Behavioral/Systems/Cognitive

Brain Bases for Auditory Stimulus-Driven Figure–Ground Segregation

Sundeep Teki,[1,2]* Maria Chait,[3]* Sukhbinder Kumar,[1,2] Katharina von Kriegstein,[1,4] and Timothy D. Griffiths[1,2]

[1]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom, [2]Newcastle Auditory Group, Medical School, Newcastle University, Newcastle-upon-Tyne NE2 4HH, United Kingdom, [3]UCL Ear Institute, University College London, London WC1X 8EE, United Kingdom, and [4]Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

Auditory figure–ground segregation, listeners' ability to selectively hear out a sound of interest from a background of competing sounds, is a fundamental aspect of scene analysis. In contrast to the disordered acoustic environment we experience during everyday listening, most studies of auditory segregation have used relatively simple, temporally regular signals. We developed a new figure–ground stimulus that incorporates stochastic variation of the figure and background that captures the rich spectrotemporal complexity of natural acoustic scenes. Figure and background signals overlap in spectrotemporal space, but vary in the statistics of fluctuation, such that the only way to extract the figure is by integrating the patterns over time and frequency. Our behavioral results demonstrate that human listeners are remarkably sensitive to the appearance of such figures.

In a functional magnetic resonance imaging experiment, aimed at investigating preattentive, stimulus-driven, auditory segregation mechanisms, naive subjects listened to these stimuli while performing an irrelevant task. Results demonstrate significant activations in the intraparietal sulcus (IPS) and the superior temporal sulcus related to bottom-up, stimulus-driven figure–ground decomposition. We did not observe any significant activation in the primary auditory cortex. Our results support a role for automatic, bottom-up mechanisms in the IPS in mediating stimulus-driven, auditory figure–ground segregation, which is consistent with accumulating evidence implicating the IPS in structuring sensory input and perceptual organization.

Introduction

Auditory figure–ground segregation—listeners' ability to extract a particular sound from a background of other simultaneous sounds—is a fundamental aspect of scene analysis. Segregation involves several processes: grouping of simultaneous figure components from across the spectral array (Micheyl and Oxenham, 2010), grouping of figure components over time (Moore and Gockel, 2002), and separation of grouped components from the rest of the acoustic scene (de Cheveigné, 2001).

Investigations of the brain bases for these aspects of scene analysis in humans and animal models have identified activation patterns that correlate with an integrated versus segregated percept in a distributed network, extending from the auditory periphery (Pressnitzer et al., 2008) to thalamus (Kondo and Kashino, 2009), primary auditory cortex (Fishman et al., 2001; Micheyl et al., 2005; Bidet-Caulet et al., 2007; Wilson et al., 2007), nonprimary auditory areas (Gutschalk et al., 2005; Alain, 2007; Schadwinkel and Gutschalk, 2010), and areas outside auditory cortex [the intraparietal sulcus (IPS) (Cusack, 2005)].

A limiting factor in understanding the computations occurring at these different levels and relating existing experimental results to listeners' experience in natural environments is that the stimuli used thus far have been rather basic, lacking the spectrotemporal complexity of natural sounds. Indeed, in contrast to the disordered acoustic environment that we face during everyday listening, most studies of segregation have used relatively simple stimuli consisting of sequentially presented, regularly alternating tones (Shamma and Micheyl, 2010) or static harmonic sounds (Alain, 2007).

Here, we developed a new stimulus—"stochastic figure–ground" (SFG stimulus) [conceptually extending Kidd et al. (1994) and Micheyl et al. (2007a)]—that incorporates stochastic variation of the signal components in frequency–time space that is not a feature of the predictable sequences used in previous work. Stimuli (see Fig. 1) consist of a sequence of chords containing a random set of pure-tone components that are not harmonically related. Occasionally, a subset of tonal components repeat in frequency over several consecutive chords, resulting in a spontaneous percept of a "figure" (a stream of constant chords) popping out of a background of varying chords (see supplemental Audio 1, available at www.jneurosci.org as supplemental material). Importantly, this stimulus is not confounded by figure and background signals that differ in low-level acoustic attributes, or by the use of a spectral "protective region" around the figure (Gutschalk et al., 2008; Elhilali et al., 2009). Here, at each point in

Received July 21, 2010; revised Aug. 25, 2010; accepted Oct. 24, 2010.

This work is supported by Wellcome Trust Grant WT061136MA awarded to T.D.G. S.T. is supported by the same Wellcome Trust grant. M.C. is supported by a Deafness Research UK Fellowship. K.v.K. is funded by the Max Planck Society. We are grateful to the Radiology staff at the Wellcome Trust Centre for Neuroimaging for their excellent technical support.

*S.T. and M.C. contributed equally to this work.

Correspondence should be addressed to either of the following: Sundeep Teki, Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK, E-mail: s.teki@fil.ion.ucl.ac.uk; or Maria Chait, UCL Ear Institute, University College London, London WC1X 8EE, UK, E-mail: m.chait@ucl.ac.uk.

DOI:10.1523/JNEUROSCI.3788-10.2011

 0270-6474/11/310164-08$15.00/0

time, the figure and background are indistinguishable and the only way to extract the figure is by integrating over time (over consecutive chords) and frequency (identifying the components that change together). Behavioral results (see below) demonstrate that listeners are remarkably sensitive to the appearance of such figures.

We used functional magnetic resonance imaging (fMRI) to study the early, stimulus-driven mechanisms involved in parsing the SFG signals. Figure salience was systematically varied by independently manipulating the number of repeating components and the number of repeats, allowing us to investigate the neural bases of the emergence of an auditory object from a stochastic background as occurs during the automatic parsing of natural acoustic scenes.

Materials and Methods

Psychophysical experiment

Participants

Ten paid subjects (5 female; mean age = 29.2 years) participated in the experiment. All reported normal hearing and had no history of neurological disorders. Experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

Stimuli

We developed a new stimulus (SFG stimulus) [as an extension of Kidd et al. (1994) and Micheyl et al. (2007a)] to model naturally complex situations characterized by a figure and background that overlap in feature space and are only distinguishable by their fluctuation statistics. Contrary to previously used signals, the spectrotemporal properties of the figure vary from trial to trial and without any spectral gap between the figure and the background.

Figure 1 presents examples of the SFG stimuli. Each stimulus consisted of a sequence of random chords, each 50 ms in duration with 0 ms interchord interval, presented for a total duration of 2000 ms (40 consecutive chords). Each chord contained a number (randomly distributed between 5 and 15) of pure tone components. Component frequencies were randomly drawn from a set of 129 values equally spaced on a logarithmic scale between 179 and 7246 Hz. The onset and offset of each chord were shaped by a 10 ms raised-cosine ramp. In half of these stimuli, several consecutive chords included components of the same frequency. In other words, a sequence of repeating tones occurred within the otherwise random background (Fig. 1*B*). When the number of consecutive chords over which tones are repeated is sufficiently large, the resulting percept is that of a "figure" (a stream of constant chords) that readily pops out of a background of randomly varying chords (see supplemental Audio 1, available at www.jneurosci.org as supplemental material, for an example of an SFG stimulus with a 2-s-long figure). In the present experiment, we used very short "figures" whose duration was determined on the basis of pilot experiments to be just sufficient for detection (see supplemental Audio 2 and 3, available at www.jneurosci.org as supplemental material). They are heard as a brief warble amid the ongoing background and require some training to be detected. We continue to refer to these signals as "figure–ground" under the assumption that the brain mechanisms involved in their processing also contribute to the perceptual pop-out of the longer-figure stimuli.

The number of consecutive chords over which tones were repeated (2–7) and the number of repeated tonal components (1, 2, 4, 6, or 8) were varied as parameters. We refer to these as the "duration" and "coherence" of the figure, respectively. The onset of the figure was jittered between 15 and 20 chords (750–1000 ms) after stimulus onset. To eliminate the confound that the appearance of the figure results in fewer background (varying) components, the appearance of the figure was realized by first generating the random background and then adding additional, repeating components to the relevant chords. To avoid the problem that the interval containing the figure might, on average, also contain more frequency components, and to prevent listeners from relying on this feature in performing the figure detection task, the remaining 50% of the stimuli

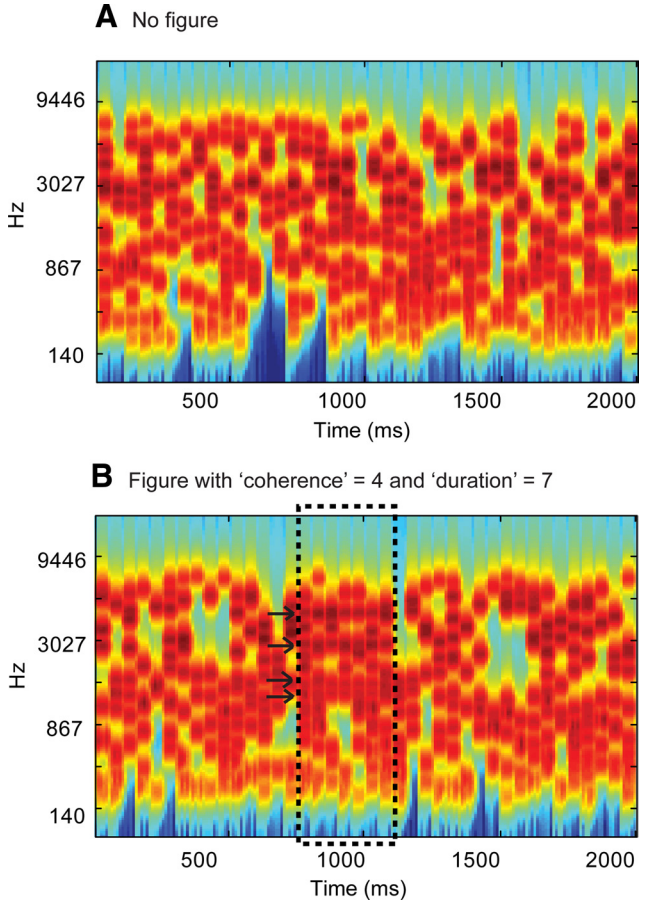

Figure 1. Examples of the SFG stimuli used. ***A***, Signals consisted of a sequence of 50-ms-long chords containing a random set of pure tone components. ***B***, In 50% of the signals, a subset of tonal components repeated in frequency over several consecutive chords, resulting in the percept of a "figure" popping out of the random noise. The figure emerged between 15 and 20 chords (750–1000 ms) after onset. The number of repeated components (the "coherence" of the figure) and the number of consecutive chords over which they were repeated (the "duration" of the figure) were varied as parameters. The plots represent auditory spectrograms, generated with a filter bank of 1/ERB (equivalent rectangular bandwidth) wide channels (Moore and Glasberg, 1983) equally spaced on a scale of ERB-rate. Channels are smoothed to obtain a temporal resolution similar to the equivalent rectangular duration (Plack and Moore, 1990).

(those containing no figure) also included additional (1, 2, 4, 6, or 8) tonal components, which were added over a variable number (2–7) of consecutive chords at the same time as when a figure would have appeared (between 15 and 20 chords after onset). But these additional components changed from chord to chord and did not form a coherent figure.

All stimuli were created online using MATLAB 7.5 software (The MathWorks) at a sampling rate of 44.1 kHz and 16 bit resolution. Sounds were delivered to the subjects' ears with Sennheiser HD555 headphones and presented at a comfortable listening level of 60–70 dB SPL (self adjusted by each listener). Presentation of the stimuli was controlled using the Cogent software (http://www.vislab.ucl.ac.uk/cogent.php). The different stimulus conditions were interleaved randomly with an interstimulus interval (ISI) randomly distributed between 500 and 2000 ms. Overall, listeners heard 20 repetitions of each of the sixty different stimulus types [5 coherence levels × 6 duration levels × 2 conditions (figure present or absent)].

Procedure

The experiment lasted ~1.5 h. Subjects were tested in an acoustically shielded room (IAC triple-walled sound-attenuating booth). They were instructed to look at a fixation cross, presented on a computer screen,

while performing a figure-detection task in which they were required to press a keyboard button as soon as they detected a figure popping out of the random tonal noise (50% of the signals). The actual experiment was preceded by a 15 min practice session during which listeners performed the task with feedback. No feedback was provided during the main experiment. The experimental session was divided into runs of ~10 min each. Listeners were allowed a short rest between runs.

Functional imaging experiment

Participants

Fourteen paid subjects (9 female; mean age = 27.4 years) with normal hearing and no history of neurological disorders participated in the experiment. None of these subjects participated in the psychophysics study. Experimental procedures were approved by the Institute of Neurology Ethics Committee (London, UK), and written informed consent was obtained from each participant. The data for one subject were excluded from analysis due to a technical problem. All listeners completed the passive listening block. A subset of seven subjects (3 female; mean age = 28.8 years) also subsequently completed an "active detection" block to assess performance on the figure-detection task in the scanner.

Stimuli

Main (passive listening) block. A key feature of the present experimental design is the brief figure duration. Whereas most previous studies used relatively long, ongoing figure–ground stimuli and, in many cases, instructed listeners to actively follow one of the components (Scheich et al., 1998; Cusack, 2005; Gutschalk et al., 2005; Wilson et al., 2007; Elhilali et al., 2009), in this imaging experiment naive listeners were presented with very short figure stimuli, embedded in an ongoing random background. Figure duration (a maximum of 6 repeating chords—300 ms) was determined by psychophysical assessment of the minimum number of repeating chords required for reliable detection. With such a design, we aimed to specifically tap the bottom-up segregation mechanisms rather than subsequent processes related to selectively attending to the figure over prolonged periods.

The stimuli were created in the same way as the psychophysical stimuli (see above) with the following differences: the results of the psychophysics study (see Fig. 2) identified two particular parameters as potentially informative to study the underlying brain mechanisms because performance on those conditions spans the range from nondetectable to detectable: (1) fixed coherence with four components and varied duration and (2) fixed duration of four chords and varied coherence. The stimulus set in the fMRI experiment therefore contained signals with a fixed coherence level of four components with five duration levels (2–6) and signals with a fixed duration level of four components with five coherence levels (1, 2, 4, 6, 8), resulting in nine different stimulus conditions. Due to considerations related to BOLD response dynamics, and the need for a larger interval between events of interest, the duration of the signals was increased (relative to the psychophysics study) to 2750 ms (55 chords), with the figure appearing between 1250 and 1500 ms (25–30 chords) after onset. Sixty-six percent of the signals contained a figure. Overall, listeners heard 40 repetitions of each stimulus type. Additionally, the stimulus set also contained a proportion (15%) of "decoy" stimuli consisting of 200 ms wide-band noise bursts (ramped on and off with 10 ms cosine-squared ramps), and interspersed randomly between the main stimuli.

To avoid effects of transition between silence and sound, and to allow for a straightforward evaluation of brain responses to the figure as opposed to the ongoing tonal background, all stimuli were presented in direct succession with no silent intervals. The resulting continuous stimulus consisted of an ongoing tonal background noise with occasional, randomly occurring figures. This ongoing signal was intermittently interrupted by brief noise bursts to which subjects were instructed to respond. Stimuli were presented via NordicNeuroLab electrostatic headphones at an rms SPL of 85–90 dB.

Active detection block. The active detection block was used to confirm that listeners perform similarly to the subjects in the psychophysical study, and are able to hear out the figures in the presence of the MRI scanner noise. Signals were identical to those in the psychophysical experiment with the following differences: as in the passive listening block above, we used signals with a fixed coherence level of four components and five duration levels (2–6) and signals with a fixed duration level of four components with five coherence levels (1, 2, 4, 6, and 8). Overall, listeners heard eight repetitions of each stimulus condition. The order of presentation of different stimulus conditions was randomized with an ISI between 500 and 1250 ms. After every eighth stimulus, the ISI was increased to 12 s (to allow analysis of the sound vs silence contrast in the fMRI activation).

Procedure

The experiment lasted ~2 h and consisted of a "passive listening" block followed by an "active figure-detection" block. Each block was divided into three runs of ~10 min. Participants completed both blocks without exiting the scanner; they were allowed a short rest between runs but were required to stay still.

In the "passive listening" block, the subjects, who were naive to the stimulus structure and aims of the experiment, were instructed to look at a fixation cross and respond as fast as possible (by pressing a response button held in the right hand) to the noise bursts (decoy stimuli) appearing within the continuous stream of the tonal background. In the "active detection" block, subjects were instructed to perform a figure-detection task by pressing the response button as soon as they detected a figure popping out of the random tonal noise (50% of the signals). Crucially, this task was explained to the participants only immediately before the "active detection" block, to ensure that they were naive to the existence of figures during the "passive listening" block. Indeed, pilot experiments suggested that while the figures are readily detectable after a short practice, naive listeners performing the decoy task remained unaware of their presence.

Before beginning the task, subjects completed a short practice session (~10 min) in quiet (but while still lying in the MRI scanner) with feedback. Time constraints prevented us from carrying out a longer practice session. To facilitate learning, feedback was also provided during the session proper.

Image acquisition

Gradient-weighted echo planar images (EPI) were acquired on a 3 Tesla Siemens Allegra MRI scanner using a continuous imaging paradigm with the following parameters: 42 contiguous slices per volume; time to repeat (TR): 2520 ms; time to echo (TE): 30 ms; flip angle α: 90°; matrix size: 64 × 72; slice thickness: 2 mm with 1 mm gap between slices; echo spacing: 330 μs; in-plane resolution: 3.0 × 3.0 mm^2. Subjects completed three scanning sessions resulting in a total of 510 volumes. To correct for geometric distortions in the EPI due to magnetic field variations (Hutton et al., 2002), field maps were acquired for each subject with a double-echo gradient echo field map sequence (short TE = 10.00 ms and long TE = 12.46 ms). A structural T1-weighted scan was also acquired for each subject after the functional scan (Deichmann et al., 2004).

Image analysis

Imaging data were analyzed using Statistical Parametric Mapping software (SPM8; Wellcome Trust Centre for Neuroimaging). The first two volumes were discarded to control for saturation effects. The remaining volumes were realigned to the first volume and unwarped using the field map parameters. The realigned images were spatially normalized to stereotactic space (Friston et al., 1995a) and smoothed by an isotropic Gaussian kernel of 5 mm full-width at half-maximum.

Statistical analysis was conducted using the general linear model (Friston et al., 1995b). Onsets of trials with fixed coherence and fixed duration were orthogonalized and parametrically modulated by different levels of duration and coherence respectively. These two conditions were modeled as effects of interest and convolved with a hemodynamic boxcar response function. A high-pass filter with a cutoff frequency of 1/128 Hz was applied to remove low-frequency variations in the BOLD signal.

A whole-brain random-effects model was implemented to account for within-subject variance (Penny and Holmes, 2004). Each individual subject's first-level contrast images were entered into second-level *t* tests for the primary contrasts of interest—"effect of duration" and "effect of

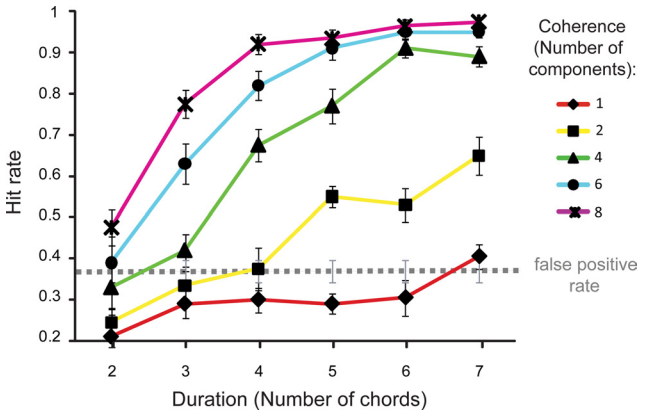

Figure 2. Results of the psychophysics experiment. Hit rate is shown as a function of figure coherence and duration. The dashed line marks the average false-positive rate. Error bars represent SE.

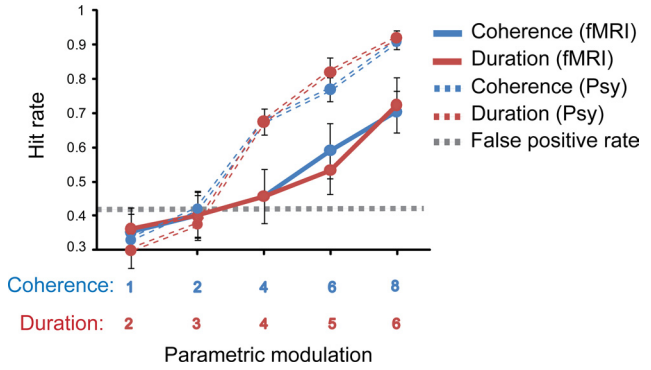

Figure 3. Comparison of behavioral performance in the psychophysics and functional imaging experiments. Behavioral performance on the figure detection task obtained in the scanner with continuous image acquisition (solid lines) presented along with data from the same stimuli obtained in quiet (dashed lines; see psychophysical study, Fig. 2). Hit rate is shown as a function of fixed coherence (4 components) and increasing duration (in red) and as a function of fixed duration (4 chords) and increasing coherence (in blue). The dashed line represents the mean false-positive rate. Error bars represent SE.

coherence." Functional results are overlaid onto the group-average T1-weighted structural scans.

Results

Psychophysics

The results of the psychophysics experiment are presented in Figure 2. The data demonstrate that listeners are remarkably sensitive to the appearance of figures in our stimuli. For coherence levels of six and eight components, ceiling performance is reached for as few as six repeating chords (300 ms). With eight coherent frequencies (coherence = 8), two repeating chords (100 ms) are sufficient for listeners' performance to emerge above the "false-positive floor" (hit rate is significantly higher than the false-positive rate, $p = 0.026$; the somewhat elevated false-positive rate likely stems from the instructions, which emphasized speed, and the fact that listeners received little training). For coherence levels of four and six components, listeners can perform the task with as few as three repeating chords (150 ms; $p < 0.001$, $p = 0.07$, respectively); with two repeating frequency components (coherence = 2), listeners require five repeating chords (250 ms) to extract the figure ($p < 0.001$).

The fact that listeners can extract the figure so efficiently suggests that the auditory system possesses mechanisms that are sensitive to such cross-frequency and cross-time correlations. The purpose of the fMRI experiment was to identify the neural mechanisms in which activity is modulated by these two parameters. On the basis of the behavioral results, we selected two types of signals for subsequent use in the fMRI study: (1) fixed coherence of four frequency components and varying duration (Fig. 3, red dashed curve) and (2) fixed duration of four chords and varying coherence (Fig. 3, blue dashed curve), as these parameters resulted in behavioral performance that monotonically spanned the range between not detectable and highly detectable. These parameters were selected to identify the brain areas in which the activity increases parametrically with an increase in the corresponding changes in coherence (while keeping duration fixed) and duration (while keeping coherence fixed), respectively.

Functional imaging

As the aim of the present study was to uncover the automatic, stimulus-driven mechanisms that subserve segregation in the SFG stimuli, we discuss functional imaging results from the "passive listening" block only.

The primary purpose of the active detection block (presented after the passive listening block) was to ensure that subjects were indeed able to detect the figures despite the loud, interfering MRI scanner noise and to compare their performance to that measured in the psychophysics study. Because of the differences in stimulus presentation between the passive and active blocks (see Materials and Methods), as well as other perceptual factors such as attentional load and focus of attention, a comparison of the activation patterns in the two blocks is not straightforward. The functional imaging results for the "active detection" block are therefore not reported.

Figure 3 presents the behavioral results recorded in the scanner ("active detection" block) along with responses to the same stimuli as obtained in the psychophysics study (see above). Overall, listeners performed worse than in the psychophysical experiment (a difference of ~20%). This may stem from several factors: (1) interfering scanner noise; (2) lack of sufficient practice—it was important to keep listeners naive for the passive part of the experiment, so all instructions related to the active detection task were delivered after the passive block, while listeners were already in the scanner; and (3) due to time constraints, the session was overall much shorter than the psychophysics experiment. Moreover, because of the experimental design, listeners also encountered overall fewer "easy" signals (those with a fixed coherence of six and eight components and long duration), which may have contributed to a smaller improvement with exposure.

Critically, the present data demonstrate that the figures are readily detectable even in a noisy scanner environment and that the parametric modulation used is effective at eliciting a wide range of figure detection performance from nondetectable to detectable.

Passive listening block

The purpose of the passive listening block was to identify the brain areas in which activity is modulated by figure salience in the SFG stimuli. We used a decoy task (noise burst detection) to assure that subjects are generally vigilant and attentive to the auditory stimuli, while distracting them from the stimulus dimension of interest. All subjects performed the decoy task at ceiling level. Because we aimed to tap predominantly bottom-up segregation mechanisms in the passive listening block (those that are independent of the attentional state of the listener), it was

Table 1. Stereotactic MNI coordinates for effects of duration and coherence

Contrast	Brain area	*x*	*y*	*z*	*t* value	*z* score
Effect of duration	Left IPS	−42	−46	64	5.14	3.67
		−48	−40	61	4.89	3.56
	Right IPS	51	−28	61	5.17	3.68
		45	−37	64	4.24	3.25
	Left STS	−57	−34	−2	4.42	3.34
	Right STS	60	−13	−11	4.06	3.16
	Right PT	60	−13	10	4.96	3.59
	Left MGB	−15	−25	−8	4.85	3.54
	Right MGB	18	−25	−8	4.92	3.57
Effect of coherence	Left IPS	−21	−73	46	4.99	3.60
		−24	−73	37	4.36	3.31
	Right IPS	27	−82	31	3.69	2.96
	Left STS	−48	−16	−5	3.43	2.81
	Right STS	39	−4	−26	3.77	3.00

Local maxima for effects of duration and coherence are shown at a threshold of $p < 0.001$ (uncorrected).

essential that subjects were naive to existence of the figures. Indeed, when questioned at the end of the block, none of the subjects reported noticing the figures popping out of the background.

We were specifically interested in testing whether activity in the primary or nonprimary auditory cortex, as well as non-auditory areas such as the IPS (Cusack, 2005), is correlated with figure processing (see Table 1 for prespecified anatomical regions for which responses were considered significant at $p < 0.001$ uncorrected). The critical contrasts included a test for the effects of increasing duration (with fixed coherence) and increasing coherence (with fixed duration) on brain responses.

Effects of duration

The analysis of parametric changes in brain responses to figures characterized by a fixed coherence and varying duration showed significant bilateral activations in the anterior IPS (Fig. 4*A*), the superior temporal sulcus (STS) (Fig. 4*B*), the medial geniculate body (MGB) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material), and the right planum temporale (PT) (Fig. 4*B*).

Effect of coherence

The analysis of the effect of increasing the coherence of the figures while keeping the duration fixed showed significant bilateral activations in the posterior IPS (Fig. 5*A*), and the STS (Fig. 5*B*).

Auditory cortex activations

A stringent analysis was performed to check for auditory cortex activations in the two contrasts of interest. Using the probabilistic cytoarchitectonic maps for primary auditory cortex—TE 1.0, TE 1.1, and TE 1.2 (Morosan et al., 2001), which are incorporated in the SPM Anatomy toolbox (http://www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox)—a volume of interest analysis was performed. We tested for auditory cortex activations but did not find any significant voxels that survived a test for multiple comparisons of $p < 0.05$ (family wise error correction) when examined with the three PAC maps.

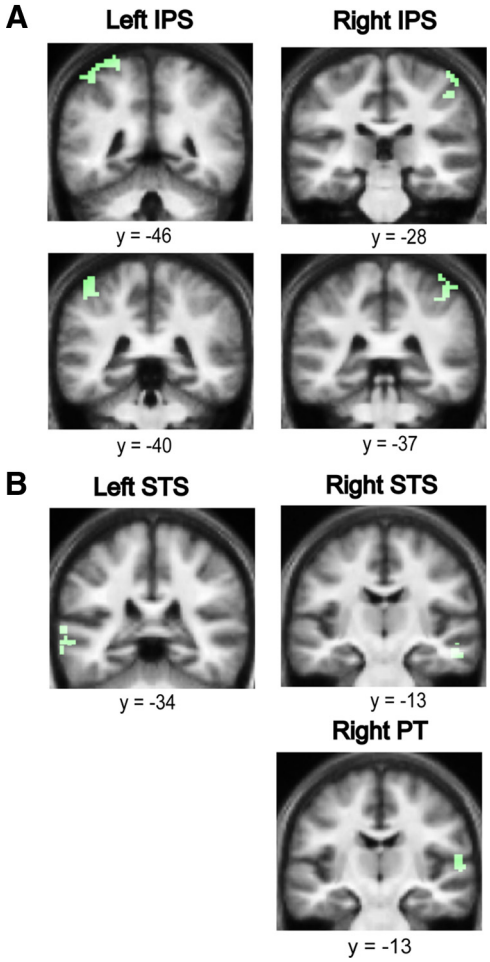

Figure 4. The effect of duration on auditory segregation. ***A***, Areas in the anterior IPS showing an increased hemodynamic response as a function of increasing duration of the figures with fixed coherence (in green). Significant clusters for the effect of duration were found in the anterior IPS bilaterally. Results are rendered on the coronal section of the subjects' normalized average structural scan, and results are shown at $p < 0.001$ uncorrected. ***B***, Areas in the STS and PT showing an increased hemodynamic response as a function of increasing duration of the figures with fixed coherence (in green). Significant clusters for the effect of duration were found in the STS bilaterally and in the right PT. Results are rendered on the coronal section of the subjects' normalized average structural scan, which is tilted (pitch = −0.5) to reveal significant clusters in the superior temporal plane, at $p < 0.001$ uncorrected.

Discussion

In this study, we developed a new SFG stimulus. Conceptually similar to the Julesz (1962) texture patterns, the figure and ground are indistinguishable at each instant and can be segregated only by integrating the patterns over time and frequency. An important perceptual characteristic of the SFG stimulus is the rapid buildup rate (the time required to segregate the figure from the background). For coherence levels of four components and above, as few as seven consecutive chords (a total of 350 ms) are sufficient to reach ceiling detection performance (Fig. 2). This is in contrast to the longer buildup time (~2000 ms) reported in many brain imaging or electrophysiological experiments of streaming (Micheyl et al., 2007b; Gutschalk et al., 2008; Pressnitzer et al., 2008; Elhilali et al., 2009), attributed to prolonged accumulation of sensory evidence, possibly requiring top-down mechanisms (Denham and Winkler, 2006). The shorter buildup times observed for SFG signals suggest that segregation may rely on partially different mechanisms from those that mediate streaming in signals commonly used in brain imaging experiments (see also Sheft and Yost, 2008). All of these features make the SFG stimulus an interesting complement to streaming signals, with which to study preattentive auditory scene analysis. Using fMRI, we identified the IPS and the STS as the primary brain areas involved in the process of automatic, stimulus-driven figure–ground decomposition in this stimulus.

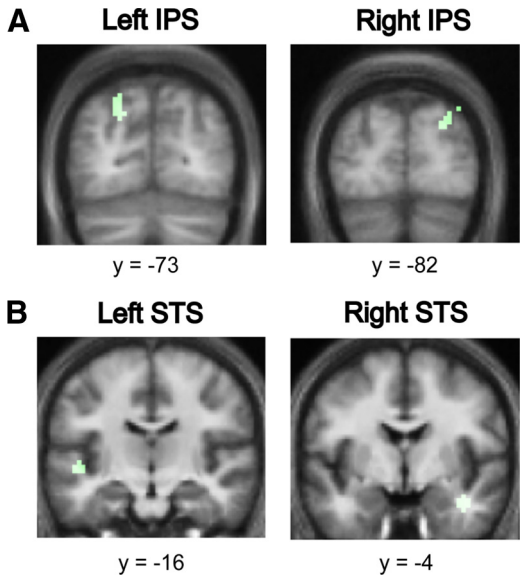

Figure 5. The effect of coherence on auditory segregation. ***A***, Areas in the posterior IPS showing an increased hemodynamic response as a function of increasing coherence of the figures with fixed duration (in green). Significant clusters for the effect of duration were found in the posterior IPS bilaterally. Results are rendered on the coronal section of the subjects' normalized average structural scan at $p < 0.001$ uncorrected. ***B***, Areas in the STS showing an increased hemodynamic response as a function of increasing coherence of the figures with fixed duration (in green). Significant clusters for the effect of duration were found in the STS bilaterally and in the right PT. Results are rendered on the coronal section of the subjects' normalized average structural scan to reveal significant clusters in the superior temporal plane at $p < 0.001$ uncorrected.

Auditory segregation

Classically, auditory segregation has been investigated using two classes of stimuli. Simultaneous organization has been studied using signals consisting of multiple concurrent components where properties such as harmonic structure (tuned vs mistuned) (Alain, 2007; Lipp et al., 2010), spatial location (McDonald and Alain, 2005), or onset asynchrony (Bidet-Caulet et al., 2007; Sanders et al., 2008; Lipp et al., 2010) were manipulated to induce the percept of a single source versus several concomitant sources. Using such signals, human electroencephalography (EEG) and magnetoencephalography (MEG) experiments have identified responses in nonprimary (Alain, 2007; Lipp et al., 2010) and primary (Bidet-Caulet et al., 2007) auditory cortex that covary with the percept of two sources.

The second class of stimuli used to study scene organization, is the streaming paradigm (van Noorden, 1975; Bregman, 1990; Shamma and Micheyl, 2010). Streaming refers to the process by which sequentially presented elements are perceptually bound into separate "entities" or "streams," which can be selectively attended to (Elhilali et al., 2009). Human EEG and MEG experiments have demonstrated a modulation of the N1 (or N1m) response, thought to originate from non-primary auditory cortex, depending on whether stream segregation takes place (Gutschalk et al., 2005; Snyder and Alain, 2007; Schadwinkel and Gutschalk, 2010). fMRI studies have additionally identified activations in earlier areas such as the MGB (Kondo and Kashino, 2009) and primary auditory cortex (Wilson et al., 2007; Deike et al., 2010; Schadwinkel and Gutschalk, 2010) that are correlated with the streaming percept, in line with neurophysiological evidence from animal experiments (Fishman et al., 2001, 2004; Bee and Klump, 2004, 2005; Micheyl et al., 2005).

Auditory cortex and segregation

Stimulus-driven segregation has been hypothesized to be mediated by basic response properties of auditory neurons: frequency selectivity, forward suppression, and adaptation, resulting in the activation of distinct neural populations pertaining to the figure and background (Micheyl et al., 2007b; Snyder and Alain, 2007; Shamma and Micheyl, 2010). Such mechanisms have been observed in primary auditory cortex (Fishman et al., 2004; Micheyl et al., 2005) and in the periphery (Pressnitzer et al., 2008). Together, human and animal work suggests that segregation occurs in a distributed network over multiple stages in the ascending (and possibly descending) auditory pathway. Interestingly, an area outside the classic auditory cortex, the IPS, has also been implicated in this process (Cusack, 2005; see below).

Consistent with results from Cusack (2005), but contrary to the studies reviewed above, we did not find activation in primary auditory cortex. This difference could be due to methodological issues (see also Cusack, 2005) or the relatively more complex nature of our stimuli. In most streaming experiments that found activity in auditory cortex to be correlated with the perception of one versus two streams, stimulus parameters were modulated to produce streaming, and any effect on primary cortex activity might be due to altered stimulus representation (but see Kondo and Kashino, 2009). On the other hand, Cusack (2005) used stimuli that produced a bistable percept, in the absence of corresponding changes in the physical properties of the stimulus. The lack of activation differences in primary auditory cortex in that experiment is consistent with sensory rather than perceptual representation at that level. In the present experiment, the lack of activation in Heschl's gyri could stem from the fact that adaptation-based mechanisms in primary auditory cortex, thought to underlie stream segregation (see above), are not (or not sufficiently) activated by the stochastic SFG stimuli. Alternatively, primary auditory cortical activation observed in previous studies could be due to active following of the figure from amid the background [indeed, see Bidet-Caulet et al. (2007) and Elhilali et al. (2009)], while our experimental design incorporated short figures and naive subjects (see Materials and Methods) to specifically eliminate such attentional processes and focus on automatic, bottom-up, stimulus-driven mechanisms.

Our results implicate the STS in the stimulus-driven partitioning of grouped components into "figure" and "ground." Previous studies suggest the involvement of STS in the perception of complex stimuli with a stochastic structure where different specific stimuli can fall into the same perceptual category. The area has been implicated in the analysis of spectral shape (Warren et al., 2005), changing spectrum over time (Overath et al., 2008), and detecting increasing changes in spectrotemporal coherence within acoustic "textures" (Overath et al., 2010). Together, these studies point to a role for STS in the abstraction of features over frequency–time space relevant to the perception of distinct categories. STS is also involved in the analysis of natural categories associated with semantic labels, such as voices (Belin et al., 2000; Kriegstein and Giraud, 2004).

Notably, while the design of these previous studies was based on sequentially presented patterns that varied in their spectrotemporal structure, the present results are based on stimuli in which certain features must be integrated over frequency–time space to create a perceptual figure, distinct from a concurrently presented stochastic background.

The IPS and auditory perceptual organization

Accumulating evidence points to the involvement of the IPS in perceptual organization (e.g., for review, see Cusack, 2005) such as encoding object representations (Xu and Chun, 2009), binding of sensory features within a modality (Friedman-Hill et al., 1995; Donner et al., 2002; Shafritz et al., 2002; Kitada et al., 2003; Yokoi and Komatsu, 2009), and across different modalities (Bremmer et al., 2001; Calvert, 2001; Beauchamp et al., 2004; Miller and D'Esposito, 2005; Buelte et al., 2008; Werner and Noppeney, 2010).

The role of the IPS in auditory perceptual organization was first suggested by Cusack (2005) in a study that measured fMRI activation during the presentation of perceptually bistable streaming sequences and correlated changes in BOLD response with listeners' percepts. The only region exhibiting significant differential activation was the IPS, showing increased activation when subjects perceived two streams as opposed to one. Consistent with these findings, we observed bilateral IPS activation that is related to preattentive, stimulus-driven figure–ground decomposition in our SFG stimuli.

We additionally found a functional dissociation within the IPS such that the anterior and the posterior IPS were activated for the effects of duration and coherence, respectively. This is consistent with previous reports of functional dissociation within the IPS (e.g., Rushworth et al., 2001a,b; Rice et al., 2006; Cusack et al., 2010).

The implication of IPS in auditory segregation has been puzzling in view of classic models of auditory scene analysis based on mechanisms within the "auditory system" (Micheyl et al., 2007b; Snyder and Alain, 2007; Shamma and Micheyl, 2010). A central issue has been whether activity in IPS is causally responsible for segregation or whether it reflects the output of perceptual organization occurring in auditory areas (Carlyon, 2004; Shamma and Micheyl, 2010). It has been suggested that IPS activation observed by Cusack (2005) may result from the application of top-down attention during a subjective task or an upstream effect of organization (e.g., shifting attention between streams).

Although IPS has been implicated in voluntary and involuntary control and shifts in auditory attention (Molholm et al., 2005; Watkins et al., 2007; Salmi et al., 2009; Hill and Miller, 2010), it is unlikely that the activation observed here relates to top-down application of attention, or the active shifting of attention between objects. Subjects in our study were naive to the existence of the figure, and, when questioned, none reported noticing the figures. Additionally, the fact that we find different parametric modulations (duration vs coherence) to engage different fields in the IPS is inconsistent with a simple account in terms of subjective attention. Our results are therefore in line with the suggestion that IPS plays an automatic, stimulus-driven role in auditory figure–ground segregation, and encourage a revaluation of the standard outlook on the brain systems involved in auditory scene analysis.

References

Alain C (2007) Breaking the wave: effects of attention and learning on concurrent sound perception. Hear Res 229:225–236.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823.

Bee MA, Klump GM (2004) Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J Neurophysiol 92: 1088–1104.

Bee MA, Klump GM (2005) Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. Brain Behav Evol 66:197–214.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O (2007) Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. J Neurosci 27: 9252–9261.

Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT Press.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. Neuron 29:287–296.

Buelte D, Meister IG, Staedtgen M, Dambeck N, Sparing R, Grefkes C, Boroojerdi B (2008) The role of the anterior intraparietal sulcus in crossmodal processing of object features in humans: an rTMS study. Brain Res 1217:110–118.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.

Carlyon RP (2004) How the brain separates sounds. Trends Cogn Sci 8:465–471.

Cusack R (2005) The intraparietal sulcus and perceptual organization. J Cogn Neurosci 17:641–651.

Cusack R, Mitchell DJ, Duncan J (2010) Discrete object representation, attention switching, and task difficulty in the parietal lobe. J Cogn Neurosci 22:32–47.

de Cheveigné A (2001) The auditory system as a separation machine. In: Physiological and psychophysical bases of auditory function (Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R, eds), pp 453–460. Maastricht, The Netherlands: Shaker.

Deichmann R, Schwarzbauer C, Turner R (2004) Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. Neuroimage 21:757–767.

Deike S, Scheich H, Brechmann A (2010) Active stream segregation specifically involves the left human auditory cortex. Hear Res 265:30–37.

Denham SL, Winkler I (2006) The role of predictive models in the formation of auditory streams. J Physiol Paris 100:154–170.

Donner TH, Kettermann A, Diesch E, Ostendorf F, Villringer A, Brandt SA (2002) Visual feature and conjunction searches of equal difficulty engage only partially overlapping frontoparietal networks. Neuroimage 15:16–25.

Elhilali M, Xiang J, Shamma SA, Simon JZ (2009) Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol 7:e1000129.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151:167–187.

Fishman YI, Arezzo JC, Steinschneider M (2004) Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116:1656–1670.

Friedman-Hill SR, Robertson LC, Treisman A (1995) Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. Science 269:853–855.

Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS (1995a) Spatial registration and normalization of images. Hum Brain Mapp 2:165–189.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995b) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci 25:5382–5388.

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6:e138.

Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cereb Cortex 20:583–590.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. Neuroimage 16:217–240.

Julesz B (1962) Visual pattern discrimination. IRE Trans Inf Theory IT-8:84–92.

Kidd G Jr, Mason CR, Deliwala PS, Woods WS, Colburn HS (1994) Reducing informational masking by sound segregation. J Acoust Soc Am 95:3475–3480.

Kitada R, Kochiyama T, Hashimoto T, Naito E, Matsumura M (2003) Moving tactile stimuli of fingers are integrated in the intraparietal and inferior parietal cortices. Neuroreport 14:719–724.

Kondo HM, Kashino M (2009) Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. J Neurosci 29:12695–12701.

Kriegstein KV, Giraud AL (2004) Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955.

Lipp R, Kitterick P, Summerfield Q, Bailey PJ, Paul-Jordanov I (2010) Concurrent sound segregation based on inharmonicity and onset asynchrony. Neuropsychologia 48:1417–1425.

McDonald KL, Alain C (2005) Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. J Acoust Soc Am 118:1593–1604.

Micheyl C, Oxenham AJ (2010) Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. Hear Res 266:36–51.

Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48:139–148.

Micheyl C, Shamma S, Oxenham AJ (2007a) Hearing out repeating elements in randomly varying multitone sequences: a case of streaming? In: Hearing—from basic research to application (Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S, Verhey J, eds), pp 267–274. Berlin: Springer.

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E (2007b) The role of auditory cortex in the formation of auditory streams. Hear Res 229:116–131.

Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 25:5884–5893.

Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ (2005) The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. Cereb Cortex 15:545–551.

Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 74:750–753.

Moore BCJ, Gockel H (2002) Factors influencing sequential stream segregation. Acta Acustica 88:320–333.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13:684–701.

Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. J Neurosci 28: 13268–13273.

Overath T, Kumar S, Stewart L, von Kriegstein K, Cusack R, Rees A, Griffiths TD (2010) Cortical mechanisms for the segregation and representation of acoustic textures. J Neurosci 30:2070–2076.

Penny W, Holmes AP (2004) Random-effects analysis. In: Human brain function (Frackowiak RS, Friston KJ, Frith C, Dolan RJ, Price CJ, eds), pp 843–850. San Diego: Academic.

Plack CJ, Moore BCJ (1990) Temporal window shape as a function of frequency and level. J Acoust Soc Am 87:2178–2187.

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128.

Rice NJ, Tunik E, Grafton ST (2006) The anterior intraparietal sulcus mediates grasp execution, independent of requirement to update: new insights from transcranial magnetic stimulation. J Neurosci 26:8176–8182.

Rushworth MF, Ellison A, Walsh V (2001a) Complementary localization and lateralization of orienting and motor attention. Nat Neurosci 4:656–661.

Rushworth MF, Krams M, Passingham RE (2001b) The attentional role of the left parietal cortex: the distinct lateralization and localization of motor attention in the human brain. J Cogn Neurosci 13:698–710.

Salmi J, Rinne T, Koistinen S, Salonen O, Alho K (2009) Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. Brain Res 1286:155–164.

Sanders LD, Joh AS, Keen RE, Freyman RL (2008) One sound or two? Object-related negativity indexes echo perception. Percept Psychophys 70:1558–1570.

Schadwinkel S, Gutschalk A (2010) Activity associated with stream segregation in human auditory cortex is similar for spatial and pitch cues. Cereb Cortex 20:2863–2873.

Scheich H, Baumgart F, Gaschler-Markefski B, Tegeler C, Tempelmann C, Heinze HJ, Schindler F, Stiller D (1998) Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. Eur J Neurosci 10:803–809.

Shafritz KM, Gore JC, Marois R (2002) The role of the parietal cortex in visual feature binding. Proc Natl Acad Sci U S A 99:10917–10922.

Shamma SA, Micheyl C (2010) Behind the scenes of auditory perception. Curr Opin Neurobiol 20:361–366.

Sheft S, Yost WA (2008) Method-of-adjustment measures of informational masking between auditory streams. J Acoust Soc Am 124:EL1–7.

Snyder JS, Alain C (2007) Toward a neurophysiological theory of auditory stream segregation. Psychol Bull 133:780–799.

van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. Eindhoven, The Netherlands: University of Technology, Eindhoven.

Warren JD, Jennings AR, Griffiths TD (2005) Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24:1052–1057.

Watkins S, Dalton P, Lavie N, Rees G (2007) Brain mechanisms mediating auditory attentional capture in humans. Cereb Cortex 17:1694–1700.

Werner S, Noppeney U (2010) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J Neurosci 30:2662–2675.

Wilson EC, Melcher JR, Micheyl C, Gutschalk A, Oxenham AJ (2007) Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. J Neurophysiol 97:2230–2238.

Xu Y, Chun MM (2009) Selecting and perceiving multiple visual objects. Trends Cogn Sci 13:167–174.

Yokoi I, Komatsu H (2009) Relationship between neural responses and visual grouping in the monkey parietal cortex. J Neurosci 29:13210–13221.