

Listening to polyphonic music recruits domain-general attention and working memory circuits

PETR JANATA, BARBARA TILLMANN, and JAMSHED J. BHARUCHA
Dartmouth College, Hanover, New Hampshire

Polyphonic music combines multiple auditory streams to create complex auditory scenes, thus providing a tool for investigating the neural mechanisms that orient attention in natural auditory contexts. Across two fMRI experiments, we varied stimuli and task demands in order to identify the cortical areas that are activated during attentive listening to real music. In individual experiments and in a conjunction analysis of the two experiments, we found bilateral blood oxygen level dependent (BOLD) signal increases in temporal (the superior temporal gyrus), parietal (the intraparietal sulcus), and frontal (the precentral sulcus, the inferior frontal sulcus and gyrus, and the frontal operculum) areas during selective and global listening, as compared with passive rest without musical stimulation. Direct comparisons of the listening conditions showed significant differences between attending to single timbres (instruments) and attending across multiple instruments, although the patterns that were observed depended on the relative demands of the tasks being compared. The overall pattern of BOLD signal increases indicated that attentive listening to music recruits neural circuits underlying multiple forms of working memory, attention, semantic processing, target detection, and motor imagery. Thus, attentive listening to music appears to be enabled by areas that serve general functions, rather than by music-specific cortical modules.

Most natural environments contain many simultaneously active sound sources, thereby presenting listeners with a complex auditory scene. As with visual scenes, in which it is important to segregate objects from one another in order to be able to orient to and interact with the visual environment, the auditory scene must also undergo extensive analysis before individual sources can be identified and segregated from others (Bregman, 1990). Natural auditory scenes vary greatly in their content and complexity, and it is often necessary to focus attention selectively on a single sound source. The most famous example of an auditory scene that humans analyze is the cocktail party in which the voice of an individual must be separated from the voices of others (Cherry, 1953). Music forms the basis for another intriguing and diverse class of auditory scenes, which are created and appreciated in all human cultures.

Polyphonic music is a composite of multiple streams, in which the aggregate can be appreciated, yet individual components can be analyzed. For example, when listening to a blues band, it is relatively easy to focus attention selectively on the drums, bass, keyboards, or lead guitar. Relevant acoustic features that facilitate stream segregation are the relative pitch height and rate at which tones are presented (Bregman & Campbell, 1971), timbre (Iverson, 1995), rhythmic patterns (Jones, 1993), and spatial location (see Bregman, 1990, for a review). These features provide a basis for selective attention in music, and they define auditory objects, sequences of which constitute auditory streams. For example, when the tones of two different melodies are interleaved temporally, a listener can discern the individual melodies, provided they differ from each other along at least one feature dimension—for example, pitch range, timbre, or loudness (Dowling, 1973).

The neural mechanisms underlying auditory attention to both single and multiple streams of auditory information have been studied extensively, using event-related potential (ERP) recordings from the human scalp (see Näätänen, 1992, for a review). With few exceptions, however, auditory streams in most ERP and psychophysical experiments have consisted of extremely simple stimuli, such as repeated pure tones with occasional pitch or loudness deviants. Experiments in which selective attention to one of two streams has been investigated generally have distinguished the streams spatially by presenting a different stimulus train to each ear (Alho, 1992; Hillyard, Hink, Schwent, & Picton, 1973; Woldorff, Hackley, & Hillyard, 1991). The neural mechanisms underlying selective attention

This research was supported by NIH Grant P50 NS17778-18 to J.J.B., the National Institute of Drug Abuse, the McDonnel Foundation, and the Dartmouth Brain Imaging Center. We thank Jeffrey L. Birk for transcribing the Schubert excerpt into MIDI format for Experiment 2, Lauren Fontein for helping us to test subjects, Matthew Brett for providing display routines, Souheil Inati for assistance in pulse sequence selection and for providing file conversion routines, and Scott Grafton for helpful comments. The manuscript benefited from the criticisms and suggestions of two anonymous reviewers. The data and stimuli from the experiments reported in this paper are available upon request from the fMRI Data Center at Dartmouth College (<http://www.fmridc.org>) under Accession Number 2-2002-112YT. Correspondence concerning this article should be addressed to P. Janata, Department of Psychological and Brain Sciences, 6207 Moore Hall, Dartmouth College, Hanover, NH 03755 (e-mail: petr.janata@dartmouth.edu).

tion to somewhat more complex streams consisting of repeating phoneme and consonant–vowel (CV) syllable tokens have also been investigated (Hink, Hillyard, & Benson, 1978; Sams, Aulanko, Aaltonen, & Näätänen, 1990; Szymanski, Yund, & Woods, 1999a, 1999b). Selective attention in the presence of more than two streams has received scant attention (Alain & Woods, 1994; Brochard, Drake, Botte, & McAdams, 1999; Woods & Alain, 2001).

Functional neuroimaging (positron emission tomography [PET] and functional magnetic resonance imaging [fMRI]) investigations of attention in auditory contexts have begun to provide converging evidence as to which general brain circuits are recruited during detection of target objects within single streams with or without the presence of secondary streams. In an fMRI experiment, Pugh et al. (1996) presented both binaural (single stream) and dichotic (dual stream) conditions while subjects performed a CV or a frequency modulation (FM) sweep detection task. The increased attentional demands in the dichotic condition resulted in increased activation in the inferior parietal, inferior frontal, and auditory association areas. In a PET/ERP experiment modeled on the typical auditory oddball paradigm used in the ERP field, Tzourio et al., (1997) presented frequent low tones and rare high tones, under passive-listening conditions and with deviance detection, randomly to the right or the left ear. Subjects attended to and reported deviants presented to one of the two ears. Both Heschl's gyrus (HG) and the planum temporale (PT) were active in the passive and attentive conditions, as compared with the rest condition, but there was no significant activation difference in these areas between the passive and the attentive conditions. However, selective attention increased activation significantly, as compared with passive listening, in the supplementary motor area (SMA), anterior cingulate, and precentral gyrus bilaterally. These results and others (Linden et al., 1999; Zatorre, Mondor, & Evans, 1999) suggest that selective attention operates on representations of auditory objects after processing in the primary and secondary auditory areas is complete, although other fMRI and ERP data suggest that selective attention operates on early auditory representations (Jäncke, Mirzazade, & Shah, 1999; Woldorff et al., 1993; Woldorff & Hillyard, 1991). However, in the experiments cited above, the stimuli were acoustically simple, the contexts were unnatural, and attention was deployed in the framework of a target detection task. Thus, the specific requirements of target detection and decision making were combined with potentially more general attentional processes.

In the music perception domain, Platel et al. (1997) performed a PET study of selective attention to musical features of short, simple tonal sequences. While listening to the same body of material repeatedly, subjects identified the category membership of individual sequences along the feature dimensions of pitch, timbre, and rhythm. The goal of the study was to identify those brain areas that uniquely represent or process these different feature dimensions. By contrasting various combinations of activation

maps—for example, pitch versus timbre or rhythm versus pitch and timbre combined—several areas in the frontal lobes were found to be differentially activated by attention to the different feature dimensions. In a recent PET study of attentive listening in musicians, Satoh, Takeda, Nagata, Hatazawa, and Kuzuhara (2001) found that listening for target pitches in the alto voice of four-part harmonic progressions resulted in greater bilateral regional cerebral blood flow (rCBF) in superior parietal and frontal areas, relative to listening for minor chord targets in the same progressions. The frontal activations appear to include regions around the inferior frontal sulcus (IFS) and the precentral sulcus (PcS), the left-SMA/pre-SMA, and the orbitofrontal cortex.

Although the use of acoustically and contextually simple stimuli is alluring from an experimental point of view, our brains have adapted to rich and complex acoustic environments, and it is of interest to determine whether stream segregation tasks using more complex stimuli exhibit brain response patterns similar to those using simpler stimuli. Therefore, just as experiments on the neural mechanisms of visual processing have largely progressed beyond the use of simple oriented bar stimuli to the use of faces and other complex objects, both natural and artificial, presented individually (Haxby, Hoffman, & Gobbini, 2000) or as natural scenes consisting of multiple objects (Coppola, White, Fitzpatrick, & Purves, 1998; Stanley, Li, & Dan, 1999; Treves, Panzeri, Rolls, Booth, & Wakeman, 1999; Vinje & Gallant, 2000), we believe that functional neuroimaging experiments using more complex acoustic stimuli embedded in natural contexts may help identify those brain regions that have adapted to the processing of such stimuli. This type of neuroethological approach has proven indispensable for identifying the neural substrates of behaviors in other species that depend on complex auditory stimuli—for example, birdsong (Brenowitz, Margoliash, & Nordeen, 1997; Konishi, 1985).

We conducted two experiments in order to identify the brain areas recruited during attentive listening to excerpts of real music and to identify the brain areas recruited during selective listening to one of several simultaneously presented auditory streams. To this end, we varied task demands, musical material, and trial structure across the experiments. In the first experiment, subjects were instructed to listen to the musical passages either globally/holistically, without focusing their attention on any particular instrument, or selectively, by attentively tracking the part played by a single instrument. Within the constraints of the fMRI environment, we sought to image blood oxygen level dependent (BOLD) signal changes as the subjects engaged in relatively natural forms of music listening. Our objective of imaging natural attentive music listening posed a significant challenge, because we did not want to contaminate “normal” listening processes with additional demands of secondary tasks, such as target detection. Auditory attention is studied almost exclusively in the context of target detection, in which a stream is monitored for the occurrence of single target events. By virtue

of being able to measure the number of targets that are detected as a function of attentional orienting, this type of design provides a powerful tool for studying attention. However, the subjective experience of listening attentively to music is not the same as performing a target detection task, and it is possible that brain dynamics vary with those different forms of attentional processing.

To further investigate this hypothesis with a second experiment, an excerpt of classical music was adapted for use in a target detection task with two levels of attentional instructions. As in the first experiment, subjects were required to orient their attention either globally or selectively. This time, the attentional orienting was coupled to detection of targets that could occur in any one of three streams (global/divided-attending condition) or in the single attended stream (selective-attending condition).

Finally, we sought to identify the cortical areas in which the BOLD signal increased reliably during attentive music perception across different experimental contexts. A conjunction analysis enabled us to identify the cortical areas that were activated in both of the experiments, thereby identifying processes common to attentive music listening despite variation in the musical material and tasks.

EXPERIMENT 1

For this experiment, we chose a musical example consisting of two simultaneous melodies (streams). Specifically, we selected excerpts from two baroque flute duets in which the melodic contour and rhythmic properties of the two voices were very similar. The feature dimension of primary interest was timbre, which we manipulated by assigning two perceptually very different timbres (Krumhansl, 1989; McAdams, Winsberg, Donnadieu, Desoete, & Krimphoff, 1995) to the melodies. We focused on timbre because it is a potent cue for auditory stream segregation and should facilitate orienting of attention to a single melody line and because we wanted to identify regions of interest for further studies of timbre processing.

Method

Subjects. Twelve subjects (7 females; mean age, 28.5 years; age range, 20–41) participated in the experiment: 8 fellows from the 2000 Summer Institute in Cognitive Neuroscience held at Dartmouth College, 3 visitors to Dartmouth, and 1 member of the Dartmouth community. All the subjects reported having normal hearing, and 11 of the 12 subjects were right-handed. The duration of the subjects' formal musical training was 8.3 ± 3.5 years, and the average duration for which they had played one or more instruments was 17.4 ± 6.9 years. All the subjects provided informed consent according to Dartmouth human subjects committee guidelines.

Stimuli and Procedure. Melodies were excerpted from the *Vivace* and *Giga* movements of the Sonata in F Major for two recorders by J. B. Loeillet (1712, op. 1, no. 4), thus providing two sets of two melodies (streams) each (Figure 1A). All of the individual melodies were stored as separate MIDI tracks, using MIDI sequencing software (Performer 6.03, Mark of the Unicorn). When stored as MIDI sequences, every note in all of the melodies could be equated in intensity by assigning the same velocity value. Each melody could be assigned to an arbitrary timbre. We used the "vibes" and "strings" timbres generated by an FM tone generator (TX802, Yamaha). These

two timbres occupy opposite corners of three-dimensional (3-D) timbral space, as defined by multidimensional scaling solutions of perceptual similarity judgments (Krumhansl, 1989; McAdams et al., 1995). The most salient difference between the two timbres is the attack time.

The assignment of timbres to melodies was completely counterbalanced so that attention was directed to each of the four melodies played by each of the two timbres over the course of the experiment (Figure 1B). Every subject was presented with a complete set of melody/timbre combinations across the two blocks. The ordering of blocks was counterbalanced across subjects. The counterbalance design allowed us to compare directly the activations caused by selective attention to the two different timbres. Stimulus files corresponding to each of the different orders were arranged in a single track of an audio file (SoundEdit16, Macromedia). Scanner trigger and timing pulses were arranged on a second track, and both tracks were written to CD.

The stimuli were delivered via ear-insert tubephones (ER-30, Etymotic Research). In order to obtain better acoustic separation of the music from the echo-planar imaging (EPI) ping, the subject wore ear muffs (HB-1000, Elvex), through which the tubephone s had been inserted. Prior to amplification, the audio signal from the CD was filtered with a 31-band 1/3 octave band equalizer (Model 351, Applied Research Technologies) to compensate for the attenuation by the tubephones of frequencies above 1.5 kHz. The stimuli were presented at ~ 102 dB SPL.

Each functional EPI run began with 30 sec of rest in order to adapt the subject to the EPI sound and to allow the global signal intensity in the EPI images to stabilize. Eight seconds prior to the onset of each task epoch, a verbal cue indicated the task: listen, attend vibes, or attend strings. Task epochs were 30 sec long and were immediately followed by a 30-sec rest period (rest condition) prior to the next verbal cue. We began and ended each block with epochs in which the subjects were asked to listen to the melodies in a holistic, integrated manner, rather than attempting to focus their attention on one timbre or the other (listen condition). The subjects were instructed to focus their attention as best they could on the cued instrument during the attend epochs. Following the experiment, the subjects completed a questionnaire and an interview about their musical training, task difficulty ratings, and the strategies they had used in performing the task.

Data acquisition. Data were acquired on a General Electric Signa Horizon Echospeed MRI scanner (1.5 T), fitted with a GE birdcage head coil. Functional images were collected with a gradient echo EPI pulse sequence with the following parameters: TR = 2 sec; TE = 35 msec; field of view (FOV) = 240×240 mm; flip angle (α) = 90° ; matrix size = 64×64 ; resolution = $3.75 \times 3.75 \times 5.0$ mm; interslice spacing = 0 mm. Twenty-seven axial slices were collected, providing whole-brain coverage. Two sets of high-resolution T1-weighted anatomical images were also obtained. The first was a set of 27 slices taken in the same planes as the functional images (coplanar) and obtained with a two-dimensional fast spin echo sequence with the following parameters: TR = 650 msec; TE = 6.6 msec; FOV = 240×240 mm; $\alpha = 90^\circ$; matrix size = 256×256 ; resolution = $0.937 \times 0.937 \times 5.0$ mm; interslice spacing = 0 mm. The second set consisted of 124 sagittal slices and was obtained with a 3-D SPGR sequence with the following parameters: TR = 25 msec; TE = 6.0 msec; FOV = 240×240 mm; $\alpha = 25^\circ$; matrix size = 256×192 ; resolution = $1.2 \times 0.937 \times 0.937$ mm.

Data processing. EPI volumes from the initial 30-sec adaptation phase of each run, during which the subjects were waiting for the task to begin, were discarded prior to the analysis. SPM99 was used for image preprocessing and statistical analyses (Friston et al., 1995). Unless otherwise specified, all algorithms ran with default settings in the SPM99 distribution (<http://www.fil.ion.ucl.ac.uk/spm>). In order to estimate and correct for subject motion, realignment parameters relative to the first image of the first run were computed for each of the functional runs. A mean functional image

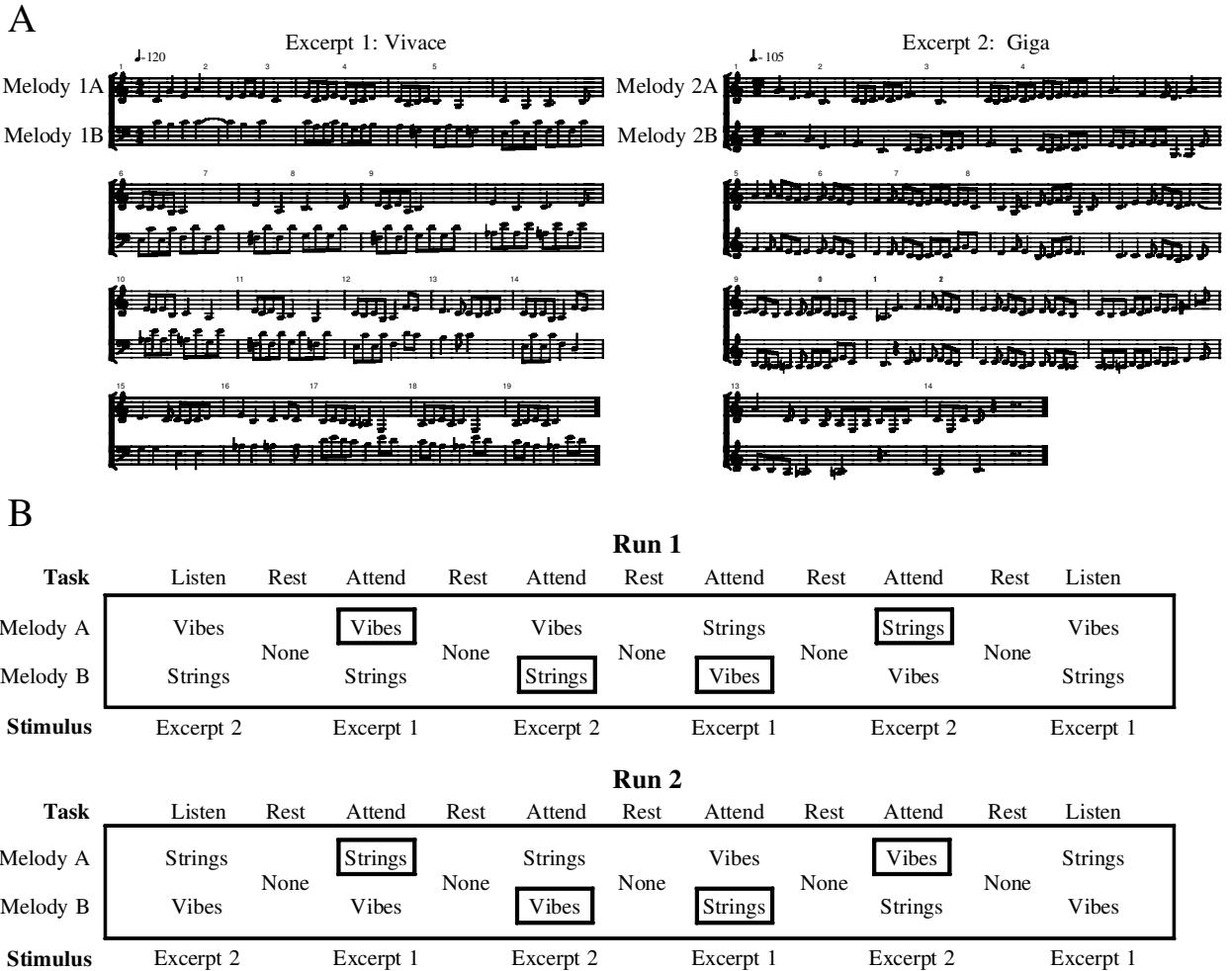


Figure 1. Design of Experiment 1. (A) The excerpts from two baroque flute duets used in the experiment. Each melody could be mapped to an arbitrary synthesized timbre, in this case “vibes” and “strings” timbres. (B) Counterbalancing scheme used in the experiment. In each run, task blocks alternated with rest blocks. The subjects heard only echo-planar imaging pinging during the rest blocks. Streams A and B refer to the melodic lines shown above—for example, Melody 1A and Melody 1B, respectively. The diagram indicates which excerpt was played and which timbre was mapped to each stream (melodic line) in each task block. For task blocks in which the subjects focused their attention selectively on a single stream, the attended stream (timbre) is highlighted with a box.

was constructed and used to coregister the functional images with the coplanar anatomical images. The mutual information algorithm implemented in SPM99 was used for coregistration (Maes, Colligon, Vandermeulen, Marchal, & Suetens, 1997). The coplanar images were then coregistered with the 3-D high-resolution images. These, in turn, were normalized to the Montreal Neurological Institute’s T1-template images that approximate Talairach space, and the normalization parameters were applied to the functional images. Default parameters were used for the normalization, with the exception that normalized images retained the original voxel size of $3.75 \times 3.75 \times 5$ mm. Normalized images were smoothed with a $6 \times 6 \times 8$ mm (FWHM) Gaussian smoothing kernel. Both individual and group statistical parametric maps (SPMs) were constructed from these images.

Regression coefficients were estimated for rest, verbal cue, listen, attend vibes, and attend strings conditions. The verbal cue onset events were modeled as consisting of an *early* and a *late* component, using a modified version of the basis function routine in SPM99. Gamma functions for the early and the late components were specified by pass-

ing shape parameter values of 2 and 3 to the gamma probability density function specified in SPM99, and these were then orthogonalized by the same procedure as the default set of SPM99 gamma functions. Similarly, the onset of each attentional condition was modeled with these gamma functions in order to emphasize the initial period of attentional orienting at the onset of the musical excerpt. The regressors for the rest, listen, and attend conditions were modeled as a boxcar waveform convolved with the SPM99 canonical hemodynamic response function (HRF). Regressors for a linear trend and estimated motion parameters were also included in the model for each subject.

In order to identify general task-related BOLD signal increases and for purposes of comparison with Experiment 2, SPMs were created for the following contrasts: *listen* minus *rest*, and *attend* (pooled across timbres) minus *rest*. These contrasts identify general regions that respond to the musical stimulus under the different attentional demands. Because the attend conditions were also separated by timbre, we compared activation maps elicited by attending to different timbres, as compared with the rest condition. In addition, the focused-attending (attend) condition was compared directly with the global-

listening (listen) condition to determine whether the different modes of attending to the music elicited different patterns of BOLD signals.

The contrast maps from each subject were then entered into a random-effects analysis, and SPMs for the group data were created. SPMs were thresholded at $p < .01$ (uncorrected). Activations significant at the cluster level (extent threshold: clusters larger than the estimated resolution element size of 19 voxels) were identified by comparing SPMs projected onto the mean T1-weighted structural image (averaged across the subjects in the study) with the atlas of Duvernoy (Duvernoy, 1999).

Results and Discussion

All the subjects reported that they performed the task as instructed (Table 1). Eleven of the 12 subjects judged the two timbres to be equally loud. However, the ratings of how easy it was to attend to each timbre were more variable. On average, each timbre could be attended to with relative ease, although the within-timbre ratings differed widely across subjects for both timbres. Some subjects had an easier time attending to vibes than to strings, whereas for other subjects the opposite was true. The difficulty ratings for the two timbres were within one point of each other for 5 of the subjects.

Significant BOLD signal increases were observed in three principal areas in the listen–rest contrast (Table 2, Figure 2A). These included, bilaterally, the superior temporal gyrus (STG) spanning from the planum polare (PP) to the PT, thus encompassing the regions surrounding and including portions of HG. The SMA/pre-SMA showed increased BOLD signals bilaterally. The third area of significantly increased BOLD signal was the right Pcs.

When the subjects were instructed to attend selectively to one instrument or the other, the pattern of BOLD signal increases was largely the same as when they were instructed to listen more holistically. However, increased BOLD signals during selective listening was noted along the left intraparietal sulcus (IPS) and the supramarginal gyrus (Figure 2A, +40 mm slice). Frontal BOLD signal increases were observed bilaterally in a swath stretching from the dorsal aspect of the inferior frontal gyrus (IFG), along the IFS to the inferior Pcs (Figure 2A, slices +20 to +60). The pre-motor areas on the right were activated during both holistic listening and selective attending, whereas the frontal and parietal regions in the left hemisphere appeared to be most affected by the selective-listening task. Figure 2B shows that a direct statistical comparison of the attend and the listen conditions largely conformed to these observations (Table 2). In addition, selective attending resulted in significantly greater BOLD signals in the posterior part of the left STG and the underlying superior temporal sulcus (STS). In

the opposite contrast, the listen condition exhibited clusters of significantly greater BOLD signals than did the selective attend condition bilaterally along the anterior calcarine sulcus and in the rostral fusiform gyrus.

Because the stimulus materials were completely counterbalanced, we could explore whether there were any differences in the responses to the two timbres. We observed no striking differences in the group images as a function of the timbre that was attended (note the overlap of contours in Figure 2A). The direct comparison of attentively listening to each timbre yielded very small foci, although neither did these fall in the auditory cortex or surrounding association areas nor were they closely apposed, as might be expected from activation of an area possessing a timbre map (data not shown). Thus, we did not pursue the issue further.

EXPERIMENT 2

Experiment 2 was performed in order to obtain objective behavioral verification, in the context of a target detection task, that subjects had oriented their attention as instructed during the fMRI scans. In addition, we used an excerpt from a Schubert trio to assess whether the activation patterns elicited with the stimulus materials of Experiment 1 could be elicited with a musical excerpt consisting of three streams. In contrast to the first experiment, Experiment 2 used a timbral deviance detection task to assess the degree to which the subjects were able to focus on a single instrument or divide their attention across instruments. Because we needed to obtain behavioral responses during divided (global/holistic) and selective attention conditions, our ability to use the musical excerpt as a control stimulus in a passive holistic-listening condition (as in Experiment 1) was compromised. To take the place of the passive-listening condition, we added a *scrambled* condition in which the subjects listened to a nonmusical stimulus that was derived from the musical excerpt by filtering white noise with the average spectral characteristics in successive 1.5-sec windows of the original Schubert trio excerpt. Thus, very coarse spectrotemporal features were preserved, but stream and rhythmic cues were removed. Owing to the change in task structure, the experimental protocol was modified from one in which 30-sec task epochs were alternated with 30-sec rest epochs to one in which 15-sec trials were interleaved with 8-sec rest periods.

Method

Subjects. Fifteen subjects (8 females; mean age, 18.73 years; range, 18–22) participated in behavioral pretesting of the stimulus material. Of these subjects, 5 (4 females; age range, 18–19 years) participated in the fMRI scanning sessions. Three additional subjects (2 females; age range, 19–20 years) participated in the fMRI scanning session following a brief training session on the task. The average amount of formal musical training in the behavioral pretest cohort was 4.4 ± 3.76 years, and for the fMRI cohort it was 6.13 ± 5.79 years. All the subjects had normal hearing.

Stimuli. The musical passage is shown in Figure 3. The instruments' parts were transcribed into MIDI sequences to allow the introduction of timbral deviants at various points in the excerpt. The

Table 1
Average Ratings of Task Difficulty in Experiment 1

	Rating	Range
Relative loudness	3.9 ± 0.3	3–4
Attend to strings	2.73 ± 1.5	1–5
Attend to vibes	2.5 ± 1.2	1–5

Note—For relative loudness, 1 = *strings*, 7 = *vibes*; for the attend conditions, 1 = *easy*, 7 = *difficult*.

Table 2
Areas of Activation in Experiment 1

Note—STG, superior temporal gyrus; STS, superior temporal sulcus; FG, fusiform gyrus; SMA, supplementary motor area; PoG, precentral gyrus; PoS, precentral sulcus; IFS, inferior frontal sulcus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; IPS, intraparietal sulcus; GP, globus pallidus.

note-on and note-off velocities were set to the same value within each part, and the relative note velocities between instruments were adjusted to achieve comparable salience of the individual parts. The standard piano notes were rendered by a Korg SG-1D sampling grand piano. Timbres for all violin and cello notes and for piano deviants were rendered by an FM tone generator (TX802, Yamaha).

In addition to the original standard sequence, 4 sequences containing deviants were constructed for each of the three instruments, resulting in a total of 12 deviant sequences. For any given deviant sequence, there was a deviant in a single instrument and one of four locations. The locations were distributed across early and late portions of the excerpt. In the case of the violin and cello, the deviant notes encompassed one measure—that is, a window of 1,000 msec—whereas for the piano, the deviant notes took up two thirds of a measure (666 msec). The deviant notes were played with a slightly dif-

ferent timbre—for example, one of the other “violin” sound patches available on the Yamaha synthesizer. The standard and deviant sequences were recorded with MIDI and audio sequencing software (Performer 6.03, MOTU). Each sequence was 15.5 sec long. The same set of sequences was used for the selective and the divided attention conditions described below.

The time courses of root-mean square (RMS) power were compared for the standard and the deviant sequences. Initially, the note velocities of the deviants were adjusted until the RMS curves overlapped for the standard and the deviant sequences. Unfortunately, deviant notes adjusted by this procedure were very difficult to detect and often sounded quieter than the surrounding standard notes. We therefore adjusted the velocities of deviants until detection performance in subsequent groups of pilot subjects rose above chance levels. This adjustment resulted in a combination of timbral and

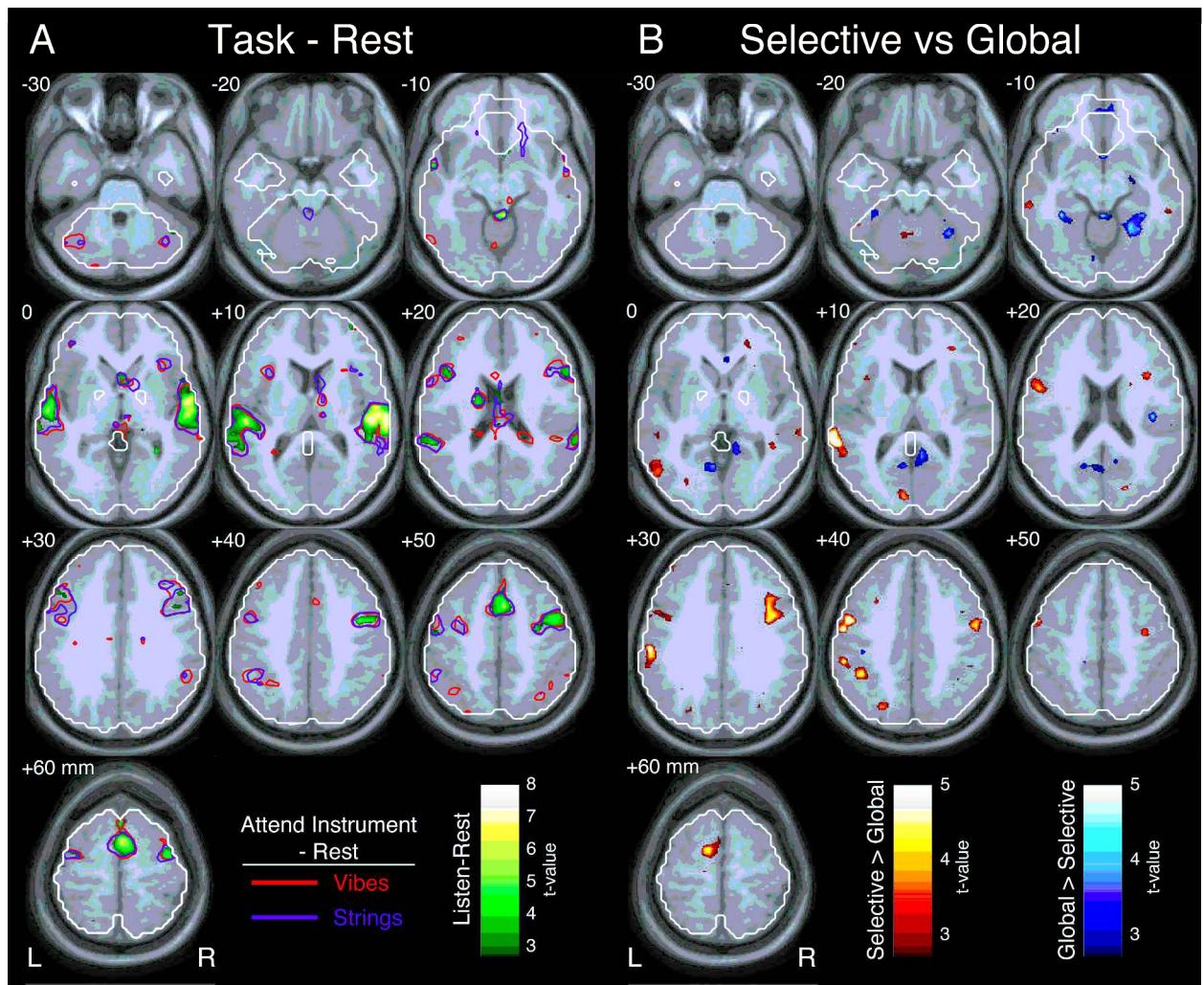


Figure 2. Group ($N = 12$) images showing significant blood oxygen level dependent (BOLD) signal increases ($p < .01$) for two sets of contrasts from Experiment 1, superimposed on the average T1-weighted anatomical image of the subject cohort. (A) Contrast of global listening (green blobs) and selective listening to each of the timbres (red and blue contours) with rest. (B) The direct contrast of selective listening with global listening. Brain areas with significantly stronger BOLD signals during selective listening, as compared with global listening, are shown in a red gradient, whereas the opposite relationship is shown in a blue gradient. The white contour line denotes the inclusion mask that shows the edges of the volume that contained data from all participants. Susceptibility artifact signal dropout in orbitofrontal and inferotemporal regions is clearly seen in the top row of images.



Figure 3. The Experiment 2 stimulus in musical notation. The arrows mark the starting and stopping points of the 15-sec excerpt used as the stimulus. From F. Schubert, "Trios für Klavier, Violine, und Violincello," op. 100, D 929, HN 193, p. 96, measures 137 through 156, Copyright 1973 by G. Henle Verlag, Munich.

loudness deviance. We must emphasize that the multidimensional aspect of the deviants is tangential to the primary goal of this experiment. In other words, we employed deviants in this task not because we were trying to investigate target detection along specific feature dimensions, but rather because we wanted objective verification that the subjects were attending to the excerpts and we wanted to create two different types of this attention with this task—divided and selective.

In addition to the musical excerpts, we synthesized a control excerpt (scrambled) designed to match the gross spectral characteristics of the original excerpt in 1.5-sec time windows. This was accomplished as follows. First, the amplitude spectra of the standard sequences were created for 64 consecutive short segments of 2^{10} samples each (23.22-msec windows). These were then averaged to yield the average spectrum of a longer segment (2^{16} samples, 1,486.08-msec windows). Next, the spectra of corresponding short segments of white noise were multiplied by the average spectrum of the longer window and were converted into the time domain. This process was repeated across the total duration of the original excerpt. In order to eliminate clicks owing to transients between successive noise segments, the beginning and end of each noise segment were multiplied by linear ramps (3.63 msec). At shorter window durations of average spectrum estimation, the filtered noise stimulus tended to retain the rhythmic qualities of the original excerpt and induced a weak sense of stream segregation between a low and a high register in the noise. Since we wanted to minimize recruitment of musical processing or attention to one of multiple streams in this control stimulus, we chose the 1.5-sec windows. We did not use envelope-matched white noise (Zatorre, Evans, & Meyer, 1994), since that type of stimulus retained the rhythmic properties of the excerpt.

Procedure. The subjects were scanned during three consecutive 10.5-min runs. Each run consisted of divided and selective attention trials. During a divided attention trial, the subject heard the verbal cue, "everything," indicating that the deviance, if it occurred, could occur in any of the three instruments. During the selective attention trials, the subjects heard the name of one of the three instruments. They were informed that if a deviant was to occur, it would occur only in the cued instrument. Deviants never occurred in the uncued instruments. The subjects were instructed to press a left button after the excerpt finished playing if they detected a deviant and a right button if they detected none. Since buttonpresses immediately following targets would have contaminated activations arising from the process of attentive listening with motor response activations, which would have introduced a requirement for additional control conditions, the subjects were instructed to withhold their buttonpresses until the excerpt had finished playing. The attentional trials (divided and selective) were interspersed with the scrambled excerpts, to which the subjects listened passively, pressing the right button after the excerpt finished playing. Verbal cues were given 6 sec prior to the onset of the excerpt. The pause between the end of the excerpt and the following verbal cue was 8.5 sec. Across the three runs, 12 deviant and 9 standard sequences were presented in each of the attentional conditions. Thus, the probability of hearing a sequence with a deviant was .57. Overall, 12 scrambled excerpts were presented. Note that in the behavioral pretest, the probability of deviants was .50, and there were no scrambled trials. The greater proportion of deviants in the fMRI experiment arose from reducing the overall length of functional scans by eliminating 3 standard sequences during each run.

Stimuli were pseudorandomized so that the number and timbral identity of deviants would be balanced across runs. Within each run, the ordering of stimuli was randomized. Event timing and stimulus

presentation were controlled by PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Stimuli were presented from the left audio channel, which was subsequently split for stereo presentation.

Data acquisition. The equipment and protocols described in Experiment 1 were used for data acquisition and stimulus presentation. A magnet-triggering pulse at the start of each run and event markers at the onsets of verbal cues and music excerpts (sent from the right audio channel of the computer running PsyScope), together with the output of the response button interface box (MRA Inc.), were mixed (MX-8SR, Kawai) and were recorded to an audio file by SoundEdit 16 running on an independent iMac computer (Apple Computer).

Data processing. The initial nine EPI volumes, during which the subjects performed no task, were discarded prior to any data analysis. Procedures for image coregistration, spatial normalization, and smoothing were the same as those in Experiment 1. A regression model was specified in which experimental conditions were modeled as either *epochs* or *events*. Conditions specified as epochs were the rest condition, during which no stimulus was presented, the acoustic control sequence (scrambled), and a total of eight music conditions reflecting the divided and the selective attention conditions and each of the four deviance types (cello, piano, violin, and none) that were present in each of the attention conditions. Verbal cues and responses were modeled as events. The onset times of verbal cues, stimulus excerpts, and buttonpresses, relative to the start of data acquisition on each run, were extracted from the audio files recorded during each run, using custom-written event extraction scripts in Matlab (Mathworks). Epochs were modeled as a 15.5-sec boxcar convolved with the canonical HRF in the SPM99 toolbox. In order to emphasize processes related to auditory stream selection at the beginning of a musical excerpt, the onsets of each epoch in both the divided and the selective attention conditions were modeled as events. The verbal cue and attentional onset events were modeled as consisting of an *early* and a *late* component, as in Experiment 1. Response events were modeled as an impulse convolved with the canonical HRF. The motion parameters estimated during EPI realignment and a linear trend were added to the model to remove additional variance.

In order to facilitate comparison with the data from Experiment 1, the beta parameter estimates from the selective attention condition regressors were combined and compared with the beta parameter estimate for the rest condition. Similarly, the beta parameter estimates from the divided attention condition regressors, which presumably promoted active holistic listening, were combined and contrasted with those for the rest condition. The response to the acoustic control stimulus (scrambled) condition was also contrasted with the rest condition. Finally, the selective and divided attention conditions were compared directly by calculating the difference between their respective beta parameter estimates. Group effects were estimated from the contrast maps for individual subjects, thresholded at $p < .01$ (magnitude) and one resolution element (24 voxels for the contrasts with the rest condition and 17 voxels for the direct comparison of attention conditions) and were projected onto the mean T1-weighted anatomical image. Anatomical regions of significant BOLD signal changes were identified as in Experiment 1.

Results and Discussion

Behavior. The best behavioral evidence in support of dissociable neural mechanisms underlying selective and divided attention in real musical contexts would be a cost in detection accuracy during divided attention relative to selective attention. The ideal distribution of responses in the target detection framework would be one in which targets occurring during divided attention would be missed—that is, detected at chance levels—and all targets in the selectively attended stream would be detected correctly. False alarms would also be lower under selective attention conditions.

In the first set of calibration experiments, in which regular headphones were used, the subjects ($n = 6$) detected deviants under the divided attention condition at chance level (49%). In the selective attention condition, detection of deviants was above chance (60%) and significantly higher than in the divided attention condition (paired t test, $t = -4, p < .01$).

The adjustment of target intensity levels proved more difficult when fMRI tubephones were used. With a group of 9 subjects, 71% of the divided attention targets were detected on average, as compared with 81% of the selective attention targets, demonstrating that targets could be reliably detected even under less than optimal listening conditions. However, the difference between attentional conditions only bordered on statistical significance ($p = .084$). In practice, we found that we were unable to adjust the amplitudes of the individual timbral deviants so that we could observe a strong and reliable dissociation between forms of attention in the detection of targets across subjects. The subjects showed different propensities in detecting deviants in the different instruments. Some detected the piano deviants better than the cello deviants, and for others the opposite was true. Similarly, the accuracy with which the subjects detected deviants at each of the four positions within the passage varied from subject to subject. The length of the passage (15 sec), the overall length of the fMRI experiment (1.5 h), and our dependence on pre-recorded material prevented us from engaging in an adaptive psychophysical procedure to determine each individual's threshold for each timbral deviant in the context of a musical piece.

Five subjects whose behavioral performance was better for the selective than for the divided attention condition during the calibration experiments participated in the fMRI scanning session, as did 3 additional subjects who received instruction and 12 training trials immediately preceding the fMRI session. Detection of deviants was significantly above chance levels in both the selective ($71\% \pm 7\%$ hits [mean \pm SEM]; $t = 2.0, p < .05$) and the divided ($73\% \pm 6\%$; $t = 2.7, p < .02$) conditions. These hit rates were not significantly different from each other (paired t test, $t < 1$). In addition, false alarm rates were low in both the selective and the divided attention conditions (11% and 8%, respectively).

Table 3 summarizes the subjects' ratings of task difficulty. Ratings were made on an integer scale from 1 (*very easy*) to 7 (*very difficult*). The subjects found attending to the violin part and detecting violin deviants relatively more difficult, as compared with the other instruments, during fMRI scanning. The subjects reported that the EPI ping-pong interfered with their ability to detect deviants in the violin stream, because the two fell into the same register. We were aware of the interaction between the magnet and the violin streams at the outset but were not concerned by it because the *de facto* task of separating the violin from the magnet was consistent with our aim of investigating processes underlying stream segregation in natural contexts. In general, both the listen condition of Experiment 1 and the divided attention condition in the present experi-

Table 3
Average Ratings of Task Difficulty

	Overall Difficulty	Attend			Detect			Detect in Divided	
		Violin	Cello	Piano	Violin	Cello	Piano	Divided	Ignore
First pretest	5.8 ± 1.2	4.0 ± 2.6	3.5 ± 1.2	5.7 ± 1.0	4.8 ± 2.4	3.8 ± 1.5	5.3 ± 1.2	6.2 ± 1.0	5.0 ± 1.6
Second pretest	4.7 ± 1.1	3.6 ± 1.6	3.7 ± 1.9	3.1 ± 1.5	3.7 ± 1.7	4.3 ± 1.5	3.0 ± 1.3	5.1 ± 1.5	4.9 ± 2.2
fMRI	4.4 ± 1.2	4.5 ± 1.4	2.6 ± 1.1	3.0 ± 1.3	4.6 ± 1.8	2.8 ± 1.3	3.2 ± 1.7	4.0 ± 1.3	5.9 ± 1.4

Note—Ratings were made on an integer scale from 1 (easy) to 7 (difficult). “Attend” columns indicate how difficult it was to maintain attention to each stream, whereas “Detect” columns indicate how difficult it was to detect deviants within each stream. “Detect in Divided” refers to difficulty of detecting targets in the divided attention condition. “Ignore” indicates how often the subjects found themselves ignoring the music altogether (1, often; 7, not often).

ment may involve a baseline level of selective attention associated with segregating the music from the EPI pinging. The overall difficulty of the task was rated as moderately difficult both outside and inside the magnet, as was the difficulty of detecting deviants when the instruction was to attend to all the instruments. Ratings of how often (1, *often*; 7, *not often*) the subjects found themselves ignoring the music altogether indicated that the subjects almost always attended to the music. These self-reports are in agreement with the observed performance in the detection task.

Physiology. In contrast to Experiment 1, the attentional load in this experiment was high whenever the subjects heard the musical excerpt. Not surprisingly, the patterns of increased BOLD signals were very similar in the divided and the selective attention conditions, as compared with the REST condition (Table 4, Figure 4A). This observation is consistent with the absence of behavioral differentiation of the two conditions. Despite the similar behavioral performance and largely overlapping activation pattern, a direct comparison of the BOLD signal elicited under the different attention conditions revealed significant differences in the activation patterns (Table 4, Figure 4B). In the comparisons with the rest condition, both conditions recruited, bilaterally, regions of the STG surrounding and including the HG, the pre-SMA, and the frontal operculum. The right-hemisphere IFG activation extended more laterally than on the left. On the right, an area around the intersection of the superior precentral and superior frontal sulci and the posterior part of the middle frontal gyrus (MFG) was activated in both conditions, although in the selective attention condition, the size of the cluster at this location was slightly smaller than the extent threshold. The right IPS and the angular gyrus were activated in both attentional conditions, whereas the left IPS was activated primarily in the divided attention condition.

The divided attention condition was associated with two additional activation foci not observed in the previous experiment. One was observed bilaterally in the ventro-rostral part of the MFG, although the cluster on the right was slightly smaller than the extent threshold used in computing the contrasts. The other was activation of the anterior cingulate gyrus just ventrolateral to the large pre-SMA activation cluster.

The direct comparison of the selective and the divided attention conditions further highlighted the activation differences between the two conditions (Table 4, Figure 4B). In particular, during divided attention, the BOLD signal

was stronger in bilateral parietal, right superior frontal, and left anterior cingulate areas. In contrast, the selective attention condition was associated with a stronger BOLD signal in the left fusiform gyrus and in a number of occipital areas bilaterally, including the lingual gyrus, the lateral occipital sulcus, and the anterior calcarine sulcus. Note that none of the latter areas showed significant activity for the selective attention condition, relative to the rest condition.

When BOLD signal increases were determined for the scrambled condition, relative to the rest condition (Figure 4A), only two areas exceeded the height and extent thresholds: the right STG in the vicinity of Heschl’s sulcus ($x, y, z = 60, -15, 10$), and the left postcentral gyrus ($x, y, z = -52, -19, 50$). The only task during the acoustic control sequences was to listen to the sound and make a response with the right button when the sound terminated.

CONJUNCTION ANALYSIS OF EXPERIMENTS 1 AND 2

Identification of Common Patterns Despite Variations in Tasks and Stimuli

Despite the variability in stimulus materials and tasks, the two experiments making up this study had two conditions in common. In one, the subjects attempted to focus their attention selectively on one of either two or three different instruments that played concurrently (attend condition in Experiment 1, selective condition in Experiment 2), and in the other, they rested while listening to the EPI pinging (rest condition in both experiments). Other conditions that were somewhat related across both experiments were those in which the subjects listened to the music in a more global, integrative, holistic fashion (listen condition in Experiment 1, divided condition in Experiment 2). However, the latter conditions were not as well equated, owing to the increased task demands in the second experiment during the divided attention condition. Given these similarities across experiments, we were interested in identifying the set of cortical areas activated by similar conditions in both experiments.

Problems of Specifying Baseline Conditions and Control Tasks in the Functional Neuroimaging of Audition

Additional motivation for a conjunction analysis stems from the observation that specification of baseline tasks

Table 4
Areas of activation in Experiment 2

Lobe	Hemisphere	Region	Divided-Rest			Selective-Rest			Divided-Selective			Selective-Divided		
			x	y	z	Z Score	x	y	z	Z Score	x	y	z	Z Score
Temporal	left	STG	-52	-19	5	4.16	-56	0	0	3.29	-26	-68	-10	3.1
	right	FG STG	-56	-38	15	3.13	-56	-22	10	4.38				
Frontal	bilateral	pre-SMA	56	-11	5	3.79	56	11	10	3.78	-26	-68	-10	3.1
			56	-22	10	4.86	52	-22	10	4.11				
Frontal	left	pre-SMA frontal operculum	68	-26	10	3.4	0	4	45	4.07	-26	-68	-10	3.1
			68	-26	10	3.4	0	19	45	4.72				
Frontal	right	CS frontal operculum	-38	19	5	4.21	-38	19	5	3.89	-4	8	50	3.34
			-38	22	0	3.78	-38	22	0	3.58				
Frontal	right	SFS superior PeS/SFS/MFG	-30	49	5	3.56	-34	-26	60	3.49	-4	55	3.91	3.20
			-30	49	5	3.56	-45	15	5	3.91				
Frontal	right	MFG PoCG	38	19	0	3.57	34	-26	60	3.49	26	19	50	3.20
			38	19	0	3.57	45	15	5	3.91				
Frontal	right	IPS	34	22	10	3.42	34	-26	60	3.49	30	-4	55	3.92
			34	22	10	3.42	45	15	5	3.91				
Parietal	left	IPS	-41	-49	50	3.93	-56	-22	50	3.09	-34	-52	40	3.41
			-41	-49	50	3.93	-56	-22	50	3.09				
Parietal	right	IPS/angular gyrus	30	-4	55	4.8	30	-4	55	3.87	-30	-64	40	3.65
			34	4	65	2.86	34	4	65	3.87				
Parietal	right	IPS	-41	-49	50	3.93	-56	-22	50	3.09	34	-49	50	3.47
			-41	-49	50	3.93	-56	-22	50	3.09				
Limbic	bilateral	anterior cingulate gyrus posterior cingulate gyrus anterior cingulate lingual gyrus	38	-56	60	4.08	41	-56	55	3.87	-8	34	25	3.28
			38	-56	60	4.08	41	-56	55	3.87				
Occipital	left	MOG/lateral occipital sulcus anterior calcarine sulcus cerebellum	-8	30	30	4.04	-8	30	30	3.09	-8	34	25	3.28
			-8	30	30	4.04	-8	30	30	3.09				
Occipital	right	MOG/lateral occipital sulcus anterior calcarine sulcus cerebellum	-11	-75	-45	3.24	-22	-68	-30	3.17	15	-82	-5	3.42
			-11	-75	-45	3.24	-22	-68	-30	3.17				
Other	left	anterior thalamus cerebellum	15	4	10	3.43	19	4	15	3.62	15	-79	-5	3.78
			15	4	10	3.43	19	4	15	3.62				
Other	right	anterior thalamus cerebellum	11	-8	10	4.65	11	-8	10	4.65	15	-60	5	3.63
			11	-15	5	3.62	11	-15	5	3.62				

Note—STG, superior temporal gyrus; FG, fusiform gyrus; SMA, supplementary motor area; MFG, middle frontal gyrus; CS, central sulcus; SFS, superior frontal sulcus; PeS, precentral sulcus; PoCG, postcentral gyrus; IPS, intraparietal sulcus; MOG, middle occipital gyrus.

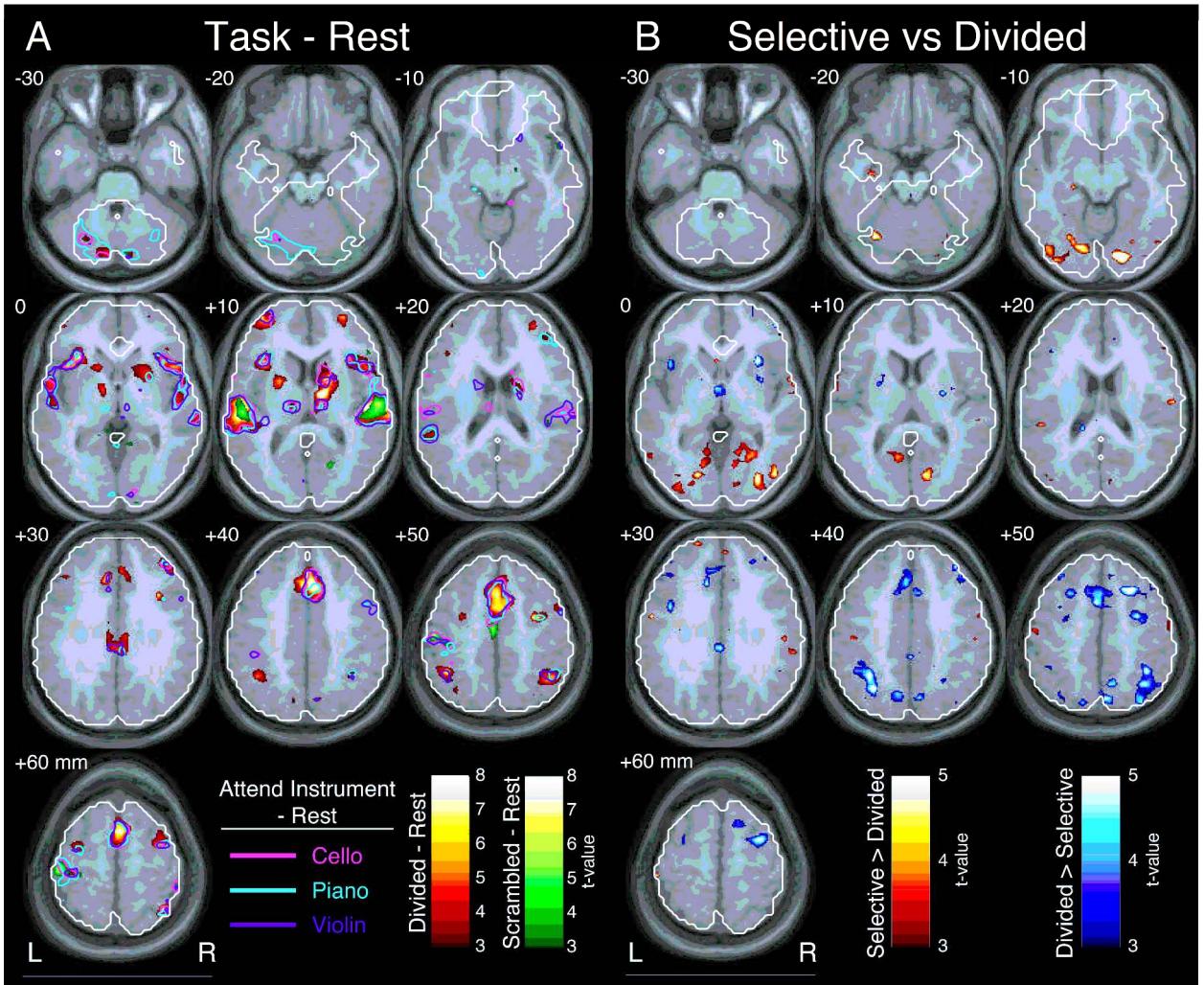


Figure 4. Group ($N = 8$) images showing significant blood oxygen level dependent (BOLD) signal increases ($p < .01$) for two sets of contrasts from Experiment 2, superimposed on the average T1-weighted anatomical image of the subject cohort. (A) Contrast of divided attention (global) listening and rest is depicted with a red gradient. The contrasts of selective listening to each instrument, relative to rest, are denoted with a contour line of different color for each instrument. The responses to an acoustically matched control stimulus are shown in green. (B) The direct comparison of selective and divided attention conditions. Areas with a stronger BOLD response during selective attending, as compared with divided attending, are shown in red, whereas areas with a stronger BOLD response during divided attending are shown in blue.

causes problems for PET/fMRI studies that try to dissociate levels of processing in audition, particularly as more acoustically complex and structured stimuli are employed. The crux of the problem is twofold: matching the acoustical complexity of stimuli across conditions while inferring/confirming the mental state of subjects across conditions in which they hear the same stimulus. Seemingly, the simplest solution to this is to present the identical stimulus in the different conditions and independently vary the task requirements—for example, passive listening versus attentive listening. This type of design has been used to describe mnemonic processing in the context of melodic perception (Zatorre et al., 1994), phonetic and pitch judgments about speech syllables (Zatorre, Evans, Meyer, & Gjedde, 1992),

and auditory selective attention to tones and syllables (Benedict et al., 1998; Jäncke et al., 1999; Tzourio et al., 1997). One argument against using *passive* listening as a control task is that the mental activity associated with it is not well defined and may vary among subjects, particularly as stimulus complexity and naturalness increase. For example, one subject may ignore melodies, speech sounds, or tones and think, instead, about the time he or she is wasting in the scanner, whereas other subjects may attempt to shadow (sing/speak along with) the acoustic material. Despite these problems—which are likely to increase the amount of variance in the data that is unaccounted for—passive-listening conditions have traditionally served as an important intermediate state between resting in the ab-

sence of a stimulus and very directed processing of a stimulus. For this reason, we used resting and passive listening as control tasks in Experiment 1. We should caution, however, that even the passive-listening condition may require a minimum of selective attention for segregating the music from the continuous pinging sound of the EPI pulse sequence. The degree to which separating the music from the pinging requires attentive mechanisms, rather than preattentive automatic mechanisms, according to principles of auditory stream segregation (Bregman, 1990) remains to be determined.

Other studies of auditory attention dispense with acoustically matched control stimuli altogether, instead comparing BOLD signal levels during attention with silent rest blocks (Celsis et al., 1999; Linden et al., 1999; Zatorre et al., 1999) or white noise bursts (Pugh et al., 1996). This approach provides a comprehensive view of auditory processing in a particular task/stimulus condition, although it precludes dissociating attentional processes from those related more specifically to sensory encoding and preattentive acoustic analysis. An additional concern is that the cognitive state of a resting subject is unknown to the experimenter and may influence the activation patterns derived from comparisons of task and rest activations, possibly even obscuring activations of interest. Despite these shortcomings, passive rest is often the common denominator across experiments. As such, it affords a reasonably constant reference against which task-specific activations can be contrasted within experiments, and these contrasts can then be entered into conjunction analyses to investigate the commonality across experiments.

Conjunction Analysis Methods

For the purposes of the conjunction analysis, the listen and the divided conditions were grouped together (labeled *global*) to reflect the holistic, global listening that was suggested or required. The attend and the selective conditions, which collectively required focusing attention on individual instrument streams, were grouped together as the selective conditions. Regions of SPMs (thresholded at $p < .05$) that overlapped in the two experiments are shown in Figure 5 and Table 5 for the global–rest and selective–rest contrasts, respectively. We applied a lower significance criterion than in the individual experiments, since the conjunction analysis was essentially a screen for regions of interest for further studies on the topic of attention in real musical contexts. The likelihood of observing a significant conjunction was 2.5×10^{-3} . We did not perform a global versus selective conjunction analysis, because within each experiment, the global versus selective contrast would reflect the difference between high-level cognitive states that were somewhat different in the two experiments. In other words, in Experiment 2, the contrast would be between two attentionally demanding states in the context of target detection, whereas in Experiment 1, the contrast would be between a demanding selective attention state (not requiring detection of single targets) and a less demanding passive-

listening state. Thus, a direct comparison of the *difference states* in a conjunction analysis would not be readily interpretable. The conjunction maps were projected onto the average T1-weighted anatomical image of the Experiment 1 subjects, and Brodmann's areas were assigned to the activations, using the atlas of Duvernoy (Duvernoy, 1999) and Brodmann's areal descriptions (Brodmann, 1909/1999).

Conjunction Analysis Results and Discussion

Figure 5 shows the results of the conjunction analysis in red. In both experiments, the requirement to focus attention on a single instrument, relative to passive resting (selective–rest), resulted in bilateral activation of the STG, pre-SMA, frontal operculum/rostral insula, superior PCs, IPS, and cerebellum (Figure 5A). In addition, there was right-lateralized activation in the thalamus and the caudate nucleus. A similar pattern was observed for the global–rest conjunction with respect to the STG, pre-SMA, superior PCs, and IPS (Figure 5B).

Also shown in Figure 5 are the activations that were unique to Experiment 1 (green patches) and Experiment 2 (yellow patches). When the activations unique to each experiment are compared with the conjoint activations, both within a contrast and across contrasts, a picture emerges that can be accounted for, tentatively, on the basis of task differences across the two experiments. Two major differences existed between the two experiments. First, Experiment 2 was a target detection task, but Experiment 1 was not. Second, owing to the different tasks, the attentional characteristics were more closely matched between experiments for the selective-listening conditions than for the global-listening conditions. Thus, areas that are very sensitive to attentional load would be expected to show a conjunction in the selective–rest contrast, whereas a conjunction would be less likely for the global–rest contrast, owing to mismatched attentional demands.

Several areas that were activated in both experiments in the selective–rest comparison were activated to a greater extent or exclusively in Experiment 2 for the global–rest comparison. These included the frontal operculum/rostral insula and the IPS bilaterally and the caudate nucleus and the thalamus on the right. In addition, the medial frontal activation extended more ventrally into the anterior cingulate sulcus in Experiment 2, and there was an activation focus in the posterior cingulate. The presence of these activation foci in the global-listening condition of Experiment 2, but not of Experiment 1, is consistent with the fact that the global-listening condition was a demanding divided attention target detection task in Experiment 2 but was a passive-listening task in Experiment 1.

Experiment 1 exhibited unique bilateral activation along the IFS extending caudally into the superior PCs in the selective–rest contrast (Figure 5A, slices +20 to +50). In addition, there was a unique parietal activation in the left supramarginal gyrus. The activation of these areas might be explained by the strategy for focusing attention on a single stream that was suggested to all the subjects in Experiment 1. The suggested strategy was to listen to each

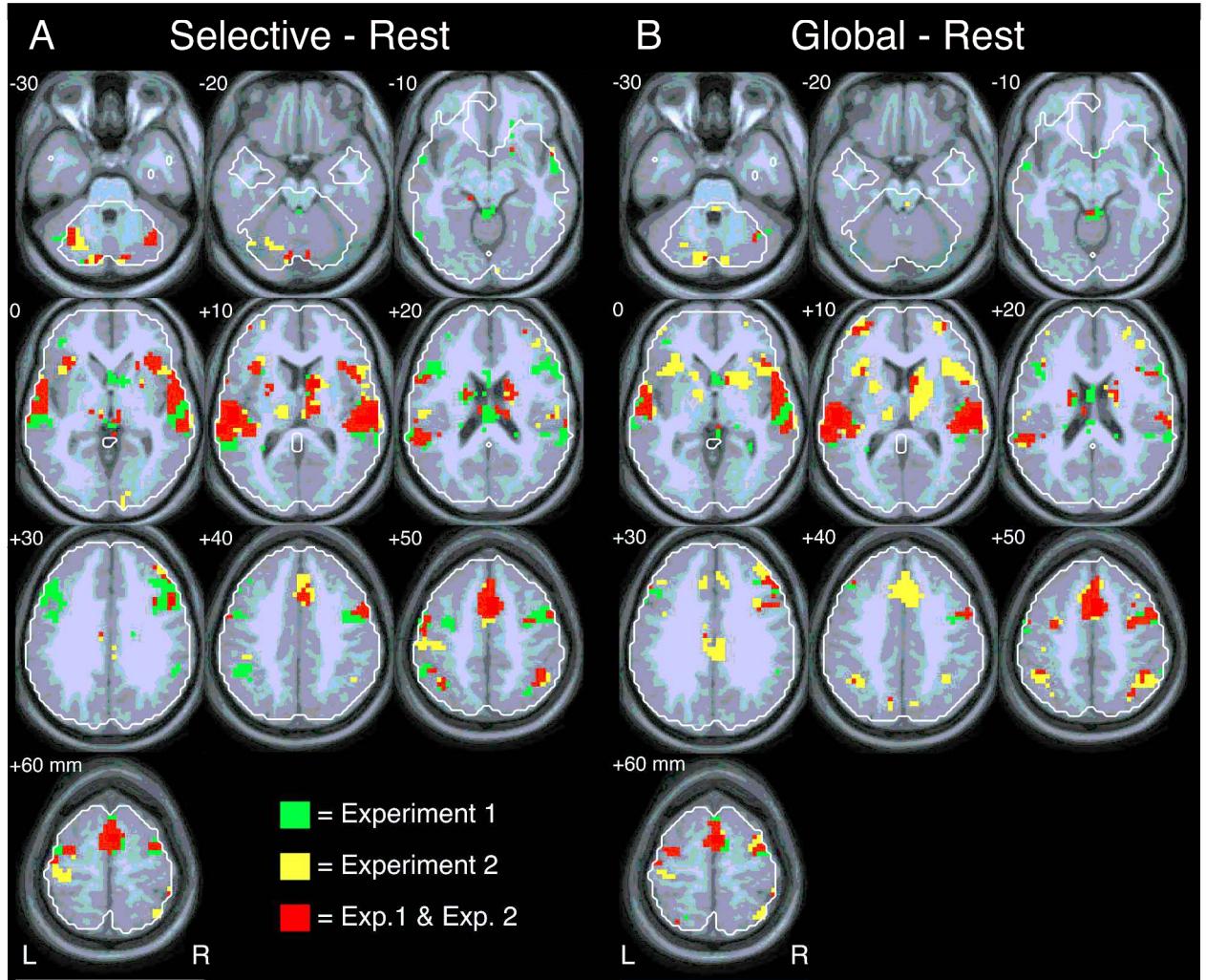


Figure 5. Conjunction analysis showing activation regions common to both experiments. The conjunctions were performed by using the contrasts of global- and selective-listening conditions relative to rest, because rest was the only condition that was identical across experiments. Contrasts from the individual experiments were thresholded at $p < .05$, and these contrast masks were superimposed to identify common voxels (shown in red). Activations for each contrast that were unique to Experiment 1 are shown in green, whereas voxels unique to Experiment 2 are shown in yellow. The white line around the edge of the brain denotes the conjunction of the data inclusion masks from the two experiments.

instrument's part as though one was trying to learn/memorize it. Given that the observed frontal areas are intimately involved in working memory functions (see the General Discussion section), the suggested listening strategy may have facilitated recruitment of these areas during the selective-listening task. Because Experiment 2 was a target detection task, the goal of detecting a deviant within a specified stream was deemed to be sufficient for orienting attention, and the listening strategy was not suggested to the subjects.

GENERAL DISCUSSION

Our experiments were designed with several questions in mind. First, we wanted to identify neural circuits that

are activated during attentive listening to music, as compared with rest, regardless of the form of attentive listening. Second, we wanted to determine whether we could dissociate activations in response to attending to a single stream from more holistic attending. Third, we were curious about how the results of our experiments using real musical stimuli would differ from several studies of auditory attention that have used simpler acoustic stimuli. Similarly, we wondered what the differences might be between attentive listening to real music in a target detection context and attentive listening in a more natural context without the constraints of target detection.

This last point presents a significant dilemma. A hallmark of cognitive psychology research is inference of the cognitive state of a subject from behavioral responses. For

instance, how well a subject is paying attention to the features of a stimulus is usually inferred from his or her detection accuracy and response times to the target features. Our concern is that requiring subjects to make decisions about stimulus features or to detect targets simply in the name of verifying compliance with attentional instructions fundamentally alters the constellations of brain areas that are engaged by the mental states we are interested in. Thus, it becomes almost impossible to study attentional states that occur outside the context of target detection or decision making. These considerations led us to vary the task demands across two experiments and compare the results via a conjunction analysis of the two experiments. In addition to the target detection accuracy data obtained in Experiment 2, ratings were obtained (in both experiments) of the difficulty each subject had orienting his or her attention as instructed to each of the instruments. Although not as compelling as reaction time or accuracy data, the ratings showed that the subjects at least attempted to orient their attention as instructed. Thus, the subjective ratings and the results of the conjunction analysis left us fairly confident that we had obtained fMRI measures of auditory attention to musical stimuli, even outside of a target detection context.

Circuitry Underlying Attention to Complex Musical Stimuli

For both the global and the selective attention conditions, we observed increased BOLD signals in the temporal, frontal, and parietal areas during attentive listening to excerpts of real music. This finding is in agreement with

recent studies that have supported the hypothesis (Zatorre et al., 1999) that a supramodal attention circuit consisting of the temporal, parietal, and frontal areas is recruited during the detection of auditory targets presented in simple acoustic contexts (Benedict et al., 1998; Kiehl, Laurens, Duty, Forster, & Liddle, 2001; Linden et al., 1999; Pugh et al., 1996; Sakai et al., 2000; Tzourio et al., 1997; Zatorre et al., 1999) and musical contexts (Satoh et al., 2001). Most selective attention experiments in audition are designed as target detection tasks in which subjects form a search image of a single target stimulus—for example, a pitch, a syllable, or a location. In contrast to these studies, our subjects were required to maintain attentional focus on a melody carried by a specific timbre. Rather than selecting a single search image from a background of stimuli within an attended stream, the subjects were required to select an entire stream in the presence of multiple streams. Despite these differences, we observed an activation pattern across both experiments that was very similar to that observed during auditory selective attention for acoustic features under dichotic listening to simple FM sweeps and syllable stimuli, relative to detection of noise bursts (Pugh et al., 1996). Common to the study of Pugh et al. and our present set of experiments were activations of the STG, the IPS (bilateral for syllables, right-lateralized for FM sweeps), the bilateral IFS, the right-lateralized PCs, the SMA/pre-SMA, and the bilateral frontal operculum.

Our observation that the set of cortical areas activated by attentive listening to music is very similar to the set that is activated during processing of simpler acoustical stimuli might be seen as problematic, given that we argue

Table 5
Areas of Activation Common to Experiments 1 and 2

Lobe	Hemisphere	Region (Brodmann Areas)	Global–Rest			Selective–Rest		
			x	y	z	x	y	z
Temporal	left	STG (41/42/22)	−55	7	−5	−54	7	−4
			−62	−42	20	−52	−44	26
Frontal	bilateral	STG (41/42/22)	55	7	0	53	16	0
			64	−34	15	64	−34	16
		STS (22)	60	−28	−5	60	−31	−5
		pre-SMA (6)	0	10	54	0	4	55
		frontal operculum (44)				−35	20	5
		IFG, p. operc. (44)				−60	7	20
Frontal	left	superior PCs (6/8)	−31	−5	51	−52	−4	48
		rostral MFG (10)	−27	56	9			
			−41	49	9			
		frontal operculum (44)	44	19	5	35	20	5
		IFS/PCs (6/8)	47	12	30	51	10	33
		superior PCs (6/8)	39	−2	60	37	−2	57
Parietal	left	IPS (7/39)				−42	−48	46
		IPS (7/39)				44	−48	50
Other	right	IPS/angular gyrus (39)	35	−60	45			
		cerebellum				−32	−65	−28
		caudate	10	9	8	11	6	12
		thalamus				11	−17	12
		cerebellum				33	−66	−25

Note—If more than one set of coordinates is given for one hemisphere of an anatomical location, the coordinates denote the extent, rather than the center, of the activation conjunctions. STG, superior temporal gyrus; STS, superior temporal sulcus; SMA, supplementary motor area; IFG, inferior frontal gyrus; PCs, precentral sulcus; MFG, middle frontal gyrus; IFS, inferior frontal sulcus; IPS, intraparietal sulcus.

that it is important to use natural stimuli. In other words, if natural stimuli are somehow superior to simple stimuli, should they not result in activation of brain areas in which simple stimuli are not capable of eliciting a response? If responses to the two classes of stimuli are not different from each other, what have we learned? Our study has been a first step in charting out the neural circuitry that is recruited as listeners guide their attention in real music. To determine conclusively whether there are brain areas specialized for the processing of real music, one would have to titrate the musical complexity of the stimulus set while holding the task and overall acoustical complexity constant. However, our study allows us to conclude that attentive processing in musical contexts engages a set of brain areas that is similar to the set engaged by attentive processing of simpler acoustic contexts. This conclusion was not evident prior to our performing these experiments, since it might have been argued that processing of simple stimuli does not need the same kind of complex network that might be required by rich natural stimuli. In this regard, our results can be seen as a validation of experiments that have used simpler stimuli. We will now relate the functional circuitry we observed for both global and selective attention conditions to that found in relation to other studies of auditory attention and to circuits for general working memory, attention, semantic processing, target detection, and motor imagery functions.

Selective Versus Global Listening

Auditory attention in the STG. Several studies have tried to determine whether attention influences activity in the primary and auditory association areas along the STG. Numerous studies from the human electrophysiological literature suggest that attention influences activity in these areas (Näätänen, 1992), whereas the functional neuroimaging literature is equivocal. Primarily because of heterogeneity across experiments in the stimuli, tasks, and specific statistical comparisons performed, the locus of attentional selection along the STG is uncertain. In several functional neuroimaging studies, it has been observed that there is very little, if any, modulation of the primary and secondary auditory cortical areas of the STG in attend–listen comparisons (Benedict et al., 1998; Tzourio et al., 1997; Zatorre et al., 1992; Zatorre, Halpern, Perry, Meyer, & Evans, 1996) or in comparisons of activation during attention to different stimulus dimensions (Benedict et al., 1998; Platel et al., 1997; Zatorre, Mondor, & Evans, 1999). However, a recent fMRI study focusing on superior temporal lobe activation suggests that although passive listening itself shows increased BOLD signal levels, as compared with rest, both primary and secondary auditory cortices are activated during target detection relative to passive listening (Jäncke et al., 1999).

In both experiments, we found the spatial extent of STG activation to be similar in global- and selective-listening conditions, as compared with rest conditions. We observed more extensive activation bilaterally in response to atten-

tive listening to the musical stimulus than in response to passive listening to the acoustically matched control stimulus in Experiment 2 (Figure 4A). A direct comparison of selective listening with passive global listening in Experiment 1 showed a significantly stronger BOLD signal during selective listening in the left posterior STG/STS (Figure 2B). This result might lead one to infer that this is the temporal lobe region in which stream selection occurs. However, the same area showed no difference between the selective and the divided attention conditions in Experiment 2. These conditions were closely matched in their attentional demands, suggesting that the left posterior STG/STS area observed in Experiment 1 is more sensitive to overall attentional load than to selection of a single stream per se. Our findings do not allow more detailed inferences about mechanisms of attentional selection of auditory streams that are characterized by a variety of features—for example, timbral, temporal, and so forth—that might influence activity patterns within the STG region.

Auditory attention beyond the STG. When we compared selective- and global-listening conditions directly, we observed significant differences in several brain regions. This was true for both experiments, although the activation patterns differed. In Experiment 1, selective listening was associated with the temporo-parieto-frontal circuitry alluded to above and described in more detail below. When the target detection conditions of Experiment 2 were compared directly, selective listening showed greater activity in a number of visual areas in the occipital lobe than in the temporo-parieto-frontal areas that we might have expected, given the results of Experiment 1. However, the divided attention condition in Experiment 2 showed stronger activation of the parietal and medial frontal structures that were more active in the selective attention condition of Experiment 1. When compared with the rest condition, both target detection conditions in Experiment 2 exhibited activation of temporo-parieto-frontal circuitry that was similar to that observed in the selective-attending condition of Experiment 1. This set of contrasts across the two experiments suggests that activation of the temporo-parieto-frontal circuitry may depend more on the degree of attentional load than on selection of single streams. Thus, we were able to delineate a cortical network associated with attentive listening to music, but we did not attain our original objective of identifying areas that are specific for selection of single acoustic streams in the presence of other, simultaneously sounding streams.

There is scant evidence in the literature for where auditory stream selection areas might exist. One other functional neuroimaging study that has compared focused- and divided-listening conditions directly (detection of target syllables in a stream of syllables accompanied by a simultaneous stream of text) found no regions in which rCBF was greater during selective listening (detection of syllables while ignoring text) than during divided listening (detection of syllables while monitoring text for content; Benedict et al., 1998). A direct comparison of di-

vided listening with focussed listening found activation of the left posterior MFG ($x = -42, y = 0, z = 48$). This region is roughly homologous to right-hemisphere activations that we found in the divided versus selective contrast of Experiment 2 (Table 4).

A recent study in which selective listening and global listening in music were compared found the parietal and frontal areas to be more active during selective listening for specific notes in the alto harmony part of four-part harmony compositions than during global listening for minor chords in the same pieces (Satoh et al., 2001). The set of regions they found to be preferentially active during selective listening to the alto part is very similar to the regions we found in Experiment 2 to be significantly more active during divided attending than during selective attending when the conditions were compared directly. This apparent discrepancy might be explained in terms of overall attentional load. Detecting targets in one of three simultaneous streams is more difficult than detecting targets in a single stream. Similarly, listening for a target that is a component note of a chord is more difficult than judging whether or not the overall chord is a minor-sounding chord, because four notes played simultaneously in the same timbre will tend to integrate into a whole from which it is more difficult to pick out individual notes.

Domain-General Activation of Fronto-Parietal Circuitry

The parieto-frontal network, consisting of the IPS, the pre-SMA, the Pcs, and the IFS, that was activated in our tasks in combination with temporal areas is also recruited by a diverse array of cognitive tasks that, at first glance, have very little in common with our stimuli and tasks. A partial listing (see Cabeza & Nyberg, 2000, for a more comprehensive compilation) includes studies of spatial attention (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; LaBar, Gitelman, Parrish, & Mesulam, 1999), temporal attention (Coull, Frith, Buchel, & Nobre, 2000; Coull & Nobre, 1998), and working memory tasks that involve memory for spatial locations, abstract and concrete objects (LaBar et al., 1999; Mecklinger, Bosch, Gruenewald, Bentin, & von Cramon, 2000), verbal and non-verbal material (Jonides et al., 1998; Kelley et al., 1998), the pitch of single notes in melodic contexts (Zatorre et al., 1994), temporal interval patterns (Sakai et al., 1999), and spatiotemporal patterns (Schubotz & von Cramon, 2001). Our results provide further support for the idea that this fronto-parietal network serves domain-general attentional and working memory functions.

In our study, the IPS was activated bilaterally in both experiments, although the extent of activation was greatest in the divided attention condition of Experiment 2 (Figure 4). In the selective attention condition of Experiment 1, we observed activation in left supramarginal and IPS areas, along with bilateral frontal activation along the caudal IFS and the Pcs. This experiment differed from Experiment 2 in that the subjects were simply requested to attend to single instruments, rather than perform a target

detection task. As was mentioned above, we suggested to the subjects that it was easier to focus attention on a single instrument by listening to it as though they were trying to learn the part. A hypothesis for further study is that the left parieto-frontal activation pattern we observed specifically reflects the attempt to encode the melodies.

Across the two experiments, we observed BOLD signal increases during selective and global attention in four frontal regions: the IFS, the frontal operculum, the Pcs, and the SMA/pre-SMA. Cortical regions along the IFS, and the portion of the IFG ventral to the IFS, particularly in the left hemisphere, respond to semantic representations of objects (see the review in Poldrack et al., 1999), selection demands (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Thompson-Schill, D'Esposito, & Kan, 1999), and phonological processing (Poldrack et al., 2001; Poldrack et al., 1999). Interestingly, BOLD signal increases are observed in these same regions during discrimination and detection of simple auditory objects that possess spectral transients (Müller, Kleinhans, & Courchesne, 2001), consistent with the region's role in phonological processing. In the latter study, activation of the dorsal IFG and the ventro-medial IFG (frontal operculum) was bilateral. Across our experiments, the most consistent activation along the extent of the Pcs and the IFS, regardless of attentional condition, was in the right hemisphere. In the global attention conditions, activation in these structures was largely absent on the left, whereas it was present on the left in the selective attention conditions in Experiment 1 (Figure 5). The overall right-hemispheric bias in the Pcs/IFS region may reflect the nonsemantic aspect of music, along the lines of the proposition that nonverbal working memory functions are biased toward the right hemisphere (Buckner, Kelley, & Petersen, 1999; Kelley et al., 1998). Although musical streams are highly structured and bound together by features along multiple dimensions—that is, timbral, rhythmic, and melodic—they possess no obvious semantic codes that would be represented in the left hemisphere. Nonetheless, the bilateral activation of the premotor/IFS region by our tasks and stimuli suggests that polyphonic music engages similar operations that integrate attention, working memory, and semantics in these regions of the prefrontal cortex. For example, the suggestion that the dorsal aspect of the left IFG selects semantic information from among competing alternatives (Thompson-Schill et al., 1997) is similar to the idea that auditory streams can be segregated and auditory targets detected in the presence of the competing streams.

The Pcs and the pre-SMA are generally implicated in organizing action plans and imagination of movements (see Grezes & Decety, 2001), including those associated with attentional and working memory tasks (Mecklinger et al., 2000; Sakai et al., 2000). These areas are also implicated in memory for temporal patterns (Sakai et al., 1999) and endogenous shifts of attention to events in time (Coull et al., 2000; Coull & Nobre, 1998), both of which are important elements of music cognition. Right-hemisphere sites appear to be recruited more strongly in the retention

of complicated timing patterns—for example, noninteger ratio rhythms (Sakai et al., 1999). Specifically with regard to music, the SMA/pre-SMA is active in a variety of tasks using melodic stimuli and may reflect processes associated with internal singing, even if there are no words associated with the musical excerpt (Halpern & Zatorre, 1999; Perry et al., 1999; Zatorre et al., 1994; Zatorre et al., 1996).

The role of the pre-SMA/SMA and premotor circuits in action planning, imagined movements, and attention to temporal patterns is of particular importance to understanding the effect of music on the brain. Music might be thought of as a collage of well-defined acoustic events that are patterned in time. Inherent to music is the concept of motion, both in terms of spectral information (melodic and harmonic motion) and temporal information (rhythm). Music is used to structure movements, be they related to playing an instrument, dancing, or simply tapping one's fingers or toes. Thus, when a subject's focus is on extracting the spectrotemporal properties of a single instrument or on appreciating the combined pattern arising from the parts of all the instruments playing, premotor circuitry may be recruited as a part of the perceptual processing. This view is consistent with proposals that the fronto-parietal circuitry underlies the conjunction of perceptual and motor representations (Grezes & Decety, 2001; Mecklinger et al., 2000; Schubotz & von Cramon, 2001), is responsible for shifting attentional focus over space, time, or cognitive domain (LaBar et al., 1999), or integrates action selection and timing (Sakai et al., 2000). The premotor and parietal sites appear to underlie a working memory system that is coupled to the IFG (also referred to as the inferior prefrontal cortex), which, in the left hemisphere, supports semantic working memory (Gabrieli, Poldrack, & Desmond, 1998) and, bilaterally, may support syntactic processing of musical sequences (Maess, Koelsch, Gunter, & Friederici, 2001; Tillmann, Janata, & Bharucha, 2002).

Conclusions

Both our results and the study of Satoh et al. (2001) suggest that what one takes as evidence for the circuits underlying the selection of single musical streams depends on the tasks being compared and the overall experimental context. Target detection tasks are immensely attractive because they provide a means of objectively verifying behavioral compliance. Unfortunately, they may establish mental states that are not necessarily characteristic of how the brain operates in more natural attentive situations. Instructing subjects to orient their attention in various ways without demanding that they make decisions or detect targets has the advantage of bringing one closer to a natural situation but has the distinct disadvantage of forcing the experimenter to rely on the subjects' self-reports as verification of their task performance.

The conjunction analysis helped us to identify areas of the brain that are activated when subjects listen to music under different attentional conditions and listening strategies. These areas include the STG extending rostrally along the PP in both hemispheres, the right superior PCs,

and the pre-SMA. The comparison of the activations common to both experiments with the activations that were unique to each experiment further suggested that listening strategies bias activation of networks involving further parietal, frontal, striatal, thalamic, and cerebellar areas. Additional experiments will be necessary to determine how stimulus and task factors bias responses to real musical contexts within domain-general cortical networks and whether there are music-selective areas in the brain.

REFERENCES

ALAIN, C., & WOODS, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics*, **56**, 501-516.

ALHO, K. (1992). Selective attention in auditory processing as reflected by event-related brain potentials. *Psychophysiology*, **29**, 247-263.

BENEDICT, R. H. B., LOCKWOOD, A. H., SHUCARD, J. L., SHUCARD, D. W., WACK, D., & MURPHY, B. W. (1998). Functional neuroimaging of attention in the auditory modality. *NeuroReport*, **9**, 121-126.

BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

BREGMAN, A. S., & CAMPBELL, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.

BRENOWITZ, E. A., MARGOLIASH, D., & NORDEEN, K. W. (1997). An introduction to birdsong and the avian song system. *Journal of Neurobiology*, **33**, 495-500.

BROCHARD, R., DRAKE, C., BOTTE, M. C., & MCADAMS, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1742-1759.

BRODMANN, K. (1999). *Brodmann's localisation in the cerebral cortex* (L. J. Garey, Trans.). River Edge, NJ: World. (Original work published 1909)

BUCKNER, R. L., KELLEY, W. H., & PETERSEN, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience*, **2**, 311-314.

CABEZA, R., & NYBERG, L. (2000). Imaging cognition: II. An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, **12**, 1-47.

CELSIS, P., BOULANOUAR, K., DOYON, B., RANJAVA, J. P., BERRY, I., NESPOULOUS, J. L., & CHOLLET, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, **9**, 135-144.

CHERRY, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.

COHEN, J. D., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments, & Computers*, **25**, 257-271.

COPPOLA, D. M., WHITE, L. E., FITZPATRICK, D., & PURVES, D. (1998). Unequal representation of cardinal and oblique contours in ferret visual cortex. *Proceedings of the National Academy of Sciences*, **95**, 2621-2623.

CORBETTA, M., KINCADE, J. M., OLLINGER, J. M., McAVOY, M. P., & SHULMAN, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, **3**, 292-297. [Published erratum appears in *Nature Neuroscience*, 2000, **3**, 521.]

COULL, J. T., FRITH, C. D., BUCHEL, C., & NOBRE, A. C. (2000). Orienting attention in time: Behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia*, **38**, 808-819.

COULL, J. T., & NOBRE, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, **18**, 7426-7435.

DOWLING, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, **5**, 322-337.

DUVERNOY, H. M. (1999). *The human brain: Surface, three-dimensional*

sectional anatomy with MRI, and blood supply (2nd ed.). Vienna: Springer-Verlag.

FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. P., FRITH, C. D., & FRACKOWIAK, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**, 189-210.

GABRIELI, J. D. E., POLDRACK, R. A., & DESMOND, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences*, **95**, 906-913.

GREZES, J., & DECETY, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, **12**, 1-19.

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.

HAXBY, J. V., HOFFMAN, E. A., & GOBBINI, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, **4**, 223-233.

HILLYARD, S. A., HINK, R. F., SCHWENT, V. L., & PICTON, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, **182**, 177-180.

HINK, R. F., HILLYARD, S. A., & BENSON, P. J. (1978). Event-related brain potentials and selective attention to acoustic and phonetic cues. *Biological Psychology*, **6**, 1-16.

IVERSON, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 751-763.

JÄNCKE, L., MIRZAZADE, S., & SHAH, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neuroscience Letters*, **266**, 125-128.

JONES, M. R. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 69-112). Oxford: Oxford University Press.

JONIDES, J., SCHUMACHER, E. H., SMITH, E. E., KOEPPE, R. A., AWH, E., REUTER-LORENZ, P. A., MARSHUETZ, C., & WILLIS, C. R. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience*, **18**, 5026-5034.

KELLEY, W. M., MIEZIN, F. M., McDERMOTT, K. B., BUCKNER, R. L., RAICHLE, M. E., COHEN, N. J., OLLINGER, J. M., AKBUDAK, E., CONNURO, T. E., SNYDER, A. Z., & PETERSEN, S. E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron*, **20**, 927-936.

KIEHL, K. A., LAURENS, K. R., DUTY, T. L., FORSTER, B. B., & LIDDLE, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: An event-related fMRI study. *Psychophysiology*, **38**, 133-142.

KONISHI, M. (1985). Birdsong: From behavior to neuron. *Annual Review of Neuroscience*, **8**, 125-170.

KRUMHANSL, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds.), *Structure and perception of electro-acoustic sound and music* (pp. 43-54). Amsterdam: Excerpta Medica.

LABAR, K. S., GITELMAN, D. R., PARRISH, T. B., & MESULAM, M. M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional MRI comparison within subjects. *NeuroImage*, **10**, 695-704.

LINDEN, D. E. J., PRVULOVIC, D., FORMISANO, E., VOLLINGER, M., ZANELLA, F. E., GOEBEL, R., & DIERKS, T. (1999). The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks. *Cerebral Cortex*, **9**, 815-823.

MAES, F., COLLIGNON, A., VANDERMEULEN, D., MARCHAL, G., & SUETENS, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, **16**, 187-198.

MAESS, B., KOELSCH, S., GUNTER, T. C., & FRIEDERICI, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, **4**, 540-545.

MCADAMS, S., WINSBERG, S., DONNADIEU, S., DESOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research-Psychologische Forschung*, **58**, 177-192.

MECKLINGER, A., BOSCH, V., GRUENEWALD, C., BENTIN, S., & VON CRAMON, D. Y. (2000). What have Klingon letters and faces in common? An fMRI study on content-specific working memory systems. *Human Brain Mapping*, **11**, 146-161.

MÜLLER, R. A., KLEINHANS, N., & COURCHESNE, E. (2001). Broca's area and the discrimination of frequency transitions: A functional MRI study. *Brain & Language*, **76**, 70-76.

NÄÄTÄNEN, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

PERRY, D. W., ZATORRE, R. J., PETRIDES, M., ALIVISATOS, B., MEYER, E., & EVANS, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, **10**, 3979-3984.

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.

POLDRACK, R. A., TEMPLE, E., PROTOPAPAS, A., NAGARAJAN, S., TALAL, P., MERZENICH, M., & GABRIELI, J. D. E. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, **13**, 687-697.

POLDRACK, R. A., WAGNER, A. D., PRULL, M. W., DESMOND, J. E., GLOVER, G. H., & GABRIELI, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage*, **10**, 15-35.

PUGH, K. R., SHAYWITZ, B. A., FULBRIGHT, R. K., BYRD, D., SKUDLARSKI, P., KATZ, L., CONSTABLE, R. T., FLETCHER, J., LACADIE, C., MARCHIONE, K., & GORE, J. C. (1996). Auditory selective attention: An fMRI investigation. *NeuroImage*, **4**, 159-173.

SAKAI, K., HIKOSAKA, O., MIYAUCHI, S., TAKINO, R., TAMADA, T., IWATA, N. K., & NIELSEN, M. (1999). Neural representation of a rhythm depends on its interval ratio. *Journal of Neuroscience*, **19**, 10074-10081.

SAKAI, K., HIKOSAKA, O., TAKINO, R., MIYAUCHI, S., NIELSEN, M., & TAMADA, T. (2000). What and when: Parallel and convergent processing in motor control. *Journal of Neuroscience*, **20**, 2691-2700.

SAMS, M., AULANKO, R., AALTONEN, O., & NÄÄTÄNEN, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience*, **2**, 344-357.

SATOH, M., TAKEDA, K., NAGATA, K., HATAZAWA, J., & KUZUHARA, S. (2001). Activated brain regions in musicians during an ensemble: A PET study. *Brain Research: Cognitive Brain Research*, **12**, 101-108.

SCHUBOTZ, R. I., & VON CRAMON, D. Y. (2001). Functional organization of the lateral premotor cortex: fMRI reveals different regions activated by anticipation of object properties, location and speed. *Brain Research: Cognitive Brain Research*, **11**, 97-112.

STANLEY, G. B., LI, F. F., & DAN, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *Journal of Neuroscience*, **19**, 8036-8042.

SZYMANSKI, M. D., YUND, E. W., & WOODS, D. L. (1999a). Human brain specialization for phonetic attention. *NeuroReport*, **10**, 1605-1608.

SZYMANSKI, M. D., YUND, E. W., & WOODS, D. L. (1999b). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *Journal of the Acoustical Society of America*, **106**, 3492-3505.

THOMPSON-SCHILL, S. L., D'ESPOSITO, M., AGUIRRE, G. K., & FARAH, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, **94**, 14792-14797.

THOMPSON-SCHILL, S. L., D'ESPOSITO, M., & KAN, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, **23**, 513-522.

TILLMANN, B., JANATA, P., & BHARUCHA, J. J. (2002). Activation of the inferior frontal cortex in musical priming. Manuscript submitted for publication.

TREVES, A., PANZERI, S., ROLLS, E. T., BOOTH, M., & WAKEMAN, E. A. (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Computation*, **11**, 601-631.

TZOURIO, N., ELMASIOU, F., CRIVELLO, F., JOLIOT, M., RENAULT, B., & MAZOYER, B. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage*, **5**, 63-77.

VINJE, W. E., & GALLANT, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, **287**, 1273-1276.

WOLDORFF, M. G., GALEN, C. C., HAMPSON, S. A., HILLYARD, S. A., PANTEV, C., SOBEL, D., & BLOOM, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, **90**, 8722-8726.

WOLDORFF, M. G., HACKLEY, S. A., & HILLYARD, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones [see comments]. *Psychophysiology*, **28**, 30-42.

WOLDORFF, M. G., & HILLYARD, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography & Clinical Neurophysiology*, **79**, 170-191.

WOODS, D. L., & ALAIN, C. (2001). Conjoining three auditory features: An event-related brain potential study. *Journal of Cognitive Neuroscience*, **13**, 492-509.

ZATORRE, R. J., EVANS, A. C., & MEYER, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, **14**, 1908-1919.

ZATORRE, R. J., EVANS, A. C., MEYER, E., & GIEDDE, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, **256**, 846-849.

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.

ZATORRE, R. J., MONDOR, T. A., & EVANS, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage*, **10**, 544-554.

(Manuscript received June 29, 2001;
revision accepted for publication April 3, 2002.)