



Published in final edited form as:

*J Neurophysiol.* 2007 March ; 97(3): 2230–2238.

## Cortical fMRI Activation to Sequences of Tones Alternating in Frequency: Relationship to Perceived Rate and Streaming

E. Courtenay Wilson<sup>1,2,3</sup>, Jennifer R. Melcher<sup>1,3,4</sup>, Christophe Micheyl<sup>2</sup>, Alexander Gutschalk<sup>2,3,4</sup>, and Andrew J. Oxenham<sup>1,2</sup>

<sup>1</sup>Speech and Hearing Bioscience and Technology Program, Harvard–MIT Division of Health Sciences and Technology, Cambridge

<sup>2</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge

<sup>3</sup>Eaton–Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts

<sup>4</sup>Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts

### Abstract

Human listeners were functionally imaged while reporting their perception of sequences of alternating-frequency tone bursts separated by 0, 1/8, 1, or 20 semitones. Our goal was to determine whether functional magnetic resonance imaging (fMRI) activation of auditory cortex changes with frequency separation in a manner predictable from the perceived rate of the stimulus. At the null and small separations, the tones were generally heard as a single stream with a perceived rate equal to the physical tone presentation rate. fMRI activation in auditory cortex was appreciably phasic, showing prominent peaks at the sequence onset and offset. At larger-frequency separations, the higher- and lower-frequency tones perceptually separated into two streams, each with a rate equal to half the overall tone presentation rate. Under those conditions, fMRI activation in auditory cortex was more sustained throughout the sequence duration and was larger in magnitude and extent. Phasic to sustained changes in fMRI activation with changes in frequency separation and perceived rate are comparable to, and consistent with, those produced by changes in the physical rate of a sequence and are far greater than the effects produced by changing other physical stimulus variables, such as sound level or bandwidth. We suggest that the neural activity underlying the changes in fMRI activation with frequency separation contribute to the coding of the co-occurring changes in perceived rate and perceptual organization of the sound sequences into auditory streams.

### Introduction

A brief sound, repeated many times, changes its perceptual character depending on the rate at which it is repeated. At slow repetition rates, each individual repetition is heard as a distinct perceptual event. At higher rates (>10–12 times/s), the individual sounds blur into a single, rough-sounding percept. At even higher rates (>30–50 times/s), the sound begins to have a tonal quality, with a pitch corresponding to the rate of repetition (e.g., Warren 1999).

The effect of repetition rate on activation in human auditory cortex was examined in functional magnetic resonance imaging (fMRI) studies using prolonged (e.g., 30 s) sequences of various stimuli, including broadband and narrowband noise bursts, tone bursts, clicks, and speech (Binder et al. 1994; Harms and Melcher 2002; Harms et al. 2005; Tanaka et al. 2000). In

---

Address for reprint requests and other correspondence: E. C. Wilson, Eaton–Peabody Laboratory, MEEI, 243 Charles Street, Boston, MA 02114 (E-mail ecwilson@mit.edu).

Present addresses: A. J. Oxenham and C. Micheyl: Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455; A. Gutschalk: Department of Neurology, University of Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany.

response to sequences with long gaps between each sound ( $\geq 200$  ms), activation amplitude generally increases as the rate of sound presentation is increased. However, when the silent gap between successive sounds is  $\leq 200$  ms, activation has a different dependency on sound presentation rate. First, the overall activation amplitude (averaged over sound duration) begins to decrease with increasing rate. Second, the time course of activation is profoundly affected: at low rates activation is mainly sustained throughout the sequence presentation, whereas at high rates it becomes more phasic, dominated by prominent response peaks just after sequence onset and offset (Harms and Melcher 2002; Harms et al. 2005).

The fMRI rate studies undertaken so far have mainly considered sequences consisting of repetitions of the same or similar sounds rather than, for instance, sequences of tones alternating in frequency (i.e., sequences of the form AB-ABAB..., where A and B are tones of different frequency). A sequence of alternating tones has the interesting property that it can be readily manipulated to form very different percepts simply by changing the frequency separation between tones. When there is no frequency separation (i.e., all tones have the same frequency), the tones are heard as a coherent sequence with a perceived rate that corresponds to the physical presentation rate of the tones. However, when a sufficiently large frequency difference is introduced, the A and B tones segregate perceptually into two independent sequences or "streams" (e.g., Miller and Heise 1950), in a phenomenon commonly referred to as "auditory stream segregation" (Bregman 1990). When this happens, the perceived rate of each stream equals that of the A- (or B-) tones alone; i.e., the perceived rate is half the physical rate of the overall ABAB sequence.

The changes in percept with increasing frequency separation of ABAB tone sequences raise the following question: Does the cortical fMRI activation produced by these sequences change with frequency separation in a manner predictable from the perceived rate? If it does, it would suggest that not only the perceived, but also the physical rate of sound may be encoded in the activity of auditory cortex. If it does not, it would indicate that the previously reported fMRI rate dependencies for auditory cortex reflect activity at a processing stage before conscious perception.

In this study, two experiments were conducted to distinguish between the two possibilities just outlined. Both examined the fMRI activation in human auditory cortex in response to sequences of pure tones. A first, preliminary experiment, established that the fMRI activation produced by alternating (ABAB) tone sequences with either a very large frequency separation or no frequency separation differed in magnitude and time course in a manner comparable to sequences of same-frequency tones presented at low and high physical rates, respectively. In the second, main experiment, sequences with no, intermediate, and large frequency separations were tested, while subjects simultaneously reported how the sequences were perceived. Psychophysical tests on the same subjects in a quiet booth helped establish that the acoustic noise produced by the scanner did not prevent the stimuli from being perceptually organized as they would be in complete quiet.

The results show that increasing the frequency separation between successive tones produces progressive changes in cortical fMRI activation as well as in subjects' perception of the stimuli and that the changes in activation and perception occur in a coordinated way. The results further show that activation changes in a manner consistent with the perceived rate of the stimuli. The findings suggest that the neural activity underlying the changes in fMRI activation may contribute to the coding of the co-occurring changes in perceptual organization and perceived rate of the sound sequences.

Portions of this work were presented at the 27th and 28th Mid-Winter Meetings of the Association for Research in Otolaryngology (2004, 2005).

## Methods

### Subjects

Ten audiometrically normal adult volunteers (ages 21–63 yr, seven female, all right-handed) with no known neurological disorders took part in this study. Seven subjects participated in our main experiment that included both psychophysical and fMRI testing. One of the seven subjects was not able to generate a complete fMRI data set because of time constraints, resulting in only two of four conditions for this subject. The three remaining subjects participated in the preliminary experiment, which involved only fMRI.

The experimental protocol was approved by the Institutional Review Boards of the Massachusetts General Hospital, Massachusetts Eye and Ear Infirmary, and Massachusetts Institute of Technology. All subjects provided their written, informed consent before testing.

### Stimuli

Both experiments used 32-s sequences of tone bursts organized temporally into a repeating ABAB... pattern, where each A and B represents a pure tone (100-ms duration, including 5-ms raised-cosine on and off ramps). There were no silent gaps between consecutive tones. For the main experiment, the A-tone frequency was always 600 Hz. The frequency of the B tone was constant within each sequence but varied from one sequence to the next, being 0, 1/8, 1, or 20 semitones above the A-tone frequency (i.e., 600, 604, 636, or 1,905 Hz, respectively). Our preliminary experiment used only two frequency separations between A and B tones (0 and 20 semitones) and decreased the B-tone frequency relative to the A tone (instead of increasing it as in the main experiment). Specifically, the A-tone frequency was fixed at 1,900 Hz and the B tone was either 0 or 20 semitones below it (i.e., 1,900 or 599 Hz, respectively). This experiment also used sequences of constant-frequency tones consisting of only the high (i.e., A) or only the low (B) tones of the 20-semitone sequence. In these “high-frequency, low-rate” and “low-frequency, low-rate” sequences, tones of either 1,900 Hz (A\_\_A\_\_...; “high-frequency, low-rate”) or 599 Hz (B\_\_B\_\_...; “low-frequency, low-rate”) were separated by a silent gap with a duration equal to that of the missing tone (100 ms).

The different sequences were always presented in random order. Each sequence was presented four times in each fMRI session and five times in each psychophysical session in the quiet test booth. The sound pressure level (SPL) of each tone was 75 dB.

### Psychophysical measurements

For our main experiment, each subject was tested psychophysically, first in a sound-treated booth and then during fMRI. In both settings, subjects were instructed to indicate whether they perceived the sequence being presented as a single stream of rapidly repeating tones (“one high-rate stream”) or as two separate streams of lower-rate tones (“two low-rate streams”), to respond as soon as possible after the onset of the sequence, and to update their response each time (and as soon as) the percept changed. In the booth, subjects pressed one of two computer keys to indicate the number of streams heard. During fMRI, subjects controlled a handheld knob to illuminate either one or two lights depending on whether they heard one or two streams. In the preliminary experiment, subjects were instructed to listen passively to the stimuli and no psychophysical measurements were made.

### Imaging

Subjects were imaged using a 3-Tesla head and neck scanner (Siemens Allegra) while sounds were presented through headphones (GEC Marconi). T2-weighted anatomical images were acquired of nine slices oriented parallel to the Sylvian fissure and covering the superior temporal lobe (slice thickness = 4 mm; gap between slice = 1 mm; in-plane resolution:  $3.1 \times$

3.1 mm; pulse sequence: gradient echo; TE = 30 ms; flip angle = 90°). Functional images of the same slices were acquired while presenting the 32-s tone sequences alternately with 32-s silent periods. Mitigation of the scanner acoustic noise during fMRI was achieved by turning the scanner coolant pump off and by using a “sparse” imaging protocol that intersperses image acquisition with long intervals free of the acoustic noise produced by the scanner gradient coils. With the coolant pump off, the ambient noise level between acquisitions was comparable to that of a quiet room (41-dB SPL A weighted at the ear).<sup>1</sup> The nine slices were imaged in brief (<1-s) clusters spaced by 8 s (Edmister et al. 1999; Hall et al. 1999). The timing of clusters was systematically staggered by 0, 2, 4, or 6 s, relative to the onset of the stimulus sequence so that the time course of fMRI activation could be determined with 2-s resolution.

### Analysis of psychophysical data

The psychophysical data were averaged across all presentations of a given stimulus sequence and expressed as a percentage of “two low-rate stream” judgments as a function of time. A time-averaged percentage of these responses was calculated by averaging over the entire sequence presentation (0–32 s). The analyses were performed separately for the booth and scanner data.

The psychophysical data collected in the scanner were further examined for possible effects of scanner acoustic noise. This involved temporally shifting the psychophysical data so that the timing of the image acquisitions coincided across presentations of a given sequence (instead of being temporally staggered). Specifically, the psychophysical data for each presentation were shifted in time by an amount equal to the time between presentation onset and the first image acquisition during the presentation. Then, for each subject, the data were averaged across presentations. In these temporally realigned psychophysical data, systematic effects of the scanner noise on the listeners' responses would manifest themselves as variations in the percentage of “two stream” judgments, with a periodicity equal to the scanner interacquisition time (8 s).

### Analysis of fMRI data

Activation was detected using a general linear model (GLM), which operated on a set of basis functions reflecting different temporal components of fMRI activation in the auditory cortex (Harms and Melcher 2003). This approach models the signal versus time within each voxel as a weighted sum of basis functions and identifies “active” voxels based on the goodness of fit of this model. Activation maps were created for each stimulus sequence by estimating (using an *F*-statistic; Fomby et al. 1984), for every voxel, whether the amplitude of any of the basis functions was significantly different from zero.

Four measures of fMRI activation were calculated:

- 1) Activation extent was calculated as the number of voxels meeting a threshold criterion of  $P < 0.001$ .
- 2) The time course of activation was calculated two ways. For the amplitude calculations below, it was calculated as a weighted sum of basis functions (weightings determined by the GLM fit and expressed as percentage change in image signal). For display in the figures, it was calculated from the raw image data by averaging across all presentations of a given sequence after accounting for the staggered timing between sequence onset and image acquisition.

---

<sup>1</sup>The level of the gradient noise, calculated over the time window of gradient activity, was about 70 dBA at the ear. Gradient and intervening ambient noise levels at the ear were estimated from measurements of unattenuated noise by correcting for the attenuation provided by the headphones (reported in Ravicz and Melcher 2001). The methods for measuring the scanner noise are described in Ravicz et al. (2000).

3) The amplitude of fMRI activation was calculated by averaging the activation over time from stimulus onset to 10 s after stimulus offset for a total time of 42 s.

4) A “waveshape index” was calculated to quantify the shape of the time course of activation. This index, calculated from the basis function amplitudes, describes time course waveshape on a continuum from completely sustained (index = 0) to completely phasic (= 1) (details in Harms and Melcher 2003).

fMRI activation was quantified in two anatomical regions of interest (ROIs), one corresponding to the posteromedial 2/3 of Heschl's gyrus (the more anterior one in hemispheres with two Heschl's gyri) and the other corresponding to planum temporale (PT). In hemispheres with two Heschl's gyri, the more posterolateral one was included in the PT ROI.

## Results

### Preliminary experiment

Figure 1 shows cortical fMRI activation from our preliminary experiment examining 1) responses to ABAB sequences with two extreme frequency separations (0 and 20 semitones) and 2) responses to the A and B tones of the 20-semitone sequence presented separately (A\_\_A\_\_ and B\_\_B\_\_). The data shown are for Heschl's gyrus only, but those for the other analyzed cortical division (PT) showed the same major trends. Data from this experiment allowed several comparisons. The first was between the ABAB, 0-semitone sequence (i.e., 1,900-Hz tones presented at a rate of 10/s) and the A\_\_A\_\_ sequence (i.e., 1,900-Hz tones presented at a rate of 5/s). Activations in response to these two conditions are shown in Fig. 1A as “high-frequency high-rate” and “high-frequency low-rate,” respectively. This comparison demonstrated consistency with previous findings in that the amplitude of fMRI activation *increased* when the physical rate of tone presentation *decreased* and the time course shifted from phasic to highly sustained (Harms and Melcher 2002). The trends were apparent in all three subjects as an increase in activation amplitude (time-averaged between 0 and 42 s) and as a decrease in a numerical index of time course waveshape (which ranges from 0 for completely sustained to 1 for completely phasic).

A second comparison was made between the two ABAB sequences with frequency separations of 0 and 20 semitones. As shown in Fig. 1B, the average fMRI activation for the 20-semitone sequence was greater in amplitude and more sustained in waveshape than the activation for the 0-semitone sequence. The same trends were apparent in each individual subject; time-averaged amplitude and the index of waveshape increased on both Heschl's gyrus and in PT in every instance. The differences in activation between the 0- and 20-semitone sequences resemble the differences between high- and low-rate sequences in Fig. 1A but, unlike Fig. 1A, occurred without any change in physical tone presentation rate. Instead, they coincided with the introduction of a frequency separation between tones, which is generally known to correspond to a change in perceived rate from high rate for the 0-semitone ABAB sequence to low rate for the 20-semitone sequence.

A third and final comparison is illustrated in Fig. 1C. This comparison was between the activation produced by the ABAB 20-semitone sequence and a superposition of the activations produced by the A tones and B tones presented alone (i.e., response to A\_\_A\_\_ plus response to B\_\_B\_\_). The summed activation exceeded the 20-semitone activation (on average and in each individual), demonstrating that the fMRI response to the ABAB 20-semitone sequence was less than the sum of its parts and superposition did not hold. This result is not surprising because response amplitude for the ABAB 20-semitone sequence (time average from 0 to 42 s) exceeded that for the A tones (A\_\_A\_\_; Fig. 1A) and B tones (B\_\_B\_\_); not shown) alone by a factor of about 1.3 and 1.8, respectively, amounts consistent with the 3 dB greater root



mean square sound level of the ABAB sequence (Hall et al. 2001; Hart et al. 2003; Sigalovsky and Melcher 2006). Thus the lack of superposition may simply reflect the compressive relationship between sound level and change in fMRI activation observed when using single sounds.

## Main experiment

**Psychophysical Responses**—Figure 2 displays the psychophysical data obtained in five of the seven subjects tested in our main experiment. The bar graphs in Fig. 2, *top* show the percentage of time that subjects reported hearing two low-rate streams time-averaged across the 32-s sequence. The hatched and solid bars indicate data taken in the sound-attenuating booth and in the scanner, respectively. As expected, the two extreme frequency separations of 0 and 20 semitones were heard overwhelmingly as a single high-rate stream and two low-rate streams, respectively. The 1/8th-semitone frequency separation resulted mostly in a “one stream” percept that was not significantly different from the responses in the 0-semitone condition. The 1-semitone separation produced mixed results: two subjects heard mostly two low-rate streams (with individual “two low-rate streams” responses of 96.4 and 77% on average); responses from the three other subjects oscillated between “one high-rate stream” and “two low-rate streams” over much of the sequence duration, resulting in intermediate time-averaged percentages of “two low-rate stream” responses (44.3, 65.4, and 51.4%).

The data from the remaining two subjects participating in this experiment were excluded from Fig. 2 because they reported an anomalously high percentage of two-stream responses (63%) in the 0-semitone condition, where no stream segregation should have occurred. These anomalous responses, reported in both the booth and scanner, suggest that these two subjects did not fully understand the instructions. In fact, after all testing was complete, one of the two subjects reported spontaneously that she had probably not performed the task correctly. Because there was no reason to doubt the validity of the fMRI data in these two subjects (i.e., their values were not significantly different from those of the other subjects), these data are included in the fMRI results below. The trends in the fMRI data were the same with or without these two subjects.

The fact that the psychophysical data obtained in the scanner were not significantly different from the data obtained in the quiet conditions of the booth (compare hatched and solid bars in Fig. 2A) suggests that the acoustic noise created by the scanner gradient coils had little or no effect on the perceived organization of the tone sequences. This conclusion is supported by the fact that plots of the percentage of two low-rate stream responses versus time (Fig. 2, *bottom*) showed the classic buildup of streaming expected for larger frequency separations (i.e., a steady increase in the seconds after sequence onset). This conclusion was further supported by an additional analysis of the psychophysical data collected in the scanner, which involved temporally aligning the subjects' responses relative to the time of image acquisition instead of sequence onset (see methods). The realigned data were then scrutinized for systematic changes in the percentage of “two low-rate stream” responses occurring around each image acquisition. None was found. The lack of effect of image acquisition for the 0-, 1/8-, and 20-semitone conditions was obvious because the responses were so stable (see above). Because of the bistability of the percept for the remaining condition (and corresponding fluctuations in response), a qualitative examination of the data could not conclusively rule out acquisition-related response changes, so the data were also examined quantitatively by comparing the time-averaged responses over the 4 s before and after each acquisition. Although some subjects showed a trend toward a reduced probability of reporting a two-stream percept after each acquisition than before, a repeated-measures ANOVA revealed no significant difference between the proportion of two-stream responses before and after the acquisition [ $F(1,4) = 5.38, P = 0.081$ ; Green-house–Geisser correction was applied wherever required]. Thus

the results failed to show a significant effect of scanner gradient noise on streaming, despite an apparent trend for some of the subjects.

**Extent of fMRI Activation**—Figure 3A shows typical activation maps from one subject for sequences with different frequency separations between the A and B tones. Here, the extent of activation was considerably greater at moderate to large frequency differences (1 and 20 semitones) than at null or small frequency differences (0 and 1/8th semitone), a pattern found in all subjects tested. The region activated by the stimuli always included both Heschl's gyri and PT. In an average across subjects, both of these areas showed an increase in activation extent with increasing frequency separation (Fig. 3B). A two-way ANOVA (region  $\times$  frequency separation) confirmed a significant effect of frequency separation [ $F(3,18) = 4.108, P = 0.04$ ]. It also showed no effect of anatomical region, indicating that the effect was similarly present in both Heschl's gyrus and PT. The difference in activation extent between the 0- and 20-semitone conditions was highly significant [ $F(1,9) = 15.713, P = 0.003$ ] and no significant difference was found between the 0- and 1/8th-semitone conditions. Activation extent differed between the 1- and 20-semitone conditions [ $F(1,5) = 8.425, P = 0.034$ ], but the difference was significant for only Heschl's gyrus [ $P = 0.018$ ;  $P = 0.920$  for PT].

**Amplitude of fMRI Activation**—Figure 4 shows the effect of frequency separation on the amplitude of fMRI activation. In time courses averaged across subjects (Fig. 4A), amplitude during and immediately after the stimulus increased progressively with increasing frequency separation. This trend, apparent in both Heschl's gyrus and PT, is captured by the time-averaged amplitude, which covers the period from the stimulus onset to 10 s after the stimulus offset (Fig. 4B). A two-way ANOVA on the time-averaged amplitude (region  $\times$  frequency separation) showed a highly significant effect of frequency separation [ $F(1.7,10.25) = 25.106, P = 0.0001$ ] and no effect of region.

**Time Course of fMRI Activation**—Figure 5 shows activation time courses after normalization to highlight waveshape instead of amplitude differences between stimulus conditions. Each trace was obtained by normalizing the time course for each subject to have a maximum value of one and then averaging across subjects. In both Heschl's gyrus (not shown) and PT, the time courses for the 1/8th- and 0-semitone conditions showed a peak just after the onset and offset of the stimulus, whereas the time courses for the 1- and 20-semitone conditions are more sustained. These time-course changes with frequency separation were reflected in the waveshape index (in both Heschl's gyrus and PT), which decreased with increasing frequency separation in five of the six subjects in whom complete data (at all four frequency separations) were obtained. In this subgroup of five listeners, the decrease in waveshape index with increasing frequency separation was statistically significant, as revealed by a two-way repeated-measures ANOVA (region  $\times$  frequency separation) [ $F(1.5,6.0) = 8.35, P = 0.022$ ] followed by a planned linear-contrast analysis across the levels of the frequency-separation factor [ $F(1,4) = 12.20, P = 0.025$ ], indicating a trend for the waveshapes to become more “sustained” with increasing frequency separation. Consistent with the observation that the trend was present in both Heschl's gyrus and PT,<sup>2</sup> there was no significant interaction between the region and frequency-separation factors, neither in the repeated-measures ANOVA [ $F(2.16,8.64) = 0.563, P = 0.601$ ] nor in the linear-contrast analysis [ $F(1,4) = 0.018, P = 0.901$ ]. These findings are qualified by the observation that the remaining listener with complete data, who was not included in the preceding analysis, showed a trend in the opposite direction (i.e., an increase in waveshape index with increasing frequency separation). Because of the relatively

<sup>2</sup>Although there was no significant difference between regions when all frequency separations were considered together, there was a difference when separations (i.e., 0 and 1/8th semitone) yielding more phasic responses (but not those yielding more sustained responses, i.e., 1 and 20 semitone) were considered alone. Specifically, stimuli yielding phasic responses on Heschl's gyrus yielded slightly more phasic responses on PT, as previously observed (Harms et al. 2005).

small sample size involved, this single-subject departure from an otherwise consistent trend was sufficient to make the outcome of the repeated-measures ANOVA nonsignificant when the data from this listener were pooled with those from the other five listeners [ $F(1.17,5.87) = 1.96, P = 0.216$ ]. Although the reasons for this departure remain unclear, the observation of a trend in the same direction in both Heschl's gyrus and PT in five of six listeners with complete data indicates that for most of the listeners tested in this study, the waveshape index decreased with increasing frequency separation between consecutive tones.

## Discussion

The present results demonstrate that the fMRI response of auditory cortex to a sequence of tones alternating in frequency (i.e., ABAB) profoundly depends on the frequency separation between tones. As the frequency separation increased, so did the extent and amplitude of activation in fMRI. Additionally, the time course of fMRI activation changed from phasic for small-frequency separations (distinct signal peaks at the onset and offset of the sequence) to more sustained for larger-frequency separations (less decline in signal after the onset and no offset peak): all three subjects from the preliminary experiment showed this trend, as did five of the six subjects (with complete data sets) in the main experiment. These trends in activation were observed for both Heschl's gyrus and PT, regions incorporating primary and nonprimary auditory cortical areas, respectively.

By measuring behavioral responses during scanning, it was established that the perceptual organization of the ABAB tone sequences also varied systematically with frequency separation. The systematic changes in percept, although expected for quiet conditions, were not a given during scanning because one could easily imagine the perception of the sequences being disrupted by the scanner acoustic noise. However, by analyzing the behavioral data collected in the scanner in multiple ways and comparing it with data collected in a quiet booth, we determined that subjects had similar perceptual experiences in the two settings. Specifically, there was a similar buildup of streaming after sequence onset. There were also similar changes in the perceptual organization of sequences with frequency separation: sequences were heard as one fast-rate stream when the frequency separation was very small or null, as two streams, each with a low repetition rate, when the separation was large, and as a percept fluctuating between these two extremes for intermediate frequency separations.

The covariation of perception and cortical fMRI activation with the frequency separation between the two tones in the sequence may be fortuitous and thus may not reflect any causal relationship. On the other hand, the covariation may reflect a relationship in which the neural activity underlying the fMRI activation changes helps give rise to the co-occurring changes in the perceived rate and number of streams. In light of the nature of the activation and perceptual changes, we are inclined to hypothesize the latter possibility. Phasic to sustained changes in activation time course, as occurred here, were shown to be highly specific to changes in sound temporal envelope characteristics such as rate; there is, for instance, little or no change in time course with sound intensity or bandwidth (Harms et al. 2005; Sigalovsky and Melcher 2006). Furthermore, the changes in time course associated with changes in perceived rate were in a direction that one would predict if we had substituted actual changes in rate for the perceived ones: time courses were more phasic when sequences were perceived as having a fast rate and were more sustained when perceived to be slow. Given the specific way that activation and perceived rate covaried, a causal relationship between the two seems likely.

### Possible neural mechanisms underlying the dependency of fMRI activation on frequency separation

Because the dependency of fMRI activation on frequency separation closely resembles the previously reported dependency of activation on the physical repetition rate of sound, the two



dependencies may reflect similar underlying neural mechanisms. In one previous study varying the physical repetition rate of noise bursts, it was proposed that the rapid decline after the initial onset peak of phasic fMRI time courses reflects forward suppression<sup>3</sup>—that is, a suppressive effect of one burst on the neural response to subsequent bursts (Harms and Melcher 2002). This interpretation was supported by comparisons of fMRI activation for small numbers (e.g., one or two) of consecutive bursts. It was further proposed that an increase in fMRI activation amplitude with decreasing rate, and the co-occurring shift from phasic to more sustained time courses, occurred because the degree of suppression lessened as the time between bursts increased (i.e., rate decreased). A similar release from forward suppression may underlie the changes that occurred here with increasing frequency separation. However, a difference compared with the rate-manipulated noise burst sequences is that, here, a release putatively occurred because of an increase in spectral rather than temporal separation between successive bursts. A remaining change in fMRI activation was the emergence of a peak after sequence offset as the frequency separation between successive bursts was reduced. This peak closely resembles the off peak that emerges when the temporal separation between bursts in a noise burst sequence is reduced (i.e., rate is increased). Based on previous experiments examining the physiological basis of off peaks in fMRI activation (Harms and Melcher 2002), we propose that the off peak seen here reflects a neural response to sequence offset.

Single-unit recordings from anesthetized cats (Brosch and Schreiner 1997, 2000; Brosch et al. 1999; Calford and Semple 1995) and awake primates (Bartlett and Wang 2005) provide evidence for forward suppression in the neural activity of primary auditory cortex. The results indicate that, under certain stimulus conditions, the response to a “probe” tone was suppressed by a preceding “masker” tone. Maximal suppression was found when the masker frequency was within the neuron's excitatory receptive field, close to that of the probe, and there was minimal or no delay between the probe onset and masker offset. The suppression usually decreased as the frequency separation and temporal delay between the masker and probe increased. For some units and stimulus conditions, responses to the probe tone were enhanced rather than diminished by the preceding “masker” (Bartlett and Wang 2005; Brosch and Schreiner 2000; Brosch et al. 1999). Although some auditory cortical neurons may well have shown a similar enhancement in the present study, this effect appears to have been overwhelmed by suppressive effects in a majority of the neurons contributing to the measured fMRI activation.

Evidence of forward suppression in auditory cortex was also previously observed in microelectrode, evoked potential, and magnetic recording studies using sequences of alternating-frequency tones (ABAB, as in the present study), as well as sequences of tone triplets (ABA\_\_), a stimulus eliciting similar changes in perceived rate and streaming with frequency separation (Bee and Klump 2004, 2005; Butler 1968; Fishman et al. 2001, 2004; Gutschalk et al. 2005; Kanwal et al. 2003; Micheyl et al. 2005). Whereas with probe/masker pairs only the response to one tone (i.e., the second of the pair) can be affected by forward suppression, with multiple-tone sequences, each tone can potentially affect the response to any subsequent tones. Thus there is the potential for an accumulation of suppression during the sequence. The evoked potential and magnetic recording results demonstrate that a net forward suppressive effect is manifested by the temporally synchronized population activity underlying evoked potential and magnetic responses from human auditory cortex (Butler 1968; Gutschalk et al. 2005), as well as by the temporally averaged neural activity reflected in fMRI activation.

---

<sup>3</sup>The previous paper (Harms and Melcher 2002) used the term “adaptation” instead of “forward suppression.” However, we prefer the latter because it does not imply a physiological process behind the suppression effects (unlike “adaptation,” which can imply synaptic depletion, for example).

## Related neuroimaging studies

Two recent studies examined fMRI responses to sound sequences similar to those of the present study. Cusack (2005) examined responses to repeating ABA triplets in a study designed to identify correlates of streaming without confounding changes in the physical stimulus. The experiments involved measuring activation during the presentation of sequences that elicited a bistable percept (i.e., spontaneously fluctuating between one and two streams) and comparing activation during the perception of one versus two streams. Cusack (2005) found differential activation, corresponding to the percepts of one and two streams, within the intraparietal sulcus, an area previously implicated in feature binding in the visual domain and in cross-modal integration (Calvert 2001; Shafritz et al. 2002). The intraparietal cortex was not fully encompassed by the scans in the present experiment and was not incorporated into our analysis. Thus we cannot say whether it was differentially activated in a manner consistent with Cusack's results.

Another finding of Cusack (2005) was a lack of differential activation in auditory cortex, based either on perception (one or two streams) or on frequency separation. There are at least two ways in which Cusack's null finding may be reconciled with our finding of both amplitude and waveshape effects in auditory cortex activation. The first relates to the perceived rate of Cusack's ABA triplets compared with our repeating AB stimuli. With repeating AB pairs, the perceived rate of the sequence halves as the percept changes from one to two streams. The relationship between perceived rate and streaming is more complex with the ABA triplets. As the percept changes from one to two streams, the perceived rate can increase, decrease, or stay the same, depending on whether the individual tones or the overall triplets are attended in the one-stream mode and whether the A or B tones are attended in the two-stream mode. Thus if changes in waveshape reflect changes in perceived rate (Harms and Melcher 2002), predictions for changes in activation in auditory cortex would be problematic in the case of the ABA triplets and may have resulted in no overall effect in the study of Cusack. The second explanation is that the amplitude and/or waveshape effects relate to within-stream temporal gaps, rather than streaming per se, such that longer gaps between successive tones within a stream lead to larger responses and more sustained activation. In the case of our alternating AB tones, the within-stream gaps increase from zero (apart from the 5-ms onset and offset ramps) to 100 ms—the duration of each tone. In the case of the ABA triplets, there is already a gap equivalent to the duration of one tone, even in the one-stream case, which remains the same in the two-stream case if the A tones are attended, leading to no predicted differential effect in the case of Cusack's stimuli.

A second fMRI study, by Deike et al. (2004), used ABAB sequences in which A and B were harmonic tones differing in spectral envelope and therefore timbre. Activation produced by these sequences (perceived as two streams) was compared with activation during fixed-stimulus sequences with the same overall rate (AAAA and BBBB, perceived as one, higher rate stream). The results showed greater activation during the two-stream condition, a finding consistent with the greater activation produced by large-frequency (20-semitone) compared with small-frequency (0-, 1/8th-semitone) separations in the present study—conditions that elicited two and one stream percepts, respectively. The finding of a large interhemispheric disparity in the magnitude of the difference for two- versus one-stream activation (highly significant in the left hemisphere, but not significant on the right) was not replicated in the present study.

## Possible neural substrates for stream segregation

A prevalent hypothesis in the recent auditory streaming literature is that the degree to which one sound affects the response to a subsequent sound determines whether the sounds are bound together in the same stream (Bee and Klump 2004, 2005; Bregman et al. 2000; Fishman et al.

2001, 2004; Kanwal et al. 2003; Micheyl et al. 2005). For the alternating tone sequences of the present study, the effect of the A tones on the neural response to the B tones (and vice versa) would presumably decrease with increasing frequency separation, leading to a two- rather than a one-stream percept. A similarly reduced interaction between tones would also presumably occur when the temporal separation between tones is increased, a manipulation that also shifts the perceived number of streams from one to two.

The magnetoencephalographic study of Gutschalk et al. (2005) provides some evidence in support of a direct relationship between forward suppression in auditory cortex and stream segregation. This study measured neuromagnetic responses to repeating pure-tone triplets (i.e., ABA\_\_) that had a close frequency separation between tones so as to elicit a percept spontaneously fluctuating between one and two streams. Selective averaging according to subjects' streaming perception revealed stronger suppression of the B-tone responses during the perception of one stream than during the perception of two streams. This suppression resembled the suppression of B-tone responses that occurs when the frequency of the A tone is brought closer to that of the B (documented in the same study) and was therefore similarly interpreted as reflecting an increased influence of the first A tone in each triplet on the response to the following B tone. Importantly, the increased influence (i.e., suppression) occurred even in the absence of physical stimulus changes, indicating that it was a direct correlate of the perceptual binding of the A and B tones into a single stream—evidence favoring the physiological basis of streaming hypothesized in the recent literature.

Most electrophysiological and neuroimaging studies concerning auditory streaming have used pure-tone stimulus sequences for which the degree of perceived stream segregation covaries with the frequency separation between tones. However, psychophysical results (not to mention daily experience in settings akin to the classic “cocktail party”; Cherry 1953) clearly demonstrate that stream segregation can also occur based on more complex stimuli and higher-level features, such as fundamental frequency (Vliegen and Oxenham 1999) and modulation rate (Grimault et al. 2000). The fMRI data of Deike et al. (2004) and Gutschalk et al. (2006) for complex tones streamed according to spectral envelope and fundamental frequency, respectively, suggest that fMRI activation effects found in the present study may generalize to stimuli other than pure tones and stimulus differences other than frequency. Thus it is possible that at least some of the fMRI activation effects seen here reflect a general neural code for auditory streaming. Additional fMRI measurements using a variety of stimuli streamed based on widely different features would provide a strong test of this hypothesis.

### Acknowledgements

The authors thank A. Dreyer and A. Lee for comments on an earlier version of this manuscript and B. Norris for assistance with the figures.

### Grants

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R01-DC-05216 to A. J. Oxenham, P01-DC-00119 to J. R. Melcher, and P30-DC-005209 to A. Gutschalk; a Hertz Foundation Fellowship to E. C. Wilson; and National Center for Research Resources Grant P41-RR-14075 and Mental Illness and Neuroscience Discovery Institute Deutsche Forschungsgemeinschaft fellowship GU 593/2–1 to A. Gutschalk.

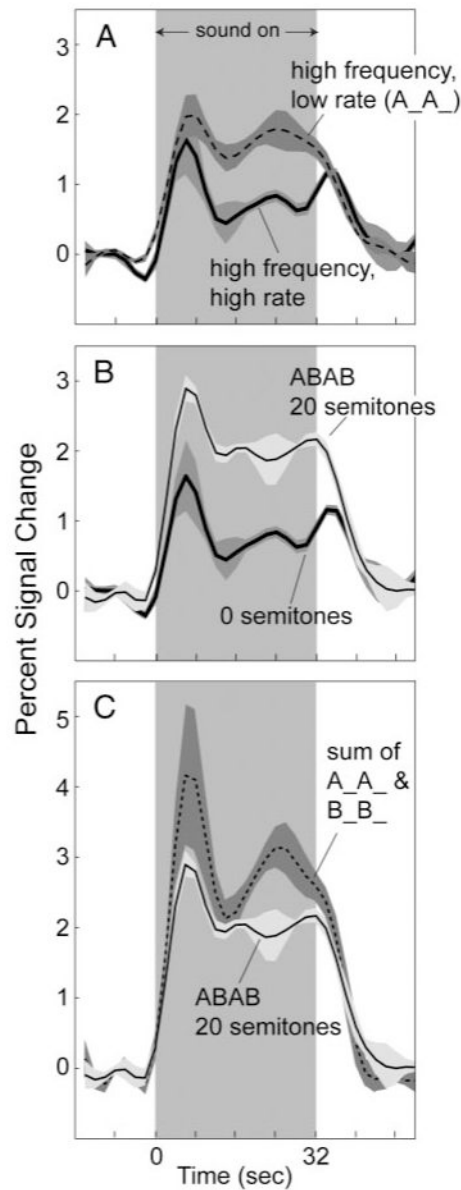
### References

- Bartlett EL, Wang X. Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* 2005;94:83–104. [PubMed: 15772236]
- Bee MA, Klump GM. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol* 2004;92:1088–1104. [PubMed: 15044521]

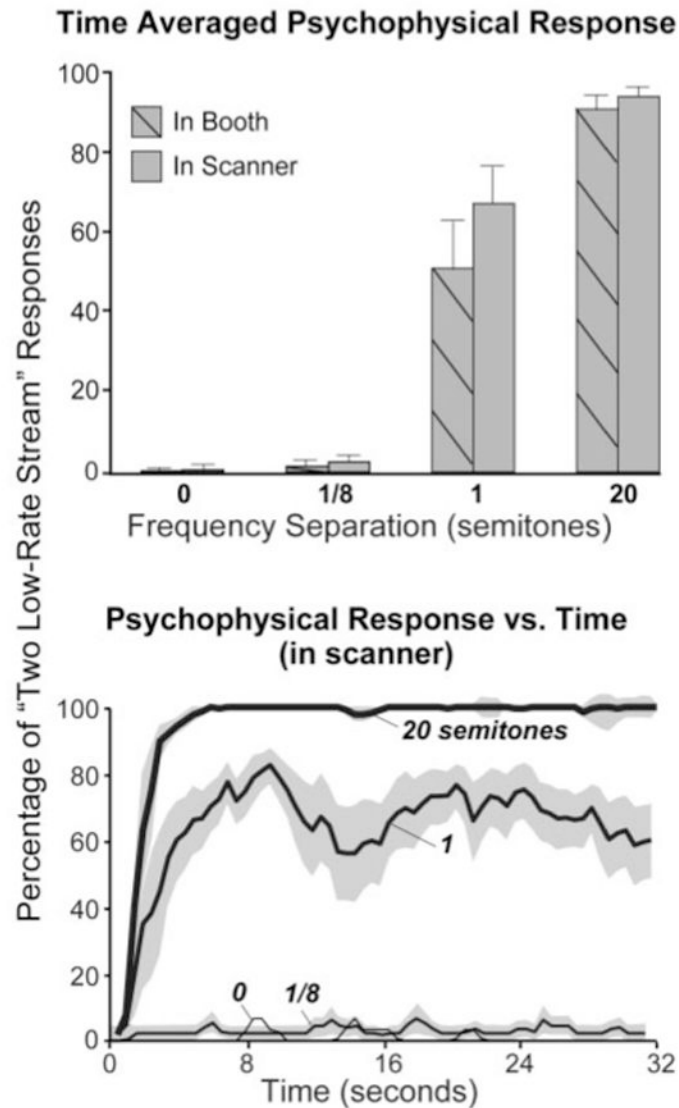
- Bee MA, Klump GM. Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav Evol* 2005;66:197–214. [PubMed: 16127270]
- Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Hyde JS. Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Res Cogn Brain Res* 1994;2:31–38. [PubMed: 7812176]
- Bregman, AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press; 1990.
- Bregman AS, Ahad PA, Crum PAC, O'Reilly J. Effects of time intervals and tone durations on auditory stream segregation. *Percept Psychophys* 2000;62:626–636. [PubMed: 10909253]
- Brosch M, Schreiner CE. Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* 1997;77:923–943. [PubMed: 9065859]
- Brosch M, Schreiner CE. Sequence sensitivity of neurons in cat primary auditory cortex. *Cereb Cortex* 2000;10:1155–1167. [PubMed: 11073865]
- Brosch M, Schulz A, Scheich H. Processing of sound sequences in macaque auditory cortex: response enhancement. *J Neurophysiol* 1999;82:1542–1559. [PubMed: 10482768]
- Butler RA. Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *J Acoust Soc Am* 1968;44:945–950. [PubMed: 5683660]
- Calford MB, Semple MN. Monaural inhibition in cat auditory cortex. *J Neurophysiol* 1995;73:1876–1891. [PubMed: 7623087]
- Calvert GA. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 2001;11:1110–1123. [PubMed: 11709482]
- Cherry EC. Some experiments on the recognition of speech, with one and two ears. *J Acoust Soc Am* 1953;25:975–979.
- Cusack R. Intraparietal sulcus and perceptual organization. *J Cogn Neurosci* 2005;17:641–651. [PubMed: 15829084]
- Deike S, Gaschler-Markefski B, Brechmann A, Scheich H. Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* 2004;15:1511–1514. [PubMed: 15194885]
- Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* 1999;7:89–97. [PubMed: 9950066]
- Fishman YI, Arezzo JC, Steinschneider M. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am* 2004;116:1656–1670. [PubMed: 15478432]
- Fishman YI, Reser DH, Arezzo JC, Steinschneider M. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 2001;151:167–187. [PubMed: 11124464]
- Fomby, TB.; Hill, RC.; Johnson, SR. *Advanced Econometric Methods*. New York: Springer-Verlag; 1984.
- Grimault N, Bacon SP, Micheyl C. Auditory stream segregation on the basis of amplitude-modulation rate. *J Acoust Soc Am* 2002;111:1340–1348. [PubMed: 11931311]
- Gutschalk, A.; Melcher, JR.; Micheyl, C.; Wilson, EC.; Oxenham, AJ. Neural correlates of streaming without spectral cues in human auditory cortex (Abstract). *Assoc Res Otolaryngol Mid-Winter Meeting*; Baltimore, MD. February 4–9, 2006;
- Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci* 2005;25:5382–5388. [PubMed: 15930387]
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 1999;7:213–223. [PubMed: 10194620]
- Hall DA, Haggard MP, Summerfield AQ, Akeroyd MA, Palmer AR, Bowtell RW. Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. *J Acoust Soc Am* 2001;109:1559–1570. [PubMed: 11325127]
- Harms MP, Guinan JJ Jr, Sigalovsky IS, Melcher JR. Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. *J Neurophysiol* 2005;93:210–222. [PubMed: 15306629]

- Harms MP, Melcher JR. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol* 2002;88:1433–1450. [PubMed: 12205164]
- Harms MP, Melcher JR. Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. *Hum Brain Mapp* 2003;20:168–182. [PubMed: 14601143]
- Hart HC, Hall DA, Palmer AR. The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones. *Hear Res* 2003;179:104–112. [PubMed: 12742243]
- Kanwal JS, Medvedev AV, Micheyl C. Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network* 2003;14:413–435. [PubMed: 12938765]
- Micheyl C, Tian B, Carlyon RP, Rauschecker JP. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 2005;48:139–148. [PubMed: 16202714]
- Miller GA, Heise GA. The trill threshold. *J Acoust Soc Am* 1950;22:637–638.
- Ravicz ME, Melcher JR. Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: examination of noise conduction through the ear canal, head, and body. *J Acoust Soc Am* 2001;109:216–231. [PubMed: 11206150]
- Ravicz ME, Melcher JR, Kiang NYS. Acoustic noise during functional magnetic resonance imaging (fMRI). *J Acoust Soc Am* 2000;108:1683–1696. [PubMed: 11051496]
- Shafritz KM, Gore JC, Marois R. The role of the parietal cortex in visual feature binding. *Proc Natl Acad Sci USA* 2002;99:10917–10922. [PubMed: 12149449]
- Sigalovsky IS, Melcher JR. Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. *Hear Res* 2006;215:67–76. [PubMed: 16644153]
- Tanaka H, Fujita N, Watanabe Y, Hirabuki N, Takanashi M, Oshiro Y, Nakamura H. Effects of stimulus rate on the auditory cortex using fMRI with “sparse” temporal sampling. *Neuroreport* 2000;11:2045–2049. [PubMed: 10884068]
- Vliegen J, Oxenham AJ. Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am* 1999;105:339–346. [PubMed: 9921660]
- Warren, RM. *Auditory Perception: A New Analysis and Synthesis*. Cambridge, UK: Cambridge Univ Press; 1999.

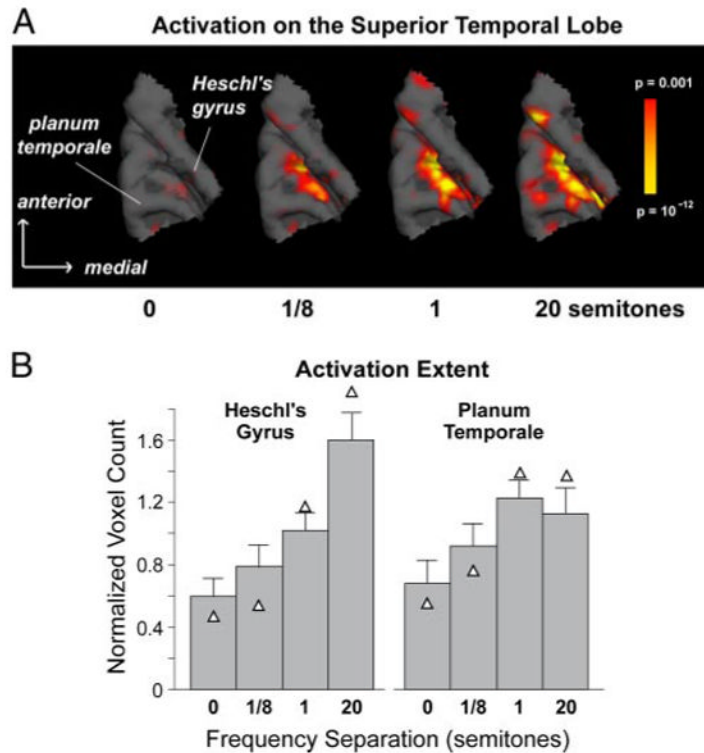


**FIG. 1.**

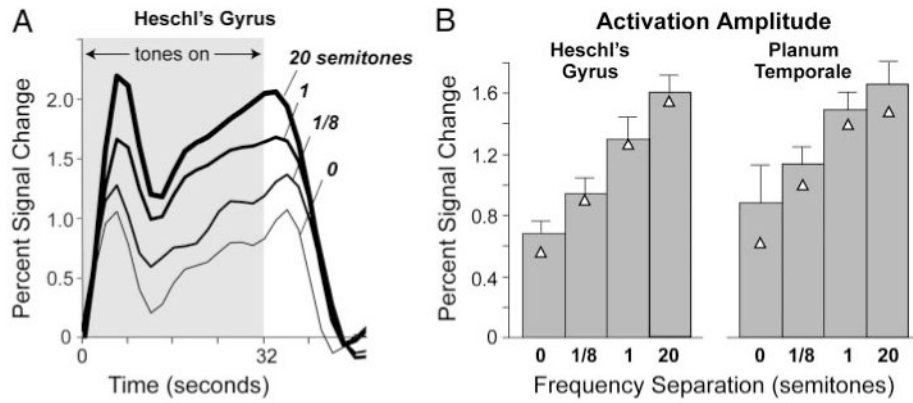
A: time course of activation on Heschl's gyrus for (A) sequences of high-frequency (1,900 Hz) tones presented at a high (10/s) and a low (5/s) rate, (B) ABAB sequences with frequency separations between A and B of 20 and 0 semitones, and (C) an ABAB, 20 semitone sequence—same as in B. C: sum of the time courses produced when the A and B tones of the 20-semitone ABAB sequence were presented separately (A\_A\_ and B\_B\_). In all panels, the traces are each an average across the 3 subjects who participated in the preliminary experiment. Time course for each subject and stimulus condition was calculated by averaging across all Heschl's gyrus voxels active ( $P < 0.001$ ) for either the 0- or 20-semitone condition. Shading indicates  $\pm 1$  SE. In calculating the error, each subject was considered a separate data point.

**FIG. 2.**

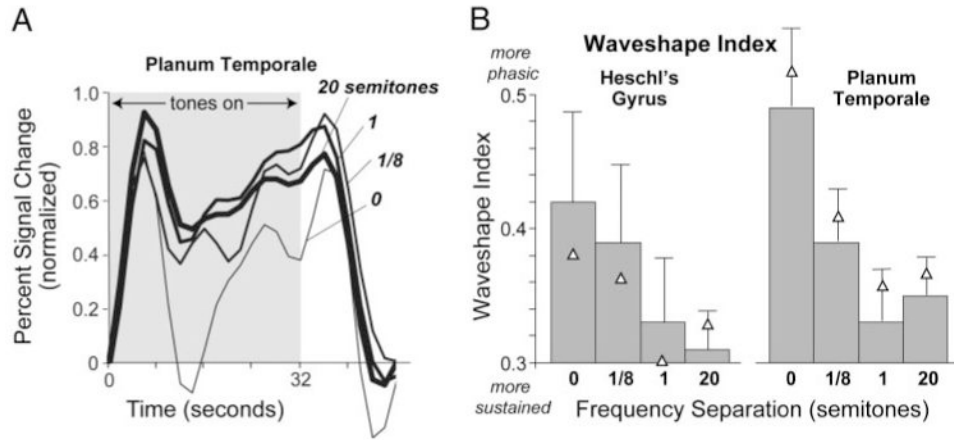
*Top:* time-averaged percentage of “two low-rate stream” responses measured in the soundproof booth (hatched bars) and during functional magnetic resonance imaging (fMRI; solid) for 32-s ABAB sequences with 4 frequency separations between A and B: 0, 1/8, 1, and 20 semitones. Responses were recorded in the scanner during fMRI. *Bottom:* percentage of “two low-rate stream” responses vs. time. Each curve or bar is an average across 5 subjects and 4–5 replications of a given sequence within each subject. Two psychophysical outliers (see text) have been excluded. Error bars indicate 1 SE ( $\pm 1$  error for the *bottom panel*). Each subject was considered a separate data point.

**FIG. 3.**

*A*: activation in auditory cortex for ABAB sequences with 4 frequency separations: 0, 1/8, 1, and 20 semitones. Activation is shown overlaid on a 3D reconstruction of the superior temporal lobe obtained from T1-weighted (nearly  $1 \times 1 \times 1$ -mm resolution) images (MPRAGE). Subject #7—activation on Heschl's gyrus for this subject was mainly in the sulcus. *B*: normalized activation extent averaged across subjects ( $n = 7$ ). Extent was quantified as the number of voxels with  $P < 0.001$ , normalized for each subject to the subject's average across stimulus conditions, and averaged across subjects. Triangles indicate the mean when the 2 psychophysical outliers (subjects 2 and 3; see text) are excluded. Error bars indicate 1 SE. Each subject was considered a separate data point.

**FIG. 4.**

*A*: time course of fMRI activation in Heschl's gyrus for ABAB sequences with 4 frequency separations. *B*: activation amplitude calculated by time-averaging fMRI time courses from 0 to 42 s. Each time course (*A*) or bar (*B*) indicates the mean across subjects ( $n = 7$ ). Triangles indicate the mean when the values for the 2 psychophysical outliers are excluded. Error bars in *B* indicate 1 SE.

**FIG. 5.**

**A:** normalized time course of fMRI activation for 4 frequency separations. For each subject, time courses were calculated by averaging over voxels active for any frequency separation ( $P < 0.001$ ). They were then normalized to have a maximum amplitude of one and, for each frequency separation, were averaged across subjects. **B:** time course waveshape quantified in terms of a waveshape index that ranges from 0 (most sustained) to 1 (most phasic). Each bar indicates an average across 7 subjects. Triangles indicate the mean when the values for the 2 psychophysical outliers are excluded. Error bars indicate 1 SE. Each subject was considered a separate data point.