Neuromagnetic Correlates of Streaming in Human Auditory Cortex

Alexander Gutschalk,[1,2,3] Christophe Micheyl,[2] Jennifer R. Melcher,[3,4,5] André Rupp,[1] Michael Scherg,[1] and Andrew J. Oxenham[2,4]

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

The brain is constantly faced with the challenge of organizing acoustic input from multiple sound sources into meaningful auditory objects or perceptual streams. The present study examines the neural bases of auditory stream formation using neuromagnetic and behavioral measures. The stimuli were sequences of alternating pure tones, which can be perceived as either one or two streams. In the first experiment, physical stimulus parameters were varied between values that promoted the perceptual grouping of the tone sequence into one coherent stream and values that promoted its segregation into two streams. In the second experiment, an ambiguous tone sequence produced a bistable percept that switched spontaneously between one- and two-stream percepts. The first experiment demonstrated a strong correlation between listeners' perception and long-latency (>60 ms) activity that likely arises in nonprimary auditory cortex. The second demonstrated a covariation between this activity and listeners' perception in the absence of physical stimulus changes. Overall, the results indicate a tight coupling between auditory cortical activity and streaming perception, suggesting that an explicit representation of auditory streams may be maintained within nonprimary auditory areas.

Key words: auditory cortex; magnetoencephalography; scene analysis; stream segregation; bistable percepts; adaptation

Introduction

Listening to one speaker or following an instrumental melodic line in the presence of competing sounds relies on the perceptual separation of acoustic sources. As acoustic events unfold over time, the brain generally succeeds in correctly assigning sound from the same source to one "auditory stream" while keeping competing sources separate in a process known as "auditory stream segregation" or streaming (Bregman, 1990).

The auditory system can segregate sounds into streams based on various acoustic features (Carlyon, 2004). A compelling demonstration (van Noorden, 1975) is provided by a sequence of repeating tone triplets separated by a pause (ABA_). When the frequency difference (ΔF) between the A and B tones is small and the interstimulus interval (ISI) is long, the sequence is generally heard as a single stream with a characteristic galloping rhythm. In contrast, when the ΔF is large and the ISI is brief, the sequence splits into two streams, the galloping rhythm is no longer heard, and instead two regular, or isochronous, rhythms are heard, one (A-tone stream) at twice the rate of the other (B-tone stream). At

intermediate ΔF and ISI values, the percept depends on the listener's attentional set as well as the duration of listening to the sequence (Anstis and Saida, 1985; Carlyon et al., 2001), and spontaneous switches in percept can occur just as they do in vision for ambiguous figures or in situations involving binocular rivalry (Blake and Logothetis, 2002).

The present study examined the neural basis of streaming using magnetoencephalography (MEG). Our experiments were motivated by a hypothesized relationship between streaming and neural adaptation (Fishman et al., 2001) that can be understood by considering the auditory evoked N_1m, a surface negative wave that occurs in auditory cortex 80–150 ms after stimulus onset. For a sequence of tones at the same frequency, the N_1m evoked by each successive tone adapts to a steady-state value that depends on the ISI between tones (Ritter et al., 1968; Hari et al., 1982; Imada et al., 1997). For a sequence of tones alternating in frequency, the N_1m amplitude and its dependence on ISI remains the same, as long as the ΔF between tones is small. However, for large ΔF, the amplitude (and hence the degree of adaptation) becomes consistent with the longer ISI between successive tones of the same frequency rather than the shorter ISI between temporally adjacent tones of different frequency, indicating that adaptation occurs selectively on the basis of tone frequency (Butler, 1968; Picton et al., 1978; Näätänen et al., 1988). This is consistent with perceptual organization in a streaming task: when all tones are perceived as one stream, the perceived rate is uniquely determined by the ISI between successive tones (plus the tone dura-

tion). At large ΔF, when the two tone frequencies are segregated into separate streams, the perceived rate is lower and depends on the ISI between tones within each stream. As such, the N_1m might increase with the perceived ISI, whereas the physical ISI remains unchanged. Thus, the N_1m may provide a physiological indicator of auditory stream segregation.

The present study tested this hypothesis by measuring the auditory evoked field (AEF) in response to triplet sequences that were heard as either one or two streams. In one experiment, the percept was altered by manipulating stimulus parameters. In another, stimulus parameters were held constant and were chosen so that the percept was bistable, inducing spontaneous switches between the perception of one and two streams.

Materials and Methods

Listeners. Fourteen listeners (seven females and seven males) participated in each experiment. The mean age was 31.4 years. All of the listeners were right-handed and had no history of peripheral or central hearing disorder. They provided written informed consent before participating in the experiments.

Experiment 1. MEG measurements were made to test the effect of the frequency difference on the AEF in an ABA_ paradigm using pure tones. The total duration of all tones was 100 ms, multiplied with 10 ms raised cosine ramps at the beginning and end. The ISIs (defined as the duration from tone offset to the next tone onset) were 50 ms within the triplet and 200 ms between triplets (supplemental Fig. 1, available at www.jneurosci. org as supplemental material). Our protocol was designed to examine responses to the B tones, rather than the A tones, because we expected larger changes with ΔF for the B tones. This is because the ISI between a B tone and the preceding tone from the same stream differs considerably between the one-stream and two-stream percepts. The relevant ISI for the one-stream percept (between a B tone and the preceding A tone) (supplemental Fig. 1, top; available at www.jneurosci.org as supplemental material) is 50 ms, whereas the relevant ISI for the two-stream percept (between successive B tones) (supplemental Fig. 1, bottom; available at www.jneurosci.org as supplemental material) is 500 ms. In contrast, the within-stream ISI for the first A tone in each triplet remains constant for the one- and two-stream percepts (200 ms) and increases only moderately for the second A tone (50 vs 200 ms). The B tone was fixed at a frequency of 1000 Hz, whereas the A tone was chosen from within the octave above the B tone in intervals of two semitones (1122, 1260, 1414, 1587, 1782, or 2000 Hz). Our protocol also included two baseline conditions, one with identical A and B tones (i.e., 1000 Hz; $\Delta F = 0$ semitones) and the other with no A tones (i.e., only B tones with an ISI of 500 ms; "no A"). Each triplet was repeated 50 times before switching to another condition. The presentation was continuous, with no interruption between consecutive sequences corresponding to different conditions. The eight conditions (seven values of ΔF and the no-A baseline condition) were repeated in a pseudorandomized order; each sequence of 50 triplets was presented four times, yielding a total of 200 B-tone presentations for each condition. The stimuli were presented binaurally using ER-3 transducers (Etymotic Research, Elk Grove Village, IL) with 90 cm plastic tubes and foam ear pieces. The stimuli were presented at 60 dB above average absolute thresholds, as determined in four listeners using a binaurally presented, long-duration 1000 Hz sinusoid. Subsequent tests showed that the thresholds for the other frequencies up to 2000 Hz were similar within a range of ±3 dB. To keep subjects vigilant, they were allowed to watch a silent movie of their own choice (no subtitles; most listeners chose "The Circus" by C. Chaplin or "Nosferatu" by F. Murnau); they were instructed to ignore the auditory stimulation.

Additional MEG measurements were made with an ISI of 200 ms within triplets to test the interaction of ΔF and ISI. The duration of the tones remained at 100 ms, so that the ISI between the ABA triplets was 500 ms. Four conditions were compared in which A was 0, 2, 4, or 10 semitones above the B tone and a fifth control condition in which single B tones were presented with an ISI of 1100 ms.

Thirteen listeners (all of the MEG subjects except one who was no

longer available), completed a psychoacoustic rating task on the stimuli used during MEG. The task was arranged in 2×2 sets. The ISI was either 50 or 200 ms. For each ISI, listeners were instructed to hold to the segregated two-stream percept in one set and hold to the coherent (single-stream) galloping rhythm in another set. Six different ΔF intervals (2–12 semitones) were presented for 10 s in randomized order, and each condition was repeated 10 times. After each presentation, listeners were instructed to rate the ease with which they were able to hold on to the percept as instructed on a continuous visual scale between impossible (0) and very easy (1). Stimuli were generated with a standard soundcard and sampled at 48,000 Hz; they were presented in a quiet room over K 240 DF headphones (AKG Acoustics, Wien, Austria).

Experiment 2. This experiment investigated the relationship between the perceptual state of the subject and the AEF in concurrent MEG and psychophysical measurements. Two conditions from experiment 1 were used, which could be perceived either as one or two streams. The B tone was fixed at 1000 Hz, whereas the A tone was either four or six semitones above the B tone (i.e., A = 1414 Hz or A = 1587 Hz). The ISI was 50 ms within triplets. In four sets, each condition was continuously presented 500 times, twice with a ΔF of four semitones and twice with a ΔF of six semitones. The listeners were instructed to indicate whether they heard one or two streams by pressing a mouse button whenever the perception switched from one to the other. They were allowed to choose whether they pressed the left or right mouse button for the one-stream galloping rhythm and were instructed to press the other mouse button when they perceived the two-stream isochronous rhythm. In half of the sets, listeners were instructed to listen to the A tones whenever they heard two streams; in the other half of the sets, listeners were instructed to listen to the B tones whenever they heard two streams.

Data acquisition. The MEG was recorded continuously with a Neuromag-122 whole-head MEG system (Elekta Neuromag Oy, Helsinki, Finland). The sampling rate was 500 Hz with a 160 Hz low-pass filter. The data were averaged off-line with BESA software (MEGIS Software, Munich, Germany); artifact-contaminated epochs were rejected by an automatic gradient criterion. The cutoff was chosen at 450 fT/cm/sample in all sensors, resulting in an average rejection rate of 5% (range, 1–13%). The epoch duration for averaging was chosen from 200 ms before to 800 ms after triplet onset (50 ms ISI) or from 400 ms before to 1200 ms after triplet onset (200 ms ISI).

T1-weighted magnetic resonance imaging (MRI) of the head (isotropic voxel size, 1 mm^3) was obtained from all listeners (except one) on a Magnetom Symphony 1.5T scanner (Siemens, Erlangen, Germany). Dipole positions were coregistered on the individual MRI and then transformed into Talairach space using Brainvoyager (Brain Innovation, Maastricht, The Netherlands).

Data analysis. Spatiotemporal dipole source analysis (Scherg, 1990) was performed using an average across all ΔF conditions of the 50 ms ISI data in experiment 1 yielding a total of ~1400 averages. The data were filtered from 3–20 Hz (zero-phase-shift Butterworth filter, 12 and 24 dB/octave). Two dipoles, one for each auditory cortex, were simultaneously fitted to the peak of the P_1m at the onset of the ABA_ triplet in an interval of 20 ms centered around the individual peak latency. The average residual variance of the dipole fits was 10% (±4% SD). The dipole model was then kept as a fixed spatial filter to derive comparable source waveforms for the single conditions in all parts of the experiments.

The P_1m dipoles were located in Heschl's gyrus in 17 of 26 hemispheres and in the anterior aspects of planum temporale in the remaining nine hemispheres. We also fitted dipoles to the P_1m and N_1m evoked by the B tones. On average, dipoles for the P_1m, fitted to the A or B tone and the N_1m were in close proximity (Table 1), and both projected to the central part of Heschl's gyrus relative to normative data (Leonard et al., 1998). Therefore, the P_1m and N_1m were not modeled with separate dipoles. The dipole sources fitted for the P_1m of the first A tone showed overall a more accurate location relative to the individual anatomy, probably because it was less variable across conditions and thus provided the better signal-to-noise ratio. The analysis was therefore based on the P_1m dipoles. However, there was no relevant difference in the results when the analysis was based on the N_1m dipoles instead. There was no evidence of contributions from sources outside of the auditory cortex either in field

Table 1. Dipole locations in the space of Talairach and Tournoux (1988)

Tone	Peak	Talairach coordinates (x, y, z; mean $\pm$ SD)	
		Left auditory cortex	Right auditory cortex
A_1	P_1m[a]	$-48 \pm 6, -21 \pm 6, 6 \pm 5$	$52 \pm 4, -15 \pm 5, 8 \pm 5$
B	P_1m[b]	$-49 \pm 7, -20 \pm 6, 5 \pm 7$	$54 \pm 7, -14 \pm 6, 11 \pm 9$
B	N_1m[a]	$-46 \pm 9, -16 \pm 5, 2 \pm 8$	$50 \pm 7, -12 \pm 7, 8 \pm 7$

[a]$n = 13$.
[b]$n = 12$.

maps or in source analysis. Source waveforms were similar in both hemispheres, and there was no significant main effect of hemisphere in the statistics, which is why source waveforms for the right and left auditory cortices were averaged for data presentation.

Peak amplitudes were determined as maximum (P) or minimum (N) in the individual source waveforms. The measurement intervals were 32–82 ms (P_1m) and 62–142 ms (N_1m). Latencies were measured relative to the middle of the ramp (5 ms after stimulus onset), subtracting a 3 ms delay of the tube system. Statistics are based on the general linear model procedure for repeated measures (SAS v.8.02; SAS Institute, Cary, NC). In experiment 1, ΔF, hemisphere (left and right), and the peak magnitude of the response evoked by the B tone (P_1m and N_1m) were modeled by separate variables. In experiment 2, additional variables were included to model the perception (one vs two streams) and the task (follow A tones vs follow B tones). When appropriate, the Greenhouse–Geisser correction was applied to the degrees of freedom; significance levels were not corrected for multiple comparisons. In experiment 2, confidence intervals of difference waveforms were derived by estimating Student's t intervals with the bootstrap technique based on 1000 resamples (Efron and Tibshirani, 1993). Latency and amplitude of the difference peaks were measured in 1000 resamples to derive SEs and t intervals using the same measurement intervals as specified for the original peaks. A wave was only regarded as a significant deflection if the t interval did not span the zero line.

Results

Experiment 1: influence of frequency separation and interstimulus interval

MEG data

Long-latency AEF source waveforms from auditory cortex are shown in Figure 1. Each waveform comprises three successive transient responses: to the first A tone, to the B tone, and to the second A tone. Each response consists mainly of the peaks P_1m (50–70 ms) and N_1m (100–120 ms; latencies are relative to each tone's onset). The P_1m and N_1m evoked by the B tone increased in magnitude as ΔF was increased (ISI = 50 ms, $F_{(6,78)} = 12.35$, $p < 0.0001$; ISI = 200 ms, $F_{(3,39)} = 38.68$, $p < 0.0001$). This effect can be appreciated when comparing the different ΔF conditions with the control conditions at the top and bottom of the figure (i.e., 0 semitones and noA). We compare first the P_1m and N_1m evoked by the B tone in the zero- and two-semitone ΔF conditions. The magnitudes are similarly small in both cases. In contrast, both the P_1m and N_1m for the 10-semitone ΔF are much more prominent, and their amplitudes are close to those of the control condition at the bottom line of the figure (noA), in which the response to the B tone alone is shown. Figure 2 shows the peak amplitudes for the B-tone responses plotted as a function of ΔF. With an ISI of 50 ms, there is a strong increase in peak magnitude between a ΔF of two and four semitones, with the N_1m saturating thereafter. In contrast, with an ISI of 200 ms, there is a prominent increase in peak magnitude for ΔF between four and 10 semitones. This interaction of ISI $\times$ ΔF was significant in the statistical analysis ($F_{(3,39)} = 3.20$; $p < 0.05$; response magnitudes normalized for each ISI separately).

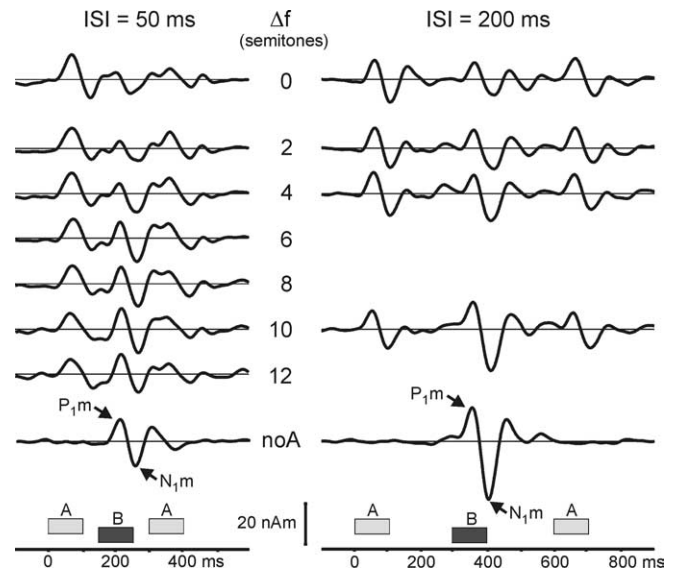**Figure 1.** Source waveforms from auditory cortex for different frequency separations (ΔF) and ISIs. The grand average over subjects ($n = 14$) and hemispheres is shown. Dipoles were fitted to the P_1m in an average across all ΔF conditions for each subject. The 3–20 Hz bandpass filter separates only the long-latency AEF components. The waveforms are shown with increasing ΔF from top to bottom. The control condition in which B was presented without A (noA) is shown. In this case, the 1000 Hz B tone was presented at an ISI of 500 ms (left) or 1100 ms (right). The peaks P_1m and N_1m belonging to the B tone are marked with arrows.
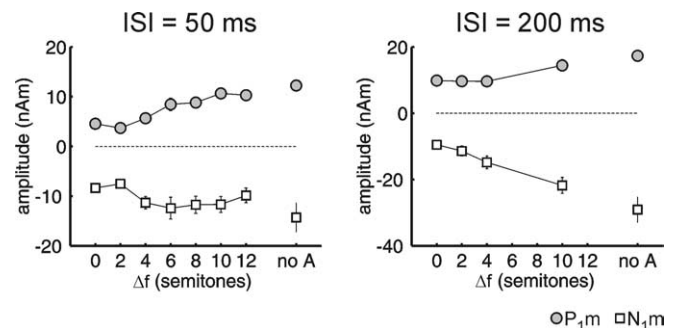

Figure 2. Amplitudes of the AEF evoked by the B tones as a function of frequency difference (ΔF). The graph includes amplitudes (mean $\pm$ SE) of the long-latency AEF peaks P_1m (gray) and N_1m (white) plotted against the ΔF in semitones. Error bars are shown only if they exceed the size of the symbol. The conditions in which the ISI was 50 ms are shown on the left, and those in which the ISI was 200 ms are shown on the right. The control condition in which B was presented without A (noA) is shown on the right in each graph. Notice the larger magnitude for P_1m and N_1m for larger ΔF or when no intervening A tones were present.

Psychophysics

In the psychophysical task, using the same stimuli as in the MEG measurements, listeners were encouraged to hear the sequence as two separate streams and rated the ease of the task (Fig. 3a). Overall, there were significant main effects of ΔF ($F_{(5,60)} = 162.87$; $p < 0.0001$) and ISI ($F_{(1,12)} = 13.50$; $p < 0.01$). As expected from previous findings (van Noorden, 1975; Bregman et al., 2000), there was also a significant interaction of ISI $\times$ ΔF ($F_{(5,60)} = 6.58$; $p < 0.01$): for the longer ISI (200 ms), a larger ΔF was required to hear two segregated streams with similar ease. Subjects also completed the inverse task, in which they were encouraged to hear the sequence as a single stream (i.e., as a coherent gallop) and again to report the ease of the task. When listeners rated the ease of hearing one stream instead (data not shown), the perception was somewhat shifted toward the perception of a
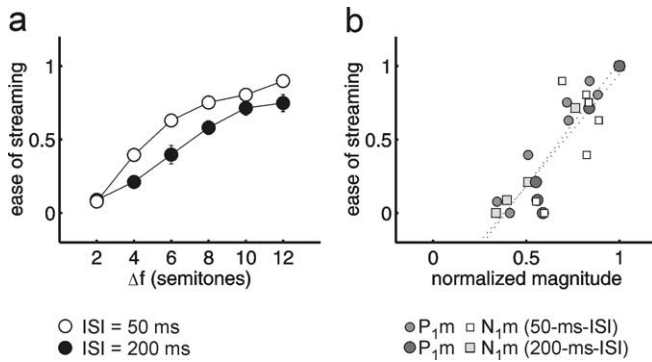
Figure 3. Stream segregation as a function of frequency difference and ISI: behavioral versus MEG data. *a*, Listeners rated the ease of holding to the two-stream perception (0, impossible; 1, very easy). The graph shows mean values ± SEs (*n* = 13; error bars are shown only if they exceed the size of the symbol). Comparison of the 50 ms ISI (open circles) with the 200 ms ISI condition (filled circles) reveals the well-known effect that larger frequency separations (Δf) are required for stream segregation as the tempo decreases. *b*, Perceptual stream segregation plotted against the normalized magnitude of P_1m (circles) and N_1m (squares) in the same listeners (*n* = 13). The data points of the 200 ms ISI condition are plotted in a darker shading than those of the 50 ms ISI. Regression lines (dotted lines) were fitted for each peak separately.

single stream (cf. van Noorden, 1975), but the overall effects of ΔF and ISI were similar.

Comparison of MEG and psychophysics
The qualitative covariations between the reported percepts and the MEG recordings are supported by a direct comparison between the two data sets. Figure 3*b* plots the psychophysical ease of streaming (same scale as in Fig. 3*a*) against the normalized P_1m and N_1m magnitudes. The amplitudes have been normalized to the magnitude of the noA control for each ISI, considering that the P_1m and N_1m amplitude difference between the two noA controls is only determined by the ISI between B tones (500 vs 1100 ms). Because the noA control represents an unambiguous B-tone stream, it should produce the maximum expected response magnitude for a given ISI. Consequently, the psychophysical value for the noA control was set to 1 (very easy to perceive as a segregated stream). Additionally, the $\Delta F = 0$ conditions have been included into this plot, and the corresponding psychophysical values were set to 0 (impossible to perceive as two streams). There was a high correlation between the normalized P_1m and N_1m magnitudes and the psychophysical judgments (Fig. 3*b*). The correlation was $r = 0.91$ ($p < 0.0001$) for P_1m and $r = 0.83$ ($p < 0.001$) for N_1m. The correlations between the inverse task (ease of hearing one stream) (data not shown) and the MEG data were also high (P_1m, $r = -0.84$; N_1m, $r = -0.71$).

Experiment 2: neural correlates of bistable auditory perception
Psychophysical data
Having established a putative correlate of streaming in P_1m and N_1m amplitude, we next tested this relationship more directly using physically invariant stimuli that evoked a bistable percept. In this experiment, the psychophysical and MEG data were acquired concurrently, and subjects were asked to attend to and report any spontaneous changes in percept during a sequence. Listeners were also asked to follow either the A tones or the B tones (in separate blocks) whenever they could hear two streams. The psychophysical results showed that a single stream and two streams were heard with nearly equal probability throughout the stimulus presentation for both values of ΔF used (four and six semitones). In terms of proportion of time spent hearing two
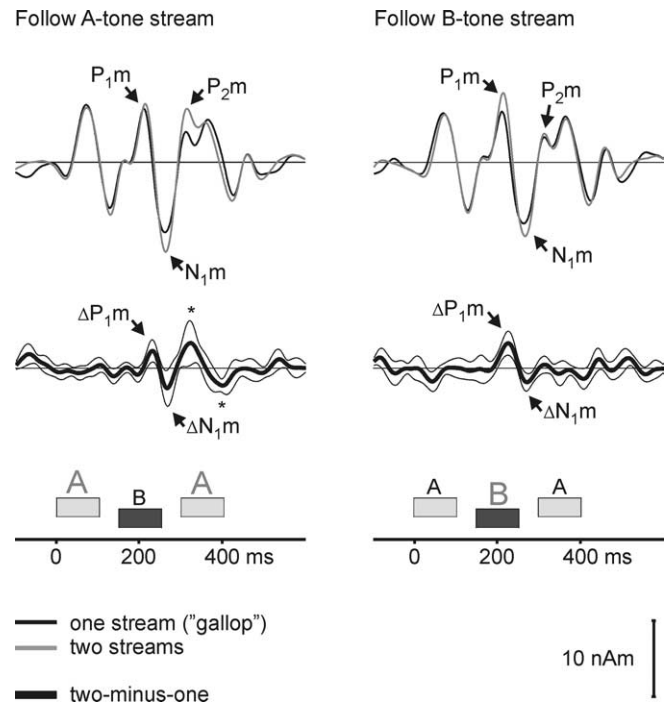

Figure 4. Source activity in auditory cortex varies with subjective perception. The grand average source waveforms (*n* = 14) averaged across hemispheres and frequency separations ($\Delta F = 4$ or 6 semitones) are shown. Waveforms were selectively averaged with reference to whether the listeners indicated that they heard one gallop rhythm (black) or two isochronous rhythms (gray). When listeners heard two streams, they followed either the faster A-tone stream (left) or the half-time B-tone stream (right). Difference waveforms (one minus two streams) are shown in the bottom panels together with bootstrap-based Student's *t* intervals. The P_1m and N_1m were significantly larger when listeners heard two isochronous streams (cf. the difference peaks ΔP_1m and ΔN_1m). This parallels the findings in experiment 1, in which the magnitude of these peaks was larger when listeners reported that holding to the isochronous rhythm was easier. When listeners followed the B-tone stream, two additional difference peaks, one positive and one negative, are observed (*). These peaks overlap with the time interval of the B-tone P_2m as well as the intervals of the P_1m and N_1m of the second A tone.

streams, as opposed to one, there were no significant differences between the four- and six-semitone conditions ($F_{(1,13)} = 0.57$, NS) or between tasks (i.e., whether subjects were instructed to listen to the A or B tones ($F_{(1,13)} = 4.30$, NS). Overall, subjects reported hearing two streams 50.7 ± 2.2% (mean ± SE) of the time. There was an initial "build up" in the tendency to perceive the sequence as segregated. However, after the first 3 s, there was no tendency for one or the other percept to be dominant during the progression of each session. The median duration of the alternating percepts was 4.8 s (or eight-tone triplets; interquartile range, 4.8 s) when listeners followed the A-tone stream and 7.2 s (or 12 triplets; interquartile range, 6.6 s) when they followed the B-tone stream (Mann–Whitney *U* test; $p < 0.001$).

MEG data
Figure 4 shows the source waveforms, averaged over both ΔF conditions and calculated separately for periods in which a single galloping stream was heard (black lines) and in which two isochronous streams were heard (gray lines). The two top panels show the raw source waveforms; the bottom two panels show the difference waveforms. The data from blocks in which listeners were instructed to follow the A-tone stream when segregated are shown on the left, and those in which listeners were instructed to follow the B-tone stream are shown on the right. In both cases, there was a response enhancement when the listeners reported

hearing two streams, as indicated by significant difference waves $\Delta P_1 m$ and $\Delta N_1 m$ following the B-tone [latencies (in ms) $\pm$ SE; follow-A task, $\Delta P_1 m = 74 \pm 3$; $\Delta N_1 m$, 109 ± 3; follow-B task, $\Delta P_1 m = 69 \pm 4$; $\Delta N_1 m$, 115 ± 4]. These observations were supported by an ANOVA on the original peak magnitudes, which showed a significant main effect of perception (one vs two streams) ($P_1 m$ and $N_1 m$, both tasks, $F_{(1,13)} = 9.75$; $p < 0.01$). In contrast, there was no significant main effect of task (follow A vs follow B; $F_{(1,13)} = 1.03$, NS) or interaction of task $\times$ perception ($F_{(1,13)} = 0.68$, NS).

A significant task $\times$ peak $\times$ perception interaction ($F_{(1,13)} = 13.11$; $p < 0.01$) is consistent with the observation in Figure 4 that the $P_1 m$ at its original peak (average latency, 61–63 ms) was only significantly enhanced when listeners perceived two streams and followed the B tones, whereas the $\Delta P_1 m$, which was significant for both tasks, had a longer latency coinciding with the falling phase of the $P_1 m$. In the condition in which listeners followed the A-tone stream, the difference waveforms showed two additional peaks following the $\Delta N_1 m$. The positive peak following $\Delta N_1 m$ had a response latency of 168 ms (± 11 ms SE) after the B-tone onset and could reflect an enhanced $P_2 m$ peak evoked by the B tone. The difference wave is rather broad based, and the significant response enhancement extends into the time interval of the $P_1 m$ evoked by the second A tone (generally, the two peaks could not be clearly separated at the single-subject level). The following negative difference wave peaks at 92 ms (± 12 ms SE) after the second A tone and, thus, most likely reflects an enhancement of the $N_1 m$ of the second A tone. However, no response enhancement was observed for the $P_1 m$ and $N_1 m$ evoked by the first A tone in either the difference waves or the ANOVA.

Discussion

These data provide converging evidence for a strong relationship between auditory cortical activity during tone-triplet sequences and the perceptual organization of these sequences. The results of the first experiment showed that manipulations of ΔF and ISI produced changes in the magnitude of the AEF that corresponded closely to the degree of perceived stream segregation as measured in the same subjects. Specifically, the $P_1 m$ and $N_1 m$ evoked by the B tones in repeating ABA_ triplets increased in magnitude with increasing ΔF between the A and B tones. A similar pattern of results was observed in the psychophysical data, such that correlations were high between the $P_1 m$ and $N_1 m$ magnitude and the perceived ease of stream segregation. The results of the second experiment showed that even when the stimulus parameters were kept constant, the $P_1 m$ and $N_1 m$ covaried with the percept in a manner consistent with experiment 1: the response was larger during the perception of two segregated streams than during the perception of a single integrated stream.

Selective adaptation and stream segregation

The increase in the magnitude of the $P_1 m$ and $N_1 m$ evoked by the B tones with increasing ΔF is consistent with reduced neural adaptation caused by preceding A tones. Butler (1968, 1972) suggested that the frequency-selective adaptation of auditory responses in EEG was a reflection of partial refractoriness along the tonotopic axis of the auditory cortex. Indeed, more recent animal models provide evidence for frequency- and rate-dependent adaptation in A1 (Calford and Semple, 1995; Brosch and Schreiner, 1997). Fishman et al. (2001) suggested that the separation of tones into separate tonotopic channels is closely related to the perception of stream segregation. When they presented alternating tone sequences (ABAB) to awake macaques and recorded at

the A-tone best-frequency site in A1, the B tones were suppressed at smaller ΔF as the ISI between tones decreased (Fishman et al., 2004). Similar findings have been obtained in bats (Kanwal et al., 2003) and birds (Bee and Klump, 2004). It is unclear, however, how these data relate to the $P_1 m$ and $N_1 m$ data in our study. When the repetition rate was fast (>20 Hz), the A-tone response increased as the B-tone frequency became more distant (Fishman et al., 2001). Assuming that, concomitantly, the B-tone response increases at its best-frequency site, one might predict an overall enhancement of the population response with increasing ΔF, which should, in principle, be detectable by MEG. However, at the slow rates and longer ISI used in our study, the data from the study by Fishman et al. (2004) did not indicate a response enhancement with increasing ΔF. The recovery time of the $P_1 m$ and $N_1 m$ appears to be longer than the one observed in monkey A1 (Fishman et al., 2001, 2004), suggesting that our data do not reflect exactly the same process. Ulanovsky et al. (2003) recently described selective adaptation of neurons in cat A1 (but not in thalamus) that involves longer time constants, which might better fit with the selective adaptation of $P_1 m$ and $N_1 m$ in our study.

Although the animal data (Fishman et al., 2001, 2004; Ulanovsky et al., 2003) reflect local recordings in A1 only, human AEFs also involve contributions from nonprimary areas of the auditory cortex. Intracranial recordings in humans have shown that the middle-latency peaks N_a and P_a are recorded from a limited area in medial Heschl's gyrus (Liegeois-Chauvel et al., 1991) and, thus, most probably in human A1 (Hackett et al., 2001). Later waves like the $P_1 m$ and $N_1 m$ with latencies of 50–80 ms ($P_1 m$) and 90–150 ms ($N_1 m$) are observed in more lateral electrode locations. These intracranial locations do not directly translate into dipole locations, because dipole source analysis does not model the spatial extent of a neural source. Based on intracranial studies and source analysis of human data (Liegeois-Chauvel et al., 1994; Gutschalk et al., 2004), $P_1 m$ and $N_1 m$ are generated mostly in lateral Heschl's gyrus, planum temporale, and the superior temporal gyrus and, thus, in nonprimary areas. These areas are considered to comprise neurons that show a higher degree of feature specificity than A1 (Tian et al., 2001; Warren and Griffiths, 2003; Gutschalk et al., 2004), and thus it would not be surprising if selective adaptation in these neurons is involved in stream segregation.

Selective adaptation by feature-specific neurons might be a general neural mechanism subserving perceptual organization, because it has also been observed in the auditory cortex for interaural phase disparity (Malone et al., 2002) and in the visual cortex for various features, such as orientation (Boynton and Finney, 2003) or color contrast (Engel and Furmanski, 2001). Selective adaptation of feature-specific neurons may help account for the streaming of complex tones (with peripherally unresolved harmonics), which cannot be explained by spectral channeling along the tonotopic axis (Vliegen and Oxenham, 1999; Cusack and Roberts, 2000; Grimault et al., 2002; Roberts et al., 2002).

Cortical activity covaried with streaming perception

The variations in the cortical responses $P_1 m$ and $N_1 m$, depending on a subject's percept of an ambiguous stimulus sequence, were similar to (although smaller than) the variations produced by physical stimulus manipulations: the responses were larger when listeners perceived two segregated streams than when they perceived one integrated stream. The fact that B-tone responses were enhanced in experiment 2, regardless of which tones (A or B) were attended, indicates that the enhancement cannot be attributed solely to selective attention, although selective attention may

have been an influencing factor. For instance, the stronger P_1m enhancement during attention to the B tones, could indicate a selective enhancement of the attended (B-tone) stream, as observed previously in dichotic listening tasks (Hillyard et al., 1973; Woldorff et al., 1993). Similarly, there was evidence of N_1m (and possibly P_1m) enhancement for the attended second A tone. However, an argument against this interpretation is that one would expect the P_1m and N_1m of the first A tone to be enhanced (because it was also attended), but they were not. Moreover, the enhancement for the second A tone may also have occurred because of the increase in within-stream ISI for the second A tone (see Materials and Methods). Another aspect of the data that is suggestive of selective attention effects is that the P_2m evoked by the B tone was only enhanced when listeners followed the A tones. It is possible that P_2m was unchanged when the B tones were followed, because the negative difference wave that often occurs during selective attention (in this case, to the B tones) (Hansen and Hillyard, 1984; Riff et al., 1991; Alain et al., 1993) canceled any streaming-related increase in P_2m.

The interaction between streaming and attention is complex and not well understood. For instance, it has been suggested that streaming occurs preattentively, because the mismatch negativity (MMN), which can occur in the absence of attention, is elicited only within streams (Sussman et al., 1999; Yabe et al., 2001). In contrast, it has been shown that the MMN can be modulated by attention in experiments comprising complex attentional loads (Alain and Izenberg, 2003). Moreover, the build up of streaming may depend on attention to the stimuli (Carlyon et al., 2001), and listeners have some control over their perception in the ambiguous ΔF range (van Noorden, 1975). In the visual domain, the alternation of some bistable percepts can be influenced by selective attention (Meng and Tong, 2004), but spontaneous switches still occur. The longer percept duration observed when listeners followed the B-tone stream compared with the A-tone stream in this study may also indicate that the mode of listening influences switches of the percept. A final delineation of streaming and attention is not possible from our data, but there is good evidence that part of the response enhancement we observed during the streaming of bistable sequences (i.e., the ΔP_1m and ΔN_1m) was independent of the focus of attention.

A recent auditory functional MRI (fMRI) study, which also used ambiguous ABA_ sequences (Cusack, 2005), did not find significant activation dependencies on percept or ΔF in auditory cortex. Instead, increased activity was found in the intraparietal sulcus when listeners reported hearing two streams. The reason for the null finding for auditory cortex in the fMRI study is not clear but may be related to the poor temporal resolution of fMRI. Conversely, our failure to find sources outside the auditory cortex may indicate that the increase in intraparietal sulcus activity found by Cusack (2005) was not time locked to single tones in the sequence, which would have been necessary for the detection of neuromagnetic activity in the present study.

Several studies have shown a covariation of visual cortex activation and the percept of ambiguous stimuli. For instance, during presentation of alternating dots, area V5 shows enhanced activity when a single moving dot is perceived instead of two separate alternating dots (Muckli et al., 2002). Similarly, modulation of V5 activity was shown to relate to ambiguous motion perception of moving grid stimuli (Castelo-Branco et al., 2000, 2002). Our results show a similar covariation between ambiguous auditory percepts and auditory cortex activation starting after ~60 ms.

References

Alain C, Izenberg A (2003) Effects of attentional load on auditory scene analysis. J Cogn Neurosci 15:1063–1073.

Alain C, Achim A, Richer F (1993) Perceptual context and the selective attention effect on auditory event-related brain potentials. Psychophysiology 30:572–580.

Anstis S, Saida S (1985) Adaptation to auditory streaming of frequency-modulated tones. J Exp Psychol Hum Percept Perform 11:257–271.

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

Blake R, Logothetis NK (2002) Visual competition. Nat Rev Neurosci 3:13–21.

Boynton GM, Finney E (2003) Orientation-specific adaptation in human visual cortex. J Neurosci 23:8781–8787.

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

Bregman A, Ahad PA, Crum PAC, O'Reilly J (2000) Effects of time intervals and tone durations on auditory stream segregation. Percept Psychophys 62:626–636.

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

Butler RA (1968) Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. J Acoust Soc Am 44:945–950.

Butler RA (1972) Frequency specificity of the auditory evoked response to simultaneously and successively presented stimuli. Electroencephalogr Clin Neurophysiol 33:277–282.

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.

Castelo-Branco M, Goebel R, Neuenschwander S, Singer W (2000) Neural synchrony correlates with surface segregation rules. Nature 405:685–689.

Castelo-Branco M, Formisano E, Backes W, Zanella F, Neuenschwander S, Singer W, Goebel R (2002) Activity patterns in human motion-sensitive areas depend on the interpretation of global motion. Proc Natl Acad Sci USA 99:13914–13919.

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

Cusack R, Roberts B (2000) Effects of differences in timbre on sequential grouping. Percept Psychophys 62:1112–1120.

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap, pp 158–159. New York: Chapman and Hall.

Engel SA, Furmanski CS (2001) Selective adaptation to color contrast in human primary visual cortex. J Neurosci 21:3949–3954.

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.

Grimault N, Bacon SP, Micheyl C (2002) Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am 111:1340–1348.

Gutschalk A, Patterson RD, Scherg M, Uppenkamp S, Rupp A (2004) Temporal dynamics of pitch in human auditory cortex. NeuroImage 22:755–766.

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.

Hansen JC, Hillyard SA (1984) Effects of stimulation rate and attribute cuing on event-related potentials during selective auditory attention. Psychophysiology 21:394–405.

Hari R, Kaila K, Katila T, Tuomisto T, Varpula T (1982) Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. Electroencephalogr Clin Neurophysiol 54:561–569.

Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the human brain. Science 182:177–180.

Imada T, Watanabe M, Mashiko T, Kawakatsu M, Kotani M (1997) The silent period between sounds has a stronger effect than the interstimulus

interval on auditory evoked magnetic fields. Electroencephalogr Clin Neurophysiol 102:37–45.

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.

Leonard CM, Puranik C, Kuldau JM, Lombardino LJ (1998) Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? Cereb Cortex 8:397–406.

Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. Brain 114:139–151.

Liegeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P (1994) Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol 92:204–214.

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.

Meng M, Tong F (2004) Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. J Vision 4:539–551.

Muckli L, Kriegeskorte N, Lanfermann H, Zanella FE, Singer W, Goebel R (2002) Apparent motion: event related functional magnetic resonance imaging of perceptual switches and states. J Neurosci 22:RC219(1–5).

Näätänen R, Sams M, Alho K, Paavilainen P, Reinikainen K, Sokolov EN (1988) Frequency and location specificity of the human vertex N1 wave. Electroencephalogr Clin Neurophysiol 69:523–531.

Picton TW, Woods DL, Proulx GB (1978) Human auditory sustained potentials. II. Stimulus relationships. Electroencephalogr Clin Neurophysiol 45:198–210.

Riff J, Hari R, Hämäläinen MS, Sams M (1991) Auditory attention affects two different areas in the human supratemporal cortex. Electroencephalogr Clin Neurophysiol 79:464–472.

Ritter W, Vaughan Jr HG, Costa LD (1968) Orienting and habituation to auditory stimuli: a study of short term changes in average evoked responses. Electroencephalogr Clin Neurophysiol 25:550–556.

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 6 (Grandori F, Hoke M, Romani GL, eds), pp 40–69. Basel: Karger.

Sussman E, Ritter W, Vaughan Jr HG (1999) An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology 36:22–34.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290–293.

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: University of Technology.

Vliegen J, Oxenham AJ (1999) Sequential stream segregation in the absence of spectral cues. J Acoust Soc Am 105:339–346.

Warren JD, Griffiths TD (2003) Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J Neurosci 23:5799–5804.

Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE (1993) Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc Natl Acad Sci USA 90:8722–8726.

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Sutoh T, Hiruma T, Kaneko S (2001) Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. Brain Res 897:222–227.