Journal of Experimental Psychology: Human Perception and Performance

An Objective Measurement of the Build-Up of Auditory Streaming and of Its Modulation by Attention

Sarah K. Thompson, Robert P. Carlyon, and Rhodri Cusack

Online First Publication, April 11, 2011. doi: 10.1037/a0021925

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/11/$12.00 DOI: 10.1037/a0021925

An Objective Measurement of the Build-Up of Auditory Streaming and of Its Modulation by Attention

Sarah K. Thompson, Robert P. Carlyon, and Rhodri Cusack
MRC Cognition & Brain Sciences Unit, Cambridge, England

Three experiments studied auditory streaming using sequences of alternating "ABA" triplets, where "A" and "B" were 50-ms tones differing in frequency by Δf semitones and separated by 75-ms gaps. Experiment 1 showed that detection of a short increase in the gap between a B tone and the preceding A tone, imposed on one ABA triplet, was better when the delay occurred early versus late in the sequence, and for Δf = 4 vs. Δf = 8. The results of this experiment were consistent with those of a subjective streaming judgment task. Experiment 2 showed that the detection of a delay 12.5 s into a 13.5-s sequence could be improved by requiring participants to perform a task on competing stimuli presented to the other ear for the first 10 s of that sequence. Hence, adding an additional task demand could improve performance via its effect on the perceptual organization of a sound sequence. The results demonstrate that attention affects streaming in an objective task and that the effects of build-up are not completely under voluntary control. In particular, even though build-up can impair performance in an objective task, participants are unable to prevent this from happening.

Keywords: auditory streaming, attention, build-up, auditory scene analysis

In everyday life we must understand sounds, such as speech and music, that occur not in isolation but in the presence of multiple competing sources. An important aspect of this task of sound segregation concerns the separation and perceptual binding of sound over time. In an early study, Miller and Heise (1950) described how a pattern of alternating high- and low-pitched tones will undergo a perceptual split into two streams. They called this the "trill phenomenon." To a listener, it is as though there are two different real-world sounds coming from two different real-world objects. The percept is very compelling and has found extensive application in music, where, for example, an instrument with a solo voice, such as a flute or violin, is capable of simultaneously carrying different melodic lines. This perceptual segregation of successive events is referred to as *sequential streaming*. The sequence has essentially been parceled up into distinct perceptual objects—the streams—by the auditory system (Anstis & Saida, 1985; Bregman & Campbell, 1971; Cusack & Carlyon, 2004; van Noorden, 1975).

Early research on streaming used subjective tasks to uncover the relationship between the physical parameters of a sequence and its tendency to split into more than one stream. One of the most influential investigations was that of van Noorden (1975), who described the perceptual effects of manipulations on the simplest of streaming sequences—alternating tone sequences. Sequences of the form "ABA-ABA-. . .," where A and B are regularly repeating tones of different frequencies, with A repeating at twice the rate of B, are known to produce a particular streaming effect: When the sequence is integrated, a characteristic "galloping" rhythm is heard, whereas when the sequence segregates, it is heard as two regular streams of tones. By asking participants to change various stimulus parameters in order to induce one or the other percept, van Noorden (1975) showed that manipulations of the frequency difference between A and B, and of the repetition rate of individual sounds, were the two prominent factors influencing stream segregation. Subsequent researchers have shown that many other stimulus parameters, including temporal envelope, pitch, and spatial differences, can also affect how streaming sequences are perceived (for a review, see Moore & Gockel, 2002). A finding that is particularly important for the current investigation is that streaming tends to "build up" over time: for a given sequence, the tendency for listeners to report hearing two streams is greater later on in the sequence than near its beginning (Anstis & Saida, 1985; Bregman, 1978).

More recently, a number of researchers have developed new techniques for studying auditory streaming, and have started to address the issue of its neural basis and of its relationship to cognitive processes such as attention. Perhaps the most important technical development has been toward objective, performance-based measures of streaming. Advantages of performance-based methods include the fact that one can exclude effects based on response biases, and that such measures can help reveal whether a particular perceptual phenomenon is "compulsory" rather than reflecting a bias toward selecting from two or more possible perceptual representations. This general approach has been used successfully for other aspects of auditory scene analysis, for example, by showing that the continuity illusion can both improve and impair performance in forced-choice tasks (Carlyon, Deeks, Norris, & Butterfield, 2002; Carlyon, Micheyl, Deeks, & Moore, 2004; Plack & White, 2000).

Sarah K. Thompson, Robert P. Carlyon, and Rhodri Cusack, MRC Cognition & Brain Sciences Unit, Cambridge, England.

Correspondence concerning this article should be addressed to Robert P. Carlyon, MRC Cognition & Brain Sciences Unit, 15 Chaucer Rd., Cambridge, England CB2 7EF. E-mail: bob.carlyon@mrc-cbu.cam.ac.uk

Performance-Based Measures of Streaming

Hartmann and Johnson (1991) asked participants to identify previously heard melodies from mixtures where successive tones came from different melodies. The melodies can only be "heard out" when the streams are segregated. They tested melody identification performance in a number of different conditions where the successive tones that formed the A and B melodies were differentiated on some particular dimension. They found that the task was performed well when the melodies differed in frequency separation, ear of presentation, or timbre, while differences in other dimensions, such as level, envelope, and duration, did not significantly aid performance. They argued that this showed that early channeling in the peripheral auditory system was the primary driver of stream segregation.

Another performance-based measure of streaming arises from the finding that listeners are poor at making judgments concerning the relative timing of sounds in different streams (Bregman & Dannenbring, 1973; Vliegen, Moore, & Oxenham, 1999; Vliegen & Oxenham, 1999; Warren, Obusek, Farmer, & Warren, 1969). For example, Cusack and Roberts (2000) presented participants with sequences of alternating A and B tones in which half of the trials were isochronous. For the other half of trials, the first eight tones in the sequence were isochronous, but then the B tones began to "slip," shifting progressively earlier or later relative to the A tones over the next 12 tones. They found that listeners were worse at identifying which trials contained this "temporal slip" when the frequency separation between the tones was greater, and also when there was a timbre difference between the A and B sounds. More recently, Roberts, Glasberg, & Moore (2008) measured the smallest detectable deviation from isochrony in a brief sequence of alternating A and B tones, and reported that thresholds could be increased by the presence of a preceding "inducing" sequence consisting only of the "A" tones (c.f. Rogers & Bregman, 1993).

The temporal discrimination task described above was easier when sequences were heard as a single stream. Micheyl and colleagues described a task, using roughly similar stimuli, that was easier when sequences were split into two streams (Micheyl, Carlyon, Cusack, & Moore, 2005). They noted that the thresholds for frequency discrimination between two target tones were greatly increased by the insertion of other irrelevant tones before and after the targets, particularly when all tones fell into a similar pitch range. They therefore reasoned that a frequency discrimination task on successive B tones in an ABA sequence would be more difficult when participants were hearing a single-stream percept, as the presence of A tones in the same stream would tend to increase this interference effect. They found that thresholds for correct detection of an upward or downward shift in the frequency of the final B tone in a sequence of ABA triplets decreased markedly with increasing frequency separation, and that thresholds were generally lower as the sequence length increased.

Relationship of Streaming to Higher-Level Cognitive Processes

A question of increasing interest concerns whether auditory streaming arises solely from automatic, low-level mechanisms, or whether it is intimately connected with higher-level processes such as attention. Such an influence of attention could occur either due to mechanisms responsible for segregation and integration operating in the central auditory system and/or due to "top-down" modulation of peripheral processes. Both modeling (Beauvois & Meddis, 1991) and physiological experiments on animals (Fishman, Reser, Arezzo, & Steinschneider, 2001; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Pressnitzer, Sayles, Micheyl, & Winter, 2008) have shown that processes such as frequency selectivity and adaptation, which are present in the earliest stages of the auditory pathway, are, in principle, capable of accounting for the effects of frequency separation and build-up on stream segregation. For example, Micheyl, Tian et al. (2005) recorded the response to repeating ABA tone triplets of cells in region A1 of the rhesus monkey auditory cortex. For cells tuned to the A tone, the response to the B tone decreased both with increasing frequency separation (Δf) and for later tones in the sequence. They successfully used these neural responses to account for the effects of Δf, presentation rate, and "build-up" observed in human listeners presented with the same stimuli. More recently, Pressnitzer et al. (2008) showed that similar findings could be obtained in the cochlear nucleus of the guinea pig. Because frequency selectivity and adaptation have been observed in anaesthetized animals, including in the study by Pressnitzer et al. (2008), these results suggest that some processes that have an impact on streaming can occur in the absence of attention.

Although processes that occur in the absence of attention have an effect on streaming, there is also, we believe, good evidence that streaming can be influenced by attention. Two findings that support this conclusion stem from a study reported by Carlyon, Cusack, Foxton, & Robertson (2001). In one experiment, they presented repeating ABA triplets to four patients exhibiting unilateral neglect to left-sided visual stimuli as well as to both brain-lesioned and healthy controls who showed no signs of neglect. When the sequences were presented to the right ears of patients, their judgments were indistinguishable from those of the control groups. In contrast, when the sequences were presented to their left ear, they made significantly fewer two-stream judgments than controls. As the deficit in neglect patients is widely believed to arise from a difficulty in attending to stimuli in the contralesional (left) side of space, this result is consistent with their attentional deficit affecting their streaming percept. In a perhaps more direct study of the effects of attention on streaming, Carlyon et al. presented 20-s sequences of repeating ABA triplets to healthy participants' left ear. In the baseline condition, no sounds were presented to the right ear, and the participants were required to make subjective streaming judgments throughout the sequence. The classic "build-up" of streaming was observed. In the experimental conditions, sequences of noise bursts were presented to the right ear for the first 10 s; these either had slow onset times with an abrupt offset, or vice versa, to give the impression of "approaching" or "departing" sounds. When participants made "approach-depart" judgments on these sequences, and then, after 10 s, switched to making streaming judgments on the tones in their left ear, these judgments resembled those made at the *beginning* of the sequences in the baseline condition. In contrast, when they ignored the noises and made streaming judgments throughout, the results were identical to those in the baseline condition. Carlyon et al.'s results suggest that either streaming had not built up in the absence of attention or, alternatively, that the act of switching attention to the sequences "reset" the streaming process (Cusack,

Deeks, Aikman, & Carlyon, 2004; Moore & Gockel, 2002). The present study does not attempt to distinguish between these two interpretations; rather, the study aims to provide an objective measure of the build-up of auditory streaming, and, importantly, of its modulation by attention.

Overall Rationale

As discussed above, a number of studies have described objective measures of the effects of frequency separation and timbre on auditory streaming. However, no strong behavioral evidence has been obtained for an objective index of how streaming builds up as the duration of a sequence of tones is increased. There are two reasons why the possible interaction between streaming build-up and attention renders this issue of theoretical interest. First, if we accept the evidence for an effect of attention on streaming, it is possible that the build-up of streaming is less automatic than the effects of, say, Δf. For example, it could be that although the effects of Δf and of repetition rate represent an early separation of the responses to A and B tones into discrete neural populations (Pressnitzer et al., 2008), the effect of build-up occurs at a later and less automatic stage of processing (Snyder, Alain, & Picton, 2006). If so, then a forced-choice task in which stream segregation is disadvantageous might reveal effects of Δf but not of build-up. Second, although Carlyon et al.'s (2001) study included a control for response biases, and subsequent experiments using different methods of responding have come to the same conclusion (Carlyon, Plack, Fantini, & Cusack, 2003), it has nevertheless been suggested that their results may have been influenced by such biases (Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). More generally, the question remains as to whether attention can have an effect on an aspect of sound segregation that arises from *obligatory* processes.

To address these issues, we first developed a performance-based measure of the build-up of streaming and compared the results to those obtained with subjective methods (Experiment 1). Experiment 2 then exploited the objective measure to see whether attention could affect streaming as measured by a forced-choice task.

Experiment 1

Rationale

Experiment 1 had two aims. The first was to develop a behavioral task that would provide an objective measure of the build-up of auditory streaming. To do this, we chose a task that should be easier for subjects to perform when a sequence of repeating ABA triplets is perceived as a single stream. The task that we used was to detect a small delay on one of the "B" tones, presented either early or late in the sequence (Figure 1; cf. Vliegen et al., 1999). Two different frequency separations (Δf) of 4 and 8 semitones were used. Our prediction was that performance would be better when the delay was imposed early, rather than late, in the sequence, and would be better at the smaller Δf. Note, however, that this outcome should only occur if the build-up of attention is outside of listeners' voluntary control. The second aim was to compare the results with those of a subjective task, obtained with the same participants and stimuli, and to check that the B-tone delay did not affect streaming—for example, by "resetting" itself after the perception of the rhythm change.

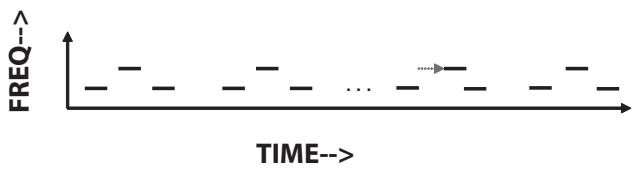

Figure 1. Schematic representations of the paradigm used in Experiment 1. Time is represented on the horizontal axis, and frequency is represented vertically. The first two triplets and the last triplet represent the standard, regular ABA- triplet, and the third represents the signal triplet, in which the B tone is delayed.

Participants

Participants were eight naïve listeners (five males, age range between 23 and 57, mean age = 36.3 years), all of whom self-reported normal hearing. They were recruited by word of mouth or from the MRC Cognition and Brain Sciences participant panel and were paid for taking part.

Stimuli and Presentation

Stimuli were sequences of repeating tones in the ABA-ABA- configuration, where A and B were 50-ms sine tones with 10-ms linear on- and off-set ramps. In the standard sequence, the tones were separated by a 75-ms silent gap. A gap of 125 ms (i.e., of equal length to another tone and gap) completed each ABA triplet, giving a total triplet length of 500 ms. Each sequence contained 25 triplets and therefore had a total duration of 12.5 s. The frequency of the B tone was roved on a trial-by-trial basis within ± 0.5 octaves of 800 Hz, while the A frequency was covaried with the B frequency and was 4 or 8 semitones lower, depending on the condition. The purpose of the frequency rove was to reduce the likelihood that the participants' percepts or responses for a given sequence would be biased by the previous sequence (Snyder, Carter, Hannon, & Alain, 2009; Snyder, Carter, Lee, Hannon, & Alain, 2008). Stimuli were generated digitally at a sample rate of 44100 Hz with 16-bit resolution. Sounds were presented through a VideoLogic Sonic Fury™ PC sound card, attenuated with TDT PA4 attenuators, and fed through a TDT amplifier to Sennheiser HD250 headphones. The stimuli were presented diotically at a level of 55 dB SPL to listeners who were seated individually in a sound-attenuating chamber.

Task

In the objective task, participants were required to detect a delay on one B tone, leading to a change described as a "skipping," irregular rhythm compared to the standard (see Figure 1). There were two levels of delay: 30 and 50 ms. Each 25-tone sequence could contain either a deviant on the 5th (*early*) or 20th (*late*) triplet (starting at 2.5 and 10 s, respectively) as well as at both or neither positions. Participants were informed that there might be zero, one, or two rhythm deviants in any given sequence. They responded to a target by using a mouse to click on a virtual button

on a computer screen. All participants completed a practice block with feedback, at $\Delta f = 4$ semitones; no feedback was provided during the main experiment. The ability of participants to perform the task was assessed by calculating a d′ measure of performance. As deviants could occur both in the early and late positions, it was necessary to assign hits and false alarms separately to the early and late categories. For sequences containing no deviants, responses occurring before 6.25 s (halfway through the sequence) were treated as "early" false alarms, and those occurring after this time were treated as late false alarms. For sequences containing a deviant, responses before 10 s—the time at which the late deviant might occur—were treated as "early" hits or false alarms (depending on whether an early deviant was present), and those occurring after this time were treated as "late" hits or false alarms (depending on whether a late deviant was present).

In the *subjective* task, participants were again asked to listen to 12.5-s sequences of repeating ABA-tone triplets, where A tones were either 4 or 8 semitones below the frequency of B, which was roved on a trial-by-trial basis within ± 0.5 octaves of 800 Hz. They were next told that the sequences could be heard in one of two ways, either as a galloping rhythm (called "horse"), where A and B tones were integrated into a single stream, or as two separate streams of low A and high B tones (this was described as the "morse" percept, as the streams sound somewhat like Morse Code; Cusack et al., 2004). Participants listened to examples of sequences at different frequency separations (including $\Delta f = 0$) in order to give them some idea of the different perceptual organizations. They were informed that their perception of one or two streams may change during the course of a sequence and were asked to track their perceptions on a moment-by-moment basis, by pressing "1" on a computer keyboard when they heard the one-stream organization and "2" when they heard two streams. There was also a short practice block for this section of the experiment, to allow participants to become accustomed to the task. No feedback was given for this, as, of course, there was no "correct" response.

During the subjective experiment, participants heard sequences that were identical to those used in the *objective* task; therefore, there could be delayed B tones in either the early or the late positions (or both, or neither). However, for this task, participants were instructed to ignore these changes if they heard them. Each participant made judgments for 112 sequences in total, which were presented as a single, self-paced block lasting upwards of 24 min. The order of presentation of trial types was fully randomized.

Responses were organized into 1-s bins. The first time bin (0–1 s) was discarded, as participants made very few responses in this range. After an initial response, streaming judgments were assumed to remain the same until a switch was recorded. (On average, participants made their first response at 2 s into a sequence.)

The *objective* task was always performed before the *subjective* task, as we did not want to draw participants' attention to the possibility of different streaming percepts. Nevertheless, verbal reports after the objective task indicated that several of the participants had "happened upon" the concept of streaming during their test. These participants commonly reported trying to "hold together" the streams when performing the *objective* task.

Results: Objective Measure

Figure 2a shows the mean detection rates across eight participants (as d′ scores) for each of the four experimental conditions ($\Delta f = 4$, *early*; $\Delta f = 4$ *late*; $\Delta f = 8$, *early*; and $\Delta f = 8$, *late*). The error bars represent between-subject standard errors. As predicted, there was a significant interaction of frequency separation ($\Delta f = 4$ or 8 semitones) and sequence position (*early* or *late*), such that the task was performed worse later in the sequence, especially at the larger frequency separation ($F_{(1,7)} = 9.482$, $p < .02$). There was also a significant main effect of Δf ($F_{(1,7)} = 17.385$, $p < .005$). The effect of sequence position also approached significance ($F_{(1,7)} = 5.141$, $p = .058$).

The results of Experiment 1 conform to the pattern of results expected from an effect of streaming on the detection of a "skip." Specifically, the main effect of Δf and its interaction with time-on-sequence are consistent with the fact that streaming is greater at wider separations and that it also builds up faster. An alternative explanation is that performance was worse later in the sequence due to some nonsensory factor—for example, fatigue—and that the interaction arose because performance at $\Delta f = 4$ semitones was close to ceiling. However, we consider this unlikely because a study using the same participants and very similar stimuli, but in which the task was to detect a change in the frequency of one of the B tones, showed that performance was *better* when the target was presented later, compared to earlier, in the sequence (Carlyon et al., 2010). Micheyl, Carlyon et al. (2005) have also reported better performance late in the sequence for a frequency-change task.

Results: Subjective Measure

Figure 2b plots the subjective streaming judgments at two frequency separations, averaged across the eight participants. The data show the characteristic "build-up" pattern of responses, where the probability of making a two-stream judgment increases with increasing sequence duration, and with this build-up being faster at the wider frequency separation (Anstis & Saida, 1985). The two arrows on the chart indicate the points at which the delayed "B" tones occurred, in those sequences that contained them.

Figure 2c shows the same data plotted separately for sequences in which a delay deviant occurred early, late, both, or not at all during the sequence. It can be seen that there is no evidence for a "resetting" of streaming following the delays. Indeed, the largest apparent discontinuity occurs at $\Delta f = 4$, 7 s into the sequence, but this occurred when the deviant was only present early in the sequence.

Experiment 2

Rationale

The aim of Experiment 2 was to apply the objective *rhythm* task used in Experiment 1 to study the effects of attention on the build-up of auditory streaming. The main comparison was between the detection of a deviant late in a sequence when participants had been attending to that sequence throughout, compared to when they had just switched their attention from a competing stimulus to the sequence.

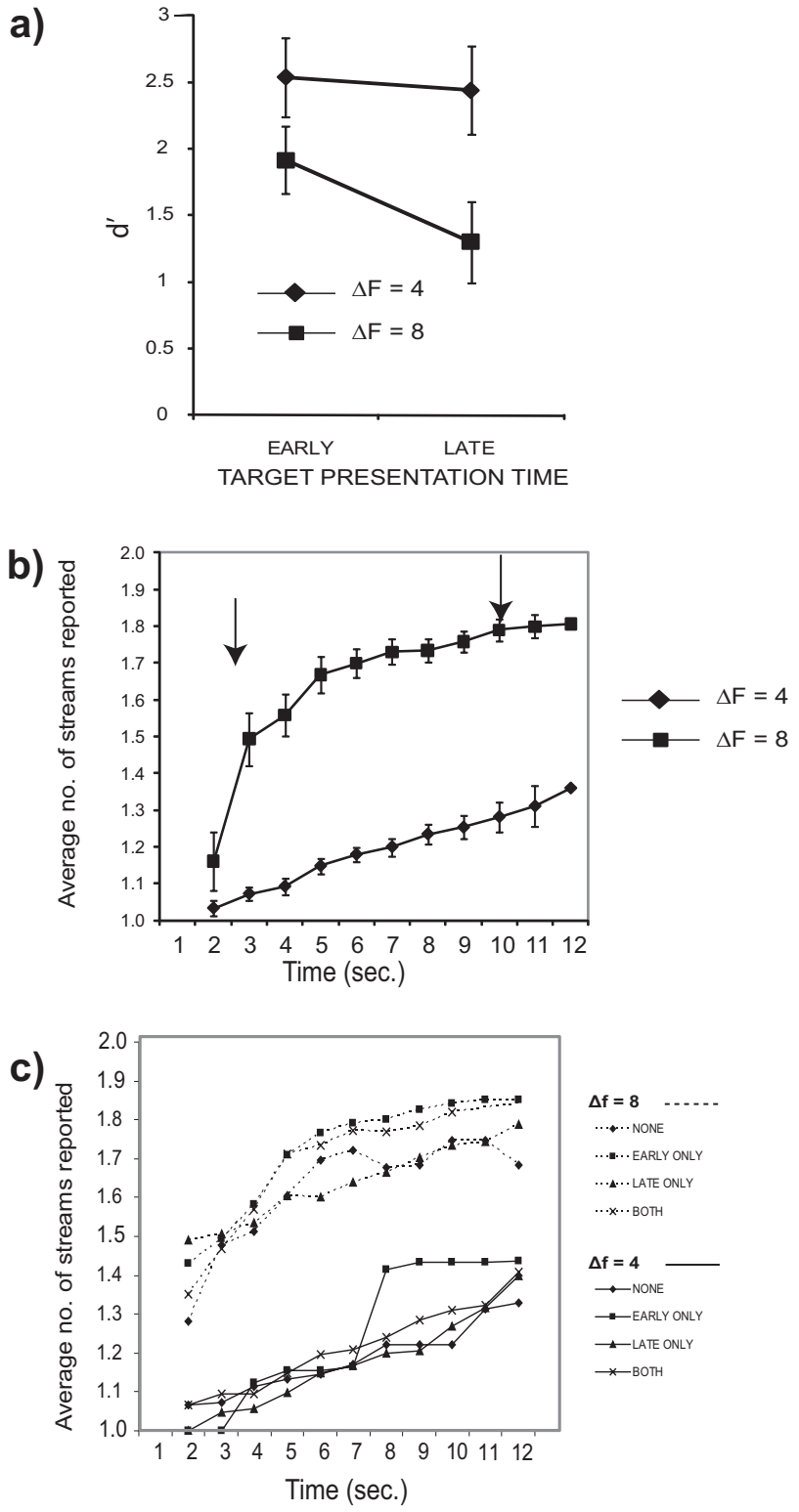

Figure 2. (a) Performance on the objective task of Experiment 1, averaged across the two delay values and across participants. Error bars show standard errors. (b) Results of the subjective streaming judgments of Experiment 1. Each line shows the average, across participants, conditions, and trials, of the number of streams heard as a function of time. Arrows show the times at which the deviants (delayed B tones) might occur. (c) Results of the subjective task shown separately for trials in which deviants occurred in the early, late, both, or no positions.

Participants

Eight participants (four women) with self-reported normal hearing took part. They were seated individually in a double-walled, sound-attenuating chamber, and listened to stimuli presented at 55 dB SPL over Sennheiser HD250 headphones.

Stimuli and Procedure

The stimuli played to the left ear were 13.5-s sequences consisting of tones in the ABA- pattern, where the A and B tones were 50-ms sinusoids with 10-ms onset ramps. The interstimulus interval (ISI) between the tones was 75 ms, with a further 125-ms silence following the second A tone, giving a total triplet duration of 500 ms. The frequency of the A tone was roved on a trial-by-trial basis within ± one-half octave of 800 Hz, and the B tone frequency was either 4 or 8 semitones ($\Delta f = 4$ or $\Delta f = 8$) higher than the A tone in any one sequence.

In the right ear, starting simultaneously with the tone sequences in the left, was a 10-s sequence of noise bursts. These were created by digitally filtering white noise between 2000 and 3000 Hz using a brick wall bandpass filter (60 dB down in stopbands). They either increased in amplitude over their 400-ms duration (*approach* noises: 350-ms linear attack ramp, 50-ms decay linear ramp) or decreased (*depart* noises: 50-ms linear attack ramp, 350-ms linear decay ramp), giving the impression of "approaching" or "departing" sounds. The noise bursts were presented at an average rate of 1 Hz, with a jitter of up to ±250 ms, ensuring that they did not match the left-ear sequence in rhythm.

On half of the trials ("attend" trials), participants were instructed to attend to the left-ear tone sequence throughout its length and detect deviants that could appear *either* early (at 2.5 s) *or* late (at 12.5 s), but never in both positions. The probability of a deviant occurring was 50%. The deviant was a B tone that occurred 50 ms later than in a standard triplet. In the other half of the trials ("switch" trials), participants were instructed to listen to the noise sequence and judge each noise burst as "approaching" or "departing." When the noise burst sequence finished, a visual cue instructed them to switch their attention to the ABA sequence and perform the delay deviant detection task.

The two types of trial were randomized within blocks, and the on-screen interface ensured that only the pertinent response type was available to the individual during the course of any particular trial. Responses were made via a computer keyboard, and the participants were instructed to use the space bar when they heard the delay deviant and to use the keys "1" and "2" on the number pad to indicate "approach" and "depart" sounds, respectively. All participants were given a practice block with feedback in order to familiarize them with the tasks. There were then two experimental blocks, with 48 trials in each block, giving a total experiment length of approximately 25 min. A hit was defined as a "yes" response following presentation of a delayed tone; any response that occurred after 2.5 s and before 12.5 s was defined as an "early" response, while any later response was presumed to be a late response. If a response occurred in a window in which no deviant had been presented, it was counted as a false alarm. Where sequences contained no deviant, any response that occurred before the midpoint of the sequence, 6.75s, was counted as an early false alarm, and any that appeared later was a late false alarm.

Results: Distracter Task

All participants could complete the approach-depart categorization task, with a mean of 76% of individual tokens being correctly classified as approaching or departing (chance = 50%). Scores ranged from 64 to 89% overall. This level of performance shows that participants were attending sufficiently to the stimuli to perform above chance but that the task was not trivially easy.

Results: Deviant Detection Task

Figure 3a shows sensitivity (d′) to the presence of a deviant that appeared in three separate contexts—at the beginning of a sequence (2.5 s; *early*), toward the end of the sequence to which attention had been paid throughout (12.5 s; *attended*), or late in a sequence where attention had been focused on a distracter sequence in the other ear (*switched*). At the narrower frequency separation (4 semitones), where we would expect the sequence to be primarily heard as one stream in both the early and late time intervals (Experiment 2), performance was good and approximately equal in all three conditions. At the wider separation (8 semitones), the results replicate the finding of experiments 1 and 2 that when participants attend to the tones throughout, the detection of a deviant is worse late than early in the sequence, consistent with streaming having built up. A two-way ANOVA was performed on the performance on those sequences where they attended throughout, with the factors of frequency separation and position in sequence (early or late). There were significant main effects of frequency separation ($F_{(1,7)} = 20.99$, $p < .01$) and the effect of position approached significance ($F_{(1,7)} = 5.542$, $p = .051$). Importantly, the interaction was highly significant ($F_{(1,7)} = 18.354$, $p < .005$), showing that the pattern of results observed in Experiment 1 also occurred when the sequences were presented monaurally and in the presence of an ignored stimulus in the other ear.

The crucial comparison was between performance later on in the sequences, when participants had switched attention to the sequence during the trial, compared to when they had been attending to the sequence throughout. When they had switched attention, detection of a late deviant was substantially better than if they had been attending throughout. A particularly compelling aspect of this result is that adding an additional task demand—switching attention midway through a sequence—produced an *improvement* in performance. This contrasts with the more usual finding that task switching produces a performance *decrement* (for a review, see Monsell, 2003).

The above trends were supported by statistical analyses. A two-way ANOVA was performed on the two late context conditions, with the factors of frequency separation and attention. This showed significant main effects of both frequency separation ($F_{(1,7)} = 19.720$; $p < .005$) and attention ($F_{(1,7)} = 7.176$; $p < .05$), as well as a significant interaction ($F_{(1,7)} = 18.931$; $p < .005$). A *t*-test performed subsequently, comparing the two late conditions only at $\Delta f = 8$ semitones, revealed a significant difference ($t_{(7)} = -4.825$; $p < .01$). At 4 semitones, there was no difference between the *attended* and *switched* conditions ($t_{(7)} = 0.019$; $p = .985$), consistent with the results of Experiment 1 (see Figure 2a); indeed, at this frequency separation, participants primarily hear a single stream even late in the sequence (see Figure 2b). Furthermore, at

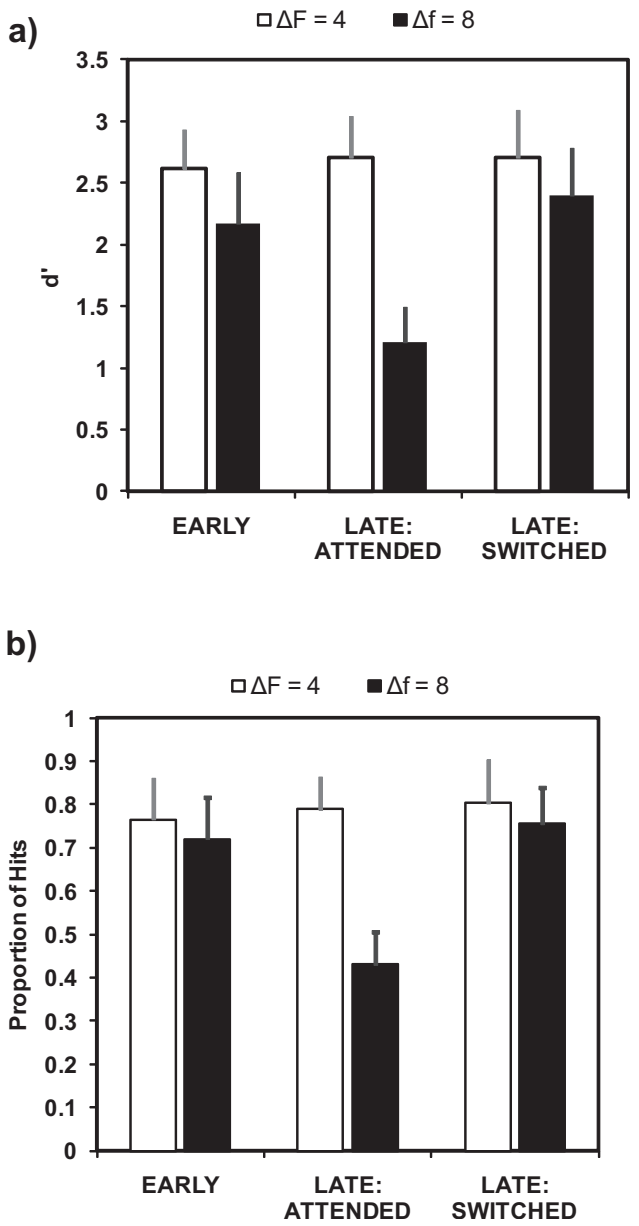

Figure 3. (a) Performance (d′) on the rhythm task, averaged across participants, for the two frequency separations and three conditions of Experiment 2. (b) Proportion of hits, averaged across participants, for the two frequency separations and three conditions of Experiment 2.

the 8 semitone frequency separation, there was no difference between the *early* (attended) condition, and the *switched* condition ($t_{(7)-} = -0.998$; $p = .352$).

One potential complication in the interpretation of the results comes from our decision to count all responses after 6.75 s, in sequences containing no deviants, as "late" false alarms. In the "switched attention" condition, participants could not make responses to the delayed-tone deviants until the noise bursts had ended, 10 s into the sequence. This could have reduced the number of false alarms to late deviants, and this, in turn, could theoretically have accounted for the higher d′ compared to the condition where participants attended to the tones throughout. To check this, we also calculated the number of hits in the various conditions, and plotted the results in Figure 3b. It can be seen that the pattern of results is the same as for the d′ values, demonstrating that the results obtained in this experiment were not strongly influenced by differences in false alarm rate between conditions. These results were supported by a two-way ANOVA on the hits obtained in the two late-context conditions, which revealed significant main effects of Δf ($F_{(1,7)} = 18.34$, $p < .01$) and of attention ($F_{(1,7)} = 8.22$, $p < .05$), as well as a significant interaction ($F_{(1,7)} = 31.6$, $p < .001$). A *t*-test performed on the two late conditions at Δf = 8 showed a significantly greater hit rate in the switched-attention condition ($t_{(7)} = 4.41$; $p < .01$).

General Discussion

Objective Measure of Streaming Build-Up

We believe that a task that requires subjects to detect timing differences between the A and B tones provides a simple objective measure not only of the effects of frequency separation on auditory streaming (Bregman & Dannenbring, 1973; Vliegen et al., 1999; Vliegen & Oxenham, 1999) but also of the build-up of auditory streaming. The paradigm used here incorporates a simple, easily learnt task that can be used to index stream segregation without the subject having to make any explicit judgments about their streaming percepts. The strong interaction between the effects of Δf and the time at which targets are presented are qualitatively consistent with the subjective measures obtained using identical stimuli and with the same participants.

A similar task was used to study temporal effects in stream segregation by Roberts et al. (2008). They used a two-interval forced choice, combined with an adaptive procedure, to measure the smallest detectable deviation from isochrony in a six-tone sequence of alternating low and high tones ("ABABAB"). Their results showed that the presence of various "inducer" sequences could increase these thresholds, and they argued that these results represented an objective measure of the build-up of auditory streaming. Their stimuli differed from ours in that the inducer sequences consisted only of the low ("A") tones, and it is possible that the "build-up" that they referred to arose from different processes than the change in streaming percept that occurs when both the A and B tones are repeated an increasing number of times. For example, it is known that prior presentation of one component of either a complex tone (Carlyon, 1994; Dannenbring & Bregman, 1976; Gockel & Carlyon, 1998) or of a sequential tone pair can "capture" that component from the mixture, and that similar effects do not occur when the whole complex is presented beforehand (Carlyon, 1994). As Roberts et al. (2008) point out, some stimulus parameters, such as interaural time differences (ITDs), appear to have much stronger effects when the inducer and test sequences differ in ITD than when the A and B tones within a single sequence differ in ITD. Furthermore, one might suspect that an inducer consisting only of the "A" tones might either bias one to attend to the novel B tones, whereas no such asymmetry would be expected as the duration of a repeating AB sequence is increased. Alternatively, or additionally, repeated exposure to the inducer might selectively adapt those neurons tuned to the A tones. However, one similarity between the two paradigms, revealed by

the results, is that both seem to index a process that is not entirely under the listener's voluntary control; although it would benefit the participant to "hang on" to the one-stream percept, it seems that this is not possible, a finding that, intuitively at least, seems slightly more surprising in the present paradigm where "capturing" effects are unlikely to take place.

Physiological correlates of streaming build-up have previously been reported in both animals and humans. As mentioned in the Introduction, single-cell recordings both from the auditory cortex of the awake rhesus monkey and from the cochlear nucleus of the anaesthetized guinea pig have shown that adaptation can influence neural responses in a way that closely mimics the way streaming builds up over time. Three electrophysiological studies in humans have also captured the effects of build-up. Sussman, Horvath, Winkler, and Orr (2007) measured the mismatch negativity (MMN)—a negative deflection to a rare deviant in a sequence of more common standards—to a tone that had a different intensity from the others in a sequence of tones having the same frequency. When this sequence was accompanied by an "interfering" sequence of tones, whose intensities varied from tone to tone, no MMN was observed if the interfering tones had a frequency close to that of the sequence containing the deviant. When the interfering tones were more distant, however, an MMN was observed when the deviant occurred toward the end of the 2.5-s sequence but not when it occurred nearer the start. Although Sussman et al. (2007) did not statistically compare the size of the MMN between the different conditions, the fact that it was influenced both by Δf and by time-in-sequence is consistent with it having indexed some processes (e.g., adaptation and frequency selectivity) that have a strong influence on streaming. We have also made preliminary progress toward an MMN measure of streaming by adapting the objective task described here. Our approach differed from that of Sussman et al. (2007) in that the MMN should be largest when the sequence is heard as a single stream. We did indeed observe that the effects of Δf and of time-in-sequence interacted, with the MMN being largest for small Δf and for deviants presented early in the sequence (Carlyon et al., 2010). Evidence using a different paradigm from the MMN was provided by Snyder et al. (2006), who measured evoked responses to 10.8-s sequences of repeating ABA triplets. They observed a number of peaks in response to each triplet, including N1 (latency $\sim$120 ms), P2 (160 ms) and N2 (200 ms). The size of the P1-N1 and of the P1-N2 deflections both increased with increasing Δf. There was also a positive deflection, peaking 150–250 ms after the onset of each triplet, which was larger for triplets later than earlier in the sequence. They concluded that these two measures indexed different aspects of auditory streaming, and, as discussed in the next subsection, noted the different effects of attention on the two measures.

The behavioral measure of streaming build-up reported here, and physiological measures from animals and humans, each have their own advantages and weaknesses. Physiological measures in animals have the capacity to identify neural processes that are likely to influence streaming and to constrain where and/or when in the auditory system they occur. Behavioral measures, and, arguably, the MMN, necessarily provide less neural specificity but more directly tap perception by demonstrating the influence of (in the present case) Δf and build-up on listeners' perceptual abilities. As noted above, a particular feature of the method reported here is that it shows a reduction in performance for targets later in the sequence, thereby showing that the streaming build-up is at least to some extent compulsory.

Attention

Perhaps the most important finding of the present study is the fact that the effects of attention on the build-up of auditory streaming can be measured using an objective task. The use of a forced-choice method rules out the possibility that these effects, previously reported using subjective measures, could arise from criterion shifts or biases at the response stage. The results also show that attention can mitigate the effects of a phenomenon—the build-up of streaming—which, at least when subjects are attending, appears to be outside of voluntary control. This finding, combined with the previous observation that build-up can be reduced or reset by silently counting backward in threes (Carlyon et al., 2003), raises an intriguing possibility. Although participants were apparently unable to "hang on" to the one-stream percept when they were attending to the tones, leading to a decrease in performance late in the sequence, they might have been able to improve performance instead by briefly performing a mentally distracting task and then returning their attention to the tones. In other words, in order to hear a one-stream percept, one should not try to do so but instead briefly divert attention elsewhere. Our results are also consistent with previous reports that neither the "bistable" nature of streaming percepts, nor their dependence on the context in which they are presented, can be eliminated by the instructions given to the participants (Pressnitzer & Hupe, 2006; Snyder et al., 2008).

One important question that is not answered by the present study is exactly how the tone sequences are represented in the brain while they are unattended. Two possibilities that have been previously mentioned are that streaming does not build up, so that the representation is of a single stream (Carlyon et al., 2001), or that streaming *does* build up but that it is "reset" when attention is reassigned to the tonal sequence (Cusack et al., 2004; Moore & Gockel, 2002; Roberts et al., 2008). A variant of this latter explanation has been suggested by Cusack et al. (2004), who reported that streaming could be "reset" equally well by introducing a silent gap of a few seconds into a tone sequence, and by leaving the sequence uninterrupted and briefly diverting the participant's attention to a competing task. In both cases, listeners' subjective streaming reports after the gap or attention switch were similar to those at the start of a sequence. Because, in the absence of a competing task, attention may be exogenously drawn to the start of a sequence, they suggested that the default representation may be for the tones in a sequence to be segregated, and that the commonly reported effects of "build-up" may be better described as a recovery from a "resetting" effect caused by attention being attracted toward the start of the sequence.

Two electrophysiological studies have investigated the modulation of measure of streaming build-up and of its modulation by attention. As noted in the previous subsection, Snyder et al. (2006) measured a positive deflection, with a latency of 150–250 ms, whose amplitude increased during a 10.8-s sequence of ABA triplets. This increase was greater when subjects attended to the tone sequence than when they watched a subtitled movie. As they pointed out, this finding is consistent with attention influencing the build-up of streaming. However, if attention to the start of a

sequence is responsible for the initial one-stream percept, then Snyder et al.'s (2006) finding could also possibly be due to the competing task reducing the participants' orientation toward the onset of each sequence. The MMN index of streaming build-up, reported by Sussman et al. (2007), has also been interpreted in terms of the effects of attention on streaming because it was obtained while participants were instructed to ignore the tone sequence and to monitor a continuous noise for a change in intensity. However, it should be noted that the detection of an intensity change is a task that may not require substantial attentional resources, as it does not show a task-sharing cost in a divided attention paradigm (Bonnel & Hafter, 1998). Furthermore, there was no comparison condition in which participants were instructed to attend to the sequences, and so, even if the competing task were attentionally demanding, one could only conclude that *some* build-up occurred without *full* attention rather than that attention had no effect in particular condition.

Another possible scenario is that the sequences are represented neither as one nor as two streams (cf. Brochard, Drake, Botte, & McAdams, 1999). For example, the percept that results from different-frequency tones presented at a very slow rate (e.g., <1 Hz) cannot really be described as integrated or segregated, but instead, each tone is perceived as a separate event. These sparse events are not perceptually bound either in time or in frequency and are thus perceived as isolated instances and not component parts of a larger whole. It may be that, in the absence of attention, even faster sequences remain "unbound" and do not conform to either a one- or a two-stream representation. The existence of this state of limbo would not conform easily to the framework adopted by physiological studies, in which streaming is represented by the degree to which neural responses are driven by one versus both tones, and where no explicit links need to be made between the responses to different tones (Fishman et al., 2001; Micheyl et al., 2007). Rather, it would have to relate to higher-level processes responsible for making those links.

Clearly, the issue of how unattended streams are represented in the nervous system remains unsolved. What the present results already make clear, though, is that attention can have a profound influence on the auditory streaming process per se, and that its effect can be measured in an objective, forced-choice task.

References

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. (1991). A computer-model of auditory stream segregation. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 43,* 517–541.

Bonnel, A.-M., & Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception and Psychophysics, 60,* 179–190.

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 380–387.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Bregman, A. S., & Dannenbring, G. (1973). Effect of continuity on auditory stream segregation. *Perception & Psychophysics, 13,* 308–312.

Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1742–1759.

Carlyon, R. P. (1994). Detecting mistuning in the presence of synchronous and asynchronous interfering sounds. *Journal of the Acoustical Society of America, 95,* 2622–2630.

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.

Carlyon, R. P., Deeks, J. M., Norris, D., & Butterfield, S. (2002). The continuity illusion and vowel identification. *Acta Acustica united with Acustica, 88,* 408–415.

Carlyon, R. P., Micheyl, C., Deeks, J. M., & Moore, B. C. (2004). Auditory processing of real and illusory changes in frequency modulation (FM) phase. *Journal of the Acoustical Society of America, 116,* 3629–3639.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception, 32,* 1393–1402.

Carlyon, R. P., Thompson, S. K., Heinrich, A., Pulvermuller, F., Davis, M. H., Shtyrov, Y., . . . Johnsrude, I. S. (2010). Objective measures of auditory scene analysis. In E. Lopez-Poveda (Ed.), *Advances in Auditory Physiology, Psychophysics, and Models* (pp. 507–520). New York, NY: Springer.

Cusack, R., & Carlyon, R. P. (2004). Auditory perceptual organization inside and outside the laboratory. In J. G. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 15–48). Amsterdam, Netherlans: Elsevier Academic Press.

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.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception and Psychophysics, 62,* 1112–1120.

Dannenbring, G. L., & Bregman, A. S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 544–555.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey (Vol 151, pp. 167, 2001). *Hearing Research, 151*(1–2), 167–187.

Gockel, H., & Carlyon, R. P. (1998). Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection. *Journal of the Acoustical Society of America, 104,* 3534–3545.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–184.

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.

Micheyl, C., Carlyon, R. P., Cusack, R., & Moore, B. C. J. (2005). Performance measures of auditory organization. In D. Pressnitzer, A. de Cheveigné, S. McAdams & L. Collet (Eds.), *Auditory Signal Processing: Physiology, Psychoacoustics and Models* (pp. 203–211). New York, NY: Springer.

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., . . . Wilson, E. C. (2007). The role of auditory cortex in the formation of auditory streams. *Hearing Research, 229,* 116–131.

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake Macaques. *Neuron, 48,* 139–148.

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America, 22,* 637–638.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7,* 134–140.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88,* 320–333.

Plack, C. J., & White, L. J. (2000). Perceived continuity and pitch perception. *Journal of the Acoustical Society of America, 108,* 1162–1169.

Pressnitzer, D., & Hupe, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology, 16,* 1351–1357.

Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology, 18,* 1124–1128.

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 992–1006.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of stream segregation. *Perception and Psychophysics, 53,* 179–189.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18,* 1–13.

Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1232–1244.

Snyder, J. S., Carter, O. L., Lee, S. K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1007–1016.

Sussman, E. S., Horvath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics, 69,* 136–152.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (Unpublished doctoral dissertation). Technische Hogeschool Eindhoven, Eindhoven, Netherlands.

Vliegen, J., Moore, B. C., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America, 106,* 938–945.

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America, 105,* 339–346.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science, 164,* 586–587.

Received February 18, 2010
Revision received July 10, 2010
Accepted September 2, 2010 ■