Behavioral/Systems/Cognitive

Human Cortical Activity during Streaming without Spectral Cues Suggests a General Neural Substrate for Auditory Stream Segregation

Alexander Gutschalk,[1,2,3] **Andrew J. Oxenham,**[4]* **Christophe Micheyl,**[4]* **E. Courtenay Wilson,**[5] **and Jennifer R. Melcher**[1,2,5]*

[1]Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, [2]Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115, [3]Department of Neurology, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany, [4]Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455, [5]Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Program in Speech and Hearing Bioscience and Technology, Cambridge, Massachusetts 02139

The brain continuously disentangles competing sounds, such as two people speaking, and assigns them to distinct streams. Neural mechanisms have been proposed for streaming based on gross spectral differences between sounds, but not for streaming based on other nonspectral features. Here, human listeners were presented with sequences of harmonic complex tones that had identical spectral envelopes, and unresolved spectral fine structure, but one of two fundamental frequencies (f_0) and pitches. As the f_0 difference between tones increased, listeners perceived the tones as being segregated into two streams (one stream for each f_0) and cortical activity measured with functional magnetic resonance imaging and magnetoencephalography increased. This trend was seen in primary cortex of Heschl's gyrus and in surrounding nonprimary areas. The results strongly resemble those for pure tones. Both the present and pure tone results may reflect neuronal forward suppression that diminishes as one or more features of successive sounds become increasingly different. We hypothesize that feature-specific forward suppression subserves streaming based on diverse perceptual cues and results in explicit neural representations for auditory streams within auditory cortex.

Key words: **auditory cortex; scene analysis; stream segregation; fMRI; MEG; adaptation**

Introduction

A prime task of the auditory system is to separate the complex mixture of sounds reaching the ears into perceptual "streams," corresponding to individual sound sources in the listener's environment (Bregman, 1990; Moore and Gockel, 2002; Carlyon, 2004). The neural underpinnings of auditory stream segregation have been investigated with intracortical recordings in animals (Fishman et al., 2001, 2004; Kanwal et al., 2003; Bee and Klump, 2004, 2005; Micheyl et al., 2005), and noninvasively in humans, using electroencephalography (EEG) (Sussman et al., 1999, Sussman, 2005; Snyder et al., 2006), magnetoencephalography (MEG) (Gutschalk et al., 2005), and functional magnetic resonance imaging (fMRI) (Deike et al., 2004; Cusack, 2005; Wilson et al., 2007). All of these studies have used pairs of sounds that differed in frequency content, and were presented alternately, or in other repeating temporal patterns. These sequences can be perceived either as a single stream of fluctuating sounds or as two segregated streams, each comprising a single repeating sound, the latter percept being fostered by larger frequency differences.

For conditions such as these, where spectral cues play a dominant role, it has been proposed that sounds segregate into different streams when they produce sufficiently different patterns of neural excitation along the frequency dimension of tonotopically organized areas of the auditory system (Hartmann and Johnson, 1991; Beauvois and Meddis, 1996; McCabe and Denham, 1997). The physiological studies cited above are generally consistent with these theoretic models. In particular, the animal studies (reviewed by Micheyl et al., 2007) have specifically demonstrated a relationship between streaming and two-tone interactions within frequency-selective neurons of primary auditory cortex.

Despite the intuitive appeal of a "tonotopic" explanation of auditory stream segregation, it is inappropriate as a general model of stream segregation. Psychophysical studies have demonstrated that sounds designed to evoke very similar tonotopic patterns of excitation in the auditory periphery can nonetheless form separate streams if they differ along other perceptual dimensions, such as pitch or perceived fluctuation rate (Vliegen and Oxenham, 1999; Vliegen et al., 1999; Grimault et al., 2000, 2002; Roberts et al., 2002). The neurophysiological substrates for streaming under such conditions are unknown.

Received May 19, 2007; revised Sept. 25, 2007; accepted Oct. 10, 2007.

This work was supported by Deutsche Forschungsgemeinschaft Grant GU 593/2-1, National Institutes of Health Grants R01DC07657 and P01DC00119, National Center for Research Resources Grant P41RR14075, the Mental Illness and Neuroscience Discovery Institute, a Hertz Foundation Fellowship (E.C.W.), and the Dietmar-Hopp-Stiftung.

*A.J.O., C.M., and J.R.M. contributed equally to this work.

Correspondence should be addressed to Alexander Gutschalk, Department of Neurology, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany. E-mail: Alexander_Gutschalk@med.uni-heidelberg.de.

DOI:10.1523/JNEUROSCI.2299-07.2007

 0270-6474/07/2713074-08$15.00/0

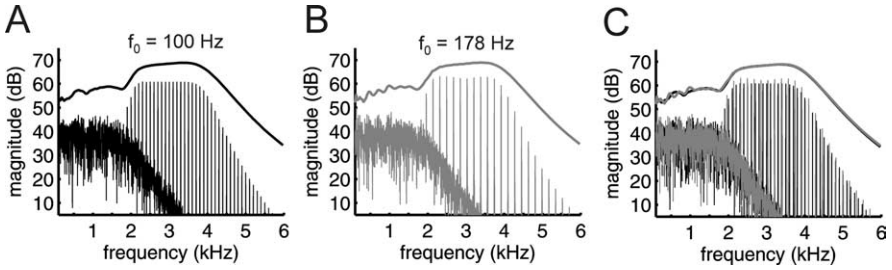

Figure 1. Similarity of spectral envelopes and excitation patterns across stimuli. Frequency spectra of two of the harmonic complexes used in this study. ***A***, ***B***, The f_0 is (***A***) 100 Hz or (***B***) 178 Hz. The f_0 difference between the tones amounts to 10 semitones, the largest Δf_0 used. ***C***, The spectra for the 100 Hz (black) and 178 Hz (gray) complex are superimposed. The low-pass noise, apparent in the low-frequency region of each panel, was designed to mask any potential auditory distortion products. The curves plotted above the spectra are excitation patterns in response to harmonic complexes (i.e., relative level of neural activity vs frequency), calculated using a model for the auditory periphery (Glasberg and Moore, 1990). For these curves, the frequency axis indicates tonotopic location along the cochlear partition. The excitation patterns are superimposed in ***C*** to illustrate that they are essentially identical.

Here, we investigate the neural basis of auditory streaming under conditions where the segregation of sounds into streams cannot be attributed to spectral differences between sounds. We used sequences of bandpass-filtered harmonic complex tones, which are perceived as either one or two streams depending on differences of fundamental frequency (f_0) between the tones. The f_0s and filtering of the tones were chosen such that the tones would evoke very similar tonotopic patterns of excitation (see Fig. 1). MEG and fMRI data, each combined with behavioral measurements, provide independent indications that streaming based on different perceptual cues may rely on similar neural mechanisms that are operative throughout auditory cortex.

Materials and Methods

Listeners. Six listeners (one female) aged between 24 and 39 years (mean 31) participated in psychophysical testing in a quiet booth, fMRI, and MEG in that order. All had normal hearing as defined by pure-tone thresholds <20 dB hearing level at octave intervals from 250 through 8000 Hz, and did not report any history of peripheral or central hearing disorders. Three additional listeners took part in pilot psychophysical measurements, the results of which were used to select suitable stimulus parameters for fMRI and MEG. A 10th subject was psychophysically tested in the quiet booth and participated in fMRI, but was excluded because of excessive head movement.

Stimuli. The stimulus sequences used in this study comprised harmonic tone complexes, each 100 ms in duration (including 5 ms raised-cosine onset and offset ramps). The f_0 of the complexes spanned a 10 semitone range, between 100 and 178.2 Hz. The tone complexes were generated digitally at a sampling rate of 24 kHz with 16 bit resolution, and were digitally bandpass filtered (fourth-order Butterworth filter with zero phase shift) between 2 and 4 kHz (3 dB cutoff frequencies). The lower cutoff frequency (2 kHz) was chosen so that the individual harmonics of the complex tones would not be resolved, or heard out individually. This is illustrated in Figure 1, which shows the frequency spectra of two complex tones used in this experiment, along with the expected patterns of excitation, based on a model by Glasberg and Moore (1990). It can be seen that changing the f_0 from 100 to 178 Hz has essentially no effect on the excitation patterns produced by the model. The model represents the excitation level as a function of tonotopic location along the sensory epithelium, and has been shown to predict data from a wide variety of experiments involving auditory masking and frequency selectivity, including experiments that specifically investigated the resolvability of tones within harmonic complexes (Bernstein and Oxenham, 2006). The limits on frequency selectivity observed behaviorally, and reflected in the model, are generally believed to reflect peripheral auditory processes within the cochlea (Evans et al. 1989; Evans 2001), and there is no evidence to suggest that selectivity is enhanced at higher stages of the auditory system. Thus, the limits of tonotopic resolution illustrated in the model likely reflect peripheral limitations that are maintained throughout the tonotopic areas of the auditory pathways. This conclusion is supported by a recent study in the cat auditory nerve, which showed that the highest resolved harmonic varied from around the fifth at an f_0 of 200 Hz to around the 10th at an f_0 of 1000 Hz (Cedolin and Delgutte, 2005). Thus, even if human tuning is sharper than that found in cat by a factor of 2 (Shera et al., 2002), within our range of f_0s (< 200 Hz) we would not expect harmonics above the 10th to be resolved in the auditory periphery or beyond. Because our highest f_0 was 178.2 Hz, the harmonics presented to the listeners in the passband were never lower than the 11th. The upper cutoff frequency (4 kHz) was selected because the transfer functions of the sound delivery systems used in the MEG and fMRI experiments exhibited significant spectral peaks and dips above that frequency, and because pitch information is generally very weak for sounds above ~4–5 kHz (Moore, 2003).

Although the pitch produced by complex tones with only unresolved harmonics is weaker than that produced by broadband complex tones (Houtsma and Smurzynski, 1990), it is still sufficiently strong to induce perceptual streaming (Vliegen and Oxenham, 1999). Because the individual harmonics are unresolved, the pitch corresponding to the fundamental frequency of these complexes is thought to be conveyed solely by temporal, rather than spectral, cues. The advantage of using these filtered complexes is, therefore, that changes in fundamental frequency lead to changes in perceived pitch without producing perceptually salient changes in the tonotopic representation of the tones.

All tones were presented at an overall sound pressure level (SPL) of 75 dB. Gaussian noise, digitally low-pass filtered (fourth order Butterworth filter with zero phase shift) at 2 kHz (3-dB cutoff), was presented continuously at an overall level of 69 dB SPL (spectrum level of ~33 dB) to mask potential distortion products and to mask portions of the stimuli that fell below the main passband. The complex tones were arranged temporally into a repeating ABBB pattern, where A and B represent tones of different f_0, with no silent gaps between consecutive tones. This pattern was chosen because it has the advantages of the triplet pattern more commonly used in streaming studies (tone triplets separated by silent gaps, BAB-BAB. . . , where the hyphen represents the silent gap), but is continuous (important for fMRI; see below). One advantage of the triplet pattern (e.g., over a simple alternating pattern of ABAB. . .) is that when two separate streams are heard, the two streams have different tempi (fast, B-B-B. . . vs slow, -A—A—A-.); this provides listeners with a clear cue for deciding whether they hear two streams (van Noorden, 1975; Carlyon et al., 2001). A second advantage of the triplet pattern is that the difference in perceived interstimulus interval (ISI) between cases where the sequence is heard as one stream and cases where it is heard as two streams is much larger for the A tones than for the B tones. In a previous MEG study (Gutschalk et al., 2005), this difference in perceived ISI allowed us to establish that the amplitudes of the P_1m and N_1m components evoked by the A and B tones increase with the ISI within each stream (i.e., the perceived ISI). The ABBB pattern used in the present study offers both advantages of the triplet pattern, because the only difference is that the silent gap between triplets has been filled with another B tone. Filling the gap was important for the fMRI part of this study, in which we examined the time course as well as the magnitude of blood oxygen level-dependent (BOLD) responses, because actual (or perceived) gaps lead to changes in BOLD time course (Harms and Melcher, 2002; Wilson et al., 2007). By ensuring no physical temporal gaps, we were able to investigate the effect perceived temporal gaps introduced when the percept shifts from one to two streams.

In our MEG measurements, we planned to focus on responses to the A tones (for which the percept-dependent differences in perceived ISI were

largest). The f_0 of the A tones was held constant at 178.2 Hz whereas that of the B tones was varied to obtain different f_0 differences (Δf_0) between A and B. The Δf_0s were introduced by lowering the f_0 of the B tones by 0, 1/2, 1, 2, 3, 4, 5, or 10 semitones. All Δf_0s were tested in the quiet booth. A subset was used in the fMRI and MEG experiments: 0 semitones (B tone frequency, 178.2 Hz), 1 (168.2 Hz), 3 (149.8 Hz), and 10 (100 Hz). The stimuli were 32 s sequences of constant Δf_0. Δf_0 was varied pseudo-randomly across sequences.

Procedures. The psychophysical session in the quiet booth began by familiarizing the listeners with the stimuli and task. The listeners received written instructions, accompanied by a schematic diagram explaining what was meant by "one stream" and "two streams," and they performed a few practice trials. During the actual measurements, the sequences corresponding to the eight different Δf_0s were presented five times each in random order. Listeners initiated the start of each 32 s sequence, and they were encouraged to take breaks between sequence presentations to ensure that they remained alert throughout the duration of the experiment. During the presentation of tone sequences, the listeners indicated whether they perceived one or two streams soon after the beginning of each sequence, and then again after each change in the percept (from one to two streams or vice versa), by pressing one of two keys, which corresponded to the two percepts. Stimuli were presented diotically (binaurally) with HD580 circumaural headphones (Sennheiser, Old Lyme, CT) in a double-walled sound-attenuating chamber.

During fMRI, the sequences were presented in 16 runs, each comprising three repetitions of one Δf_0 condition (0, 1, 3, 10 semitones). There were four runs for each Δf_0. A blocked stimulus design was used, in which each presentation of a 32 s sequence (and simultaneous noise masker) was preceded and followed by 32 s, during which the tone sequence was turned off, and only the noise masker was present. The time between runs was usually 30–60 s. Sound presentation was diotic via piezo-electric headphones (GEC Marconi, Towcester, UK). To reduce extraneous acoustic noises from the scanner, the scanner's coolant pump was switched off. Listeners were instructed to focus on the tones and indicate their percept (one or two streams) using a hand dial that controlled two lights visible to the subject through a mirror. The subject was asked to illuminate one light whenever they perceived one stream and two lights whenever they perceived two streams.

During MEG, sounds were produced by a speaker system located outside the magnetically shielded room (ADU1b; Unides, Helsinki, Finland), and they were conveyed via 3-m-long plastic tubes before being delivered to the ear canals via foam earpieces. The 32 s sequences were separated by 3 s interruptions, during which only the continuously running noise masker was present. In addition to the four Δf_0 conditions, a control condition was included, during which only the A tones of the ABBB sequence were presented. Each of the five resulting conditions was presented 20 times, in random order, yielding a total of 1600 repetitions of the A tones in each sequence (because each 32 s sequence contained 80 ABBB quadruplets). Listeners reported their percept continuously throughout each sequence by pressing one of two buttons.

Data acquisition. A 3T scanner (Magnetom Allegra; Siemens, Erlangen, Germany) and standard, bird-cage head coil were used for MRI. First, structural, magnetization prepared rapid gradient echo (MPRAGE) images of the whole head were acquired (sagittal in-plane resolution 256 × 256; slice thickness 1.3 mm). Based on these images, the volume for functional imaging was chosen as 11 near-coronal slices (perpendicular to the Sylvian fissure) that covered the auditory cortex from the posterior end of planum temporale to the anterior aspect of the superior temporal gyrus (STG), including the complete Heschl's gyrus (first and second when there were two) in both hemispheres. The volume was placed to include the inferior colliculus in the third or fifth image (counted from posterior). For coregistration, T2-weighted structural images were obtained for the same volume with a high in-plane resolution (384 × 384). Functional imaging was performed using an echo-planar imaging sequence [gradient echo; echo time (TE), 30 ms; flip angle, 90°; in-plane resolution, 64 × 64; slice thickness, 4 mm; gap, 1.32 mm]. Images of the 11-slice volume were acquired in brief clusters separated by an 8 s quiet interval to decrease the interference of the imager noise with the auditory stimulation (Edminster et al., 1999; Hall et al., 1999). To allow reconstruction of the BOLD signal time course with 2 s resolution, the stimulus presentation was delayed relative to image acquisition by varying degrees from run to run (i.e., by 0, 2, 4, or 6 s for each of the four runs of a given Δf_0). In the first three subjects (of whom one was excluded), peripheral gating was used to improve the detection of brainstem activation (Guimaraes et al., 1998) so the time between clusters [repetition time (TR)] was approximately, rather than exactly, 8 s. Even with the improvements in activation detection afforded by the gating technique, the complex tone sequences (and simultaneous masking noise) contrasted with the masker alone did not show significant activation in brainstem centers. The absence of activation did not reflect a technical problem (in one experiment, epochs without any sound stimulation were included and a contrast between masker and silence showed robust activation). Instead, it likely reflects how potently the noise masker by itself activated auditory brainstem centers, perhaps leaving little room for an additional response to the tones (Melcher et al., 2000; Hawley et al., 2005). Because there was little information to be obtained concerning subcortical activation, gating was not used during fMRI of the remaining subjects.

MEG data were acquired with a Neuromag Vectorview system (Elekta Neuromag Oy, Helsinki, Finland) in a six-layer magnetically shielded room. This system comprises 204 planar gradiometers and 102 magnetometers at 102 positions, equally spaced around the head. The data were sampled continuously at 600 Hz with a 160 Hz low-pass filter and a 0.01 Hz high-pass filter. Before imaging, four position indicator coils were fixed to the subject's head and their position was digitized relative to landmarks on the head surface. The position of these coils relative to the MEG coils was determined at the beginning of the session. Eye movements were recorded from four electrodes on the outer canthi and above and below the left eye.

Data analysis. fMRI activation was detected using a multivariate, linear regression analysis and published basis functions (Harms and Melcher, 2003). Activation maps were derived by contrasting the four Δf_0 conditions (collectively) with the noise-masker baseline. The maps were coregistered with the whole head MPRAGE data, which were processed with Freesurfer (Cortechs, Charlestown, MA) to create an inflated projection of the cortical surface. Patches of the superior temporal plane were computationally "snipped" from the inflated surface and flattened for display (Fischl et al., 1999). The analysis was restricted to the auditory cortex (AC), including Heschl's gyrus (HG), planum temporale (PT), the circular sulcus, and the STG. Three regions of interest (ROIs) covering the entirety of AC were defined on the inflated brain (Fig. 3): (1) the medial HG (medial half of first HG, including common stem duplicatures), (2) the anterior AC (anterolateral HG, circular sulcus, and STG anterior to its intersection with HG), and (3) the posterior AC (PT, including complete duplicatures of HG, and the posterior STG). The medial-HG ROI was chosen to cover the region of the human analog of primary AC (A1) (Braak, 1978; Hackett et al., 2001), and the anterior ROI covered the region where the human analog of monkey area R is expected (Formisano et al., 2003). The exact extent of these areas cannot be derived from macroscopic boundaries, and the anterior ROI in particular is expected to extend into anterior belt regions (Braak, 1978; Galaburda and Sanides, 1980; Rivier and Clarke, 1997). The posterior ROIs, including PT and posterior STG, was intended to mainly comprise the areas of the posterior auditory belt cortex. Because there are no anatomical landmarks available among these areas or for the parabelt, they were all pooled into a single ROI.

Activation time courses were calculated as a weighted sum of basis functions and converted to percent signal change vs time (Sigalovsky and Melcher, 2006). The time course for a given ROI and subject was calculated as an average over the surface vertices (in the inflated cortical projection) showing significant activation in the "all conditions versus baseline" contrast described above ($p < 0.001$; F statistic, not corrected for multiple comparisons). Any vertex showing signal changes exceeding ± 8% were considered artifactual and discarded.

The MEG data were averaged over all 1600 ABBB quadruplet repetitions that occurred during each condition. The averaged time intervals began 100 ms before and ended 100 ms after each quadruplet. Intervals with raw signal amplitudes >5000 fT/cm^2 in any of the gradiometers

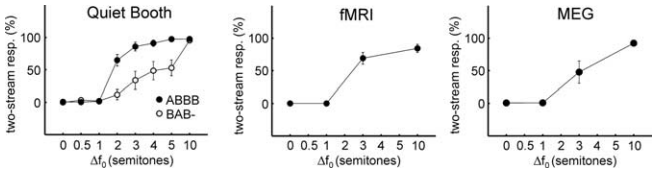

Figure 2. Behavioral streaming data. Behavioral data acquired in a quiet booth (left) and during fMRI (middle) and MEG (right). All data are for the same six listeners. In the booth, sequences comprising two patterns were compared: the ABBB pattern used during fMRI and MEG (filled circles), and the more standard BAB (open circles). On average, a substantially larger Δf_0 was required to induce stream segregation for the BAB pattern (Δf_0, $F_{(7,35)} = 74.21$, $p < 0.0001$; sequence, $F_{(1,5)} = 31.65$, $p < 0.01$; Δf_0 by sequence, $F_{(7,35)} = 13.80$, $p < 0.001$). Error bars indicate the SEM and are only shown if they exceed the size of the circle indicating the mean.

were considered artifactual and discarded. A baseline, calculated in an interval of 50 ms before the onset of the A tone, was subtracted from the response. A source analysis was performed separately for each subject. It used a boundary element model of the inner skull surface (Hämäläinen and Sarvas, 1989) and assumed two dipole sources (one in each auditory cortex). The diploes were fit to the P_1m elicited by the A tone (averaged across the 3- and 10-semitone, and the A-only conditions). Dipoles were fit simultaneously, with no further constraints, over a 20 ms interval covering the rising flank up to the peak. The dipoles were then used as a fixed spatial filter to derive source waveforms (Scherg, 1990) across all five conditions. Low-frequency, external artifacts in the magnetometers were corrected with a subspace projection, including three orthogonal topographies. These topographies were also included in the inverse matrix of the multiple dipole model.

For statistical analysis, the MEG amplitudes were submitted to the general linear model procedure for repeated measures (SAS, v.9.1; SAS Institute, Cary, NC) with the independent variables Δf_0 and hemisphere. For analysis of the fMRI data, anatomical ROI and waveform component (onset peak, trough, and sustained interval) were used as additional independent variables. Linear contrasts were calculated for Δf_0 (0, 1, 3, and 10 semitones). Where appropriate, p values were adjusted according to the Greenhouse-Geisser sphericity correction.

Results

Psychophysics

Behavioral data were obtained for two types of tone sequences: an ABBBABBB. . . sequence (where A and B designate tone complexes with different f_0) and a more commonly used BAB-BAB. . . configuration (where the hyphen represents a gap equal in duration to one tone). Figure 2 shows the percentage of "two streams" responses as a function of Δf_0, the f_0 difference between the A and B tones (expressed in semitones). This percentage represents an average across all trials, listeners, and time (from the first button press to the end of the 32 s sequence). The left panel in Figure 2 shows the data acquired in the quiet booth, in which both ABBB sequences and BAB sequences were used. The middle and right panels show the psychophysical data collected during fMRI and MEG, using just the ABBB sequences. In all three settings (booth, fMRI, MEG), the proportion of "two streams" responses increased with increasing Δf_0, as expected. For the ABBB sequence (filled circles), two streams were perceived more than half the time when Δf_0 was two semitones or greater. For the Δf_0 conditions used during fMRI and MEG, the percept was almost constantly one stream ($\Delta f_0 = 0$, 1 semitone) throughout the duration of each sequence or changed only during the first seconds of the sequence (~2 and ~8 s for 10 and 3 semitones, respectively) before reaching a response plateau comprising 85–95% two-stream responses in the quiet booth. Thus, the number of perceived streams was largely constant during the sequences used for fMRI and MEG.

During MEG, the 3-semitone Δf_0 condition was perceived as two streams somewhat less frequently than in the quiet booth and fMRI settings. However, this difference was not significant (MEG vs quiet booth, $t = 2.52$, $p > 0.05$; MEG vs fMRI, $t = 2.07$, $p > 0.05$), and was mostly caused by two listeners, who rarely indicated hearing two streams for this condition during MEG, but frequently in the booth and during fMRI. This difference in perception might have occurred because of contextual differences in stimulus presentation between the MEG, fMRI, and booth measurements. Because the 3-semitone condition was the most ambiguously perceived of the conditions used for MEG and fMRI, it may have been most susceptible to context differences. One possibility is that the presentation of the A tones alone in the MEG session (not done in fMRI or in the booth) provided a distinct percept of an A-tone-only stream, which might have biased the perception of the ambiguous stimuli toward one stream.

fMRI

Figure 3 shows areas of activation in the auditory cortex ($p < 0.001$) broken down by anatomically defined ROIs for each subject (S1–S6). Activation was detected by contrasting the four Δf_0 conditions used during fMRI (0, 1, 3, 10 semitones) against epochs where only the noise masker was present. In all listeners, large regions of the auditory cortex, including Heschl's gyrus and the planum temporale, were activated by the tone sequences, and in most cases, activity in the circular sulcus, anterior to the first Heschl's gyrus, was also observed.

Figure 4*A* shows the time course of the BOLD response, averaged across subjects and ROIs. The four traces correspond to the four Δf_0s, from 0 semitones at the top to 10 semitones at the bottom. In all four conditions, the BOLD response starts with a transient response that peaks ~6 s after sequence onset. Thereafter, for the 0- and 1-semitone Δf_0, the average signal falls slightly below the baseline level and then increases again gradually. For a Δf_0 of 3 and 10 semitones, the onset transient is greater in amplitude. Additionally, there is now considerable sustained activity throughout the 32 s block, and the signal does not fall below the baseline level after the onset response.

The effects apparent in Figure 4*A* were quantified by measuring the amplitude of the BOLD response at the onset peak, at the trough after the onset peak, and during the most sustained part of the response (time window 16–32 s). Amplitudes were determined separately for each auditory cortex ROI. For all three ROIs, response amplitude, by any of the three measures, increased significantly over the tested range of Δf_0s (general effect, $F_{(3,15)} = 16.19$, $p < 0.01$, contrast analysis with Δf_0s 0, 1, 3, and 10; linear, $F_{(1,5)} = 35.93$, $p < 0.01$; quadratic, $F_{(1,5)} = 12.10$, $p < 0.05$). Averaged over the whole auditory cortex, and most prominent in planum temporale and the posterior part of STG, the increase in response strength appears as a step between the Δf_0s of 1 and 3 semitones, similar to the increase in perceived stream segregation observed in the psychophysical data. In medial Heschl's gyrus, the response increase is more gradual. However, in the statistical analysis, this difference between Heschl's gyrus and the other auditory cortex regions produced no significant interaction of ROI × Δf_0 ($F_{(6,30)} = 2{,}51$; $p > 0.1$). All of the above effects were observed in both hemispheres and did not differ significantly between hemispheres.

Supplemental measurements were made to check that the increases in BOLD response shown in Figure 4 occurred because of the increase in Δf_0 and not because of the decrease in average f_0 that occurs because the $\Delta f_0 = 0$ condition comprises only tones with an f_0 of 178.2 Hz whereas the $\Delta f_0 = 10$ semitones condition

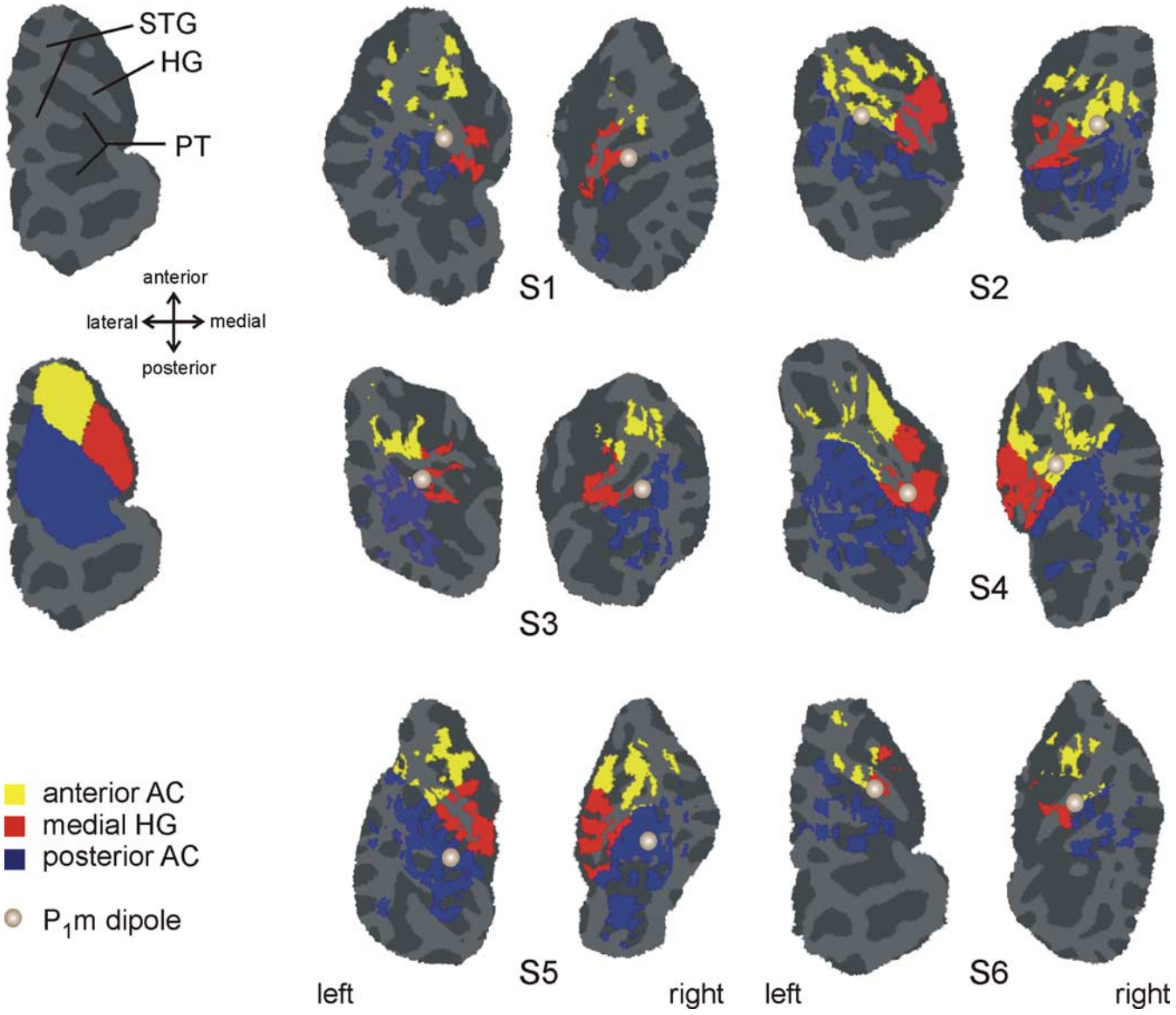

Figure 3. Auditory cortex activation produced by sequences of complex tones. The fMRI data for each subject (S1–S6) are displayed on flattened patches containing left or right auditory cortex. Color indicates significant activation in a contrast between the four Δf_0 conditions and epochs of noise masker only (cutoff $p < 0.001$, no correction for multiple comparisons). The particular color conveys the anatomically defined ROI within which the activation lies (yellow, anterior AC; red, medial HG; blue, posterior AC). The light and gray parts of each patch correspond to gyri and sulci, respectively. The top left patch indicates the location of HG, PT, and STG for left auditory cortex of subject S6. The bottom left patch shows the three anatomically defined regions of interest in their entirety for the same subject. The position of the P_1m dipole in each auditory cortex is indicated by a gray circle. The right dipole in subject 2 and the left dipole in subject 3 were elevated by ~5 mm to lie within the gray matter for display. All other dipoles were mapped to the gray matter without any correction.

comprises tones with an f_0 of 178.2 and 100 Hz. The supplemental measurements, in two listeners, used sequences comprising complex tones with a fixed f_0, specifically 178.2 and 100 Hz, the highest and lowest f_0s used in this study, which were presented in an AAA, as well as an AAAA pattern. The reduction in f_0 from 178.2 to 100 Hz generally produced a small but consistent decrease in sustained BOLD response amplitude, and therefore cannot explain the increase that would be needed to account for the amplitude trends in Figure 4.

MEG

A P_1m was consistently elicited by the A tones of ABBB sequences with Δf_0s of 3 and 10 semitones, and during a control condition in which only the A tones were presented (A—). The later waves were more variable. The location of the dipoles that were fitted to the P_1m are shown as gray circles superimposed on the activation maps in Figure 3. In most cases, the dipoles were located in the first or second Heschl's gyrus, except for subject 5, for whom they were located in planum temporale in both hemispheres. Relative to the ROI definitions used in fMRI, the dipole locations were scattered around the common border of the three ROIs. The grand-average source waveforms corresponding to the indicated dipoles are shown in Figure 5*A*. A schema of the stimulus (ABBB tone pattern) is shown at the bottom. It can be seen that in the condition where the B tones were absent, the A tones evoke a prominent P_1m, at a latency of 81 ms (SE, 6 ms). In the conditions where the B tones were present, the amplitude and latency of the P_1m were found to vary with Δf_0. The P_1m was largest at the 10-semitone Δf_0, attenuated but still detectable at the 3-semitones, and not statistically detectable at the 0- and 1-semitone Δf_0 (based on bootstrap based *t* intervals (two-tailed, $p < 0.05$) calculated for the whole source waveform). Its latency

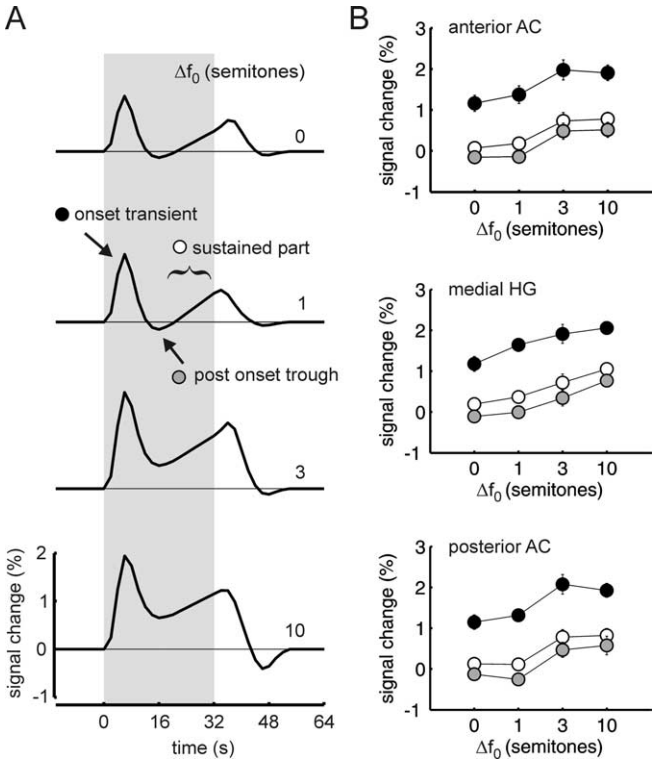

Figure 4. Increasing fMRI activation with increasing Δf_0. ***A***, Time course of fMRI activation in auditory cortex for $\Delta f_0 = 0$ semitones (top), 1, 3, and 10 (bottom). Each waveform is an average across subjects, ROIs, and hemispheres. ***B***, Amplitude of the onset transient (closed circles), postonset trough (gray), and sustained part (open) of the fMRI time courses plotted versus Δf_0 in semitones. Each point is an average across subjects. Bars indicate ± one SEM. Amplitude of the sustained part of the time courses is an average from 16 to 32 s. Amplitudes are separately plotted for the three ROIs (anterior AC, medial HG, and posterior AC) (compare Fig. 3).

was 99 ms (SE, 4 ms) at the 10-semitone Δf_0 and 109 ms (SE, 7 ms) at the 3-semitone Δf_0 ($t = 4.21$; $p < 0.0084$). Note that these effects cannot be caused by absolute changes in f_0, because the f_0 of the B tone remained constant throughout. Figure 5*B* plots the P_1m amplitude, measured in a fixed interval from 80 to 110 ms after A-tone onset, against Δf_0. The increase in amplitude with increasing Δf_0 was statistically significant (general effect, $F_{(3,15)} = 6.36$, $p < 0.05$; linear contrast analysis with Δf_0s 0, 1, 3, and 10, $F_{(1,15)} = 6.90$, $p < 0.05$) and not significantly different between the two hemispheres.

Two subjects showed an N_1m with a peak latency ~150 ms, which increased in magnitude with increasing Δf_0. In three other subjects, a second positive wave was observed at the 3- and 10-semitone Δf_0s; that wave was smaller in the condition where the B tones were absent, suggesting that it may have included contributions from the P_2m evoked by the A tone and from the P_1m evoked by the immediately following B tone. The field topography of the N_1m and the P_2m was somewhat different from that observed for the P_1m in the same subject. Because responses other than the P_1m evoked by the A tone were much less systematic, they were not analyzed in detail.

Discussion

The results demonstrate that complex tones played in a repeating sequence (ABBB) elicit both a significantly different percept and level of auditory cortical activity depending on the f_0 separation between tones. When f_0 separation was increased, the dominant percept changed from one to two sound streams, and the magnitude of both fMRI and MEG responses arising from cortex increased. Because the A and B tone complexes used in the present study were designed to have unresolvable differences in spectral fine structure and to produce almost identical tonotopic patterns of excitation in the auditory periphery, it is unlikely that either the perceptual or cortical activity changes observed with increasing f_0 separation occurred because of spectral differences between tones. Instead, the changes are likely attributable to the Δf_0-related differences in the temporal properties of the A and B tones. This dependence on temporal rather than spectral differences between A and B represents a crucial difference compared with previous imaging studies of streaming, all of which have used repeating sequences of pure tones (Gutschalk et al., 2005; Snyder et al., 2006; Wilson et al., 2007) or complex tones with gross spectral differences (Deike et al., 2004).

Given this fundamental difference in paradigm, it is noteworthy that the dependence of cortical activity on independent variable (either f_0 or pure tone frequency) is quite similar in the present and previous studies. This similarity suggests that the same neural mechanism(s) may underlie the cortical activity changes in both cases. Here, MEG component P_1m for the A tone (the tone held fixed as f_0 was varied) increased as the f_0 separation between tones was increased. Similarly, P_1m and N_1m (or their electric analogs) increased with increasing frequency separation in analogous pure tone stimulus paradigms (ABA-) (Gutschalk et al., 2005; Snyder et al., 2006) (note that an N_1m was not consistently observed in the present study because it was small at the relatively fast rate used and, in addition, was possibly canceled by overlapping positive responses). In fMRI, auditory cortical activation increased with increasing difference in f_0 in the present study, just as activation increased with increasing frequency difference in pure-tone paradigms (ABAB) (Wilson et al., 2007). A mechanism previously proposed to underlie the increases in cortical activity seen in pure-tone studies is neural "adaptation" or "forward suppression" in which (1) the neural response to one tone suppresses the responses to subsequent tones, and (2) as successive tones are made increasingly different in frequency (or the time interval between tones is increased), the degree of suppression decreases (Gutschalk et al., 2005; Wilson et al., 2007). This frequency-selective forward suppression model maps straightforwardly to the data of the present study if it is assumed that the strength of suppression in some auditory cortical neurons depends on the f_0 difference between tones (i.e., is f_0 selective). Such neurons must be distributed widely throughout auditory cortex to account for the fMRI data, which showed evidence of decreased suppression with increasing f_0 difference in both the core region of auditory cortex overlapping Heschl's gyrus (approximated by the medial HG ROI) and the surrounding nonprimary areas of the cortical belt (anterior and posterior ROIs).

Microelectrode studies in animals provide clear evidence for forward suppression at a neuronal level in auditory cortex (Calford and Semple, 1995; Brosch and Schreiner, 1997; Fishman et al., 2001, 2004; Kanwal et al., 2003; Ulanovsky et al., 2003; Bee and Klump, 2004, 2005; Wehr and Zador, 2005). Much of the data illustrates frequency-dependent suppression, with a greater effect of one tone on subsequent ones when the tones are close in frequency than when they are far apart. However, there are also demonstrations of suppression sensitive to other, "higher-level" sound features such as the combinatorial effects of the carrier and envelope of amplitude modulated (AM) tones (Barlett and Wang, 2005) or the relative phases of sounds presented to the two ears (Malone et al., 2002).

Neurons exhibiting Δf_0 dependent suppression, as suggested

by our fMRI and MEG data have not been reported in single-unit studies, although the "pitch-selective" neurons discovered previously by Bendor and Wang (2005) are likely candidates for showing such effects. However, such neurons are unlikely to account completely for effects reported here, because they appear to be localized to the abutting low-frequency ends of A1 and R, an area that likely maps to lateral HG in human AC (Formisano et al., 2003). The present findings indicate a more broadly distributed sensitivity to f_0 that may reflect neuronal sensitivity to the differing temporal properties of tones of different f_0, rather than a sensitivity to pitch per se.

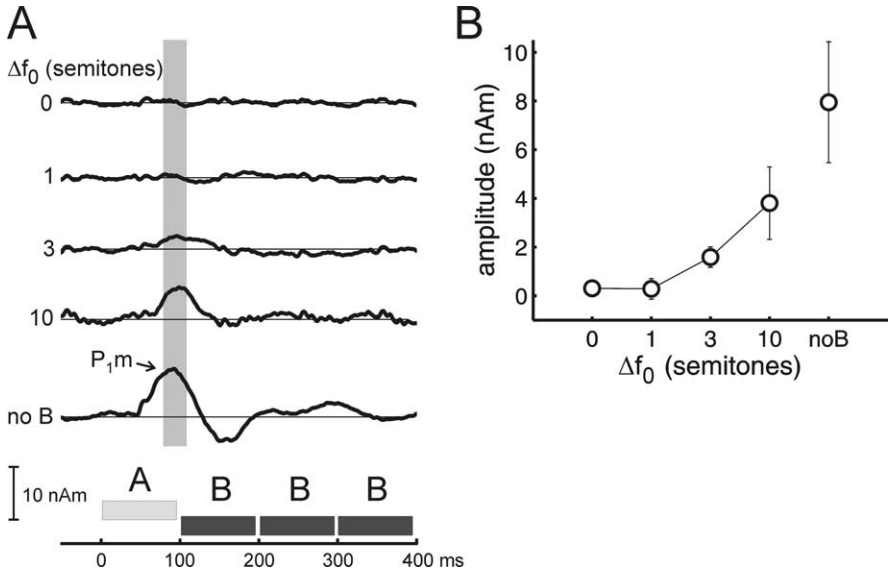

Figure 5. Increasing MEG source amplitude with increasing Δf_0. ***A***, Source waveforms for each semitone difference, averaged across subjects and hemispheres. ***B***, Average amplitude of the $P_1m \pm$ one SE measured in the fixed interval from 80 to 110 ms post stimulus onset (compare gray box in ***A***; latency corrected for air conduction delay in the MEG sound delivery system and adjusted to the middle of the 5 ms onset ramp).

Frequency-selective forward suppression of neural responses in the auditory cortex (or the equivalent in nonmammalian species) has been proposed to play an essential role in auditory streaming (Fishman et al., 2001, 2004; Kanwal et al., 2003; Bee and Klump, 2004, 2005; Micheyl et al., 2005, 2007). Specifically, it has been suggested that whether a sequence of pure tones is perceived as a single coherent stream or as two separate streams depends on the degree to which one tone influences the responses to subsequent tones. Whether the cortical activity resulting from such across-tone influences directly underlies the conscious perception of one vs two streams, or whether it represents processing at a preconscious stage remains unclear, as there is evidence in both directions. Using fMRI, and a physically stable but perceptually bistable stimulus sequence, Cusack (2005) showed that activity in the intraparietal sulcus, but not in the auditory cortex, differed during the perception of one vs two streams. However, Gutschalk et al. (2005), also using bistable sequences, did identify a neural correlate of the perception of one versus two streams in auditory cortex. Using MEG, they showed an increased response suggestive of reduced forward suppression when two streams were perceived. Similarly, the EEG study of Snyder et al. (2006) found correlates of the build-up of streaming (i.e., the increased tendency to hear two streams with increased time of exposure to the stimuli) in auditory cortex as did Micheyl et al. (2005) in their analysis of single-unit activity in monkey primary auditory cortex. One possibility that partly reconciles the existing data are as follows: In the case of bistable tone sequences, the conscious perception of distinct streams may be mediated directly by a small subset of temporally synchronized neurons whose activity in fMRI is swamped by contributions from other, unsynchronized neurons, but is detectable in MEG and EEG because of the sensitivity of these techniques to temporally synchronous neural activity.

In summary, the present findings suggest a general forward suppression or neural adaptation mechanism in auditory cortex, whereby responses to consecutive sounds that excite widely distributed and largely overlapping neural populations tend to be suppressed along multiple sound dimensions, including frequency and f_0. We hypothesize that this suppression mechanism shapes auditory cortical activity that underlies perceived stream segregation.

References

Barlett EL, Wang X (2005) Long-lasting modulation by stimulus context in primate auditory cortex. J Neurophysiol 94:83–104.

Beauvois MW, Meddis R (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. J Acoust Soc Am 99:2270–2280.

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.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165.

Bernstein JG, Oxenham AJ (2006) The relationship between frequency selectivity and pitch discrimination: effects of stimulus level. J Acoust Soc Am 120:3916–3928.

Braak H (1978) The pigment architecture of the human temporal lobe. Anat Embryol (Berl) 154:213–240.

Bregman AS (1990) Auditory scene analysis. Cambridge, MA: MIT.

Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77:923–943.

Calford MB, Semple MN (1995) Monaural inhibition in cat auditory cortex. J Neurophysiol 73:1876–1891.

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

Carlyon RP, Cusack R, Foxton JM, Robertson IH (2001) Effects of attention and unilateral neglect on auditory stream segregation. J Exp Psychol Hum Percept Perform 27:115–127.

Cedolin L, Delgutte B (2005) Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. J Neurophysiol 94:347–362.

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

Deike S, Gaschler-Markefski B, Brechman A, Scheich H (2004) Auditory stream segregation relying on timbre involves left auditory cortex. NeuroReport 15:1511–1514.

Edminster WB, Talavage TM, Ledden PL, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisition. Hum Brain Mapp 7:89–97.

Evans EF (2001) Latest comparisons between physiological and behavioural frequency selectivity. In: Physiological and psychophysical bases of audi-

tory function (Breebaart J, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R, eds), pp 382–387. Maastricht, The Netherlands: Shaker.
Evans EF, Pratt SR, Cooper NP (1989) Correspondence between behavioural and physiological frequency selectivity in the guinea pig. Br J Audiol 23:151–152.
Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II. Inflation, flattening and a surface-based coordinate system. NeuroImage 9:195–207.
Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40:859–869.
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.
Galaburda A, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. J Comp Neurol 190:597–610.
Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138.
Grimault N, Micheyl C, Carlyon RP, Arthaud P, Collet L (2000) Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. J Acoust Soc Am 108:263–271.
Grimault N, Bacon SP, Micheyl C (2002) Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am 111:1340–1348.
Guimaraes AR, Melcher JR, Talavage TM, Baker JR, Ledden P, Rosen BR, Kiang NY, Fullerton BC, Weisskoff RM (1998) Imaging subcortical auditory activity in humans. Hum Brain Mapp 6:33–41.
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.
Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol 441:197–222.
Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223.
Hämäläinen MS, Sarvas J (1989) Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. IEEE Trans Biomed Eng 36:165–171.
Harms MP, Melcher JR (2002) Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. J Neurophysiol 88:1433–1450.
Harms MP, Melcher JR (2003) Detection and quantification of a wide range of fMRI temporal responses using a physiologically motivated basis set. Hum Brain Mapp 20:168–183.
Hartmann WM, Johnson D (1991) Stream segregation and peripheral channeling. Music Percept 9:155–184.
Hawley ML, Melcher JR, Fullerton BC (2005) Effects of sound bandwith on fMRI activation in human auditory brainstem nuclei. Hear Res 204:101–110.
Houtsma AJ, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. J Acoust Soc Am 87:304–310.
Kanwal JS, Medvedev AV, Micheyl C (2003) Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. Network 14:413–435.
Malone BJ, Scott BH, Semple MN (2002) Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques. J Neurosci 22:4625–4638.
McCabe SL, Denham MJ (1997) A model of auditory streaming. J Acoust Soc Am 101:1611–1621.
Melcher JR, Sigalovsky IS, Guinan Jr JJ, Levine RA (2000) Lateralized tinnitus studied with functional magnetic resonance imaging: abnormal inferior colliculus activation. J Neurophysiol 83:1058–1072.
Micheyl C, Tian B, Carlyon BP, Rauschecker JP (2005) Perceptual organization of sound sequences in the auditory cortex of awake macaques. Neuron 48:139–148.
Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Wilson EC (2007) The role of auditory cortex in the formation of auditory streams. Hear Res 229:116–131.
Moore BC (2003) An introduction to the psychology of hearing, Ed 5. London: Academic.
Moore BC, Gockel H (2002) Factors influencing sequential stream segregation. Acta Acust United Acust 88:320–333.
Rivier F, Clarke S (1997) Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. NeuroImage 6:288–304.
Roberts B, Glasberg BR, Moore BCJ (2002) Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. J Acoust Soc Am 112:2074–2085.
Scherg M (1990) Fundamentals of dipole source analysis. In: Auditory evoked magnetic fields and electric potentials, advances in audiology, Vol VI (Grandori F, Hoke M, Romani GL, eds), pp 40–69. Basel: Karger.
Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci USA 99:3318–3323.
Sigalovsky IS, Melcher JR (2006) Effect of sound level on fMRI activation in human brainstem, thalamic and cortical centers. Hear Res 215:67–76.
Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18:1–13.
Sussman E (2005) Integration and segregation in auditory scene analysis. J Acoust Soc Am 117:1285–1298.
Sussman E, Ritter W, Vaughan Jr HG (1999) An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology 36:22–34.
Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398.
van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. Eindhoven, The Netherlands: University of Technology.
Vliegen J, Oxenham AJ (1999) Sequential stream segregation in the absence of spectral cues. J Acoust Soc Am 105:339–346.
Vliegen J, Moore BC, Oxenham AJ (1999) The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. J Acoust Soc Am 106:938–945.
Wehr M, Zador AM (2005) Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47:437–445.
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.