-hand button if the sequence of notes brings to mind, through the rhythm and/or the melody, something that you know, and the right-hand button if it doesn’t remind you of anything’.

(ii) For timbre: ‘Press the left-hand button if the sequence is played with two different timbres and the right-hand button if all the notes of the sequence are played with the same timbre’.

(iii) For pitch: ‘Press the left-hand button if there are one or more notes with a different pitch and the right-hand button if all the notes of the sequence are played at the same pitch’.

(iv) For rhythm: ‘Press the left-hand button if the lengths of the intervals and the notes in the sequence are irregular and the right-hand button if the lengths of all the intervals and notes are regular’.

Thus, the paradigm employed was such that the motor response was identical across the four tasks.

**Validation of the tapes**

For both the training and the definitive tapes, the subjects’ responses were recorded and classified as either correct or incorrect. The number of errors were used during the recruitment stage to see whether certain tasks seemed more difficult than others or whether certain of the sequences were significantly more difficult. Following a pre-test run using 10 subjects, the definitive tape was modified, on the basis of the results obtained, with new patterns, a slower rhythm and two new, more easily differentiated instrument timbres. The speed of sequence presentation and the time allowed for subject’s response in the final tape were adjusted from the mean errors and the response time during the test sessions, and from the evaluation of subjects themselves about the level of mental focusing required for each of the tasks. However, it proved impossible to adjust the tasks so that both the error rates and the reactions times were perfectly matched (see Results).

**The subjects**

The subjects were six young French males aged between 20 and 30 years, right-handed (according to the Edinburgh scale) and with a comparable educational level. They completed a medical questionnaire to exclude any psychiatric or organic pathology, particularly neurological, and none had hearing difficulties. They were not undergoing psychotropic treatment and all were ‘musical individuals’ according to the Wertheim and Botez (1959) classification. None of them could read music.

At the first interview they were given mock tests and only those who had a success rate of ≥80% in each were retained.

The volunteers were made aware of the meaning and purpose of the study and of their role in it, and gave their written and informed consent.

This research was approved by the Hammersmith Hospital Ethics Committee and the permission to administer radiation was given by the Administration of Radioactive Substances Advisory Committee of the British Department of Health.

**PET methodology**

Each subject carried out each task three times, which corresponded to three series of four tasks, i.e. 12 PET scans for each of the subjects. Because we wished to avoid as far as possible automatic semantic processing during the timbre, pitch and rhythm tasks, the task concerning ‘semantic anchorage’ was always carried out first. In addition to varying the arrangements for the sequences, and to avoid any order effect and limit learning effects, the three other tasks (corresponding to the pitch, timbre and rhythm) were carried out in a different order for each series.

Regional cerebral blood (rCBF) was studied by recording the distribution of cerebral radioactivity after intravenous injection with water molecules marked with oxygen-15 (\(^{15}\)O), which diffuses freely in the brain. An increase in rCBF results in an increase in the level of local radioactivity (Mazziotta et al., 1985; Fox and Mintun, 1989). Each rCBF study began with a background scan lasting 30 s, followed by the injection of labelled water (H\(_2^{15}\)O for 20 s). A 2-min scan of radioactivity in the cerebral tissues was started as soon as H\(_2^{15}\)O appeared in the brain. The presentation of the stimuli commenced at the beginning of the scan.

We used a CTI 953B PET scanner (CTI Inc., Knoxville, Tenn., USA) in 3D mode, with the inter-detector collimating septa removed (Watson et al., 1993). The emission data were
corrected for the attenuating effects of tissues by means of transmission scans.

**PET image analysis**

The corrected emission data were reconstructed as 31 axial planes by filtered back projection with a Hanning filter (cut-off frequency 0.5 cycles/pixel). The resolution of the resulting images was $8.5 \times 8.5 \times 6.0 \text{ mm}^3$ (full width half maximum) (Spinks et al., 1992).

The images were reformatted into 43 planes with bilinear interpolation with each plane displayed in a $128 \times 128$ pixel format (each pixel $= 2 \times 2 \text{ mm}^2$).

The planes obtained were corrected for head movement between scans, using the AIR (Automated Image Registration) software specifically developed for the purpose (Woods et al., 1992). The realigned images from each subject were averaged and the bi-commissural line (linking the anterior and posterior commissures) was identified with the aid of a computer (Friston et al., 1989). All the images from all subjects were then transformed into the standard anatomical space of the stereotaxic atlas of Talairach and Tournoux (1988). Images were smoothed using a Gaussian filter, 10 pixels (20 mm) wide. This filtering was carried out in order to attenuate the high frequency noise and to allow constructive interference between scattered homologous functional foci from different individuals during averaging (Friston et al., 1991).

All calculations and image manipulations were carried out on SPARC workstations (Sun Computers Europe Inc., Surrey, UK) using ANALYZE version 6 software (BRU, Mayo Foundation, Rochester, Minn., USA; Robb and Hanson, 1991), and PROMATLAB (Mathworks Inc., Natick, Mass., USA). Statistical maps of rCBF changes were calculated using SPM software (Wellcome Department of Cognitive Neurology, London, UK; Frackowiak and Friston, 1994).

The activity during each task was then averaged voxel by voxel following normalization for differences in flow (to a mean of 50 ml/dl/min) which was achieved by analysis of covariance (ANCOVA), with global flow as the covariate. This procedure permits the calculation of an estimate of the adjusted error variance associated with the measurement of mean task-related flow in each voxel. Task-specific brain activation was assessed by statistical comparisons with appropriate linear contrasts (weighting of the four condition means) using the $t$ statistic (Friston et al., 1991). This analysis was performed for all voxels in parallel and for each planned comparison. The resulting sets of $t$ values constitute statistical parametric maps (SPM(t)).

The subsets of voxels exceeding a threshold of $P < 0.001$ (or $P < 0.05$ with correction for multiple non-independent comparisons) were represented as statistical parametric maps rendered onto three orthogonal projections ($x$, $y$ and $z$). This threshold of significance was chosen, as empirical studies have shown that it protects against false positives (Bailey et al., 1991).

**Between-task comparisons**

The comparisons among tasks were designed to account for the cognitive structure of the tasks proposed in our paradigm. This structure is summarized in Fig. 1, in which the cognitive components and attentional levels involved for each of the tasks are shown. We assume that the components related to acoustic processing, working memory and motor decision are equivalent across all four tasks, and thus any cerebral activity related to these components would be cancelled out by the comparisons.

The semantic task (familiarity) involved attention which was divided between pitch and rhythm, these being the two psychoacoustic parameters which make it possible to have a sense of familiarity for the heard sequences; timbre does not affect performance in this task. The distribution of cognitive components shown in Fig. 1 helped limit the number of

---

**Fig. 1** Synoptic graph representing the cognitive components assumed to be involved in each of the four tasks in the present experimental paradigm. In this model, three levels of attentional weight are given to each component, and are shown as a grey–white scale. In this cognitive model, we assume motor processing, working memory, and acoustic processing are identical among the four tasks (see Material and methods). Note that the familiarity task involves divided attention between pitch and rhythm processing, in addition to semantic processing.
Table 1 Areas with significant rCBF activation *

<table>
<thead>
<tr>
<th>Areas activated</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarity versus (pitch and rhythm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left inferior frontal gyrus</td>
<td>-46</td>
<td>20</td>
<td>-12</td>
<td>47</td>
</tr>
<tr>
<td>Left superior temporal gyrus</td>
<td>-52</td>
<td>4</td>
<td>-8/0</td>
<td>22</td>
</tr>
<tr>
<td>Right superior temporal gyrus</td>
<td>52</td>
<td>-22</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Left middle occipital gyrus</td>
<td>-44</td>
<td>-70</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>Left anterior cingulate gyrus</td>
<td>-6</td>
<td>26</td>
<td>-4</td>
<td>24</td>
</tr>
<tr>
<td>Right internal pallidum</td>
<td>12</td>
<td>-6</td>
<td>-4</td>
<td></td>
</tr>
<tr>
<td>Timbre versus (pitch and rhythm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right superior and middle frontal gyrus</td>
<td>-20</td>
<td>30</td>
<td>36</td>
<td>32/8</td>
</tr>
<tr>
<td>Right middle frontal and precentral gyri</td>
<td>-40</td>
<td>-10</td>
<td>48</td>
<td>4/6</td>
</tr>
<tr>
<td>Left precuneus</td>
<td>-15</td>
<td>-65</td>
<td>40</td>
<td>7/19</td>
</tr>
<tr>
<td>Left middle occipital gyrus</td>
<td>-45</td>
<td>-70</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>Pitch versus (timbre and rhythm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left precuneus/cuneus</td>
<td>-14</td>
<td>-70</td>
<td>28</td>
<td>18/19</td>
</tr>
<tr>
<td>Left superior frontal gyrus</td>
<td>-30</td>
<td>42</td>
<td>32</td>
<td>9/8</td>
</tr>
<tr>
<td>Rhythm versus (pitch and timbre)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left insula</td>
<td>-36</td>
<td>12</td>
<td>0</td>
<td>13–16</td>
</tr>
<tr>
<td>Left inferior Broca</td>
<td>-38</td>
<td>8</td>
<td>8</td>
<td>44</td>
</tr>
</tbody>
</table>

Coordinates x, y and z are from the most significant voxel in each cerebral region, relative to standard stereotactic space (Talairach and Tournaux, 1988). *P < 0.001.

comparisons to those strictly necessary to obtain the relevant information.
Because the familiarity task involved two parameters with divided attention, this task was contrasted with the aggregate of pitch and rhythm. In the same way, specific activation results for pitch, timbre and rhythm tasks were identified by comparing pitch with rhythm and timbre; timbre with rhythm and pitch; and rhythm with pitch and timbre. Such comparisons assume equal weighting for each component on each side of the comparison (Frackowiak and Friston, 1994).
In addition to the above prospective strategy, and to allow a further understanding of the specific contribution of each task to the results from the aggregate-comparison strategy, we also performed a set of systematic task-by-task comparisons, which are meant to provide results of descriptive value only.

Results
Subjects’ performances in the tasks
We counted the number of errors made by subjects in the four tasks. We also separated the sum of the errors according to the order in which the tasks were carried out, as each subject completed each task three times. Overall, >80% of all tasks were completed correctly, except for the pitch task during the first presentation (error rate 25%).
The task involving timbre resulted in fewer errors than the others. The difference between the number of errors for this task and those for the pitch and rhythm tasks was significant P < 0.001 (t test); it was not significant between the pitch and familiarity tasks. These results were in line with those obtained at the experimental validation of the definitive tape with a larger sample of subjects (n = 48). Repetition of tasks caused a small decrease in the number of errors for some of the tasks (Fig. 2). The calculations of response time, on the other hand, were stable over task repetitions (Fig. 3). The pitch task showed the shortest reaction time and the highest average number of errors of the four tasks.

Activation results
There was no significant increase in relative blood flow (not even in the form of a tendency) in the temporal regions of the primary auditory cortex nor in the primary motor regions, for any of the comparisons, indicating good control for early auditory and motor processing across comparisons. Statistically significant activations were few and clearly differed by task. The vast majority of significant activations were found in the left hemisphere, except in the timbre task (Table 1).
Figures 4–7 illustrate the localization of activation foci for each task obtained, using the planned comparisons. These results can be summarized as follows.

Familiarity versus (pitch and rhythm) (Fig. 4)
We found activation foci in the left inferior frontal gyrus; in the left and right superior temporal gyri; in the left middle occipital gyrus; in the left anterior cingulate gyrus; and in the right internal pallidum.

Timbre versus (pitch and rhythm) (Fig. 5)
Activation foci were found in the right superior and middle frontal gyrus, in the depth of the right middle frontal and precentral gyri (although two apparently distinct foci appear on Fig. 5, they were in fact in very close proximity—see coronal projection—and actually merged in raw z score images; coordinates of the main peak were +40,-10,+48); in the left precuneus; and in the left middle occipital gyrus.

Pitch versus (timbre and rhythm) (Fig. 6)
Activation foci are localized in the left precuneus/cuneus; in the left superior temporal gyrus; and the left superior frontal gyrus.

Rhythm versus (pitch and timbre) (Fig. 7)
We found one large activation focus in the left insula merging with the left inferior Broca’s area.
An overview summary of the results obtained for task-by-task comparisons is shown in Table 2. This table is provided for descriptive purposes only.
Fig. 2 Mean percentage of errors made by the six subjects for each of the four tasks in the successive presentation of the stimulus tape.

Fig. 3 Mean response time for the six subjects (see Fig. 2 for explanations).

Discussion

General remarks

The response-time and number-of-errors curves show a trend toward better performance with increasing numbers of trials. These differences are slight and non-significant, indicating that they do not constitute a learning but a ceiling effect. This is particularly important for the memory task in which there is an order effect due to repetition of stimuli. However, an analysis of reaction times as function of order of presentation and SPM contrasts, using exclusively first and last presentations failed to show any significant effect of order.

The paradox observed in the case of the pitch task, concerning the number of errors and the time response, is easily explained. In a series of 10 notes, a difference of pitch is sometimes detectable from the second note which may result in a very short response time. Looking at the pattern of errors and response times in Figs 2 and 3, it appears that subjects were too confident on the first two trials answering too quickly and making numerous errors. On the third presentation of the stimulus, their uncertainty and concentration had both increased, which resulted in a longer reaction time and fewer errors.
The subjects performed the tasks in the PET scanner well, and relatively homogeneously, and results were in line with those obtained at the selection visit, although they had never heard the sequences used in the definitive tape prior to the PET session. The aim was that subjects would carry out the tasks as well as possible, that they would not find themselves in a stressful situation occasioned by excessive difficulty, but also that the difficulty would be sufficient to ensure that attention was maintained throughout. The high level of difficulty was also needed to minimize the effects of implicit processing of irrelevant characteristics in a given task, though implicit, or covert, processing was itself largely controlled for in the subtractive design of the experiment. An appropriate level of difficulty was attained as judged by the stability of reaction times and performance indicators across tasks. Moreover, we asked subjects to estimate their performance in the different tasks. Though many subjects realized that the task involving timbre was easier, most of them said that this was only relative and that the task demanded attention throughout scanning. The pitch task was felt by all the subjects to require the most mental focusing relative to the other tasks. Rhythm and familiarity tasks seemed to demand
fewer attentional resources. The subjects reports are consistent with the observed results and confirm that pitch and familiarity tasks are not identical in terms of cognitive processing.

**Activation**

The activation foci in the planned comparisons were, as predicted, few and focal. The profile of small activation areas and limited spatial extent is directly attributable to the experimental paradigm, the objective of which was to isolate areas associated uniquely to specific functions. That is not to say that other areas, common to all processes, may not be involved; indeed they may even be critical to the proper performance of the targeted process. However, the identified areas have relatively pure functional attributes in that they are associated with one particular cognitive component and not the others.

In what follows we will discuss the main activation foci specific for each task separately, and then treat the additional activation foci in the frontal and parietal areas across the four tasks.

**The familiarity task (Fig. 4 and Table 1)**

In keeping with the neuropsychological literature concerning recognition of tunes (Eustache et al., 1990; Lechevalier et al., 1995), the largest and most significant activations for this task were found in the left hemisphere.
Table 2  Descriptive results for all comparisons

<table>
<thead>
<tr>
<th></th>
<th>Pitch (-T)</th>
<th>Timbre (-R)</th>
<th>Rhythm (-F)</th>
<th>Familiarity (-P)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left hemisphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broca’s area (BA 44/6)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Insula (BA 13–16)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Superior temporal gyrus (BA 38/22)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Precuneus/cuneus (BA 18/19/31)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Inferior parietal lobule (BA 40)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Middle temporal gyrus (BA 21)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Caudate nucleus</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Superior and middle frontal gyri (BA 8/9/32)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Inferior frontal gyrus (BA 47)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Anterior cingulate gyrus (BA 24/32)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Middle occipital gyrus (BA 39/19)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Pallidum: ((x, y, z) = (0, 34, -12))</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td><strong>Right hemisphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caudate nucleus</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Superior temporal gyrus (BA 22)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Superior and middle frontal gyri (BA 32/8)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Inferior frontal and precentral gyri (BA 6/4)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Inferior parietal lobule (BA 40)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
<tr>
<td>Middle and temporo-occipital gyri (BA 19)</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
<td>. . . . . .</td>
</tr>
</tbody>
</table>

**T**, **R**, **F** and **P** = comparisons with timbre, rhythm, familiarity and pitch, respectively. ** = Significant activation at $P < 0.001$; $t$ = sub-threshold activation, i.e. $0.001 < P < 0.01$.

Activation of the left inferior frontal gyrus [Brodmann area (BA 47)], has been reported previously with different paradigms. In the work of Petersen et al. (1988), this region was active for the subtraction ‘verb generation versus repetition’. Zatorre et al. (1992) also found activation of this region in the subtraction ‘Speech versus Noise’. In the latter study, the ‘speech’ task corresponded to passive listening to phonemes with alternating motor responses, and the ‘noise’ task corresponded to passive listening to white noise. Thus, in both studies the material was verbal and the subtractions were such as to isolate possible semantic components and phonological processing, respectively. Mazoyer et al. (1994) also showed activation in the left inferior frontal region caused by listening to a list of words and intelligible sentences, and on the right by listening to ‘sentences’ composed of meaningless words. Although our material was not verbal and the impression of familiarity was sufficient for a positive response to the stimulus, there can be no doubt that this task, more than any of the other three, may bring to mind verbal material and induce subjects into lexico-semantic search. In other words, this task is the only one where subjects could explicitly ‘search for a name’ when they heard musical notes, and therefore recall semantic and possibly biographical information. The inferior frontal region is one of the critical areas damaged in cases of Broca’s aphasia. It probably plays an important role not only in speech output but also in lexico-semantic processing. It is also closely connected with the primary auditory areas (Pandya, 1995; Liegeois-Chauvel et al., 1993), via the superior temporal gyrus, which is an auditory association area (Mazoyer et al., 1994; Zatorre et al., 1992).

In our study, the most anterior part of the left superior temporal gyrus was also clearly activated. This focus, which is anterior and far remote from Wernicke’s area, may thus have an important role in the processing of non-verbal semantic information. In other words, it may well be that ‘melodic representations’ are stored in this part of the superior temporal gyrus. Activation of the right superior temporal gyrus was also present. Although not exactly at the same coordinates as the left-sided counterpart, bilateral homologous activations, with left-sided dominance, have been regularly noted in language tasks (Wise et al., 1991; Weiller et al., 1995). This pattern of activation was specific for familiarity as it exceeded that of pitch and rhythm combined, i.e. the two components that contribute to feelings of familiarity. Based on the comparisons among all tasks (Table 2), the pattern of activations found for the familiarity task partly overlaps with that obtained with pitch, but not that obtained with rhythm, consistent with the fact that melodic information predominates over rhythm in the semantic processing of occidental music.

The familiarity task was the only one associated with small specific activation foci in the left cingulate and middle occipital gyrus, and the right pallidum. We cannot exclude that, as suggested by others (Price et al., 1994), these results reflect a ‘first-scan’ time effect, since the familiarity task was always presented first in our paradigm.
The timbre task (Fig. 5 and Table 1)

Of the three psychoacoustic parameters studied, the timbre task was that with the most prominent activations of the right hemisphere. This is consistent with experimental and neuropsychological literature (Auzou et al., 1995; see also the Introduction). Previously, Mazziotta et al. (1982), using a rudimentary [18F]fluorodeoxyglucose activation method, reported an increase in glucose metabolism in the superior and posterior temporal regions (more marked in the right hemisphere) for a timbre discrimination test (Seashore test). The activation foci in the precentral and the inferior frontal gyri of the right hemisphere appeared very limited in extent. However, examination of single comparisons (Table 2) indicates a broad rostro-caudal activation in BA 6 over 10 mm (z = 20/30), extending throughout BA 4 (particularly in the ‘timbre versus familiarity’ comparison). Even though these areas in the left hemisphere are recognized as fundamental to language output, very little is known of their functional attributes in the right hemisphere. Attention to differences in timbre would, on the basis of our data, be one function attributable to these areas of the right hemisphere. The finding that these activations mainly reflect the ‘timbre versus familiarity’ comparison (see Table 2) suggests some participation of the same right hemispheric areas in pitch and rhythm processing, though to a lesser extent than by timbre. In other words, the timbre, pitch and rhythm tasks of our protocol appeared to engage closely neighbouring regions in the right hemisphere, which in turn explain why the specific activations associated with the timbre task were weaker than in the other conditions. We cannot offer at present a firm interpretation for the two small left-sided posterior activations observed in the timbre task, relative to the other two tasks. However, one could argue that automatic pitch analysis occurred during this task (which required specific training to distinguish changes in timbre from changes in pitch), implying some recourse to visual mental imagery (see below). This idea is reinforced by the lack of significant left precuneus activation in the ‘timbre versus pitch’ comparison, but its presence in the ‘timbre versus rhythm’ comparison (Table 2).

The pitch task (Fig. 6 and Table 1)

The activation pattern seen with pitch processing was, at first glance, the most unexpected in the light of the available literature, but the characteristics of this task clarify, for the most part, these results. First, all the statistically significant activation foci were in the left hemisphere (Fig. 6), whereas previous experimental work on melodic perception suggests that this is a right hemisphere function (see the Introduction). Furthermore, the PET studies of Zatorre et al. (1992, 1994) support the idea of specialization in right cortical regions (temporal and fronto-temporal) in the processing of pitch, while in the present study, the most statistically significant activation was in the left cuneus/precuneus, which has little apparent connection with auditory processing. Our results presumably have something to do with the particular nature of the task. What was asked of the subjects was not to make a comparison between two melodies, but to assess the pitch of all elements of the sequences in order to determine whether there were one or more elements of a different pitch. Peretz and Babai (1992) make a distinction between two types of processing involved in listening for changes of pitch in musical sequences: melodic outline (a harmonic profile of a musical sequence) and processing of intervals (which is carried out by a comparison of notes, taken out of their context, with their respective fundamental pitches). By studying dissociations in brain-damaged subjects, they showed that distinguishing melodic outlines relies mainly on the right hemisphere, while processing intervals has more to do with the left. Hemispheric specialization can thus be interpreted as the result of a strategy of global listening in the processing of outlines, and of analytical listening to intervals. This ‘hemispheric’ effect, linked to a cognitive strategy, has already been suggested with PET by the work of Mazziotta et al. (1982) and Phelps and Mazziotta (1985). These authors showed that activations during the comparison of note sequences (Seashore test) were localized on the left (superior temporal areas) when the subject’s strategy was ‘analytic’, and on the right (inferior parietal and tempo-occipital areas) when the subject’s strategy was ‘passive’. We believe that the task involving pitch used in our study is closer to the processing of intervals than of the melodic outline of musical sequences. As the task was carried out with the least success, it is also possible that difficulty may induce a more analytical cognitive strategy, which would preferentially involve areas of the left hemisphere.

The above reasoning makes it possible to suggest another interpretation for the specific activation of the cuneus/precuneus in our task. Being close to the primary visual areas, the activation of this region is often interpreted as reflecting visual ‘mental imagery’ (Demonet et al., 1992; Grasby et al., 1993; Fletcher et al., 1995). The contribution of the cuneus/precuneus to tasks involving visual material has been demonstrated (Corbetta et al., 1993). In their study, Sergent et al. (1992a) found activation foci, close to ours, in the left cuneus/precuneus areas for the reading of a musical score by musicians (‘reading score versus visual fixation’ comparison). What is the evidence that allows us to think that a visual mental strategy could be induced by this task? Probably more than any other task in our protocol, the pitch task tapped into visual imagery in terms of ‘high’ and ‘low’ in relation to a notional base line (a ‘mental stave’), and this idea was suggested by the subjects themselves during debriefing. Additional factors in favour of the use of a ‘visual representation’ type strategy are that all subjects had their eyes closed and found this test the most difficult. These findings are specifically in connection with the musical material used here and do not relate to the treatment of information about the pitch of the sounds themselves, but to a cognitive strategy used by subjects in order to improve
their performance. Thus, these results may be thought to adduce evidence in favour of a relationship between certain forms of image representation and the precuneus/cuneus.

The rhythm task (Fig. 7 and Table 1)

All the activation foci for this task were located in the left hemisphere, consistent with the neuropsychological literature, which attributes the processing of sequential and temporal material to the left hemisphere (see Introduction). Of all our tasks only that involving rhythm caused statistically significant activation of Broca’s area (i.e. BA 44 and 6). Broca’s area is, of course, known for its close link with speech output but, as pointed out by Demonet (1994) it has been shown to be activated in various tasks which do not involve oral expression, such as silent generation of verbs linked semantically with auditorily-presented nouns (Wise et al., 1991), acoustic analysis of sequences of phonemes (Démonet et al., 1992), categorization of consonants according to their phonological representation (Sergent et al., 1992a), rhyming task with pairs of syllables presented by auditory means (Zatorre et al., 1992), and memorizing a sequence of consonants presented visually (Paulesu et al., 1993a). At present these results are interpreted in terms of inner speech, of an automatic, articulatory recapitulation strategy which subjects employ to carry out the tasks. For Demonet (1994) these results argue in favour of the motor theory of speech perception put forward by Liberman in the 1960s (see Liberman and Mattingly, 1985), which postulated a strong functional connection between language perception and the articulation of verbal sounds. The temporal analysis of a sequence of sounds may call for an inner speech strategy related to phonological transcoding (the ‘recapitulation phenomenon’).

However, a different, though related, cognitive mechanism may have been operating. Thus, the activation of Broca’s area for the rhythm task may reflect the processing and organization of sequential auditory stimuli, verbal or otherwise. This interpretation would perhaps account for the fact that the rhythm task also activated the adjacent left insula. Indeed, this cortical area is known to have extensive connections with primary and secondary auditory regions (Lieggeois-Chauvel et al., 1993). Its activation was mentioned by Paulesu et al. (1993a) during a task involving phonological processing. Similarly, Zatorre et al. (1994) observed activation of the insula with the ‘First, last judgement versus passive melodies’ comparison, which consisted of deciding whether the first and last notes of a sequence were played at the same pitch. In these two studies, the functional effects of both working and short-term memories were specifically investigated. Thus, activation of the left insula in the rhythm task confirms the importance of this region in the mnesic processing of sound material of a sequential character. More specifically, this region may possess a role of regulator during the short-term memorizing auditory information.

Frontal activation

In addition to the above discussion on left inferior frontal activation in the familiarity task, both our significant findings (Table 1) and our more descriptive results (Table 2) suggest a number of additional frontal activation foci, whose interpretation is of a more speculative nature. These frontal activations consisted of both right and left superior middle frontal activations (BA 8, 9 and 32), which were found to involve the left side for comparisons concerning the pitch task and the familiarity task (‘familiarity versus timbre’), and essentially symmetrical regions in the right hemisphere in all the comparisons carried out with the timbre task.

Our interpretation would relate to the neuroanatomical work of Pandya (1995), who observed that BA 8 and 9 are among the preferred projection fields for the primary and tertiary auditory areas, which would fit with the idea of a network processing auditory musical stimuli that would include these frontal sites.

Parietal activation

Although not revealed by our prospective strategy (Table 1), uni- or bilateral activations (or non-significant trends) of the inferior parietal regions (BA 40) were observed in several descriptive comparisons (Table 2), notably involving the pitch and, less so, timbre tasks (left sided and bilateral, respectively). Sergent (1993) described bilateral parietal activation in professional musicians, only during simultaneous score reading and music listening, suggesting that these cerebral regions are involved in visual–auditory interfacing and so contribute to the coding of written musical notes into sound information, much like the auditory–visual transcoding referred to above, as an explanation of cuneus/precuneus activation during the pitch task. A different interpretation of these parietal activations relates to auditory short-term memory, as previous activation studies have suggested that BA 40 may be needed in short-term processing of visual and verbal material (Jonides et al., 1993; Paulesu et al., 1993b). As the rhythm task did not activate these areas, our findings would be consistent with a specific working-memory component, possibly involving visuo-spatial processing on pitch and timbre material.

Conclusion

Music appreciation is not exclusively subtended by a single neural-based network. As shown by our findings, there exists a degree of functional independence among the musical processing components investigated here. Thus, the pattern of activations observed in the familiarity task highlight the important role of anterior parts of the left hemisphere (notably the neighbouring inferior frontal and superior temporal areas) in semantic processing of music. Activations during the timbre task show the expected relative right-hemisphere dominance. Attention to changes in rhythm induce a specific
activation of neighbouring Broca’s and insular areas in the left hemisphere, indicating a key role for this brain region in dealing with the sequencing of auditory stimuli. Finally, we show functional evidence for cognitive strategies used by our subjects, such as the specific activation of left prefrontal–cuneus regions in the pitch task, that presumably reflect recourse to visual mental imagery.

Our findings also confirm that selective attention (the process of orienting one’s attention) to different components of a single stimulus (here the 30 sequences of notes) can radically alter the brain activation pattern, allowing isolation of specific functional responses. The results obtained using an activation paradigm which does not include an absolute reference task make it possible to show a part of what distinguishes and differentiates between the neurocognitive processes involved in complex tasks. Our results lead us to believe that this type of paradigm may be particularly helpful for the isolation of higher-order cognitive processes.

Acknowledgements
We wish to thank the staff of the MRC Cyclotron Unit for scanning facilities, and D. Luet and B. Landeau (Cyceron, INSERM Caen) for help with illustrations. R.S.J.F., R.J.S.W. comments. Nature 1993; 363: 623–5. Comment in Nature 1993; 366: 258–60.

References


Received July 2, 1996. Revised October 24, 1996. Accepted November 1, 1996