

Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation

Joel S. Snyder¹, Claude Alain^{1,2}, and Terence W. Picton^{1,2}

Abstract

■ A general assumption underlying auditory scene analysis is that the initial grouping of acoustic elements is independent of attention. The effects of attention on auditory stream segregation were investigated by recording event-related potentials (ERPs) while participants either attended to sound stimuli and indicated whether they heard one or two streams or watched a muted movie. The stimuli were pure-tone ABA– patterns that repeated for 10.8 sec with a stimulus onset asynchrony between A and B tones of 100 msec in which the A tone was fixed at 500 Hz, the B tone could be 500, 625, 750, or 1000 Hz, and – was a silence. In both listening conditions, an enhancement of the auditory-evoked response (P1–N1–P2 and N1c) to the B tone varied with Δf and correlated with perception of

streaming. The ERP from 150 to 250 msec after the beginning of the repeating ABA– patterns became more positive during the course of the trial and was diminished when participants ignored the tones, consistent with behavioral studies indicating that streaming takes several seconds to build up. The N1c enhancement and the buildup over time were larger at right than left temporal electrodes, suggesting a right-hemisphere dominance for stream segregation. Sources in Heschl's gyrus accounted for the ERP modulations related to Δf -based segregation and buildup. These findings provide evidence for two cortical mechanisms of streaming: automatic segregation of sounds and attention-dependent buildup process that integrates successive tones within streams over several seconds. ■

INTRODUCTION

Making sense of the acoustic environment requires parsing sounds that originate from different physical objects and grouping together sounds that emanate from the same object. These processes play a critical role in a listener's ability to identify and recognize complex acoustic signals such as speech and music. The collection of internal processes that segregate and group sounds to form representations of auditory objects is called *auditory scene analysis* (Bregman, 1990). Without auditory scene analysis, forming accurate representations of the external world would fail, especially in complex situations wherein multiple objects produce similar sounds. For example, a listener at a cocktail party must process speech from one person while other speakers are talking (Cherry, 1953). A similar situation arises when a listener focuses on a single musical instrument in an ensemble.

One popular paradigm for studying auditory scene analysis presents low tones (A), high tones (B), and silences (–) in a repeating ABA– pattern (see Figure 1). When the difference in frequency between the A and B tones is small and the repetition rate of the sequence is slow, listeners hear a single stream of tones in a galloping

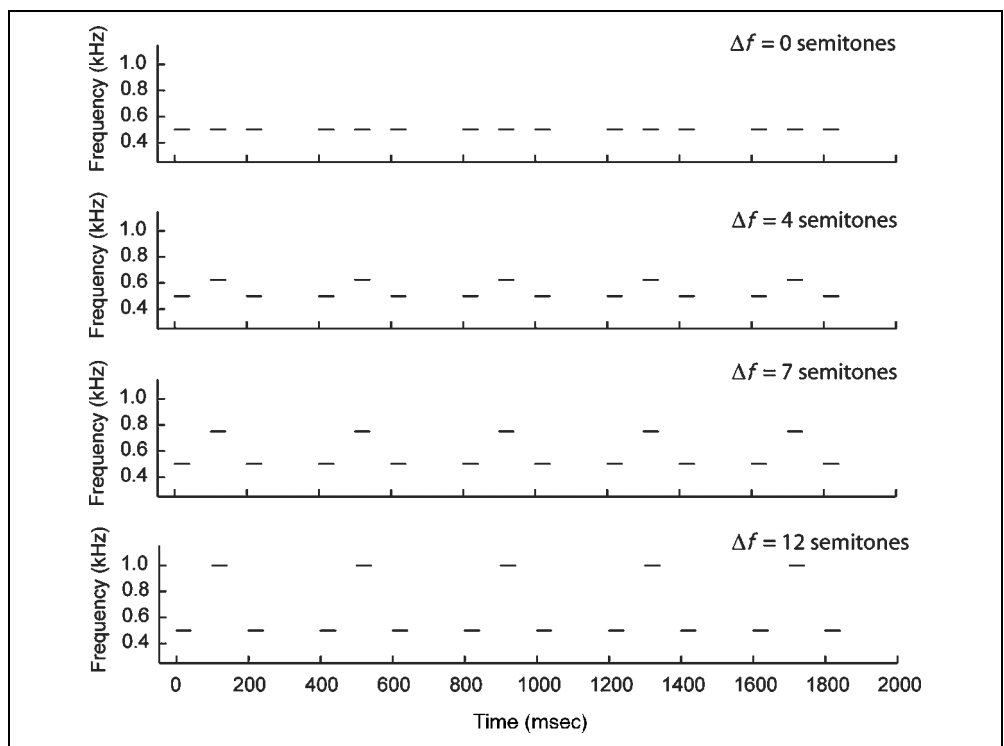
rhythm. When the frequency difference is large and the repetition rate is fast, listeners report the sequence splitting into two streams of tones, each in a metronome-like rhythm.

According to the “peripheral channeling hypothesis,” the most powerful cues for stream segregation are those that lead to two or more nonoverlapping activations in the cochlea (Hartmann & Johnson, 1991) such as pure-tone frequency and ear of stimulation. This type of place-based segregation is likely to be carried up through the ascending auditory pathway to the tonotopic fields of the auditory cortex (Kaas & Hackett, 2000). A recent study (Fishman, Arezzo, & Steinschneider, 2004) supported this idea by presenting A and B tones that alternated in frequency while recording multiunit activity from macaque monkey primary auditory cortical neurons that were maximally responsive to the A tones. As the presentation rate and pitch separation of A and B tones increased, the firing rate increased in response to the A tones. Another study reported similar findings from the primary auditory cortex of bats (Kanwal, Medvedev, & Micheyl, 2003). Bee and Klump (2004) measured multiunit activity in the auditory forebrain (homologous to the mammalian auditory cortex) of starlings that showed correspondence with behavioral data from starlings and humans.

Despite the power of the peripheral channeling hypothesis, a number of cues besides those based on

¹Baycrest Centre for Geriatric Care, Toronto, ²University of Toronto

Figure 1. Five cycles of stimuli used during attend and ignore conditions. Actual trials were composed of 27 cycles. Each bar represents a pure tone in a galloping rhythm.



peripheral segregation can lead to streaming (for a review, see Moore & Gockel, 2002), implying that centrally computed features contribute to stream segregation. Further supporting the existence of central mechanisms for streaming is the finding that perception of streaming does not occur immediately but takes several seconds to build up (Anstis & Saida, 1985; Bregman, 1978). Additionally, the effect of a biasing sequence that increases perception of streaming also lasts for several seconds (Beauvois & Meddis, 1997), with a longer time constant for musicians than nonmusicians (Beauvois & Meddis, 1997). Despite this slow buildup and decay for streaming, transient events such as a brief silence in the ABA– pattern or an attention shift can almost completely reset the buildup process (Cusack, Deeks, Aikman, & Carlyon, 2004). The long time constants for buildup and decay of streaming and the influence of musical experience further suggest that critical aspects of streaming occur at higher levels of the auditory system. In particular, the long time constants would be consistent with neuro-magnetic correlates of echoic memory in the auditory cortex (Lü, Williamson, & Kaufman, 1992), and computational modeling of streaming that use inhibitory time constants typical of the auditory cortex (Kanwal et al., 2003; McCabe & Denham, 1997). Thus, although it is clear that low-level aspects of stream segregation operate at low-level stages of the auditory system, other aspects of streaming likely require computations in the auditory cortex and other cortical structures.

Evidence of attentional effects on the buildup of streaming has further implicated higher-level influences on stream segregation (Carlyon, Cusack, Foxton, &

Robertson, 2001). When participants ignore the ABA– pattern presented to one ear by listening to sounds presented to the other ear and then switch their attention to the ABA– pattern, the buildup process of streaming is diminished compared to when participants simply attend to the ABA– patterns for the whole trial. The apparent diminishment of streaming when ignoring the ABA– pattern, however, might be in part due to the process of switching attention rather than an actual effect of ignoring the sounds (Cusack et al., 2004). Further casting doubt on the influence of attention, ignored ABA– patterns that would be perceived as streaming result in a reduction in interference in a visual memory task, compared to patterns that would not be perceived as streaming (Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). These behavioral studies thus lead to an uncertain conclusion about whether, to what extent, and at what level of processing listeners' attention affects streaming. Event-related potentials (ERPs) might help provide a clearer answer to these issues by isolating neural events that correspond to distinct aspects of streaming. Furthermore, it is possible to record ERPs when participants are actively listening as well as when they are ignoring sounds, providing a simple means for evaluating the effects of attention on streaming.

Such an approach has been used in previous studies to understand the perception of concurrently presented auditory objects rather than simultaneously unfolding auditory streams. As with stream segregation, perception of multiple concurrent sounds is promoted by increased pitch separation. For example, if one shifts the frequency of a partial in a multicomponent stimulus, all

the other frequencies of which derive from a single fundamental, this shifted component is heard as a separate tone (Moore, Glasberg, & Peters, 1986). ERP research on this perceptual phenomenon has identified a negative peak at ~ 150 msec following presentation of the complex sound called the “object-related negativity” (ORN). The ORN amplitude varies in direct proportion with perception of two simultaneous auditory objects (Alain, Arnott, & Picton, 2001) and is not affected by selective attention (Alain & Izenberg, 2003).

Earlier ERP research on sequential stream segregation used randomly presented tones during selective attention (Alain & Woods, 1994; Alain, Achim, & Richer, 1993). For example, Alain and Woods (1994) presented concurrent interleaved tone sequences of different frequencies and showed behavioral and ERP evidence that it was easier to selectively process particular pitches when the other task-irrelevant tones were grouped together. This suggested that perceptual grouping overrides the effects of physical similarity during selective attention and that auditory attention, like visual attention, may be allocated to objects (Alain & Arnott, 2000).

Studies by Winkler et al. (2003) and Sussman, Ritter, and Vaughan (1999) used the mismatch negativity (MMN) to study streaming. The MMN is a negative ERP component that peaks ~ 150 msec following a deviant stimulus in a sequence of homogeneous auditory events (for a review, see Picton, Alain, Otten, Ritter, & Achim, 2000). Studying streaming with the MMN requires events that can only be perceived as deviants if streaming has occurred. For example, Winkler et al. presented simultaneous sequences, one in which the intensity was constant except for occasional deviants and one in which the intensity varied constantly. When the two sequences overlapped in frequency, the constant intensity variations in one sequence obscured the occasional intensity deviants in the other sequence and no MMN to the occasional deviants occurred. However, when the two sequences were widely separated in frequency, an MMN occurred to the occasional deviants, suggesting that streaming had occurred prior to detection of intensity deviants. Although the MMN indicates that stream segregation has occurred, it reveals little about the neural mechanisms underlying streaming because it does not track ongoing processing of tone patterns.

The current study uses a more direct paradigm measuring brain activity that tracks the ABA– pattern as streaming occurs, in hopes of addressing some of these limitations. Using 10.8-sec trials helped to determine whether buildup of neural activity mirrors behavioral buildup of streaming (Anstis & Saida, 1985; Bregman, 1978). To avoid motor contamination of the neural measurements, participants were asked to indicate at the end of the sequence whether they heard one or two streams, enabling us to establish a relationship between

ongoing patterns of neural activity and whether streaming had occurred. To identify segregation processes, we compared activity for trials with different frequency separations (Δf) between A and B tones. To identify buildup processes, on the other hand, we compared activity at different 2-sec time bins within the 10.8-sec trials. In one session, we collected ERP data while participants made streaming judgments after the end of each trial. In a separate session, we presented identical sound patterns and asked the same participants to watch a muted subtitled video of their choice and to ignore the auditory stimuli. This manipulation was designed to test whether stream segregation as indexed by ERPs depends on focused attention, thereby allowing us to clarify the stage at which attention affects streaming. The use of muted subtitled movies is important because the text dialogue effectively captures attention while not interfering with auditory processing (Pettigrew et al., 2004).

Based on place models of streaming (McCabe & Denham, 1997; Beauvois & Meddis, 1996; Hartmann & Johnson, 1991), we expected increases in activity as a function of Δf corresponding to decreased overlap between activations corresponding to the A and B tones. We hypothesized that segregation processes would function independent of attention, whereas neural buildup processes would be affected by attention as suggested by behavioral studies (Cusack et al., 2004; Carlyon et al., 2001).

RESULTS

Behavioral Data

Figure 2 shows the mean proportion of trials heard as streaming for each of the Δf conditions. As expected, the likelihood of reporting perception of two streams increased with Δf . At 0 semitone (1 semitone = 1/12 octave), participants rarely reported hearing streaming, whereas at 12 semitones, participants almost always heard streaming by the end of the 10.8-sec trial. At

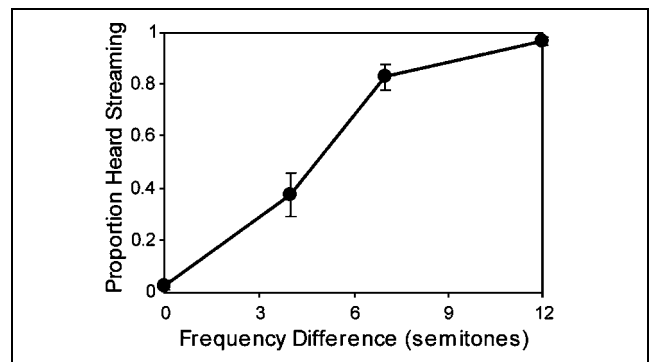


Figure 2. Group mean ($\pm SE$) proportion of trials heard as streaming across participants ($n = 10$) for the four Δf levels.

intermediate levels (4 and 7 semitones), participants sometimes heard streaming and sometimes heard the galloping pattern for the whole trial. There was a significant main effect of Δf , $F(3,27) = 88.02$, $p < .001$, with all adjacent levels of Δf differing from each other ($p < .05$).

Neural Activity Reflecting Frequency-based Segregation

Figure 3A shows ERPs elicited by the onset of the sequence for attend and ignore conditions, collapsed across Δf . The ERPs comprised P1 (~60 msec), N1 (~120 msec), and P2 (~160 msec) waves that were maximal at frontocentral scalp regions (see Figure 3B). There was also a clear negative peak at ~200 msec referred to as the N2 wave that was present only when the A and B tones differed in frequency. Following the transient responses, there was a sustained potential (SP) that was negative and maximal over the frontal regions. The N1 and SP showed larger amplitude during active than passive listening, $F(1,9) = 43.59$ and 70.82 , respectively, $p < .001$. The effect of attention was not significant for the P1, P2, and N2 waves. As shown in Figure 3C, the effect of Δf on the P1, N1, and P2 waves

was not significant. However, the N2 showed a significant amplitude increase as a function of Δf , $F(3,27) = 12.29$, $p < .001$, with a marginal effect for SP, $F(3,27) = 2.95$, $p = .054$. There was no significant interaction between attention and Δf for any ERP deflections elicited by the onset of the sequence.

A close examination of the SP revealed periodic fluctuations in amplitude that corresponded closely with rate of stimulus presentation. These smaller fluctuations were more easily assessed in smaller epochs. Figure 4A shows 2-sec ERPs in the attend condition with all four levels of Δf superimposed at FCz. In Figure 4B are single cycles of the ERPs at FCz and the left and right temporal sites (T7 and T8). The neural activity associated with increasing Δf is best illustrated by subtracting ERPs elicited by stimuli with constant frequency (i.e., 0 semitone condition) from those obtained when the A and B tones differed in frequency.

This subtraction procedure isolated a series of time-locked ERP waves elicited by the B tone, which included P1 (~60 msec), N1 (~115), and P2 (~175 msec) deflections at frontocentral scalp regions and an N1c (~160 msec) that was maximal over the right temporal electrode (i.e., T8). The results of this subtraction are shown in Figure 5A for attend and ignore conditions.

Figure 3. Group mean ERPs to the beginning of the trial. (A) ERP response to the first 2 sec of each trial for the attend and ignore conditions collapsed across Δf at the midline frontocentral (FCz) electrode. Horizontal bars above the time scale represent pure tones in the stimulus pattern. (B) Scalp distribution of voltage for the P1, N1, P2, N2, and SP waves at 72, 124, 168, 220, and 620 msec, respectively in the attend condition collapsed across Δf . Darker regions indicate more activity, with polarity labeled by + and - signs. Isocontour lines represent $0.4 \mu\text{V}/\text{step}$ for P1, N1, P2, and N2 and $0.8 \mu\text{V}/\text{step}$ for SP. (C) Same as (A) for the four Δf levels in the attend condition.

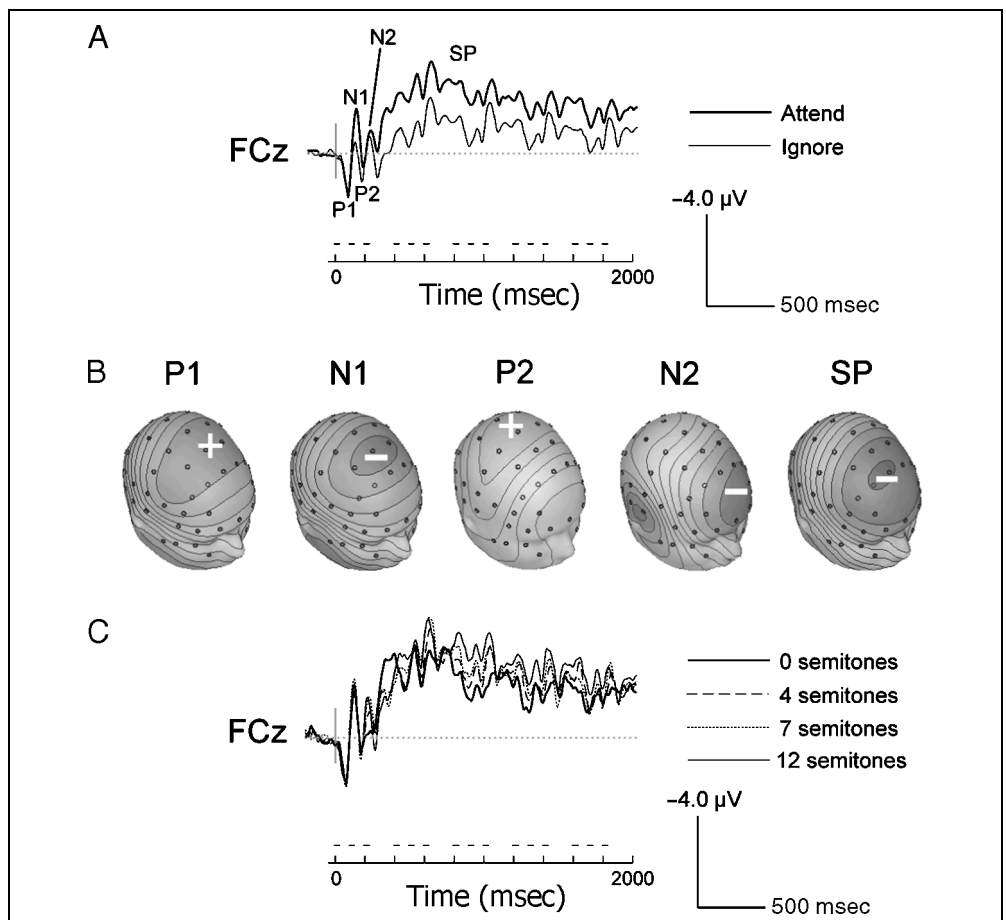
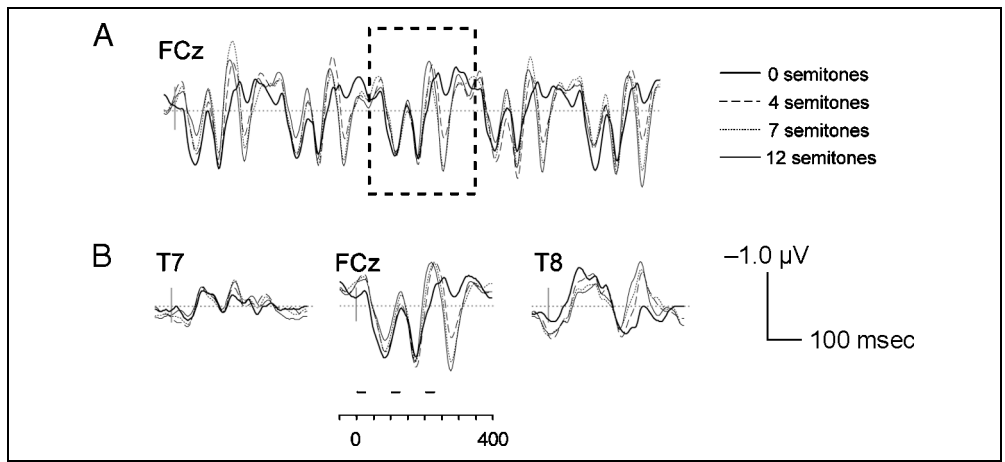


Figure 4. ERP time course as Δf increased in the attend condition. (A) ERP response at FCz to 5 cycles (2000 msec) of the ABA- pattern with a box around a single repetition of the ABA- pattern for the four Δf levels. (B) Single-cycle ERPs at T7 (left temporal), FCz (frontocentral midline), and T8 (right temporal) showing the effect of varying Δf . Horizontal bars above the time scale represent pure tones in the stimulus pattern.

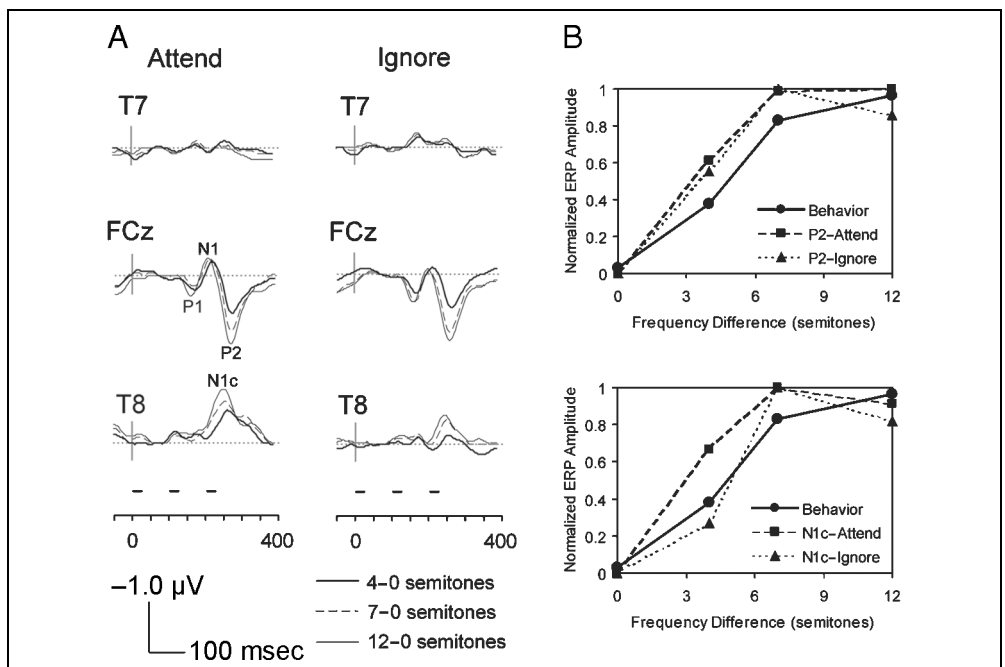


The effects of Δf and attention were examined on the P1, N1, and P2 peak latencies at the nine frontocentral electrodes (Fz/1/2, FCz/1/2, Cz/1/2). We also quantified P1-N1 and N1-P2 peak-to-peak amplitudes, allowing us to examine transient changes in neural activity while controlling for other changes in sustained activity that may overlap with the P1, N1, and P2 waves. The means across the nine frontocentral electrodes were entered into analyses of variance (ANOVAs) to test for effects of attention, Δf , and time. We quantified the peak latency and amplitude of the N1c at the left and right temporal electrodes (T7 and T8). Quantifying the N1c, which arises from current sources in the auditory cortex with a radial orientation (Picton, Alain, Woods, et al., 1999), allowed us to test for hemispheric differences in Δf -based segregation processing. We also examined

whether the latency and amplitude of these responses varied as a function of time by dividing the 10.8-sec sequences into five 2-sec periods and averaging the responses within that period. This allowed us to examine whether the neural activity associated with the processing of Δf varied as a function of time.

The P1, N1, and P2 latencies decreased, $F(2,18) = 11.09, 17.21, \text{ and } 13.44, p < .005$, and the P1-N1 and N1-P2 amplitudes increased, $F(2,18) = 6.87 \text{ and } 18.75, p < .025$, as a function of Δf . All linear trends of Δf on peak latencies and peak-to-peak amplitudes were significant ($p < .025$). There were significantly longer latencies for the N1 and P2, $F(1,9) = 5.16 \text{ and } 13.24, p < .05$, and a larger P1-N1 amplitude, $F(1,9) = 6.08, p < .05$, when participants attended the stimuli. A significant Attention \times Δf interaction occurred only for P2 latency,

Figure 5. Difference waves between ERPs elicited by 0 semitone Δf and those elicited by 4, 7, and 12 semitone Δf for attend and ignore conditions. (A) Difference waves to the 0.4-sec ABA- pattern averaged at T7, FCz, and T8 for attend and ignore conditions. Horizontal bars above the time scale represent pure tones in the stimulus pattern. (B) Normalized average amplitude in the P2 time region (244-300) across nine frontocentral channels (top) and the N1c time region (232-300) at T8 (bottom) in the attend and ignore conditions plotted against Δf along with the behavioral data from Figure 2.



$F(2,18) = 4.30, p < .05$, with a larger decrease in latency for the ignore condition than in the attend condition. There were no other main effects of attention or interactions between attention and Δf , suggesting that ignoring the stimuli had minimal effects on Δf -based segregation processing. A significant main effect of time occurred for the P1 latency, $F(4,36) = 5.24, p < .01$, with later peaks as time progressed within the trial. There were no other effects of time on latency or peak-to-peak amplitudes. Finally, a three-way interaction between attention, time, and Δf occurred for P1 latency, $F(4,36) = 3.09, p < .05$, with larger increases over time for the attend condition for only the 4 and 7 semitone conditions.

We also quantified latency and amplitude of the N1c at T7 and T8 to test for hemispheric difference in processing Δf as a function of attention and buildup. As for the N1 at frontocentral sites, the N1c latency decreased, $F(2,18) = 3.90, p < .05$, and its amplitude increased, $F(2,18) = 5.83, p < .025$, as Δf increased. The linear trends of Δf on N1c latency and amplitude were also significant ($p < .025$). The N1c amplitude was larger at T8 than at T7, $F(1,9) = 21.11, p < .001$, and it increased more with larger Δf at T8 than T7, $F(2,18) = 7.52, p < .01$. N1c was larger when participants attended the stimuli, $F(1,9) = 5.58, p < .05$, and this attention-related increase was larger at T8 than at T7, $F(1,9) = 12.84, p < .01$. There was no interaction between attention and Δf , again suggesting that attention did not influence Δf -based segregation processing. The N1c results are thus consistent with the P1–N1–P2, additionally showing a right-hemisphere (RH) dominance for Δf -based segregation.

As shown on the top of Figure 5B, behavioral judgments of streaming during the attend condition correlated significantly with P2 amplitude in both attend, $r(2) = .87, t(9) = 21.63, p < .001$, and ignore conditions, $r(2) = .86, t(9) = 17.67, p < .001$. P1 amplitude also correlated significantly with behavioral judgments, but only for the passive condition, $r(2) = .59, t(9) = 3.98,$

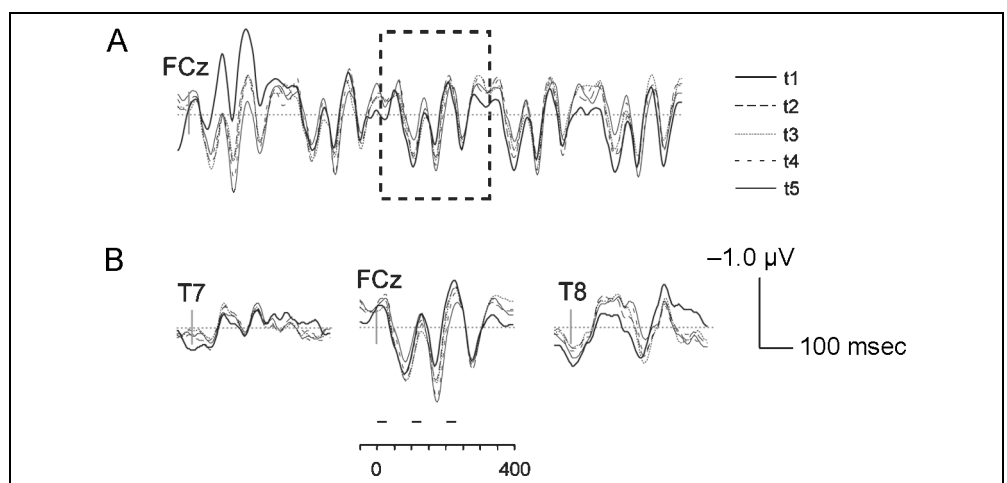
$p < .005$. The N1 amplitude did not correlate with behavioral judgments in either the active or passive conditions, likely due to relatively poor signal-to-noise ratio. As shown on the bottom of Figure 5B, behavioral judgments of streaming correlated significantly with N1c amplitude at T8 for both attend, $r(2) = -.72, t(9) = -4.95, p < .001$, and ignore conditions, $r(2) = -.57, t(9) = -3.72, p < .005$.

Neural Activity Reflecting Buildup of Streaming

To isolate buildup of neural activity as a function of time during the 10.8-sec trials, we measured the ERP at five different 2-sec time bins, collapsing across all five repetitions of the ABA– pattern within each bin (excluding the first and last ABA– cycle of each trial). Figure 6A shows the 2-sec ERPs for each time bin. Figure 6B shows the activity collapsed across all five repetitions within the 2-sec time bins. As shown in Figure 7, subtracting the t1 condition from each of the other conditions isolated buildup-related activity. In contrast to the effect of Δf , the effect of time during the trial is a temporally broad positive enhancement at FCz that peaked between 150 and 250 msec after the ABA– onset, reversing in polarity at the right temporal electrode (T8) for both attend and ignore conditions.

The mean activity across the 150–250 msec time range was quantified at the nine frontocentral electrodes and the mean across these electrodes was entered into an ANOVA to test for effects of attention, Δf , and time. There was a main effect of time, $F(4,36) = 23.96, p < .001$, and attention, $F(1,9) = 13.66, p < .005$. The interaction between time and attention showed a non-significant trend, $F(4,36) = 2.68, p < .10$. The interaction between time and Δf was significant, $F(12,108) = 4.96, p < .005$, reflecting less buildup in the 0 semitone Δf condition. There was also a significant Attention \times Δf interaction, $F(3,27) = 9.28, p < .001$, due to an in-

Figure 6. ERPs at different time bins within the trial in the attend condition. (A) ERP response at FCz to 5 cycles (2000 msec) of the ABA– pattern with a box around a single repetition of the ABA– pattern for the five time bins (t1–t5). (B) Single-cycle ERPs at T7, FCz, and T8 showing the effect of time. Horizontal bars above the time scale represent pure tones in the stimulus pattern.



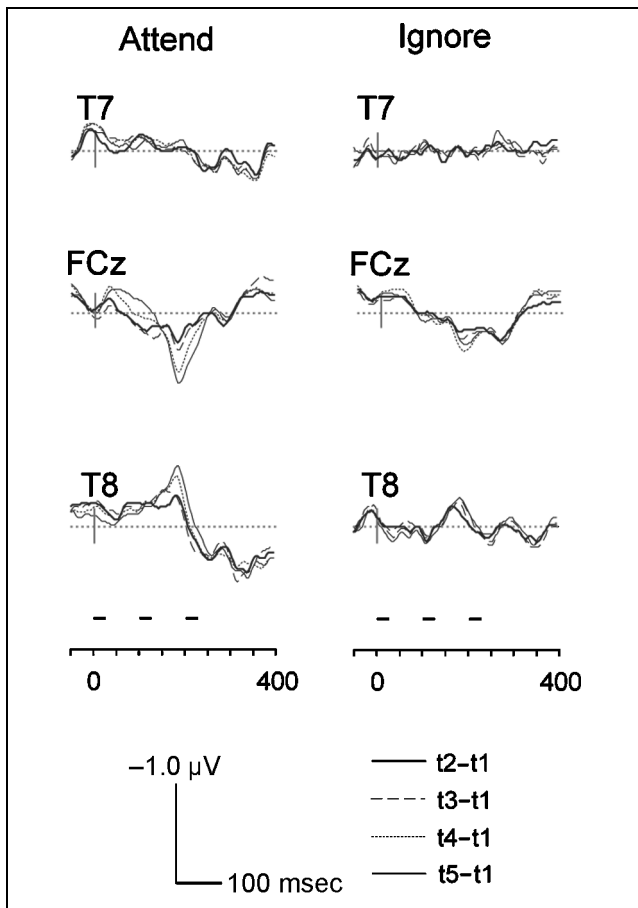


Figure 7. Buildup difference waves for attend and ignore conditions at T7, FCz, and T8. Horizontal bars above the time scale represent pure tones in the stimulus pattern.

creasing negativity as Δf increased that was larger in the attend condition.

As described earlier, the first ABA– pattern of each trial was excluded in order to control for transient ERPs that occurred at the beginning of the trial. To further rule out influences of transient responses, we repeated the previous ANOVA excluding the first time bin (i.e., the first six ABA– patterns at the start of the trial). Time, $F(3,27) = 9.54, p < .005$, attention, $F(1,9) = 10.41, p < .025$, and the Attention $\times \Delta f$ interaction, $F(3,27) = 7.28, p < .005$, remained significant and the interaction between attention and time, $F(3,27) = 4.17, p < .05$, became significant. The interaction between time and Δf was no longer significant, $F(9,81) = 2.29, p < .1$. Thus, the main effects of time and attention and their interaction cannot be attributed to transient activity at the beginning of the trial.

We also quantified buildup-related activity from 150 to 250 msec at T7 and T8 to test for effects of hemisphere, attention, Δf , and time. There was a negative displacement in the ERP as a function of time, $F(4,36) = 5.37, p < .05$, and attention, $F(1,9) = 8.62, p < .025$, but no interaction between time and attention. The negative responses related to buildup were larger at T8 than at

T7, $F(1,9) = 16.01, p < .005$, and there was a larger negative increase as a function of time at T8 than at T7, $F(4,36) = 7.27, p < .005$. This is therefore consistent with the results at frontocentral sites, with the additional finding that buildup processing showed RH dominance.

Brain Electrical Source Analysis

We used brain electrical source analysis (BESA 5.0) to determine how well sources in the primary auditory cortex (i.e., Heschl's gyrus) could account for Δf -based segregation and buildup activity measured at the scalp in attend and ignore conditions. For the Δf -related activity, we modeled the difference waves at all 65 electrodes collapsed across participants and Δf from 140 to 400 msec (P1–N1–P2 and N1c modulations). Using a broad time window was justified because the major peaks in this interval had similar source locations when modeled separately. For the buildup-related activity, we modeled the difference waves at all 65 electrodes collapsed across participants and Δf from 150 to 250 msec. Collapsing across participants and Δf levels enhanced the signal-to-noise ratio. The analysis assumed a four-shell ellipsoidal head model with relative conductivities of 0.33, 0.33, 0.0042, and 1 for the head, scalp, bone, and cerebrospinal fluid, respectively, and sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness). As an initial step, two symmetrical regional sources were placed at the Talairach coordinates of Heschl's gyrus in the RH and left hemisphere (LH) ($x = \pm 47, y = -26, z = 13$). Each source contained three orthogonal dipoles representing three dimensions of current flow at the source location (tangential, radial, and anterior/posterior). Maintaining the locations and orthogonality of the three dipoles in each regional source, the orientation of the first dipole was aligned with the maximum direction of activity.

Figure 8A and B (top) shows the RH and LH locations and orientations for the Δf -related and buildup-related sources, respectively, separately for attend and ignore conditions. The models yielded residual variances of 3.94% and 3.09% for the Δf -related activity in attend and ignore conditions and 8.98% and 6.71% for the buildup-related activity in attend and ignore conditions. Figure 8A and B (bottom) shows the source activity for the Δf - and buildup-related sources, respectively. For the Δf -related activity, the tangential sources accounted for most of the P1–N1–P2 peaks, with an additional contribution of the anterior/posterior source for the P1. The radial source accounted for the N1c in the attend condition and was larger in the RH than the LH, whereas in the ignore condition the radial source did not strongly reflect the N1c. For the buildup-related activity, the tangential and radial sources accounted for most of the activity, with a later peak of activity for the radial source. Similar results were obtained for the buildup activity

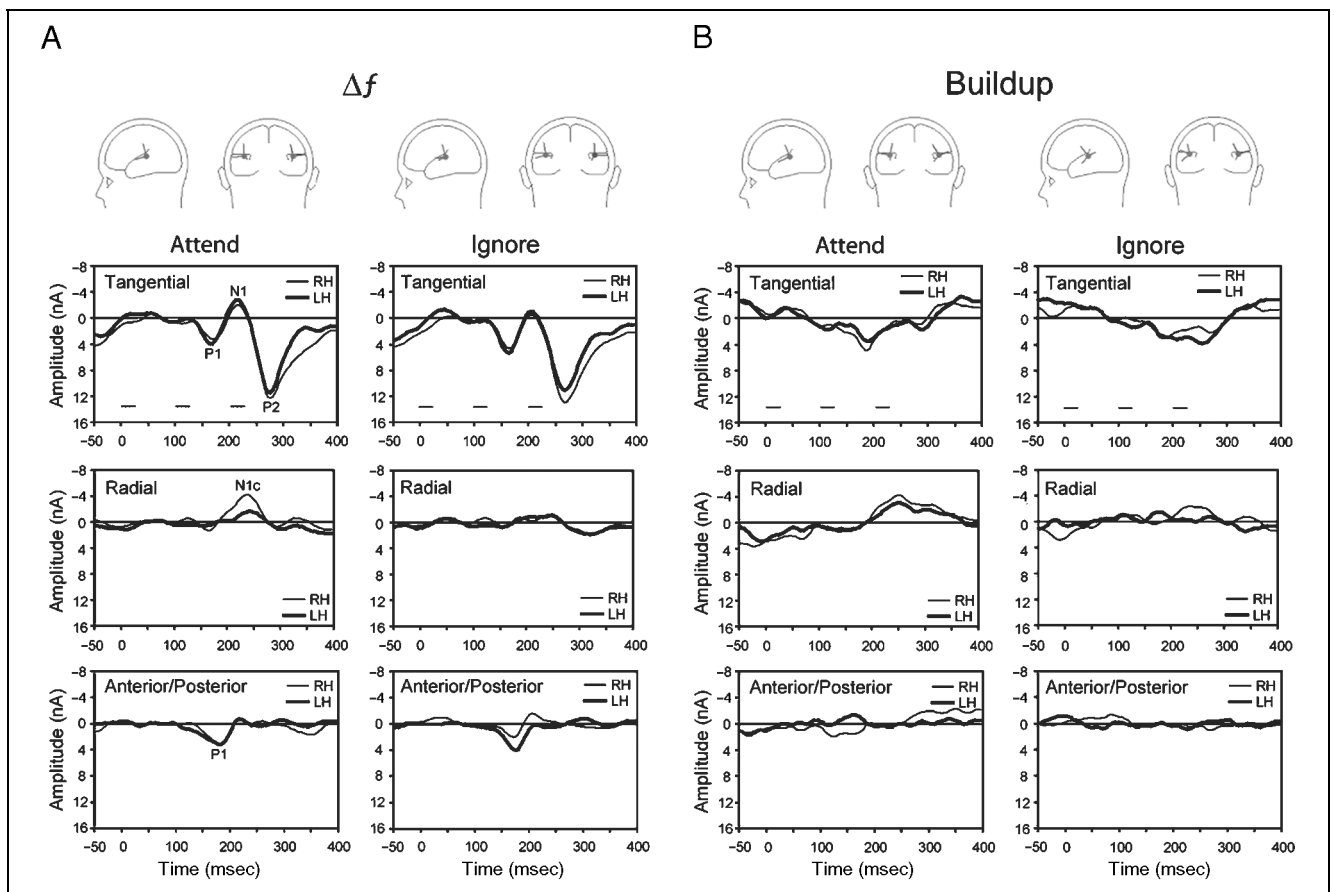


Figure 8. Brain electrical source analysis. (A) Symmetrical intracerebral sources of difference ERPs for Δf -related activity in Heschl's gyrus in attend (left) and ignore (right) conditions. Left and back views of the head showing source locations of orthogonal dipoles with tangential, radial, and anterior/posterior orientations relative to the temporal portion of the scalp. Source amplitude time courses of the tangential, radial, and anterior/posterior dipoles for Δf -related activity in the left hemisphere (LH) and right hemisphere (RH). Horizontal bars above the time scale in the tangential panel represent pure tones in the stimulus pattern. (B) Same as (A) but for the buildup-related activity. Note that effects of attention were mainly reflected in the radial sources.

when waveforms from the 0 semitone Δf condition were excluded. The radial sources accounted for the effects of attention more than the tangential sources for both the Δf - and buildup-related activity.

DISCUSSION

The likelihood of reporting two streams in the repeating tone patterns increased as Δf increased, consistent with previous studies (for a review, see Moore & Gockel, 2002). At the beginning of each trial, a transient neural response occurred, followed by an attention-dependent sustained negative potential that was similar to the response that occurs for long-duration sounds (Picton, Woods, & Proulx, 1978) and trains of repeated sounds (Picton, Campbell, Baribeau-Brown, & Proulx, 1978). The N2 was larger with increasing Δf likely due to a larger N1 to the second tone of the ABA- pattern. Similarly, the sustained response was larger with increasing Δf likely due to increased activity required to monitor two sustained streams, or to broader activation

of auditory cortices when the stimuli spanned a wider frequency range.

Superimposed on the SP was activity that tracked the rhythm of the ABA- patterns. Neural activity following onsets of the B tone showed a P1-N1-P2 enhancement as Δf increased, which was successfully modeled as bilateral sources in Heschl's gyrus. The source analysis likely reflects a large area of activation in the auditory cortex centered near Heschl's gyrus. This is consistent with previous studies showing enhanced neural responses following changes in pure-tone frequency (Näätänen et al., 1988; Picton, Woods, et al., 1978; Butler, 1968) and fundamental frequency and timbre of musical instrument tones (Jones, Longe, & Vaz Pato, 1998). A similar process may be responsible for the N2 modulation that was present in the transient response at the beginning of the trial. The particularly large P2 modulation in the current study correlated strongly with behavioral judgments of streaming. An additional modulation of the N1c that correlated with behavior occurred at the right temporal electrode but not at the left temporal electrode, suggesting an RH dominance

for Δf -based stream segregation. This asymmetry also occurred in the source analysis waveforms (Figure 8). The Δf -related modulations occurred even when participants ignored the sounds, implicating the auditory-evoked response as an index of automatic Δf -based segregation.

Behavioral studies have shown that streaming takes several seconds to buildup (Anstis & Saida, 1985; Bregman, 1978) and that this process is sensitive to top-down controlled processes (Cusack et al., 2004; Carlyon et al., 2001). An ERP wave from 150 to 250 msec following the onset of the ABA- cycle that was larger for non-0 Δf increased with time over the 10.8-sec trial. Although we did not measure perceptual buildup concomitantly with the ERPs, the timing of this ERP change is consistent with the timing of perceptual buildup of streaming that occurs over several seconds. Unlike the Δf -related activity, this neural activity was substantially reduced when participants ignored the ABA- patterns, consistent with behavioral studies showing effects of attention on perceptual buildup of streaming (Cusack et al., 2004; Carlyon et al., 2001).

The buildup activity we observed began its slow time-course shortly after the first A tone. This suggests that the buildup-related activity might respond to the onset of the repeating ABA- pattern, perhaps reflecting an increase in the likelihood of hearing two streams or a decrease in the likelihood of hearing a galloping rhythm. A negative polarity reversal of the positive frontocentral buildup occurred at the right temporal electrode as a function of time but not at the left temporal electrode, suggesting an RH dominance for stream formation. Another possible explanation for the buildup activity is that a negative difference (Nd) wave (Hansen & Hillyard, 1980) related to attending was present at the beginning of the 10.8-sec trial and as perceptual buildup of streaming took its course, participants no longer attended as much to the ABA- pattern. This would suggest that the positive difference waves we observed were actually the inverse of an Nd wave. Further experiments would be necessary to determine whether the buildup-related activity reflected an actual positive wave or the inverse of a negative wave. The source modeling of buildup-related activity was consistent with bilateral generators in or near Heschl's gyrus. The relatively high residual variances might reflect in part the activation of multiple sources over a relatively wide area of the superior and lateral temporal surfaces.

Support for a Place Model of Stream Segregation

The central idea of a place model is that when participants hear two streams of tones, spatially distinct populations of neurons are activated (Hartmann & Johnson, 1991). In the present study, increases in ERP amplitude with larger Δf could have arisen from the segregation of distinct activations corresponding to the A and B tones

in tonotopically organized structures. As two distinct populations of active neurons become farther apart in their frequency tuning, the less they will interact with each other, leading to a larger summed activation at the scalp. This interpretation is consistent with studies of streaming using single- and multi-unit activity (Bee & Klump, 2004; Fishman et al., 2004; Kanwal et al., 2003) and computational models (McCabe & Denham, 1997; Beauvois & Meddis, 1996).

Additional support for the involvement of low-level mechanisms in segregation processes come from behavioral studies in infants as young as 3 days old (McAdams & Bertoncini, 1997), and nonhuman animals (e.g., MacDougall-Shackleton, Hulse, Gentner, & White, 1998). One study that tested European starlings' perception of tone sequences provided evidence for perception that closely corresponded to adult human perception of streaming (MacDougall-Shackleton et al., 1998). The birds were first trained to peck one key when listening to a constant frequency ABA- tone pattern in a galloping rhythm (similar to the 0 semitone Δf condition in the current study), and to press a different key when listening to a single stream of tones either at the tempo of the A tones (i.e., A-A-...) or at the tempo of the B tones (i.e., -B-...). When presented with ABA- tone patterns, the birds were more likely to press the key corresponding to the streaming patterns when Δf increased. This mirrors human perceptual reports of the change in rhythm that accompanies streaming. These behavioral data were recently correlated with the difference in multi-unit responses to A and B tones in starlings' auditory forebrains (Bee & Klump, 2004). The current data and those of previous studies thus support stream segregation as a basic function of auditory processing across species.

Evidence for Distinct Mechanisms of Stream Segregation

ERP modulations related to Δf and buildup in the current study were differentially affected by attention, providing evidence for distinct mechanisms. Cusack et al. (2004) came to a similar conclusion in postulating low-level automatic segregation processes and separate buildup processes that are involved in the formation of perceptual objects and streams. Despite the evidence for dissociation between segregation and buildup processes, it remains unclear what neural structures and mechanisms are responsible for these two types of process. According to the peripheral channeling hypothesis, segregation of tone patterns depends on activation along a tonotopic representations in the cochlea and other subcortical auditory structures (Hartmann & Johnson, 1991). Given that tonotopic representations are retained up to the level of the primary auditory cortex (Kaas & Hackett, 2000), it is likely that segregation of tone sequences in the cochlea is transferred up

the ascending auditory system. Additional inhibitory processes between neurons responsive to different frequencies in the auditory cortex have been proposed to enhance frequency-based segregation (Bee & Klump, 2004; Fishman et al., 2004; Kanwal et al., 2003; McCabe & Denham, 1997), and recent evidence shows that inhibition operates on the N1 (Pantev et al., 2004; Sable, Low, Maclin, Fabiani, & Gratton, 2004). Inhibition between neurons that are tuned to different frequencies in the auditory cortex might enhance frequency-based segregation by sharpening the tuning curves in individual neurons, thus leading to less overlap in the populations of neurons that are responding to the alternating A and B tones.

The neural mechanisms underlying buildup of streaming (Anstis & Saida, 1985; Bregman, 1978) and the neural buildup observed in the current study are not as well understood as the mechanisms of frequency-based segregation. Given the strong influence of attention observed in the current study and previous behavioral studies (Cusack et al., 2004; Carlyon et al., 2001), it is likely that buildup of streaming is cortical in nature. This is consistent with an account of object perception in which subcortical and early cortical stages of processing extract features whereas later stages build more complex object representations. Further study is necessary to provide a deeper understanding of such higher-level aspects of auditory scene analysis.

Relation to Other Neurophysiological Paradigms in Humans

The current study provided information showing auditory cortical activity over time during streaming, demonstrating distinct neural processes related to frequency-based segregation and buildup of streaming. In addition to the source analysis performed in the current study, investigations using functional neuroimaging techniques might allow us to define networks of brain areas that are active during stream segregation. For example, a study using functional magnetic resonance imaging (fMRI), a measure of cerebral blood flow that correlates with neural activity, showed that posterior regions of the left auditory cortex were modulated by listening to alternating organ and trumpet tones when compared to a single stream of either organ or trumpet tones presented at the same rate (Deike, Gaschler-Markefski, Brechmann, & Scheich, 2004). This is consistent with the current study showing ERP modulations near the auditory cortex although the right temporal lobe appeared to be dominant in the current study. Another fMRI study using stimuli similar to those in the current study showed differential activity in the intraparietal sulcus depending on whether participants heard one or two streams (Cusack, 2005). Activations in such nonauditory regions might index higher-level

processes such as object formation or auditory attention to objects.

Conclusions

The present study demonstrated enhancements of the P1–N1–P2 and N1c components of the auditory-evoked potential that correlated with behavioral reports of stream segregation. An additional modulation reflected an increase in neural activity as a function of time while listening to the extended ABA– patterns that showed a similar time course as the buildup of streaming reported in behavioral studies (Anstis & Saida, 1985; Bregman, 1978). These two modulations were differentially affected by attention with a stronger reduction of the buildup-related activity while ignoring the sounds, compared to the frequency-related activity. These findings provide neurophysiological evidence for distinct mechanisms of streaming, one related to frequency-based segregation of tones in the auditory cortex and another related to the process of forming auditory objects.

METHODS

Participants

Ten young adults (6 men and 4 women, age range = 23–38 years, mean age = 29.5 years) participated after giving written informed consent according to the guidelines of the Baycrest Centre for Geriatric Care and the University of Toronto. All participants were right-handed except one, all had normal pure-tone thresholds (<20 dB HL from 250 to 8000 Hz in both ears), and all were screened for neurological and psychiatric illness. Five of the 10 participants were musically experienced (mean = 9.4 years formal training).

Materials and Procedure

As shown in Figure 1, stimuli were pure-tone patterns of alternating A and B tones with every other B tone omitted (taking the form ABA–ABA–ABA–...). These stimuli were presented binaurally through Sennheiser HD 265 headphones (Sennheiser Electronic, Old Lyme, CT) at 85 dB SPL. Within each trial, the A-tone frequency was always 500 Hz and the B-tone frequency was 500, 625, 750, or 1000 Hz. This corresponds to Δf levels of 0, 4, 7, and 12 semitones. Tone duration was 20 msec with 5.0-msec rise and fall times. The stimulus onset asynchrony (SOA) was 100 msec between adjacent A and B tones within each ABA– cycle. The silent duration (–) between ABA triplets was 100 msec (10 Hz). The A tones repeat every 200 msec (5 Hz) and the B tones repeat every 400 msec (2.5 Hz).

On each trial, participants were presented with 10.8 sec of the ABA– pattern (27 ABA– repetitions). Within a

block of trials, 80 trials were presented in which Δf varied pseudorandomly from trial to trial during electrophysiological recording (20 per Δf level). In the *attend* condition, participants indicated at the end of the sequence by pressing a button if they heard the pattern as one stream and another button if they heard the pattern as splitting into two streams by the end of the trial. The next trial began 2000 msec after the response. Participants were instructed to focus on the rhythm as a cue (i.e., not galloping or galloping) to indicate whether the pattern had split into two streams. They were also instructed to let their perception take a natural time course rather than biasing themselves towards hearing the patterns in one way or another. Each participant performed four blocks for a total of 320 trials in each experimental condition (80 per Δf level). Prior to the experiment, participants completed eight practice trials with two examples of each Δf level. In the *ignore* condition, the procedure was the same except participants watched a muted movie of their choice presented on a computer monitor with English subtitles. Instead of each successive trial being activated by a button press, each trial began 2000 msec following the completion of the previous trial. For all participants, the ignore condition took place at least 1 month after the attend condition. Each session lasted around 75 min.

Electrophysiological Recording and Analysis

In the attend condition, the electrophysiological responses were continuously collected and digitized (250 Hz sampling rate; band-pass filtered 0.05–50 Hz) from an array of 64 electrodes using NeuroScan Syn-Amps (Compumedics USA, El Paso, TX) and stored for off-line analysis. Eye movements were monitored with electrodes placed at the outer canthi and at the superior and inferior orbit. During recording, all electrodes were referenced to Cz but they were re-referenced to an average reference for off-line analysis. ERP recording in the ignore condition was identical to the attend condition except that the electrophysiological responses were digitized at a 1000-Hz sampling rate (band-pass filtered 0.05–200 Hz). The ERP data for the ignore condition were decimated to 250 Hz following averaging and were otherwise processed identically to the attend condition.

ERP activity was analyzed in the following three ways. The first analysis included 200 msec prior to the onset of the 10.8-sec trial and 2000 msec after the onset of the trial to test for effects of Δf and attention on the transient auditory-evoked response. There were 80 such 2200-msec epochs for each Δf level for each participant. The second and third analyses used epochs that included 48 msec before and 400 msec after the onset of each ABA– pattern, excluding the first and last ABA– patterns of each trial. These 448-msec epochs were sorted into five time bins taken from successive 2-sec periods within

the 10.8-sec trial, each bin containing five repetitions of the ABA– cycle. For each time bin, there were thus 400 epochs (80 trials \times 5 ABA– cycles) for each Δf level for each participant. We examined the effects of Δf in each time bin by subtracting the 0 semitone Δf condition from the non-0 semitone conditions at each time bin. Finally, to test effects of time, we compared activations at different time bins.

Trials contaminated by excessive peak-to-peak deflection ($\pm 150 \mu\text{V}$) at the channels not adjacent to the eyes were automatically rejected before averaging. ERPs were averaged separately for each level of Δf and electrode site for each participant. For each individual average, ocular artifacts (e.g., blinks, saccades, and lateral movements) were corrected by means of ocular source principal components using BESA 3.0 (Picton, van Roon, et al., 2000; Berg & Scherg, 1994). ERPs were digitally filtered to attenuate frequencies outside 0.1–30 Hz for the 2200-msec epoch and 1–20 Hz for the 448-msec epoch.

For the first epoch corresponding to the onset of the trial, ERP amplitude was measured relative to a 200-msec baseline. For the 448-msec epoch, Δf -related modulations were quantified using a baseline of 48 msec prior to the B tone because this tone was the one varying in frequency. Because of the ongoing nature of the ABA– patterns, ERP amplitude related to buildup was measured relative to the mean activity across the entire 448-msec epoch. The epochs were further analyzed by quantifying peak amplitudes at nine frontocentral electrodes (Fz, F1/2, FCz, FC1/2, Cz, C1/2) and the left and right temporal electrodes (T7 and T8).

Statistical Analysis

The proportions of trials in which participants reported hearing two streams in the attend condition were analyzed using a repeated-measures ANOVA with Δf (0, 4, 7, and 12 semitones) as the lone factor. ERP latency and amplitude averaged across nine frontocentral electrodes were analyzed using three-factor repeated-measures ANOVAs with Attention (attend vs. ignore), Δf , and Time bin (t1–t5) as factors. ERP latency and amplitude at temporal electrodes (T7 and T8) were analyzed using four-factor repeated-measures ANOVAs with Hemisphere (left and right), Attention, Δf , and Time bin as factors. When appropriate, the degrees of freedom were adjusted with the Greenhouse–Geisser epsilon (ϵ). All reported probability estimates are based on the reduced degrees of freedom although the original degrees of freedom are reported. Post hoc comparisons were performed using a Bonferroni correction for multiple comparisons. We report p values less than .05 as significant. ERP peak amplitudes were related to behavioral judgments of streaming by simple correlations for individual participants in each condition. These simple correlations were submitted to one-sample t tests, testing the

hypothesis that the correlations were different than 0, with Bonferroni corrections for multiple tests.

Acknowledgments

This research was funded by grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, and the Premier's Research Excellence Award. We thank Chenghua Wang, Jimmy Chen, Yu He, and Kelly McDonald for technical assistance.

Reprint requests should be sent to Joel S. Snyder, Department of Psychiatry - 116A, VA Boston Healthcare System, Harvard Medical School, 940 Belmont St., Brockton, MA 02301, or via e-mail: joel_snyder@hms.harvard.edu.

REFERENCES

- Alain, C., Achim, A., & Richer, F. (1993). Perceptual context and the selective attention effect on auditory event-related brain potentials. *Psychophysiology*, *30*, 572–580.
- Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, *5*, D202–D212.
- Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1072–1089.
- Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, *15*, 1063–1073.
- Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics*, *56*, 501–516.
- Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 257–271.
- Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America*, *99*, 2270–2280.
- Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics*, *59*, 81–86.
- Bee, M. A., & Klump, G. M. (2004). Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *Journal of Neurophysiology*, *92*, 1088–1104.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*, 229–241.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 380–387.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge: MIT Press.
- Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of human vertex potential. *Journal of the Acoustical Society of America*, *44*, 945–950.
- Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 115–127.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, *17*, 641–651.
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 643–656.
- Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *NeuroReport*, *15*, 1511–1514.
- Fishman, Y. I., Arezzo, J. C., & Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. *Journal of the Acoustical Society of America*, *116*, 1656–1670.
- Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, *49*, 277–290.
- Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, *9*, 155–184.
- Jones, S. J., Longe, O., & Vaz Pato, M. (1998). Auditory evoked potentials to abrupt pitch and timbre change of complex tones: Electrophysiological evidence of 'streaming'? *Electroencephalography and Clinical Neurophysiology*, *108*, 131–142.
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences, U.S.A.*, *97*, 11793–11799.
- Kanwal, J. S., Medvedev, A. V., & Micheyl, C. (2003). Neurodynamics for auditory stream segregation: Tracking sounds in the mustached bat's natural environment. *Network: Computation in Neural Systems*, *14*, 413–435.
- Lü, Z. L., Williamson, S. J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science*, *258*, 1668–1670.
- MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *Journal of the Acoustical Society of America*, *103*, 3581–3587.
- Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 43–51.
- McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America*, *102*, 2945–2953.
- McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *Journal of the Acoustical Society of America*, *101*, 1611–1621.
- Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, *80*, 479–483.
- Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, *88*, 320–333.
- Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, *69*, 523–531.

- Pantev, C., Okamoto, H., Ross, B., Stoll, W., Ciurlia-Guy, E., Kakigi, R., & Kubo, T. (2004). Lateral inhibition and habituation of the human auditory cortex. *European Journal of Neuroscience*, *19*, 2337–2344.
- Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Kei, J., Chenery, H. J., & Alku, P. (2004). Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. *Journal of the American Academy of Audiology*, *15*, 469–485.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neurotology*, *5*, 111–139.
- Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., & Trujillo, N. J. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurotology*, *4*, 64–79.
- Picton, T. W., Campbell, K. B., Baribeau-Brown, J., & Proulx, G. B. (1978). The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, NJ: Erlbaum.
- Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, *111*, 53–65.
- Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials: I. The nature of the response. *Electroencephalography and Clinical Neurophysiology*, *45*, 186–197.
- Sable, J. J., Low, K. A., Maclin, E. L., Fabiani, M., & Gratton, G. (2004). Latent inhibition mediates N1 attenuation to repeating sounds. *Psychophysiology*, *41*, 636–642.
- Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*, 22–34.
- Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., Näätänen, R., & Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences, U.S.A.*, *100*, 11812–11815.