Neuropsychologia 42 (2004) 1281–1292

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Behavioral and neural correlates of perceived and imagined musical timbre

Andrea R. Halpern[a,*], Robert J. Zatorre[b], Marc Bouffard[b], Jennifer A. Johnson[b]

[a] *Psychology Department, Bucknell University, Lewisburg, PA 17837, USA*
[b] *Montreal Neurological Institute, McGill University, Montreal, Que H3A 2B4, Canada*

Received 5 June 2003; received in revised form 3 October 2003; accepted 23 December 2003

Abstract

The generality of findings implicating secondary auditory areas in auditory imagery was tested by using a timbre imagery task with fMRI. Another aim was to test whether activity in supplementary motor area (SMA) seen in prior studies might have been related to subvocalization. Participants with moderate musical background were scanned while making similarity judgments about the timbre of heard or imagined musical instrument sounds. The critical control condition was a visual imagery task. The pattern of judgments in perceived and imagined conditions was similar, suggesting that perception and imagery access similar cognitive representations of timbre. As expected, judgments of heard timbres, relative to the visual imagery control, activated primary and secondary auditory areas with some right-sided asymmetry. Timbre imagery also activated secondary auditory areas relative to the visual imagery control, although less strongly, in accord with previous data. Significant overlap was observed in these regions between perceptual and imagery conditions. Because the visual control task resulted in deactivation of auditory areas relative to a silent baseline, we interpret the timbre imagery effect as a reversal of that deactivation. Despite the lack of an obvious subvocalization component to timbre imagery, some activity in SMA was observed, suggesting that SMA may have a more general role in imagery beyond any motor component.

Keywords: fMRI; Auditory imagery; Music; Auditory cortex

1. Introduction

Imagining a favorite tune is a common experience; this imagination can be so vivid and durable that songs can get "stuck" in the head (an "earworm", in German). People have the impression that this mental experience captures a number of aspects of real music, such as pitch, tempo, and temporal extension. And indeed, using a variety of experimental methods, Halpern (see Halpern, 1992 for a review) found that people react to melodic and temporal aspects of music in a similar way whether the tune is imagined or perceived. For instance, the preferred tempo of a familiar song that is actually being played and the preferred tempo of an imagined familiar song are quite similar over a wide range of base tempos (Halpern, 1988).

Another aspect of real music is timbre, or the sound quality of different musical instruments or voices. In many styles of music, the contrast and similarity of timbre is an essential aspect of the musical composition. A famous example is Ravel's *Bolero*, in which the same theme is repeated many times, but with different instrumentation each time. Timbre perception has been studied extensively (Risset & Wessel, 1999), but much less is known about timbre imagery. Introspection suggests that timbre could be represented in auditory images, along with pitch and tempo. The reader is invited to imagine a song, perhaps *Happy Birthday,* first as played by a piano, and then by a trumpet.

An objective demonstration of auditory imagery for timbre was offered by Crowder (1989) who presented two tones and asked listeners to say whether the pitches were the same or different. If the two tones were played in different timbres (guitar and flute), subjects were slower to say that same-pitch pairs were indeed the same pitch, compared to tones played in the same timbres, showing an interference effect of timbre on pitch judgments. In a second experiment, Crowder created an imagery version of the task by presenting a sine wave, accompanied by instructions to imagine the sound as being played by a guitar, flute, or trumpet; the second tone was then sounded in one of the three timbres. As in the perception task, when the imagined timbre was different from the perceived sound, same–different judgments were slowed.

[*] Corresponding author. Tel.: +1-570-577-1295; fax: +1-570-577-7007.
E-mail address: ahalpern@bucknell.edu (A.R. Halpern).

0028-3932/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2003.12.017

This similarity in results was interpreted as evidence for the existence of imagery for timbre, following the reasoning that if timbre had not been imagined, the cross-timbre condition would not have slowed reaction time in the pitch matching. Pitt and Crowder (1992) replicated this result with different timbres from the same instrument. They also found some limits on the ability to represent timbre imagery, in that spectral but not dynamic timbre differences could produce an interference effect. They therefore concluded that spectral aspects of a musical sound but not dynamic aspects could be represented in the auditory image.

As far as we can tell, these two studies encompass the literature on timbre imagery in adults. The results give us confidence that timbre imagery is a real phenomenon, but they also leave a number of questions open. Only one technique was used, which does rely on somewhat indirect reasoning to demonstrate the existence of timbre imagery. Also, only a limited number of timbres were tested across the two studies. Therefore, one aim of the current study was to expand our knowledge of timbre imagery by using a different technique, in this case asking for similarity judgments of a range of real and imagined instrumental sounds.

Our second aim was to investigate the neural substrate of timbre imagery. Little is known about the neural regions involved in timbre perception or discrimination, and even less about the areas that might be involved in timbre imagery. Samson and coworkers tested right and left temporal lobectomy patients on timbre discrimination (Samson & Zatorre, 1994) and timbre dissimilarity rating procedures (Samson, Zatorre, & Ramsay, 2002). Over several tasks, right temporal patients evinced much greater difficulty in responding to timbre in a coherent way, compared to left temporal patients, a finding corroborating Milner's earlier observations (Milner, 1962). These findings suggest that, akin to a number of other musical perceptual processes (Zatorre, Belin, & Penhune, 2002), structures in the right temporal lobe are important in processing timbre.

A few other studies have looked at neural activation patterns during processing of timbre, which again implicate the right hemisphere. Jones, Longe, & Vaz Pato (1998) played listeners a stream of tones that changed in pitch or timbre. Auditory evoked potentials were stronger over the right than left hemisphere for timbre as well as pitch. Platel et al. (1997) presented various kinds of musical or tonal sequences to listeners, who were scanned with PET while performing assorted tasks. Their timbre task involved presentation of a sequence whose tones were either all in the same timbre, or where the notes alternated between two similar timbres. The task was to indicate whether the sequence contained one or two timbres. Comparing cerebral blood flow in this task with a pitch/rhythm task revealed significant activation in several areas of the right frontal lobe.

Our particular interest in neural substrates of timbre imagery derives from our previous studies on auditory imagery for melodies. When we began this series of studies, several researchers had been investigating the locus of activity during visual imagery tasks, finding that a number of visual areas seemed to be active even when observers were only imagining visual stimuli (for review see Farah, 1988; Kosslyn, Ganis & Thompson, 2001). These observations in the visual domain lent credence to the proposition that imagery and perception shared representational characteristics, so that imagery could in a real sense be considered a quasi-perceptual phenomenon. However, whether this same framework could be applied to auditory imagery remained unanswered.

In our first PET study to investigate auditory imagery (Zatorre et al., 1996), we used a mental pitch comparison task. In the perception condition, subjects heard a familiar sung tune at the same time two words from separate parts of the tune were presented on the screen. The task was to indicate, without singing or humming, whether the pitch of the second lyric was higher or lower than the pitch of the first lyric. In the imagery condition, the task was presented without the song actually being played. The perception and imagery conditions (minus a visual baseline control) revealed remarkably similar patterns of cerebral blood flow, including auditory association areas bilaterally, as well as several frontal areas. The activity in auditory association areas was notable because the relevant tasks (imagery and baseline) were both performed in silence. One additional area of increased activity was the supplementary motor area (SMA), which was especially prominent in the imagery task. Although participants were not allowed to make overt vocalizations, we speculated that subjects were subvocally rehearsing the lyrics to the melodies, or the melodies themselves, which might account both for SMA activation and activation in the left temporal lobe (which is associated with verbal tasks).

To help clarify this issue, our next PET study (Halpern & Zatorre, 1999) used an auditory imagery task that did not require reference to any words. Familiar nonverbal tunes, such as movie themes and motives from classical music, comprised the stimuli. In the task of main interest, the opening few notes of the tune were played and subjects were asked to imagine the continuation of the phrase, pressing a button when this was accomplished. A positive correlation of latencies to press the button with the length of the actual melody assured us that instructions were being carried out. Several baseline conditions controlled for auditory and motor aspects of the tasks.

Our results were as expected, in that we once again saw activation of an auditory association area during the imagery condition, but this time the activation was right-lateralized, in line with many prior studies (Zatorre et al., 2002). Also once again we saw activation of the SMA, both in the main condition, and in another condition where people were asked to mentally rehearse just-presented unfamiliar note sequences. As neither the main task nor the mental rehearsal task involved any kind of verbal referent, we concluded the SMA activation in the Zatorre et al. (1996) study was probably not due to the mere presence of words, but was more likely linked to subvocal processes.

Both our previous studies, although they used different paradigms and stimuli, were similar in that the tasks required access to two or more pitches. Pitch is vocaliz*able*, even if not vocal*ized*; therefore, imagery for pitch might call upon subvocalization, which in turn may be related to motor imagery mechanisms that are known to involve the SMA (Lotze et al., 1999; Stephan et al., 1995). Alternatively, we cannot rule out the possibility that SMA is generally involved in auditory imagery, whatever the attribute being represented. By the same token, we cannot say for sure whether other structures identified from our previous tasks, such as those in the right frontal and temporal areas, are generally involved in auditory imagery for music, or in pitch tasks specifically.

Accordingly, we turned to timbre as a way of helping to answer these questions. We presented evidence above that timbre can be represented in an auditory image. But unlike pitch, people cannot easily vocalize, or even subvocalize the sounds of various instruments. Crowder (1989) offered timbre tasks as a stronger test case of the existence of musical imagery than the mental pitch comparison task originally used by Halpern (1988) and concluded that indeed "... an appeal to sensory rather than motor imagery is justified" (p. 478). Similarly, timbre does not require processing of a sequence of events. So on two dimensions, tasks requiring mental processing of timbre should be quite different than those requiring mental processing of pitch. Thus commonalities in neural processing across these tasks should point to aspects of auditory imagery that occur regardless of motoric or sequential demands.

The task we chose was a similarity comparison of perceived and imagined musical instrument sounds, a technique commonly used to investigate perceptual representation of sounded timbre (Grey, 1978; McAdams et al., 1995). In the study most relevant to ours, Kendall, Carterette, & Hajda (1999) asked listeners to rate synthetic and natural sounded timbres for similarity. Multidimensional scaling analysis was used to characterize the ratings over all the listeners. Across a variety of training levels and conditions of playback, the multidimensional scaling solution for those ratings was coherent. Two dimensions corresponding to *nasality* and *brilliance* captured 0.95 of the variance in the MDS solution. This suggested to us that a simple but effective timbre task would be to ask people to rate the similarity of sounded instruments as well as the similarity of imagined instruments. We predicted that the pattern of similarity judgments would be similar in the two conditions, validating our comparison of the neural activation in both conditions.

Our working hypothesis, based on literature reviewed above, was that generation of timbre should involve secondary auditory areas, on the right more than the left, if auditory areas are involved in evoking of musical imagery in general. Inferior frontal areas were also expected to be active, as they were in our previous studies, as all imagery tasks have some memory component associated with them. We did not expect to find activation of the SMA, as the timbre task should involve neither subvocalization nor other sequencing of events such as the SMA might be expected to mediate.

Our previous activation studies used PET, which has the virtue of allowing a silent testing environment. Another purpose of the current study was to see if auditory imagery effects could be captured during fMRI scanning, which has many advantages to PET as far as invasiveness and spatial and temporal resolution, but which is quite noisy. To overcome the latter problem we use a sparse-sampling technique (Belin, Zatorre, Hoge, Evans, & Pike, 1999).

Yoo, Lee, & Choi (2001) recently conducted a study somewhat relevant to the goals of this study. They asked subjects to learn a single computer-generated note, and to evoke that one imagined sound when prompted to do so by a tactile cue. In an event-related fMRI paradigm, significant activation, peaking at 5 s post-cue, was found in auditory cortical areas. This suggests that fMRI may be sensitive to the activation caused by at least simple internally generated sounds. However, this study had no control conditions or task validation, and did not isolate timbre, leaving our major questions of interest still open.

In summary, participants in the present study were asked to compare pairs of sounded instruments and rate them for similarity. They were asked to do the same for imagined sounds of instruments. Visual and auditory input, as well as visual imagery, were controlled by means of several secondary tasks, and fMRI was used to examine brain activation associated with timbre imagery.

2. Methods

2.1. Participants

Volunteers were 10 healthy right-handed young (mean age 24.3 years) adults (half of each sex) all of whom had had a minimum of 5 years of formal musical training. All of them were screened to ensure that they were familiar with the instrumental timbres used. We also administered an auditory imagery questionnaire to make sure all participants were able to generate auditory images. Subjects read 10 vignettes involving imagined auditory stimuli (voices, music, environmental sounds), and rated the vividness of the resulting imagery experience on a scale of one (no image) to seven (very vivid). Subjects assigned an average of 5.09, suggesting that auditory imagery is available in the mental representation.

2.2. Stimuli

Eight instrument sound files of 1.5 s duration were created from digitized samples of real musical instruments, excerpted from the McGill University Master Samples (http://www.music.mcgill.ca/resources/mums/html/mums.html). The instruments were: bassoon, clarinet, flute, French horn, oboe, tenor saxophone, trumpet, and violin. These instruments were determined by pilot work to be highly recognizable and to contain some similar and some different-sounding pairs. In addition, each had been used in the study by Kendall et al. (1999) that elicited highly stable similarity judgments. Each instrument was sounded as either a B flat or F in the instrument's middle range. The choice of note was determined by which seemed to most distinctly capture the characteristic sound of the instrument.

For the noise condition, 1.5 s matching "noise envelopes" were created by modulating a white noise stimulus to take on the amplitude envelope characteristics of its matching instrument stimulus. Most of these sounded like hissing, with some oscillation in samples matched for instruments with vibrato, such as the violin. All auditory stimuli were equalized for energy content. Examples of sounds may be heard at: http://www.zlab.mcgill.ca.

For the visual imagery control, words were chosen that matched each instrument name in length and number of syllables, and that matched stress pattern wherever possible. The referent for each word had a recognizable shape but no obvious auditory association. For instance, "yo-yo" was the match for "oboe", and "balloon" the match for "bassoon". Several words referred to objects with round shapes, such as the ones just mentioned, and several referred to objects with long shapes, such as "stick" (match for "flute") and "string bean" (match for "French horn"); other shapes were unique. We expected same-shape words to elicit higher similarity ratings than different-shape words.

2.3. Conditions

Each subject participated in five conditions: silence, noise, visual imagery control, perception, and timbre imagery. Every condition involved presentation of a stimulus pair during the relatively silent inter-acquisition interval (see further). The first stimulus of the pair was presented 1 s after the end of the previous fMRI volume acquisition; this was followed by 2 s of silence, the second stimulus then followed, and a response was then made. Scanning commenced 2 s after the offset of the second stimulus. Visual stimuli were presented on a screen placed at the rear of the scanner bore, and viewed via a mirror. Auditory stimuli were presented binaurally via Koss ESP900 electrostatic headphones calibrated at 80 dB SPL(A) using a continuous noise stimulus. All responses involved a button press with the right hand from a five-button scanner-compatible response pad.

In the silence condition, both Stimulus 1 and 2 were a single row of X's presented in the center of the screen. Subjects pressed a button at random for their response. This condition controlled for visual input and motor response. In the noise condition both stimuli consisted of a noise burst presented simultaneously with the X's. Subjects pressed a button at random for their response. This condition controlled for auditory input, in addition to aspects controlled by the silent baseline.

In the visual imagery control condition, the two stimuli were taken from the list of words matched to the instrument names. Subjects were instructed to imagine the shape represented by each word, and to rate their similarity on a scale of one (high similarity) to five (low similarity). We instituted a visual imagery task as the main control task for two reasons. First, imagining the timbres of real instruments would likely also evoke a visual image of the associated instrument, as indeed was reported by our pilot subjects. The visual control task therefore allows us to control for visual imagery and isolate the task component of interest (i.e. auditory imagery). Second, the visual imagery control task required participants to keep two items in working memory, make a comparison, make a decision via use of a rating scale, and select an appropriate response from among the five-button response pad. Thus, the control task should allow us to control for a number of non-specific and procedural aspects of our main task, leaving the central cognitive components for us to examine. In the perception condition, the two stimuli consisted of a sounded musical instrument, accompanied by the name of the instrument on the screen. Subjects were instructed to listen to the sounds and to rate their similarity on the one to five scale described above. The condition of greatest interest was the timbre imagery condition, in which stimuli were the names of musical instruments, not accompanied by a sound. Subjects were asked to imagine the sounds of the instruments named, and to rate their similarity on the one to five scale.

Before entering the scanner, all participants were familiarized with the tasks and given sample trials. They listened to the instrument sounds until they could easily name them. Each of two runs in the session consisted of 128 trials. For the timbre perception and timbre imagery conditions, all possible pairs of the eight stimuli were presented (total of 28 trials) with the order of each instrument within a pair chosen at random. For the other conditions, 24 trials were presented. Within a run, conditions were presented in interleaved blocks of six or seven trials, in a different order for each subject. This interleaving was reversed for the second run in a session.

2.4. fMRI scanning protocol and analysis

Scanning was performed on a 1.5 T Siemens Vision imager. First, a high resolution anatomical scan was obtained for each subject (voxel size: 1 mm × 1 mm × 1 mm). Second, two series of 128 echo-planar images of blood-oxygenation-level-dependent (BOLD) signal, an indirect index of neuronal activity, were acquired, each in 20 slices using a 64 × 64 × 20 matrix (voxel size 5 mm × 5 mm × 5 mm) aligned with Sylvian fissure and covering the whole brain. Clustered volume acquisitions were used, such that each volume was acquired in 2.04 s, and the repetition time (TR) was set at 10 s. The long inter-acquisition time ensured low signal contamination by noise artifacts related to image acquisition (Belin et al., 1999). BOLD signal images were smoothed using a 6 mm gaussian kernel, corrected for motion, and transformed into standard stereotaxic space using in-house software (Collins, Neelin, Peters, & Evans, 1994). For region of interest analyses, BOLD signal values were extracted after smoothing with a 14 mm filter in order to capture a wider sample of the activated cluster.

The statistical analysis of the fMRI data was based on a linear model with correlated errors and implemented in a suite of Matlab programs (Worsley et al., 2001). In a first step, 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. The two runs for each subject were then combined using the effects and standard deviations from the previous analysis. In a final step, these results were combined yielding the group statistical maps (all 10 subjects) for each contrast. The threshold for significance was set at $t = 4.97$ for a whole-brain search, and at 4.37 for a directed search within predicted areas of the superior temporal gyrus (search volume: 50 cm^3), based on 375 degrees of freedom, an 8 mm^3 voxel, smoothness of 6 mm, and a corrected significance level of $P < 0.05$ (Worsley et al., 1996).

3. Results

3.1. *Behavioral findings*

In order to verify that the participants were able to judge the timbre similarity in a coherent way, we subjected the similarity ratings to a multidimensional scaling analysis. The data from the perception condition returned a good fit in two dimensions, with the first and second dimensions accounting for 64 and 35% of the variance, respectively. As Fig. 1 shows, the solution shows a circular distribution of the instruments. The x-axis seems to correspond roughly to a "brilliance" dimension and the y-axis to "nasality", similar to dimensions found by Kendall et al. (1999). Thus, the oboe is nasal and brilliant, the tenor sax nasal and not brilliant, the violin not nasal but somewhat brilliant, and the clarinet neither nasal nor brilliant. The French horn was rated more highly on the brilliance dimension than intuition would suggest but otherwise most instruments seem to be placed in a logical ordering.

The analysis of the imagery ratings also shows a roughly circular pattern (Fig. 1), with most instruments being positioned in a similar relative position to that seen in the perception solution (note, the axis values in the two solutions are internal to the MDS analysis and are not directly comparable). The same two dimensions seem to describe the data reasonably well (again accounting for 64 and 35% of the variance), with brilliance on the x-axis and nasality on the y-axis. Among the differences between the solutions in imagery and perception, French horn in the imagery solution is rated as less brilliant relative to other instruments than it was in perception (and more nasal), whereas the tenor sax was remembered as being more brilliant relative to other instruments than when it was heard.

To get a more quantitative index of the similarities in the solutions, we correlated the interpair model distances with one another. These were significantly positively related, $r(26) = 0.63$, $P < 0.001$. A more direct way to consider how similarly people were processing the timbres in imagery and perception is to look at the average similarity ratings for each instrument pair in the imagery and perception conditions. The mean rating over all instrument pairs in each condition were very similar: 3.29 (0.77) in perception and 3.23 (0.74) in imagery. The correlation between ratings in each condition was quite high, $r(26) = 0.84$, $P < 0.001$, see Fig. 2. Only one pair of the 28 elicited a notable difference in the two conditions: flute–violin was rated as more similar in imagery than perception. This may be due to the fact that flute and violin were the only sounds that had vibrato in our samples, which may have made that one comparison a bit unusual. Removing that one pair results in $r(25) = 0.90$, $P < 0.001$.

Because some instruments were sounded as an F and some as a B flat, we checked to make sure that the pitch equality or difference was not driving the pair ratings. Fifteen pairs had different pitches and 13 pairs had the same pitch. Average similarity ratings for the different-pitch pairs were 3.36 (0.77) for perception and 3.20 (0.81) for imagery; the same figures for same-pitch pairs were 3.20 (0.81) and 3.26 (0.67). These means are all nearly identical. In addition, the perception–imagery correlations were highly significant for both types of trials: $r(13) = 0.82$ for different-pitch trials and $r(11) = 0.90$ for same-pitch trials.

We also checked the similarity ratings in the visual imagery control task. The three pairings of the three round objects (yo-yo, balloon, amoeba) and the one pairing of the two long objects (string bean and stick) elicited a mean similarity rating of 1.86 (highly similar); the average similarity rating of the remaining pairs was 4.06 (highly dissimilar). This suggests that subjects understood the control task and made appropriate shape similarity ratings.

3.2. *Functional activation findings*

Comparing the active conditions (perception, timbre imagery, and visual imagery control) to the silent baseline condition yielded widespread increases in BOLD signal throughout auditory and visual cortices, as well as bilateral inferior and dorsolateral frontal areas likely associated with active task performance. This pattern reflects the nonspecific nature of the comparisons. Therefore, in order to isolate the functional activity of interest, we compared the perception and timbre imagery conditions to the visual imagery control condition, since the latter contains similar visual input, as well as cognitive and motor response components as the perception and timbre imagery conditions. In these comparisons, many of the widespread changes dis-

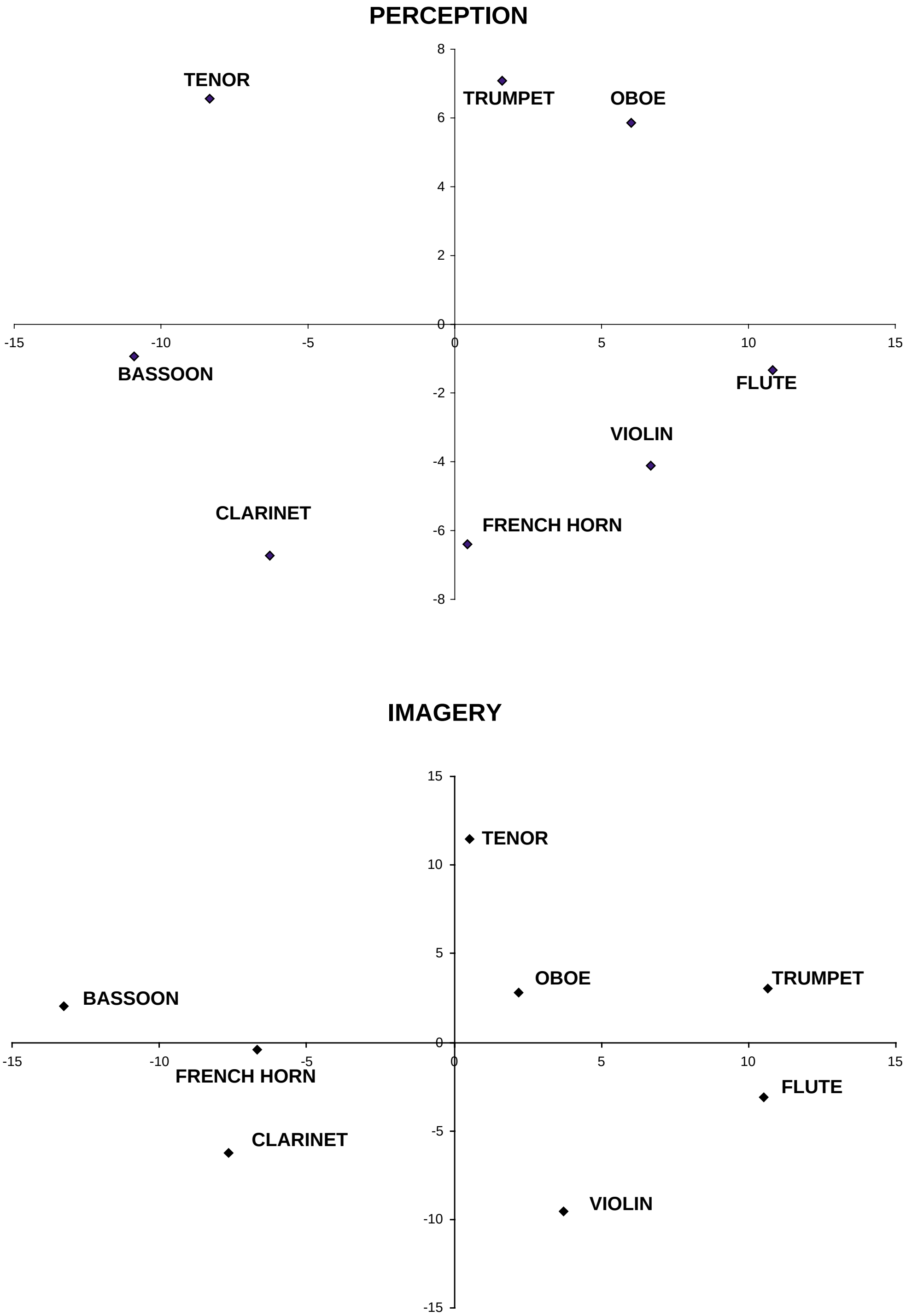

Fig. 1. Multidimensional scaling solution for similarity ratings of perceived (top) and imagined (bottom) timbre pairs.

appear, suggesting that the visual imagery control task is an adequate baseline for these purposes.

In the perception–visual imagery control comparison, highly significant BOLD signal increases were seen within the superior temporal gyrus, largely within primary and adjacent auditory cortices bilaterally, comprising portions of Heschl's gyrus and the planum temporale (PT). In addition, a retrosplenial area was also active. There was an asymmetry in the response of the superior temporal areas, with the greatest hemodynamic change observed within a more anterior location on the right side which was not matched by a corresponding peak on the left side (Fig. 3). The reverse subtraction (visual imagery control − perception) did not reveal any significant BOLD signal changes.

Comparing the timbre imagery to the visual imagery control condition also revealed increased BOLD signal within

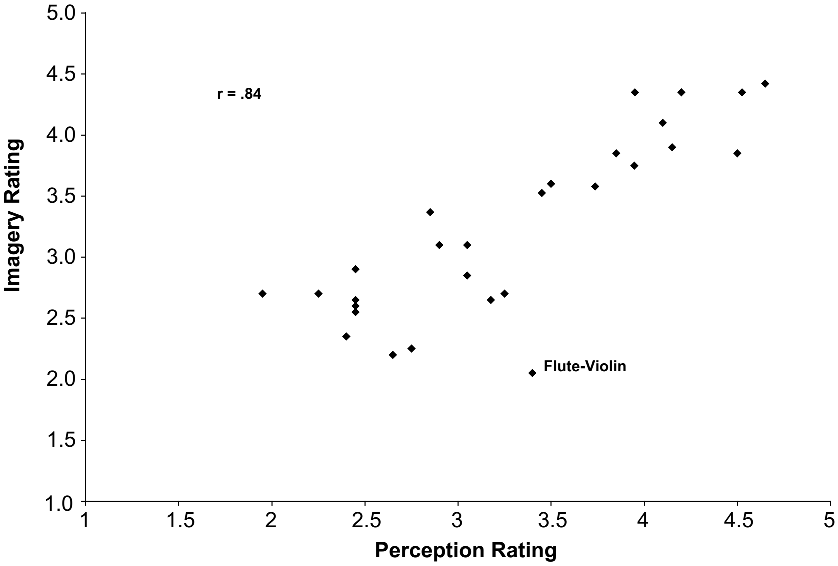

Fig. 2. Scatterplot of mean similarity ratings for each instrument pair in perceived and imagined conditions.

the superior temporal gyrus but it was much weaker than that seen in the perception condition, reaching significance using the predicted threshold values (*t*-values between 4.4 and 4.9). In addition, the location of the changes was largely posterior to the areas seen in the perception condition, in areas close to the posterior border of the PT as determined from anatomical probability maps of this structure (Westbury, Zatorre, & Evans, 1999). For this analysis, data were taken from the 14 mm filtered set in order to obtain a larger sample of the voxels activated in the region of interest. The distribution of this activity was weakly asymmetrical, with 7 of 10 subjects showing greater change on the right than on the left at the peak voxel location for each hemisphere (left: −42, −36, 16, $t = 4.27$; right: 48, −46, 20, $t = 6.10$). Extracting the BOLD signal change (relative to visual imagery control) and directly comparing the values at peak locations in left and right yielded a significant difference favoring the right side ($t = 1.97$, $P = 0.04$ one-tailed). In addition to these temporo-parietal areas, the contrast of timbre imagery versus visual imagery control also showed increased BOLD signal in the retrosplenial region, in a similar location to that seen in the perception condition. The reverse subtraction (visual imagery control − timbre imagery) did not reveal any significant BOLD signal changes, indicating once again that the visual perceptual and visual imagery components of the two tasks were well matched.

In order to determine whether the distribution of signal change in perception and timbre imagery conditions overlapped, we conducted a conjunction analysis, searching for voxels in common between the two volumes whose activity exceeded the threshold in each. This analysis yielded three sites of overlap: one within the posterior aspect of the left PT, the other just behind the posterior border of the right PT, and the third in the retrosplenial region. Thus, the conjunction analysis confirms the findings from the individual contrasts of significant overlap in posterior portions of the superior temporal gyrus across conditions, as well as in the retrosplenial cortex.

The presence of activation in the right auditory cortex during both timbre imagery and perception was further examined by extracting the BOLD signal values obtained at the peak voxel location for the timbre imagery–visual imagery control subtraction on the right, and comparing the value at this location across all five conditions. The result (Fig. 4) shows the expected increase in this auditory cortical area in response to real sound stimulation, since both the noise and perception conditions resulted in higher signal as compared to silence. However, the timbre imagery condition does not differ significantly from the silence condition. BOLD signal was significantly higher for timbre imagery compared to the visual imagery control; but this appears to be due to a relative *decrease* in hemodynamic response in the auditory areas during the visual imagery control condition relative to silence. A very similar pattern was seen for the peak voxel location in the left auditory cortex. Indeed, whole-brain voxel-wise analysis of the visual imagery control task versus silence does demonstrate significant negative values bilaterally in the PT (right: 58, −26, 20, $t = 5.20$; left: −60, −30, 14; $t = 4.63$).

Finally, we examined the activation in the SMA. In the analysis contrasting the timbre imagery condition to the visual imagery control condition, we observed sub-threshold signal change within the SMA; this was not observed in the perception condition. Although this area was not predicted to be engaged during the imagery task, we wanted to ascertain whether it might have shown increased activity. To test this we compared the imagery and perception conditions directly to one another, which yielded a subthreshold area

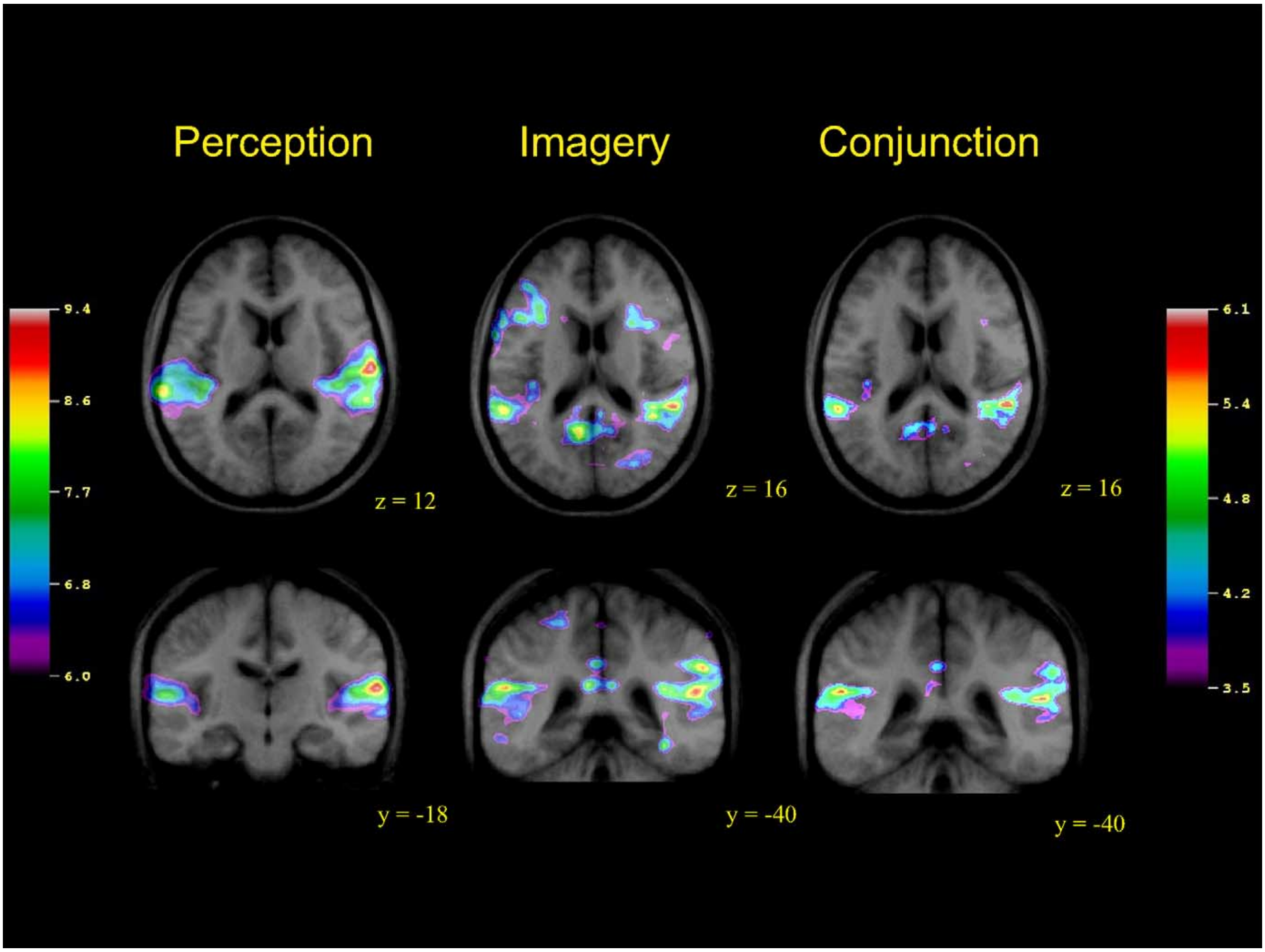

Fig. 3. Statistical parametric maps of fMRI signal changes (t-values, in color scale) superimposed upon group-average anatomical MR image at selected locations. In all panels, the top image illustrates a horizontal section taken through the temporal lobes (position indicated by stereotaxic z-coordinate), and the bottom panel illustrates a coronal section taken close to the peak location of signal change (position indicated by stereotaxic y-coordinate). Left panel: data from the timbre perception vs. visual imagery control condition (data referred to left color bar). Significant BOLD signal changes are evident in primary and adjacent auditory cortices, with an asymmetric response in the right planum temporale. Middle panel: data from the timbre imagery vs. visual imagery control condition (data referred to right color bar). BOLD signal changes are evident within posterior portions of secondary auditory cortices. Right panel: data from the conjunction analysis of perception and timbre imagery (data referred to right color bar). Significant overlap in posterior auditory areas is seen.

of activation within the SMA (−6, −2, 60; $t = 4.55$). No other differences emerged in this direction between these two conditions. As expected, however, much more hemodynamic change was evident in the auditory cortices in the perception condition as compared to the imagery condition.

4. Discussion

The multidimensional scaling results and the correlational analysis confirm that people can compare imagined timbres in a similar way as they do perceived timbres (see Figs. 1 and 2). The principal neuroimaging findings of this study confirmed predictions that activity in the auditory association cortices accompanies timbre perception and imagery. The activity during timbre imagery was seen relative to a visual imagery control task, but not relative to a silent baseline. SMA activity was present in the imagery condition, contrary to predictions, although at a subthreshold level.

4.1. Behavioral data

The MDS results for perceived timbres are quite coherent and returned a good fit. The dimensions are similar to those found by Kendall et al. (1999) even though they used a larger set of instruments and the precise nature of the sound samples differed in the two studies. Our two-dimensional solution was roughly circular in shape, showing a wide distribution of our instruments over the space defined by the two dimensions. Our solution seems to be captured by label-

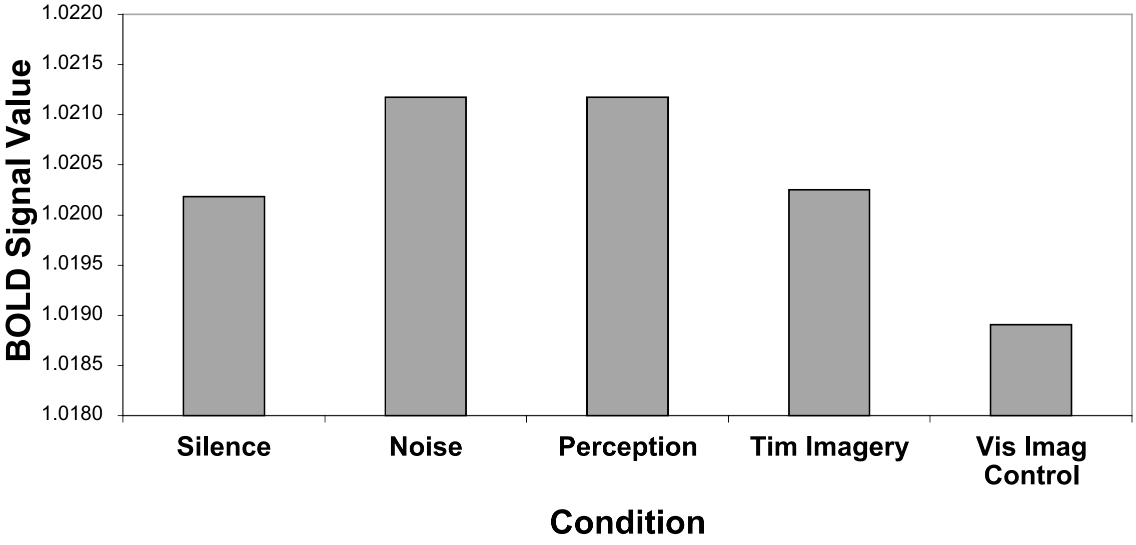

Fig. 4. BOLD signal in each condition at the most active voxel in the timbre imagery–visual imagery control comparison (48, −46, 20; right secondary auditory cortex). Standard error = approximately 0.014 for all conditions.

ing the *x*-axis as the "brilliance" dimension and the *y*-axis as the "nasality" dimension. This result provides validation that subjects were performing the task as intended, since data were acquired on line.

The imagery MDS data represent a completely new finding, and thus are interesting in their own right. The solution once again is coherent and allows interpretation on the same two dimensions as the perception task. The overall solution is quite similar in the two tasks; for example, in Fig. 1, the flute, violin, clarinet, bassoon are placed in similar relations to the axes. The positive correlation of the interpair distances in the solutions suggests that subjects are accessing similar mental representations of timbre space in perceptual and imagery tasks.

A different way of viewing these is data is to consider the correlation between the average similarity ratings for each pair in perception and imagery. As shown clearly in Fig. 2, the similarity ratings in one task strongly predicted the similarity ratings in the other task. This occurred regardless of whether the pitch of the instruments in the pair was the same or different. All these behavioral findings support our contention that the timbre rating task was meaningful to subjects and captured a critical aspect of processing timbre. Thus, we can be confident that the neuroimaging results reflect brain activity during the specific mental processes of interest in the current study.

4.2. Neuroimaging results

In order to interpret the findings from the principal tasks, it is necessary to ensure that the active control task was appropriate. We chose a visual imagery task as the control because of the likelihood that imagining the sound of a musical instrument would also entail visual imagery of the instrument. The visual imagery control task required a shape comparison for objects whose names were carefully matched to the names of the musical instruments, and a rating on a five-point scale. Thus, the control task was similar to both perception and imagery in visual input (i.e. word length), cognitive demands, and motor output.

That the visual imagery control task served its purpose is supported by the following lines of evidence. First, the behavioral data suggest that subjects were using visual imagery to perform the task in that they rated similarly shaped objects as more similar than dissimilarly shaped objects. Second, when comparing the two timbre tasks to the visual imagery control task there was no evidence of the widespread frontal cortical activity which was evident when any of the active tasks were compared to the no-task silent baseline. This is evidence that the visual imagery control task was similar to the two timbre tasks in its recruitment of these brain regions, which are likely related to working memory, decision processes, and other cognitive components. Third, subtraction of the perception condition from the visual imagery control condition did not reveal any differences, indicating that the two tasks are well matched in their visual perception and cognitive demands. Specifically, this lack of difference implies that the perception task did involve some visual imagery, as we had anticipated. However, the experiment was not designed to address whether visual imagery evokes visual cortical activation since there is no appropriate control condition to answer this point.

Having established the validity of the control task, we turn to the principal findings that secondary auditory cortical areas show significant hemodynamic increases during both timbre perception and timbre imagery, as compared to the visual imagery control condition. This result is therefore consistent with predictions derived from our previous work (Halpern & Zatorre, 1999; Zatorre & Halpern, 1993; Zatorre, Halpern, Perry, Meyer, & Evans, 1996) that neural processing within these areas of auditory cortex underlie perception and imagery. This result generalizes previous findings to a completely different type of task and probed musical dimension. In particular the task used in the present study required mental generation of only a single tone. Our prior studies used familiar melodies, which are extended in time, involve rhythm, and have semantic associations that are strong enough to override their sound qualities in an open-ended similarity judgment (Halpern, 1984) The relative similarity of the cortical sites across studies suggests that these areas mediate many aspects of musical imagery.

As expected, the magnitude of the hemodynamic response in auditory cortical areas was much weaker and less extensive during the timbre imagery as compared to the perception task. This is similar to our earlier findings (Zatorre et al., 1996), and comparable to effects seen in visual imagery experiments (Le Bihan et al., 1993). Given that the timbre imagery condition involved generating an internal representation in the absence of any physical stimulation (i.e. in silence), any activation in sensory regions is notable. By the same token, weaker effects are consistent with the phenomenological and empirical differences between perception and imagery. In the present study, subjects rated their auditory images as being less vivid than hearing real sounds. Performance on imagery tasks is often slower and less accurate than on equivalent perception tasks. Crowder (1989) found that responses in the imagined version of his experiment were about 500 ms slower than in the perceived version; Zatorre and Halpern (1993) saw a reduction from 85% correct in their perception task to 73% correct in the imagery counterpart to that task.

A point of interest concerns the fact that activation in timbre imagery was observed in posterior temporal cortex, and not in primary auditory cortical regions. This was also the case in our previous studies (y-values ranged from −42 to −48 in the current study; the closest comparison conditions in our other studies showed activation in similar posterior areas with y-values from −37 to −42), but could have reflected the nature of those tasks which involved complex relational information in melodies. In contrast, the simple stimuli of the current task might have provided a good opportunity to observe any primary activation. The consistent extra-primary location of the activated sites in auditory imagery contrasts with findings in visual imagery (Kosslyn et al., 1999) indicating recruitment of V1 under certain conditions. Yoo et al. (2001) did report primary auditory cortex activity in an fMRI paradigm, but inspection of the coordinates from this study indicate that the site of activity was likely within the PT based on anatomical probability maps of Heschl's gyrus and the PT (Penhune, Zatorre, MacDonald, & Evans, 1996; Rademacher et al., 2001; Westbury et al., 1999), and therefore outside of primary cortical regions. This body of evidence therefore suggests that on the one hand, neural substrates of auditory and visual imagery are similar to the extent that both depend on secondary cortical areas reliably. On the other hand, it remains to be seen whether the difference in contribution of primary sensory cortex in the two modalities reflects fundamental modality differences or is related to task parameters.

Table 1
Stereotaxic coordinates of peak voxel locations in each of three analyses

x	y	z	t	Area
Perception minus visual control				
64	−18	12	8.40	R PT
64	−8	6	7.49	R PT
44	−28	12	6.34	R PT
66	−30	6	6.34	R PT
60	−38	10	6.89	R posterior STG
52	−20	10	6.88	R PT/HG border
66	−18	−4	4.99	R STS
−38	−30	14	6.30	L HG
−54	−22	8	6.24	L PT/HG border
−48	−26	10	5.78	L PT/HG border
−66	−32	12	7.26	L PT
−56	−42	20	6.14	L PT
−42	−36	14	6.09	L PT
−60	−18	8	6.07	L PT
−4	−50	28	5.35	Retrosplenial
−4	−62	26	5.06	Retrosplenial
Imagery minus visual control				
54	−42	16	4.66	R posterior STG
40	−48	16	4.40	R posterior STG
40	−42	16	4.11	R posterior STG
38	−62	24	4.93	R TPO
−56	−44	22	4.97	L PT
−6	−2	60	4.08	SMA
−6	−56	20	5.88	Retrosplenial
Conjunction analysis				
54	−42	16	4.66	R posterior STG
−56	−44	22	4.98	L PT
−4	−62	24	5.02	Retrosplenial

Note: coordinate values correspond to locations in standardized Talairach/MNI space. STG: superior temporal gyrus; PT: planum temporale; HG: Heschl's gyrus; STS: superior temporal sulcus: TPO: temporo-parieto-occipital; SMA: supplementary motor area.

The response within auditory regions was largely bilateral in both the perception and timbre imagery conditions, but, as predicted, asymmetry favoring the right was observed. In the perception task, the activity within primary cortex was essentially bilateral, but an additional focus of hemodynamic response on the right side only was present lateral to Heschl's gyrus, on the anterior border of the PT (Table 1, Fig. 3). In the imagery task we also observed a greater response from the right superior temporal region. This effect is consistent with previous findings implicating this region in melodic imagery (Halpern & Zatorre, 1999). We interpret these findings as reflecting a relative advantage of right temporal regions in processing many aspects of both perceived and imagined musical information.

One difference between the current results and our previous work is the absence here of frontal activation during the timbre imagery task when compared to the perception task.

Earlier we made the point that the visual imagery control task was appropriate partly because comparisons between it and the other two active tasks encompassed many of the same cognitive operations. The absence of frontal activations in the current timbre imagery task is most likely due to similar reasons. Compared to the perception task, the timbre imagery task did not involve heavy loads on working memory, as images of only two tones needed to be generated and compared. This contrasts with our earlier paradigms in which long sequences of tones needed to be generated, from either a verbal (Zatorre et al., 1996) or tonal (Halpern & Zatorre, 1999) cue. In addition, the semantic retrieval demand in our previous tasks was higher than the current task, in that retrieval of a familiar song (probably including many associations) was required.

One point that needs to be addressed is the fact that the timbre imagery condition displayed significant secondary auditory cortex activation compared to the visual imagery control task, but not compared to the silent baseline (Fig. 4). The figure also shows that the visual imagery control task elicited less BOLD signal than the silent baseline in this area. In other words, the visual imagery control task reflects a deactivation of this area compared to silence, and the timbre imagery task reflects a reversal of this deactivation. Considerable evidence exists that primary sensory and association cortex is deactivated during presentation of information in other modalities. For instance, Laurienti et al. (2002) and Lewis et al. (2000) found that presentation of visual stimuli in some conditions produced BOLD signal decreases in auditory cortical regions. The same phenomenon explains why we observed a signal decrease in our visual imagery control task in auditory areas.

We had initially predicted that auditory cortex activation in the timbre imagery condition would exceed that of the silent baseline. However, the effect we observed is likely due both to the deactivation referred to above, and also to the general problem of interpreting activity relative to a silent, no-task baseline. As Stark and Squire (2001) point out, silent baselines may involve no external input, but at the same time, they are unconstrained and in fact can elicit considerable cognitive activity. Therefore, we consider the comparison between timbre imagery and the active control task of visual imagery to be the most appropriate one. We note that in our case, the “visual” part of the visual imagery control task included both word presentation, as well as a strong visual imagery component which might have been the primary cause of the deactivation. Teasing these apart is a subject for future research.

Another finding which may be related to deactivations occurring in baseline conditions is the retrosplenial change reported in the tables for both timbre imagery and perception conditions, relative to the visual imagery control. Gusnard and Raichle (2001) provide a review of many studies showing decreases in hemodynamic signal in retrosplenial areas comparing a variety of active judgment tasks to passive “rest” conditions. In the present study, this trend was also seen, as signal decreases were obtained in this region in both visual imagery control (6, −48, 34; $t = -5.07$) and timbre imagery tasks (8,−48, 34, $t = -4.18$) as compared to silence. It is possible therefore that the apparent activity seen in this area in the contrast of timbre imagery versus visual imagery control reflects greater suppression of this region during the visual task, rather than a true recruitment of this area for timbre imagery.

Our final observation concerns our prediction about SMA activation. We had speculated that by removing a subvocalization and sequencing component to the generated image, we would eliminate the contribution of SMA to this task. In fact, we continued to find some contribution of this area in the timbre imagery task, albeit at a subthreshold level. Two possible explanations for this occur to us: First, although subvocalizing the timbre of an instrument is difficult, the timbre was accompanied by pitch, which itself is easily vocalizable. Second, SMA may be involved in some more general aspect of auditory imagery, such as image generation or preparation, regardless of any potential subvocal contribution to the image. Resolution of this question requires tasks that systematically control potential subvocalizable aspects of the imagined stimuli. We are planning to explore auditory imagery in both musical and nonmusical domains to allow us to distinguish what imagery processes might be specific to music or generalizable to any imagined sound.

Acknowledgements

We gratefully acknowledge financial support for this project from the McDonnell-Pew Program in Cognitive Neuroscience, and the Canadian Institutes of Health Research. We also thank Christopher Plaue for assistance in setting up the behavioral study, and the staff of the McConnell Brain Imaging Center for their technical assistance and support.

References

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *NeuroImage*, *10*, 417–429.

Collins, D., Neelin, P., Peters, T., & Evans, A. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*, 192–205.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 472–478.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, *95*, 307–317.

Grey, J. M. (1978). Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America*, *64*, 467–472.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*, 685–694.

Halpern, A. R. (1984). The organization of memory for familiar songs. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 496–512.

Halpern, A. R. (1988). Mental scanning in auditory imagery for tunes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 434–443.

Halpern, A.R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697–704.

Jones, S. J., Longe, O., & Vaz Pato, M. (1998). Auditory evoked potentials to abrupt pitch and timbre change of complex tones: electrophysiological evidence of 'streaming'. *Electroencephalography Clinical Neurophysiology*, *108*, 131–142.

Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1999). Natural vs. synthetic instrument tones. *Music Perception*, *16*, 327–364.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*, 635–642.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., Ganis, G., Sukel, K. E., & Alpert, N. M. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, *284*, 167–170.

Laurienti, P., Burdette, J., Wallace, M., Yen, Y. F., Field, A., & Stein, B. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, *14*, 420–429.

Le Bihan, D., Turner, R., Zeffiro, T. A., Cuénod, C. A., Jezzard, P., & Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the USA*, *90*, 11802–11805.

Lewis, J., Beauchamp, M., & DeYoe, E. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, *10*, 873–888.

Lotze, M., Montoya, P., Erb, M., Hulsmann, E., Flor, H., Klose, U., Birbaumer, N., & Grodd, W. (1999). Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an fMRI study. *Journal of Cognitive Neuroscience*, *11*, 491–501.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychololgical Research*, *58*, 177–192.

Milner, B.A. (1962). Laterality effects in audition. In V. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance* (pp. 177–195). Baltimore, MD: Johns Hopkins Press.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, *6*, 661–672.

Pitt, M. A., & Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 28–38.

Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S. J., Lechevalier, B., & Eustache, F. (1997). The structural components of music perception: a functional anatomical study. *Brain*, *120*, 229–243.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., & Zilles, K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage*, *13*, 669–683.

Risset, J.C., & Wessel, D.L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 113–169). San Diego, CA: Academic Press.

Samson, S., & Zatorre, R. J. (1994). Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia*, *32*, 231–240.

Samson, S., Zatorre, R. J., & Ramsay, J. (2002). Deficits in musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, *125*, 511–523.

Stark, C. E. L, & Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences of the USA*, *98*, 12760–12766.

Stephan, K. M., Passingham, R. E., Silbersweig, D., Ceballos-Baumann, A. O., Frith, C. D., & Frackowiak, R. S. J. (1995). Functional anatomy of the mental representation of upper extremity movements in healthy subjects. *Journal of Neurophysiology*, *73*, 373–386.

Westbury, C. F., Zatorre, R. J., & Evans, A. C. (1999). Quantifying variability in the planum temporale: a probability map. *Cerebral Cortex*, *9*, 392–405.

Worsley, K., Marrett, S., Neelin, P., Vandal, A., Friston, K., & Evans, A. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *NeuroImage*, *4*, 58–73.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., & Evans, A. (2001). A general statistical analysis for fMRI data. *NeuroImage*, *15*, 1–15.

Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport*, *12*, 3045–3049.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.

Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, *31*, 221–232.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8*, 29–46.