Journal of Experimental Psychology:
Human Perception and Performance
2004, Vol. 30, No. 4, 643–656

0096-1523/04/$12.00 DOI: 10.1037/0096-1523.30.4.643

Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis

Rhodri Cusack, John Deeks, Genevieve Aikman, and Robert P. Carlyon
Medical Research Council Cognition and Brain Sciences Unit

Often, the sound arriving at the ears is a mixture from many different sources, but only 1 is of interest. To assist with selection, the auditory system structures the incoming input into *streams,* each of which ideally corresponds to a single source. Some authors have argued that this process of streaming is automatic and invariant, but recent evidence suggests it is affected by attention. In Experiments 1 and 2, it is shown that the effect of attention is not a general suppression of streaming on an unattended side of the ascending auditory pathway or in unattended frequency regions. Experiments 3 and 4 investigate the effect on streaming of physical gaps in the sequence and of brief switches in attention away from a sequence. The results demonstrate that after even short gaps or brief switches in attention, streaming is reset. The implications are discussed, and a hierarchical decomposition model is proposed.

The sound arriving at a listener's ears is often a mixture from many different sources, but usually only one will be of interest at any one time. Several different processes might help listeners in the task of selecting the sound from the source(s) in which they are interested. If listeners have a clear idea of the properties of the sound they wish to listen to, then they might be able to selectively attend and devote greater processing resources to sounds with characteristics like the target source. This selection might be based on primitive acoustic properties that are known to be resolved early in auditory processing, such as place of excitation on the cochlea, and/or on higher representations, such as pitch and timbre. However, there are many circumstances in which listeners do not know a priori the exact acoustic properties of the source they wish to attend to or in which there are other distracting sources that have the same set of possible acoustic qualities (e.g., trying to hear a stranger's voice at a cocktail party).

As well as using prior knowledge of the target source, the human auditory system can exploit regularities in natural sounds to perceptually group the incoming sound into a number of *streams,* each of which would ideally correspond to a single source. For example, sound sources often elicit a pitch that changes only slowly over time, and this cue is used by the auditory system to perceptually group sequential sounds together (Bregman, 1990; Bregman & Campbell, 1971; van Noorden, 1975; Vliegen & Oxenham, 1999). Similarly, sequential sounds from the same source tend to contain similar spectrotemporal patterns and evoke similar timbres, and this is also used as a cue to perceptual grouping (Culling & Darwin, 1993; Cusack & Roberts, 1999, 2000; Dannenbring & Bregman, 1976; Singh & Bregman, 1997; Smith, Hausfeld, Power, & Gorta, 1982). Naturally, in this account, time is important: Changes in pitch or timbre that occur rapidly in time are more likely to cause segregation into multiple streams than are those that occur slowly. Many more perceptual grouping rules, each exploiting a different regularity, have been identified (for reviews, see Bregman, 1990; Darwin & Carlyon, 1995).

Selective attention and perceptual grouping differ in character. To be able to select, listeners need knowledge of the characteristics of the target of interest and, indeed, to know which target they are interested in. It is a top-down process, primarily under conscious control of a listener. It is context dependent in that performance is dependent on the degree of experience the listener has in selecting sounds from a particular source and on many other factors. A simple schematic of this arrangement is shown in Figure 1a. Models like this have been proposed either as a complete account of auditory selection (M. R. Jones, 1976) or as one mechanism at a listener's disposal (schema-driven selection; Bregman, 1990). Perceptual grouping, however, might be considered an invariant process, exploiting probabilities of co-occurrences in natural sounds to group acoustic elements together into streams, regardless of the particular context of the listener. Once streams have been formed, they might then be selected or rejected as a whole. A schematic of this is shown in Figure 1b. A popular view, argued by Bregman (1990), is that perceptual grouping processes are automatic and unlearned.

The idea of a dichotomy between automatic, invariant low-level grouping processes and consciously controlled selective attention processes is appealing in its conceptual simplicity. However, there is evidence that the situation in hearing may be more complicated. Carlyon, Cusack, Foxton, and Robertson (2001) showed that the application of attention affects properties previously considered to be automatic grouping processes. Using a stimulus presented by van Noorden (1975), they played repeating sequences of high- and low-frequency tones, as shown in Figure 2. When the sequence is heard as one stream, a characteristic galloping rhythm is heard, but this is not the case when it is heard as two streams. This salient change in rhythm allows listeners to judge whether they hear one stream or two. It is well established that when sequences with a

Rhodri Cusack, John Deeks, Genevieve Aikman, and Robert P. Carlyon, Medical Research Council (MRC) Cognition and Brain Sciences Unit, Cambridge, United Kingdom.

Correspondence concerning this article should be addressed to Rhodri Cusack, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom. E-mail: rhodri.cusack@mrc-cbu.cam.ac.uk

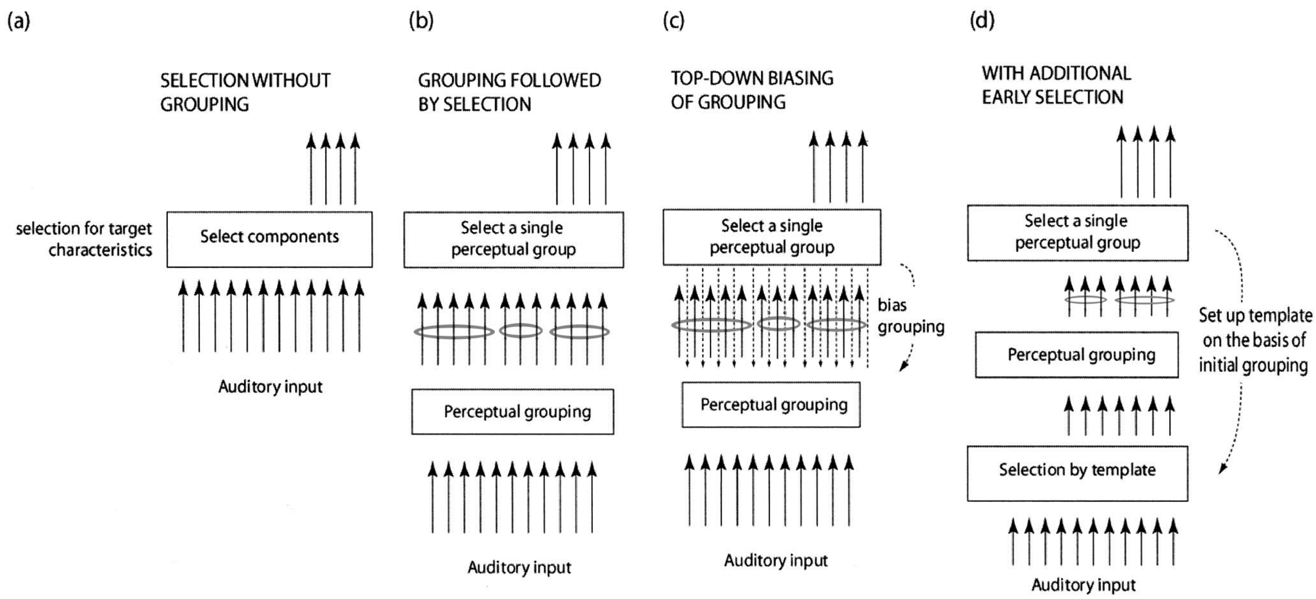

Figure 1. Four models of the selection of relevant sounds from the mixture arriving at the ears. Ellipses around arrows represent perceptual grouping.

moderate frequency separation are played, stream segregation builds up over time so that they are likely to be heard as a single stream near the beginning but more likely to be heard as two streams after a few seconds (Anstis & Saida, 1985). This is often considered to be the result of an accumulation of evidence by the perceptual grouping system that there are two distinct sources of sound present. Using healthy volunteers, Carlyon et al. (2001) played tone sequences that were 20 s long to one ear, during the first 10 s of which a sequence of noise bursts was presented to the other ear. In one condition, listeners were asked to rate the degree of streaming of the tones throughout. As expected, it was found that at intermediate frequency separations, streaming built up over the first 5–10 s. In another condition, listeners were asked to perform a task on the noise bursts for the first 10 s and then switch attention and rate the degree of stream segregation of the tones for the following 10 s. In this critical condition, it was found that listeners' streaming judgments resembled those made at the beginning of the sequences in the first condition, even though the tones had been presented (but ignored) for 10 s. A further experiment on healthy volunteers controlled for possible biasing of responses due to task switching. Two other experiments, with people with brain lesions leading to a bias in spatial attention (unilateral neglect), showed that streaming was reduced on the less attended side. These experiments demonstrated that an aspect of grouping previously thought to be low level and automatic is affected by whether the sounds are being attended to or not. An interactive model of grouping and selective attention is shown schematically in Figure 1c.

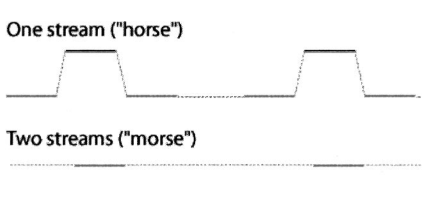

Figure 2. A schematic of the two possible percepts for the repeating triplet tone sequences in Experiment 1. The abscissa represents time, and the ordinate represents frequency. The one-stream percept is labeled *horse* because it is characterized by a galloping rhythm; the two-stream percept is labeled *morse* because it sounds like Morse code.

In the current study, we further investigate the effect of selective attention on perceptual grouping. In the first two experiments, we investigated the domain over which buildup takes place. In the previous experiments of Carlyon et al. (2001), the sounds for the distracting task were presented in a different frequency region and ear from the tone sequences. Here, we have considered three possible domains. It might be that the buildup of stream segregation happens in the ear (or perhaps at the location) that is currently being attended. Alternatively, a fundamental feature of the organization of early auditory processing in humans and animals is that sounds are broken down into frequency channels. Several authors have proposed models of streaming that are based on the separation of tones into such channels (Anstis & Saida, 1985; Beauvois & Meddis, 1996; Hartmann & Johnson, 1991). Although these *place of excitation* cues are not essential for streaming to take place (e.g., Cusack & Roberts, 2000; Vliegen & Oxenham, 1999), there is evidence that streaming is strongest when they are present (Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2000). It might therefore be that the buildup of segregation takes place in the frequency region currently being attended. Finally, it might be that a stream has been formed that contains the sounds relevant to the current task (e.g., the tones), and other sounds (e.g., the noises) are allocated to other streams, and then only the attended streams fragment further. Put another way, if one considers the auditory scene as a hierarchy, then unattended branches of the hierarchy are not elaborated: The higher and lower frequency tones are not segregated unless the tones as a whole are attended to. The first two experiments investigated these hypotheses.

In the third and fourth experiments, we looked at the effect of the temporal dynamics of the stimulus and of the task. Bregman (1978) found that turning a sequence off for 4 s reset streaming. Experiment 3 investigated how auditory stream segregation decays when gaps of various durations are introduced into sequences. Experiment 4 asked whether stream segregation decays when attention is briefly removed from a sequence, even if the sounds are continuous throughout. The results of these experiments allow us to investigate three hypotheses. First, it might be that although attention facilitates the buildup of stream segregation, attention is not required for its maintenance. In this case, if attention is withdrawn for a short time, the percept may continue as if the withdrawal had never happened. Second, switching attention away from and back to a sequence—or perhaps simply turning it off and on again—will instantly reset the percept to that of one stream. Alternatively, withdrawing attention may cause stream segregation to decay at a rate with a time constant similar to its buildup.

Experiment 1

Method

Carlyon et al. (2001) found that when participants ignored a tonal sequence for 10 s and performed a distracting task on sounds to the other ear, subsequent streaming judgments revealed less stream segregation of the previously ignored tones than they did when participants had been attending to them throughout. The aim of the current experiment was to replicate and extend this finding by testing whether buildup is observed when the sounds on which the distracting task is performed are in the same ear.

Eight participants who reported normal hearing were tested in a double-walled sound-attenuating chamber. There were two different types of stimuli, and two different tasks that were performed on each, leading to four conditions in total. In each trial, a sequence of tones of 20 s in duration was presented to a listener's left ear only. This sequence comprised 40 repetitions of a pattern ABA__, where A represents a low-frequency tone, B represents a high-frequency tone, and __ represents a silent interval. The tones were 125 ms in duration, with 20-ms linear attack and decay ramps, and they were presented at 55-dB sound pressure level (SPL). The duration of the silent interval was also 125 ms, thus leading to two isochronous streams if the A and B tones segregated. The frequency of the low-frequency tones was 400 Hz, and that of the high-frequency tones was 4, 6, 8, or 10 semitones higher (504, 566, 635, or 713 Hz, respectively). As well as the tones, a set of random noises was also presented for the first 10 s. The noises were either in the other (right) ear only or also in the left ear, depending on the condition. They were created by digitally filtering white noise using a brick wall bandpass filter between 2–3 kHz (60 dB down in stopbands), which was in a different frequency region from the tones. The noises had a 52-dB SPL, and 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). Ten noises were presented, with their onsets separated by a mean of 1,000 ms but jittered in time to reduce any rhythmic interference. If the onset of the first tone was given by $t = 0$, then the time of the ith noise, in milliseconds, was $t_i = i \times 1{,}000 + \Delta t_i$, where the Δt_is were chosen randomly from a uniform distribution between 0 and 250 ms.

For each of the two different types of trial (tones left and noises right, tones and noises left), listeners followed two different procedures. In the *one-task* conditions, they were asked to ignore the noises, attend to the tones, and perform the stream-segregation judgment throughout. Participants were asked to listen for the galloping rhythm that is characteristic of a single perceptual stream (see Figure 2, top panel) to determine whether they heard one stream (labeled *horse* in the instructions to listeners) or two streams (labeled *morse* because it sounded a little like Morse code). At the start of a sequence, as soon as listeners could determine whether they were hearing one stream or two, they indicated what they heard by pressing one of two buttons on a keyboard. Then, subsequently, they were asked to press one of two buttons every time they heard a change in percept. In the *two-task* conditions, for the first 10 s, participants were required to listen to the noises, determine whether each noise was increasing or decreasing in amplitude (approaching or departing), and indicate their response by pressing one of two keys. For the final 10 s, they were required to do the horse–morse stream-segregation discrimination task on the tones, as described above. Response requirements were communicated by messages displayed on the computer screen and by the labeling of the response buttons (also displayed on the screen).

Results

Figure 3 shows the mean number of streams heard as a function of time for each combination of condition and frequency separation. At times, for trials in which no response had yet been received, the data were not entered into the mean. For the purposes of display, data are not shown where fewer than 25% of responses across all trials and listeners had been received. From the two left-hand panels of Figure 3, it can be clearly seen that when the task was to perform the streaming judgment throughout, stream segregation built up steadily from the start of the sequences. If the buildup of stream segregation were independent of attention, then we would expect the curves in the two right-hand panels in Figure 3 to look like the latter half of the curves on the left, because in both cases the stimuli had been presented for 10 s previously. If, however, attention is important for the buildup of streaming, then the two curves on the right might be expected to look more similar to the first half of the curves on the left: Buildup only begins when the stimuli are attended to. To look at this in more detail, we generated a summary measure of streaming in each of the conditions during specific time windows. Two measures summarized the degree of streaming *early* (1.5–4.5 s) and *late* (11.5–14.5 s) in the one-task sequence. Another summarized the degree of streaming in the two-task condition shortly after the switch to the streaming task (11.5–14.5 s). These summary measures are shown in the left panel of Figure 4.

Consider first the conditions in which the noises were in the opposite ear from the tone sequences. As would be expected (e.g., van Noorden, 1975), there was greater streaming when there was a larger frequency separation between the high and low tones. When there was one task and the 20-s sequences were fully attended throughout, there was substantially more stream segregation later than earlier in the sequences. However, when participants started responding later on in the two-task condition, the buildup of streaming was much less than when the sequences had been attended to throughout. A very similar pattern of results was seen when the noises were in the same ear.

These findings were confirmed statistically using a repeated measures analysis of variance (ANOVA) with a Huynh–Feldt correction for sphericity. The mean response of each listener was divided into 1-s time bins. This size was used to provide adequate temporal resolution while keeping the analysis size manageable. The first two bins were discarded to allow listeners time to make

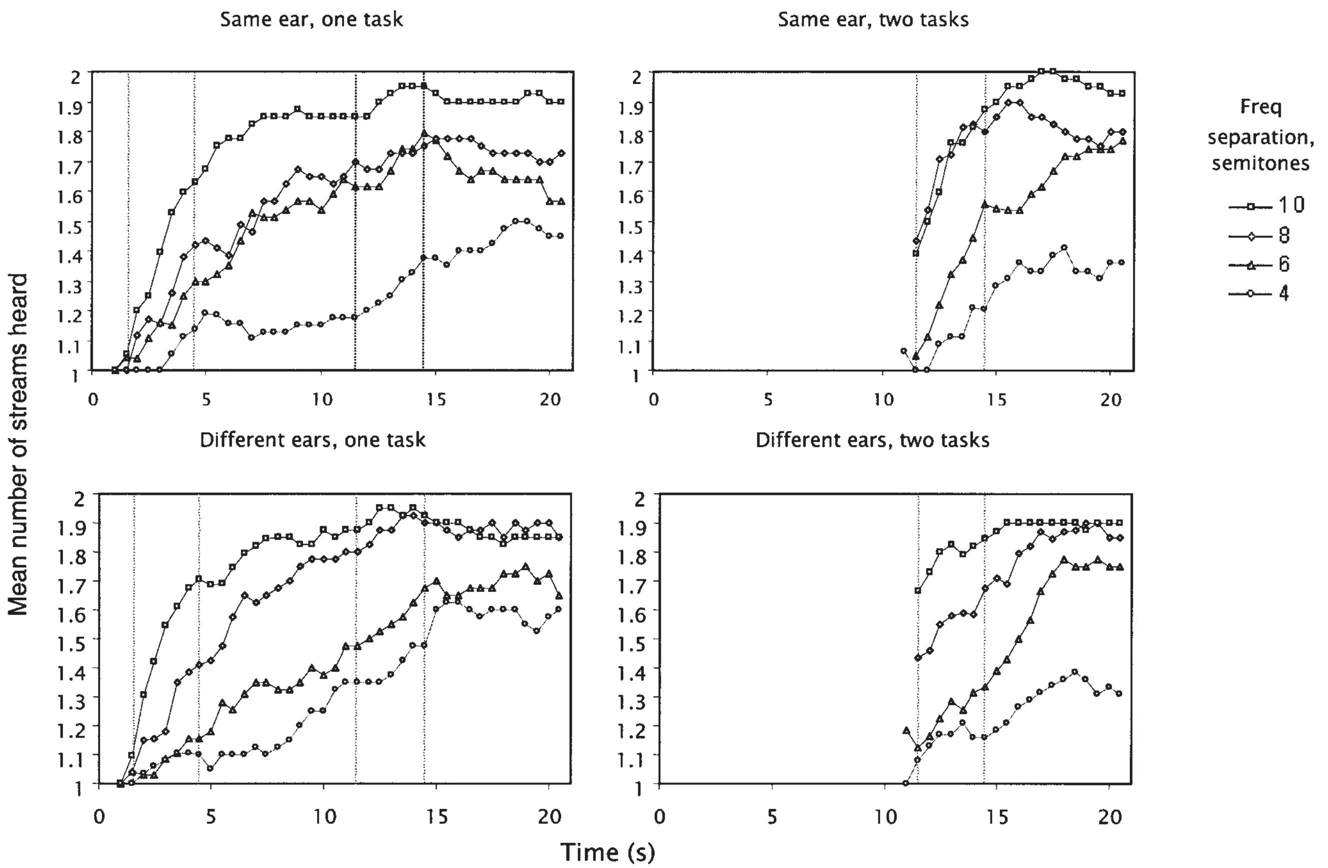

Figure 3. Number of perceptual streams reported as a function of time in Experiment 1, as measured by the mean of all responses received across listeners at a set of time points 0.5 s apart. The vertical dotted lines mark off early (1.5–4.5 s) and late (11.5–14.5 s) time periods. Freq = frequency.

an initial response, leaving eight time points across each 10-s period (2–3 s, 3–4 s, . . . 9–10 s). By 2 s, listeners had responded in 67% of trials, and by 3 s, they had responded in 84% of trials. Averages were taken of all of the data points that were available at each time. Three conditions were compared: early in the one-task sequences (2–10 s into the trial), late in the one-task sequences (12–20 s into the trial), and late in the two-task sequences (12–20 s into the trial). Also entered into the ANOVA was binaural stimulus configuration (noises and tones in the same or in different ears) and frequency separation (4, 6, 8, or 10 semitones). As expected, there were effects of frequency, $F(3, 15) = 63.9$, $p < .0001$, and time, $F(7, 35) = 44.9$, $p < .0001$. There was no effect of ear, $F(1, 5) = 0.671$, but there was an effect of condition, $F(2, 10) = 21.1$, $p < .001$, and a Condition × Time interaction, $F(14, 70) = 4.90$, $p < .05$. These main findings were qualified by further interactions—Frequency × Condition, $F(6, 30) = 3.39$, $p < .02$; Condition × Time, $F(14, 70) = 4.90$, $p < .05$—but all other interactions failed to reach significance—Frequency × Configuration, $F(3, 15) = 0.29$; Frequency × Time, $F(21, 105) = 1.77$; Configuration × Time, $F(7, 35) = 0.42$; Frequency × Condition × Configuration, $F(2, 10) = 1.82$; Frequency × Configuration × Time, $F(21, 105) = 1.66$; Condition × Configuration × Time, $F(14, 70) = 1.68$; Frequency × Condition × Time × Configuration, $F(42, 210) = 0.93$.

To investigate whether there was any reduction in the amount of stream segregation in the absence of attention, we entered the two-task (late) and one-task (late) conditions into a separate ANOVA. This showed both a main effect of condition on the amount of segregation, $F(1, 6) = 9.83$, $p < .05$, and a Condition × Time interaction, $F(7, 42) = 5.88$, $p < .05$, reflecting different patterns of buildup. To investigate whether there was any buildup in the two-task condition, we entered the two-task (late) and one-task (early) conditions into a separate ANOVA. There was some effect of buildup apparent in the main effect of condition, $F(1, 5) = 14.6$, $p < .05$, although there was no Condition × Time interaction, $F(7, 35) = 1.53$.

The results of the approach–depart task performed on the noises are shown in the upper panel of Figure 5. Performance was good in all listeners.

Discussion

When attention is focused on a different set of sounds, there is substantially less buildup in auditory streaming. This was the case whether the alternative sounds were in the same ear or in a different ear, replicating and extending the results of Carlyon et al. (2001). These results show that buildup can be inhibited even when the sounds are in the same ear as that being attended,

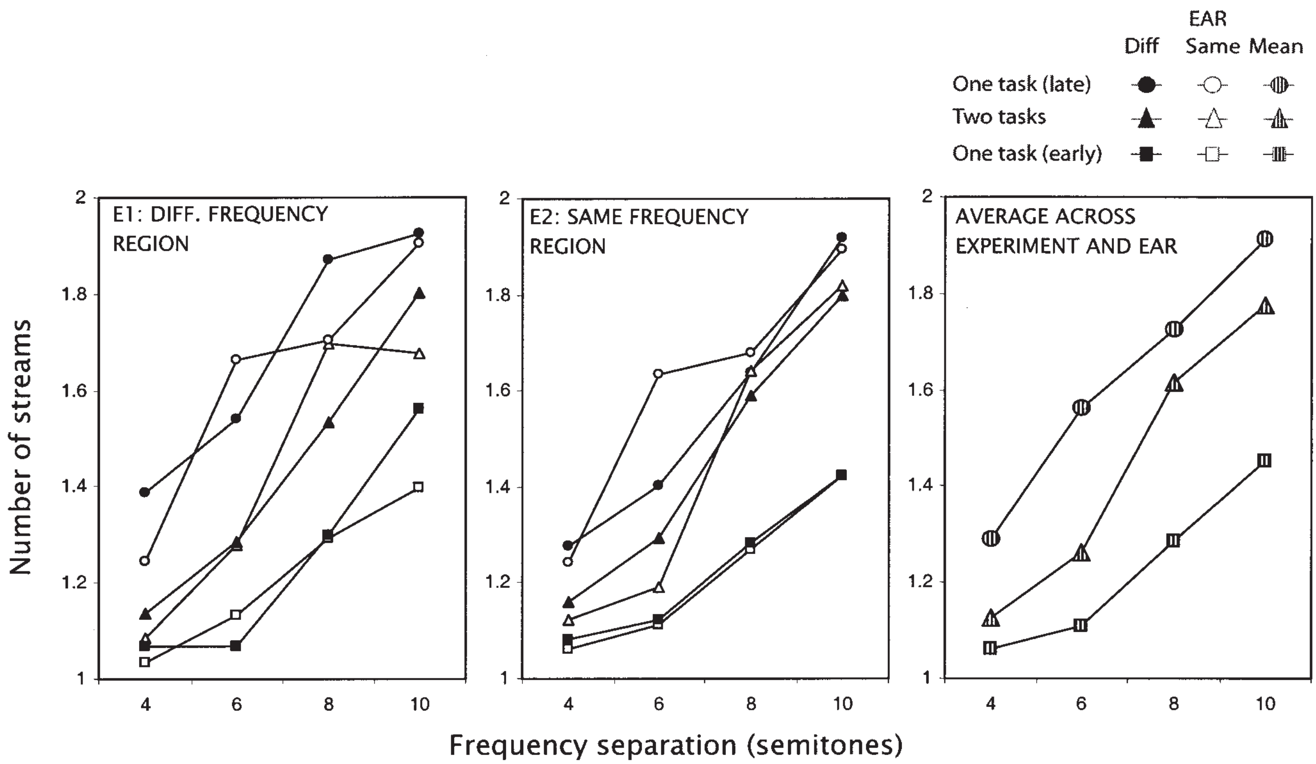

Figure 4. A summary of performance in Experiment (E) 1, in Experiment 2, and averaged across both experiments and ear as a function of frequency separation. Diff = different.

showing that inhibition is not due to a general suppression of one side of the ascending auditory pathway, of ear-specific parts of auditory cortex, or of all sounds at a particular location.

Although there was no effect of the ear that the alternate task was in, overall we did see a small but significant degree of segregation in the tones, even just after a period in which the alternate noise sounds had been attended. This may indicate either that listeners were imperfectly focusing their attention on the noise sounds and did allocate some to the tones during the competing task or that some streaming can occur in the absence of attention. Finally, it is worth noting that the one-stream judgments observed when listeners switched their attention to the tones were not simply due to a response bias whereby listeners defaulted to reporting a single stream whenever they started making streaming judgments about a sequence (Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). To control for this, Carlyon et al. (2001) played listeners a 20-s ABA__ sequence of amplitude-modulated tones in which the rate of modulation switched from a slow to a fast rate every few seconds. When listeners judged the rate of modulation for 10 s and then switched to making streaming judgments, it was observed that the streaming had built up. Carlyon et al. (2001) concluded that the buildup of streaming is unaffected by the response requirements of the task, provided that listeners attend to some feature of the to-be-streamed tones throughout. Further evidence against a response-bias interpretation is provided in the present Experiment 4 and other recent results (Carlyon, Plack, Fantini, & Cusack, 2003).

The Frequency × Condition interaction probably reflects the greater buildup for large frequency separations in the two-task

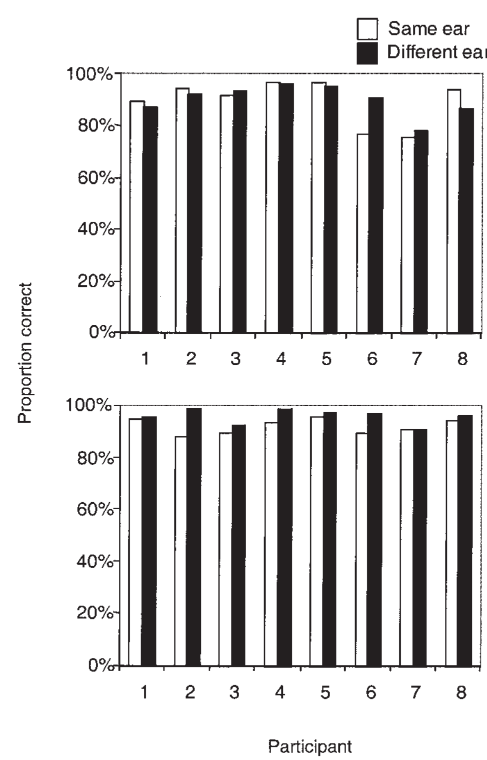

Figure 5. Performance in the approach–depart noise task in Experiment 1 (upper panel) and Experiment 2 (lower panel).

condition. The frequency difference between the noises and tones was substantial (well above an octave in all conditions), so it seems unlikely that the frequency separation was affecting the degree of selective attention. Another possibility, considered further in the General Discussion, is that at larger frequency separations, there was some automatic stream segregation.

Experiment 2

Method

The aim of Experiment 2 was to test the hypothesis that the effect of attention on the buildup of stream segregation is dependent on listeners attending to a different frequency region from the tones. To test this, we used the same structure and procedure as in Experiment 1 but modified the stimuli so that the bandwidth of the noises encompassed the same frequency region as the tones. Again, we manipulated whether the noises were in the same ear as or different ear from the tones. Eight participants who reported normal hearing were tested. All aspects of the experiment were identical to Experiment 1, except the noise sounds on which the distracting task was performed were filtered into a range that overlapped with the tones (300–900 Hz).

Results

The number of streams heard is plotted as a function of time in Figure 6. As in Experiment 1, the buildup of streaming was retarded when listeners' attention was directed to the distracting task. A summary measure was calculated by averaging over 3-s periods in the same way as in Experiment 1. These results are shown in the center panel of Figure 4. The pattern of results is generally similar to that shown in the left-hand panel (see Experiment 1), and the same general conclusions apply. Again, there was greater stream segregation when there was a larger frequency separation between the lower and higher frequency tones. When a single task had to be performed, there was greater stream segregation later rather than earlier in the sequence.

Although there was some buildup of streaming when the sequence was presented but not attended to (triangles are above squares) there was less buildup than when the tone sequence was attended (triangles are below circles). This pattern was confirmed with a repeated measures ANOVA with the Huynh–Feldt correction identical to that used in Experiment 1. Again, there were main effects of frequency, $F(3, 18) = 98.5$, $p < .001$, and time, $F(7, 42) = 14.7$, $p < .001$. There was no effect of binaural configuration, $F(1, 6) = 0.031$, but once more we observed a main effect of condition, $F(1, 6) = 19.1$, $p < .001$, and a Condition $\times$ Time interaction, $F(14, 84) = 3.14$, $p < .05$. There were a few other significant interactions—Frequency $\times$ Condition, $F(6, 36) = 4.34$, $p < .005$; Frequency $\times$ Ears, $F(2, 18) = 3.52$, $p < .05$; Frequency $\times$ Time, $F(21, 126) = 2.44$, $p < .05$; Frequency $\times$

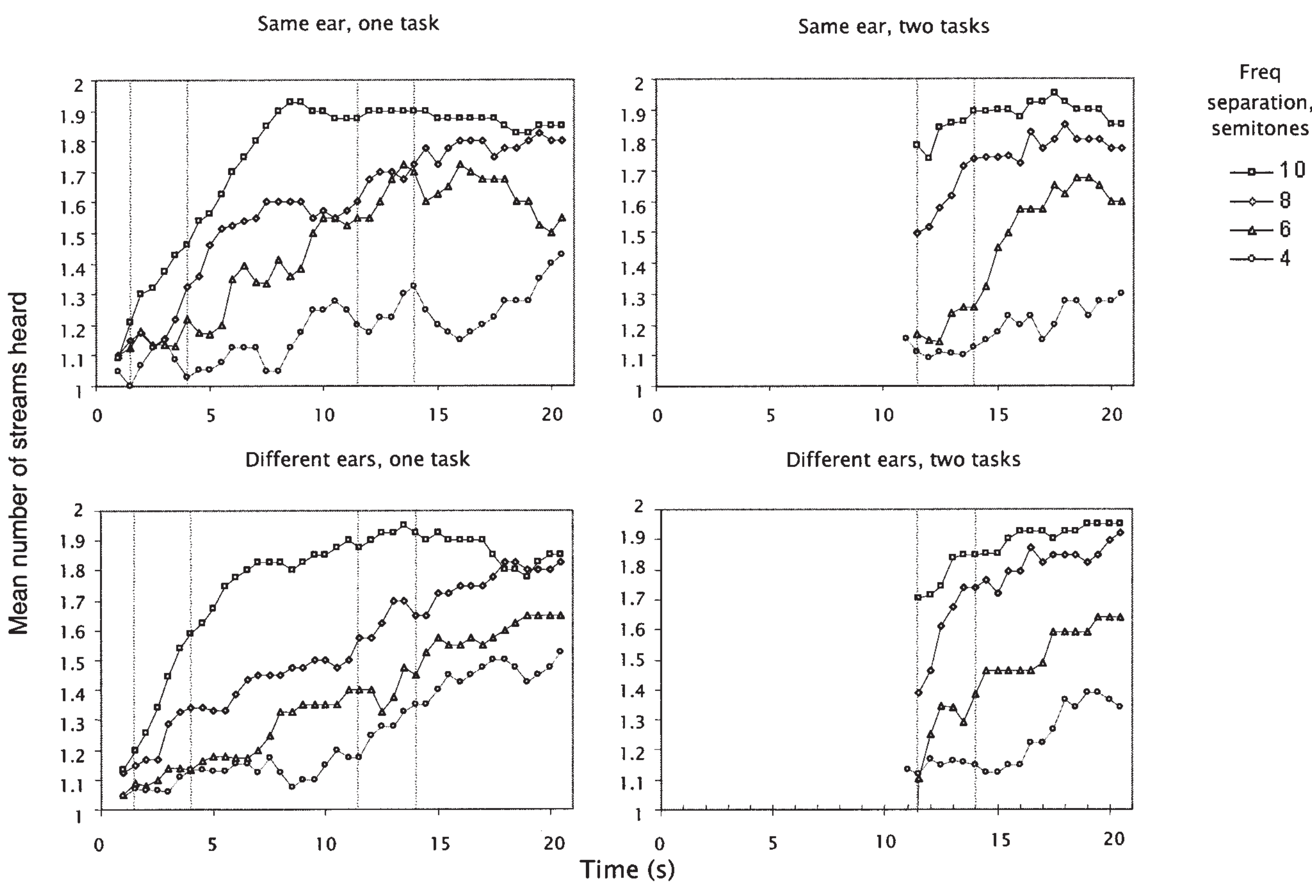

Figure 6. Number of perceptual streams reported as a function of time in Experiment 2, as measured by the mean of all responses received across listeners at a set of time points 0.5 s apart. The vertical dotted lines mark off early (1.5–4.5 s) and late (11.5–14.5 s) time periods. Freq = frequency.

Condition × Time, $F(42, 252) = 2.84$, $p < .004$; Condition × Configuration × Time, $F(14, 84) = 2.94$, $p < .02$—but all others failed to reach significance—Condition × Configuration, $F(2, 12) = 1.17$; Frequency × Condition × Configuration, $F(6, 36) = 1.11$; Configuration × Time, $F(7, 42) = 0.32$; Frequency × Configuration × Time, $F(21, 126) = 1.08$; Condition × Time × Frequency × Ears, $F(42, 152) = 1.33$.

To investigate whether there was any effect of attention on the buildup of streaming, we entered just the one-task (late) and two-task conditions into a separate ANOVA. There was not a main effect of condition, $F(1, 7) = 1.6$, but there was a significant Condition × Time interaction, $F(7, 49) = 3.42$, $p < .05$, reflecting differing patterns of buildup. This effect of condition was not qualified by any other interactions. Finally, as in Experiment 1, to investigate whether there was any buildup in the two-task condition, we entered the two-task (late) and one-task (early) conditions into a separate ANOVA. There was some effect of buildup apparent in the main effect of condition, $F(1, 6) = 40.2$, $p < .001$, although there was no Condition × Time interaction, $F(7, 42) = 0.23$.

To directly test for differences between the two experiments, we performed a combined ANOVA on the data from both. As would be expected, the main effects remained as found in the individual experiments. There was only a single term containing an effect of experiment that was significant: an Experiment × Frequency Separation × Configuration × Time interaction, $F(21, 105) = 2.03$, $p < .05$.

The results of the approach–depart task performed on the noises are shown in the lower panel of Figure 5: Performance was good in all participants. Figure 7 shows a summary of the degree of streaming collapsed across Experiments 1 and 2, frequency, time, and binaural configuration for each of the three measures.

Discussion

The buildup of stream segregation was retarded when attention was directed to a different set of sounds in the same frequency region, even if the sounds were in the same ear. The reduction in buildup of streaming was not significantly different when the sounds were in the same ear than when they were in a different ear. In the condition in which the noises and tones were in the same ear and frequency region, there would have been substantial overlap in the auditory filters that they excited. Some authors have argued that streaming is primarily determined by the segregation of sounds into different channels at an early stage in the auditory system: the *peripheral channeling* hypothesis (Anstis & Saida, 1985; Hartmann & Johnson, 1991). However, other authors have since argued that strong segregation can exist without differences in peripheral channeling (Cusack & Roberts, 2000; Vliegen & Oxenham, 1999). The results here extend this demonstration to show that the effect of selective attention on stream segregation can also be seen even when there is strong overlap between the peripheral channels that are attended and the sounds on which streaming is measured.

We should add a caveat to these conclusions. Although the noises overlapped the frequency region of the tones, because they had a wider bandwidth, there will have been auditory filters that were excited primarily by the noises alone. It would therefore be possible that when listeners were performing a task on the noises, they were attending to these channels and not to those containing the tones. What we can conclude is that attention to a competing sound can affect subsequent judgments of auditory streaming of a tone sequence even when that sound excites all frequency channels stimulated by the tones.

As in Experiment 1, the Frequency × Condition interaction might reflect some automatic stream segregation for wider frequency separations. In this experiment, another possibility is that the overlap in frequency range of the tones and noises may have altered the ease with which the tones could be ignored and the noises attended to. However, the combined analysis did not show an Experiment × Frequency × Condition interaction, suggesting that the cause in the two experiments was similar. Furthermore, there was no effect of binaural configuration in either experiment. If the listeners were having difficulty ignoring the tones when the noises were in the same ear, putting them in different ears would be expected to make this easier. There was no effect of binaural configuration in either experiment, suggesting that selective attention was easy in all conditions. This interaction is considered further in the General Discussion.

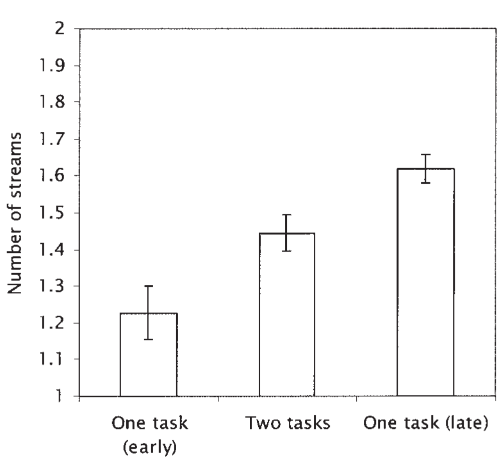

Figure 7. Number of streams perceived collapsed across Experiments 1 and 2, frequency, time, and binaural configuration. Error bars represent plus or minus one standard error.

Experiment 3

Method

The aim of Experiment 3 was to investigate the effect of gaps in a sequence on stream segregation. One possible hypothesis is that streaming might decay with a similar time constant to its buildup: This would be predicted from a *leaky integrator* model of streaming, such as that of Beauvois and Meddis (1996). There is some evidence for such a model. Beauvois and Meddis (1997) used induction sequences, which biased the perceptual grouping of subsequent test sequences of tones. They manipulated the silent interval between the induction and test sequence and found reduced bias with increasing duration. Alternatively, some studies have suggested that streaming can be reset by abrupt changes. Bregman (1978) showed that a 4-s gap would reset the perceptual grouping of a sequence to one stream.[1] In Experiment 3, we presented a set of 10-s sequences separated by gaps of either 1, 2, 5, or 10 s. If the leaky integrator hypothesis is correct, we might expect to see only a small decay in the amount of streaming after a 1-s gap but a larger amount after a longer gap. The resetting hypothesis, however, predicts that the percept is reset to one stream after even a fairly short gap.

Thirteen participants who reported normal hearing were tested. One participant was rejected because he did not respond for several seconds in the majority of trials (only 15% by 2.5 s, compared with an average of 78% for the other participants). Ten-second sequences of sounds in an ABA__ pattern, similar to that used in Experiments 1 and 2, were presented, and the same task was used to measure stream segregation. At the start of a block, a 10-s sequence was presented. There was then a gap, then another sequence, and a gap, another sequence, and so on. In each block, there were 8 gaps each of 1, 2, 5, and 10 s, presented in random order, making a total of 32 gaps and, hence, thirty-three 10-s sequences. The rating results from the first 10-s sequence were discarded, and the results of the remainder were classified according to the length of gap that preceded each sequence.

Within each block, the frequencies of the A and B tones were fixed. To each participant, four blocks were presented—two with a frequency separation of 4 semitones (A: 400 Hz; B: 504 Hz) and two with a frequency separation of 6 semitones (A: 400 Hz; B: 566 Hz). These were presented in counterbalanced order.

Results

Figure 8 shows the degree of streaming as a function of the time since the onset of each 10-s sequence. The parameter is the duration of the silent gap preceding that sequence. The responses were summarized by calculating the mean number of streams at times 0.5 s apart throughout the 10-s sequences. All listeners had responded at least once by 2 s, and so the 17 time points from 2–10 s inclusive were used for a balanced repeated measures ANOVA with the Huyhn–Feldt sphericity correction.

As would be expected, stream segregation built up over time, $F(16, 176) = 15.0$, $p < .0005$, and there was more streaming at wider frequency separations, $F(1, 11) = 8.41$, $p < .02$. Although there was a possible trend toward main effects of gap length, $F(3, 33) = 2.98$, $p = .07$, and linear contrast, $F(1, 11) = 3.86$, $p = .08$, there was no Gap Length × Time interaction, $F(48, 528) = 0.58$, or any other interaction—Frequency × Gap Length, $F(3, 33) = 2.04$; Frequency × Time, $F(16, 176) = 2.09$; Frequency × Gap Length × Time, $F(48, 528) = 0.858$.

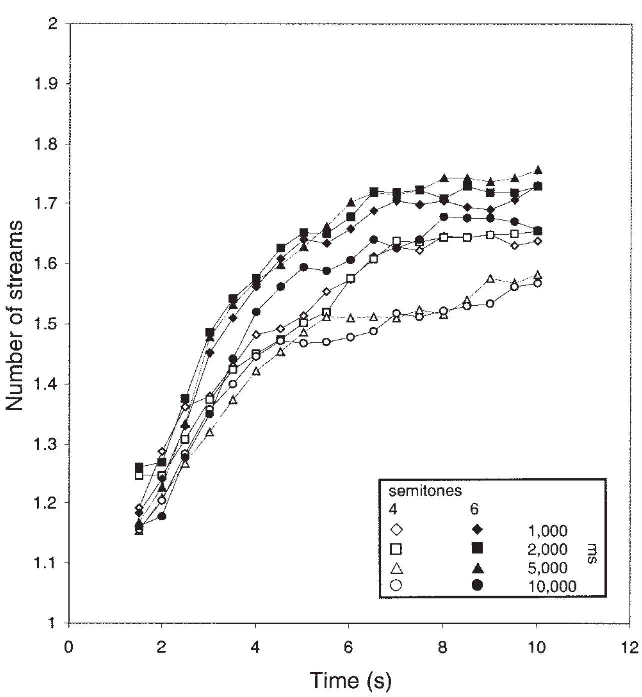

Figure 8. Number of streams as a function of time for two different frequency separations and after gaps of four different lengths since the end of the last sequence in Experiment 3.

Discussion

Even following a short 1-s gap, the percept was reset and, predominantly, one stream was heard. There was almost no trace of whether the tones had stopped just 1 s or 2, 5, or 10 s before. These results favor a resetting explanation and are not consistent with stream segregation decaying with a similar time constant to its buildup, which takes around 10 s to reach asymptote. This resetting contrasts with the findings of Beauvois and Meddis (1997), who found a gradual decay in bias due to an induction sequence. The difference between our results and theirs suggests that mechanisms in addition to those involved in streaming may be involved in the biasing effect of induction sequences.

A description that has been put forward for the relatively slow buildup of stream segregation is one of a buildup of evidence that there are two sound sources present (Bregman, 1990). If this is indeed what is happening, then the results of this experiment would force us to conclude that the evidence is thrown away when there is even a short gap in the sequence. Possible reasons for this are given in the General Discussion.

Experiment 4

Method

In Experiment 3, we found that a short gap in a sequence resets the percept to that of one stream. In Experiment 4, we investigated whether withdrawing attention from a sequence for a short time affects the percept when attention is returned to it. Three hypotheses were considered. It might be that although attention is important for the buildup of stream segregation, it is not needed for stream segregation's maintenance. In this case, if attention is withdrawn for a short time, the percept may continue as if the withdrawal had never happened. A second possibility is that in the absence of attention, stream segregation will decay at a rate with a time constant similar to its buildup. A third possibility is that, as with a true gap in the sequence (Experiment 3), the withdrawal of attention briefly will reset the percept to that of one stream.

[1] Rogers and Bregman (1998) concluded that a similar effect can be obtained by an abrupt change in loudness or location between an *induction sequence*, whose purpose was to produce a buildup of streaming, and a test sequence. They also concluded that gradual changes have much smaller effects. However, in their experiments, the change in location was confounded by differences in the average location of the inducing sound. For example, in their *sudden change* condition, the inducing sound was in the opposite side of the head throughout, whereas when the sound changed gradually, it was on average more central. Similar arguments apply to their experiments on loudness change.

There were three conditions, which are illustrated schematically in Figure 9. Visually presented instructions indicated the task and response possibilities at each moment. In all conditions, to the left ear we presented sequences of tones similar to those in Experiments 1–3, and to the right ear we presented a set of noises. The noises were present continuously throughout all conditions. In the *gap* condition, the tone sequences were 10 s long and separated by 5-s gaps. In the *continuous* and *switch* conditions, the tones were present throughout, although the task was only performed for 10 s, followed by a 5-s interval in which no streaming judgments were made. In the switch condition, listeners were asked to perform a task on the noises in the other ear for the 5-s intervals in which they were not performing the streaming task. In the continuous condition, listeners ignored the noise and were instructed to continue listening to the tones, even though the response buttons on the computer monitor were dimmed (and disabled) during these 5-s periods.

From Experiment 3, we might have expected streaming to be reset by the introduction of the 5-s gaps in the gap condition. In the continuous condition, we expected listeners' attention to remain on the tone sequences throughout, and so streaming might have been expected to reach a steady state and stay there. The critical condition was the switch condition. The diversion of attention away for 5 s might have no effect if attention is not necessary for the maintenance of segregation, a small effect if segregation decays slowly while attention is elsewhere, or a large effect if a switch of attention resets streaming.

As in Experiments 1–3, the tones were in an ABA__ pattern, with the same timings. Within a sequence, the frequencies of the A and B tones were fixed. Across sequences, two frequency separations of four semitones (A: 400 Hz; B: 504 Hz) and six semitones (A: 400 Hz; B: 566 Hz) were presented in random order. The noises were an average of 1 s apart but jittered by an amount chosen randomly, with uniform distribution from between 0 and 250 ms, to prevent rhythmic interference with the tone sequence. The stimuli (ramped noises filtered into the range of 2–3 kHz) and task (approach vs. depart) were identical to those in Experiment 1. The experiment was divided into blocks, in which there were nine 10-s periods in which the stream-segregation rating task was performed alternating with eight 5-s condition-dependent periods. Each listener performed 4 blocks of each condition, making 12 blocks in total. The conditions were presented in random order, with the instructions given prior to each set of 4 blocks. The streaming task was performed in the same way as for Experiments 1–3.

Results

Figure 10 shows the mean number of streams heard as a function of time. As would be expected, when the tone sequences were presented continuously, there was substantial segregation throughout and little change as a function of time, because each sequence was preceded by a period in which streaming had already built up (top panel). As in Experiment 3, when the 10-s sequences were preceded by a 5-s gap, streaming was reset, with the percept nearly always being heard as one stream and later being heard as two streams much more often (middle panel). The critical condition, in which the tones were continuous, but attention was shifted away for 5 s beforehand, is shown in the bottom panel. It can be clearly

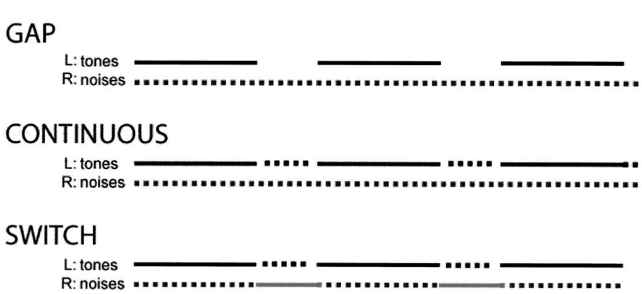

Figure 9. A schematic of the stimuli and task structure in the gap, continuous, and switch conditions of Experiment 4. The upper line in each pair shows when the tone sequences were presented and when a task was (solid line) and was not (dotted line) performed on them. The lower line in each pair shows similar information for the noises. L = left; R = right.

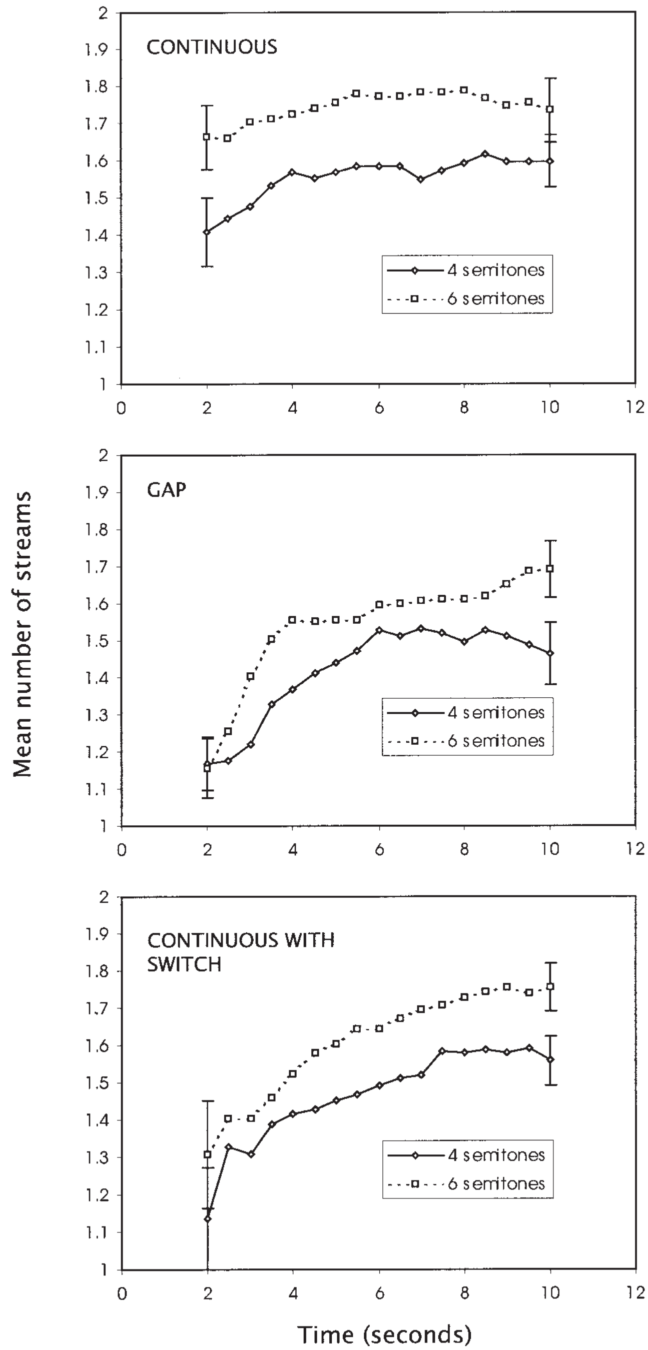

Figure 10. Mean number of streams heard as a function of time in the continuous, gap, and continuous with switch conditions of Experiment 4. Error bars represent plus or minus one standard error.

seen that the brief removal of attention reset the predominant initial percept to one stream to a similar extent as did a physical gap.

As in Experiment 3, snapshots of the responses were taken at 0.5-s intervals and entered into a repeated measures ANOVA. Again, 17 time points between 2 and 10 inclusive were used and the Huynh–Feldt correction applied. There were main effects of frequency, $F(1, 5) = 22.4$, $p < .005$, and time, $F(16, 80) = 10.9$, $p < .001$. There was no main effect of condition, but there was a substantial Condition × Time interaction, $F(32, 160) = 3.18$, $p < .005$. The other two-way interactions were not significant—Condition × Frequency, $F(2, 10) = 0.19$; Frequency × Time, $F(16, 80) = 0.72$—although there was a Condition × Frequency × Time interaction, $F(32, 160) = 2.75$, $p < .003$.

To assess the size of the critical Time × Condition interaction between pairs of conditions, each pair was entered into a separate ANOVA. As would be expected from visual inspection of the data, there was a strong significant difference between the continuous condition and each of the other two conditions—Time × Condition interactions in the ANOVA comprising just continuous and gap conditions, $F(16, 80) = 4.94$, $p < .001$, and in the ANOVA comprising just continuous and switch conditions, $F(16, 80) = 8.47$, $p < .0005$—but no significant difference between the switch and gap conditions, $F(16, 80) = 0.885$. Performance on the approach–depart task in the switch condition is shown in Figure 11: In all listeners at both frequency separations, it was good.

Discussion

A brief switch in attention had an effect on perceptual grouping that was similar to the stimuli being switched off. This suggests either that the maintenance of a segregated percept requires attention or that the act of switching back to the tones resets their perceptual grouping. It should be noted that although we did not specifically manipulate attention during the 5-s periods of the continuous condition, the results suggest that listeners did indeed continue to pay attention to the tones throughout, because the pattern was very different from that in the switch condition. Another interesting aspect of the results of the continuous condition is that listeners did not appear to default to making one-stream responses whenever they started making streaming judgments about a sequence. Rather, it was necessary for attention to be diverted away from the sequence (as in the switch condition) for streaming to be reset. This result provides further evidence that the effect of attention on streaming is not a result of listeners making one-stream responses whenever they start to evaluate the degree of streaming within a sequence (Carlyon et al., 2001, in press; Macken et al., 2003).

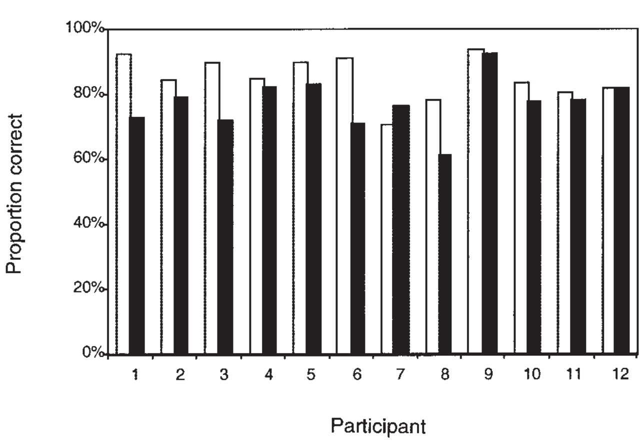

Figure 11. Performance in the approach–depart task in the switch condition of Experiment 4 for tones with four-semitone (white bars) and six-semitone (black bars) separations.

General Discussion

The experiments described here confirm and extend Carlyon et al.'s (2001) earlier finding that attention can play a crucial role in auditory stream segregation. In the next two subsections, we briefly summarize two lines of evidence that lead to a different conclusion and suggest ways in which the apparent conflict between the different approaches may be resolved.

Comparison With Previous Research: The Irrelevant Sound Effect

D. M. Jones and his colleagues (D. M. Jones, Alford, Bridges, Tremblay, & Macken, 1999; D. M. Jones & Macken, 1995; Macken et al., 2003) have used auditory streaming paradigms in their studies of working memory, and they have concluded that streaming occurs even when attention is directed elsewhere. Their experiments focused on the *irrelevant sound effect* (ISE), in which unattended auditory stimuli interfere with the serial recall of visually presented items. Repetition of a single sound appears not to disrupt performance, and these authors have argued that "changes in state" from one sound to the next are crucial for the effect—so that, for example, two alternating tones produce more disruption than a single repeated tone. The paradigm is relevant to the current debate because D. M. Jones and colleagues have shown that when a sequence splits into two streams such that each stream consists of a repetition of a single sound, disruption is also reduced in the case in which listeners hear a single stream of alternating tones (D. M. Jones et al., 1999; D. M. Jones & Macken, 1995; Macken et al., 2003). Because listeners were instructed to ignore the tones, the authors argued that the streaming must have occurred in the absence of attention.

The results presented here may be reconciled with the ISE literature in one of two ways. First, what that ISE literature shows is that some streaming can take place without complete attention being paid to the tones. However, our research reveals that streaming is reduced when a demanding competing task is applied. In some of our experiments, even following such a task, however, streaming was still somewhat greater than when the sequence had just started (Experiments 1 and 2). It is likely that the effect of a competing task on the buildup of streaming is not an all-or-none effect but varies according to the demands of the task. Comparison of the results of Experiments 1, 2, and 4 provides some support for this. In Experiment 4, the distracting task was just performed in 5-s bursts, and the additional demands of switching to this task may well have required, on average, a greater use of resources than was required during the 10-s task periods in Experiments 1 and 2. As would be expected if this difference in competing task demands affected streaming, in Experiment 4 the inhibition of buildup was complete, whereas in Experiments 1 and 2 there was some buildup. To examine this further, future experiments might specifically investigate the effect of manipulating attentional load.

Second, an alternative explanation is suggested by the results of Experiment 4, indicating that streaming may be reset when listeners are briefly given a competing task to perform. What we do not know is whether this resetting occurred as soon as attention was diverted from the sequence, which was therefore internally represented as one stream during the unattended interval, or when attention was switched back to the sequence. In the latter case, the two-stream representation would have persisted even when listeners were not attending to the sequence. It may be that when attention is drawn to a sequence, either because listeners have been so instructed or simply because it has just been turned on, it is represented in some neural structure as a single stream. Subsequently, whether or not listeners continue to attend to the sequence, this resetting effect wears off, and the resulting stream segregation can influence performance on the serial recall task. Note that this explanation implicitly assumes that the default organization is for streaming to occur and that the buildup commonly observed (Anstis & Saida, 1985; Bregman, 1978; Carlyon et al., 2001) might be more properly thought of as a recovery from an active attentional influence.

Comparison With Previous Research: Mismatch Negativity

The above two explanations are also relevant to studies of streaming using the *mismatch negativity* (MMN), most notably by Sussman, Ritter, and Vaughan (1998, 1999). The MMN is an electroencephalogram response that can be evoked to deviant stimuli among a series of standards even when listeners are instructed to ignore the sounds. Sussman et al.'s (1998, 1999) aim was to create stimuli such that particular sounds would stand out as deviants only when the sequences had segregated into separate streams. An MMN response was observed under these conditions even when listeners were not required to focus their attention. As with the ISE literature, the competing task was in the visual modality, but it was even less demanding, with listeners being asked simply to ignore the tones and read a book. Hence, listeners may have been allocating some portion of attention to the tones. It is also possible that, as described above, the effect of attention is primarily to cause a sequence to be heard as a single stream and that some segregation persists even when attention is subsequently diverted elsewhere. Finally, it is worth noting that Sussman, Winkler, Huotilainen, Ritter, and Naatanen (2002) have recently shown that attention can modify the MMN response, suggesting that, contrary to previous belief, the MMN response does not entirely reflect invariant preattentive processes.

One further point, specific to the use of physiological measures of streaming, is worth making. With the use of MMN or a similar tool to investigate the processing of sounds in the absence of attention, it is important to consider what the default organization is. Sounds are spread out over time and broken down by frequency region in the cochlea, and so in early auditory processing stages, the representation is a completely fragmented one. It would not necessarily be a sign of stream segregation if some neural marker showed that information in a particular frequency region was being processed independently of what was happening in another frequency region. Conversely, there are higher order stages in auditory processing that combine information across many auditory channels, and so, again, if some other neural marker indicated that across-channel information was being processed, this would not necessarily imply that perceptual grouping had occurred. For good evidence of grouping, one would have to show that the same neural marker reflected one stream for one stimulus or task condition but two streams for another stimulus or task condition.

Interactions Between Streaming and Attention: The Hierarchical Decomposition Model

The first two experiments investigated whether the buildup of stream segregation happens around the currently attended location, the currently attended frequency region, or only to the currently attended stream. It was found that there was similar buildup whether the tone sequence and the distracting noises were in the same or in different ears, that there was little effect of whether they were in the same frequency region, and there was no interaction between these factors. A parsimonious explanation for these results is that there is some automatic segregation, and then the further buildup of streaming is prevented outside of the stream that is the current focus of selective attention. When a task is being performed on noise bursts, these form the focus of attention, and buildup of segregation is prevented on the tones.

Preventing the further perceptual fragmentation of unattended parts of the auditory scene may provide some ecological advantage. If, for example, a listener is attending to a speaker against a background of music and some traffic noise, then it may not be advantageous for the auditory system to segregate the two backing singers or the sounds of different car engines (see Figure 12a). Instead, the results of our experiments suggest a hierarchical decomposition model (see Figure 12b). A hierarchy of perceptual grouping has been proposed before (e.g., Bregman, 1990; Darwin & Carlyon, 1995)—what we are adding here is that unattended branches are not elaborated: When one is attending to speech, one is not fragmenting the music. One straightforward advantage of not doing so is simply to preserve processing capacity. However, even if perceptual grouping (as an early process) is not capacity limited, it is known that later stages of processing are. It would be extremely confusing if on a switch of attention from a voice, the immediately available auditory objects were small components of the other sounds (e.g., the fret noise from the guitar) rather than the larger scale object of *the music.* The hierarchy in Experiments 1, 2, and 4 is shown in Figure 12c. The hierarchical decomposition model contrasts with a more general interpretation of Bregman (1990) and Rogers and Bregman (1998), who argued that unattended streams were well formed and that you do not "wrap up all your garbage in the same bundle" (Bregman, 1990, p. 193). However, this model is in line with the conclusions of Brochard, Drake, Botte, and McAdams (1999), who found that when listeners were presented with many subsequences but asked to attend to only one, the number of unattended subsequences had no effect on performance. They interpreted this as evidence that the unattended sequences were not segregated into separate perceptual groups. To reconcile these results, it seems sensible and intuitive to conclude that there is some organization in unattended streams—for example, the music, speech, and traffic in Figure 12a may be segregated

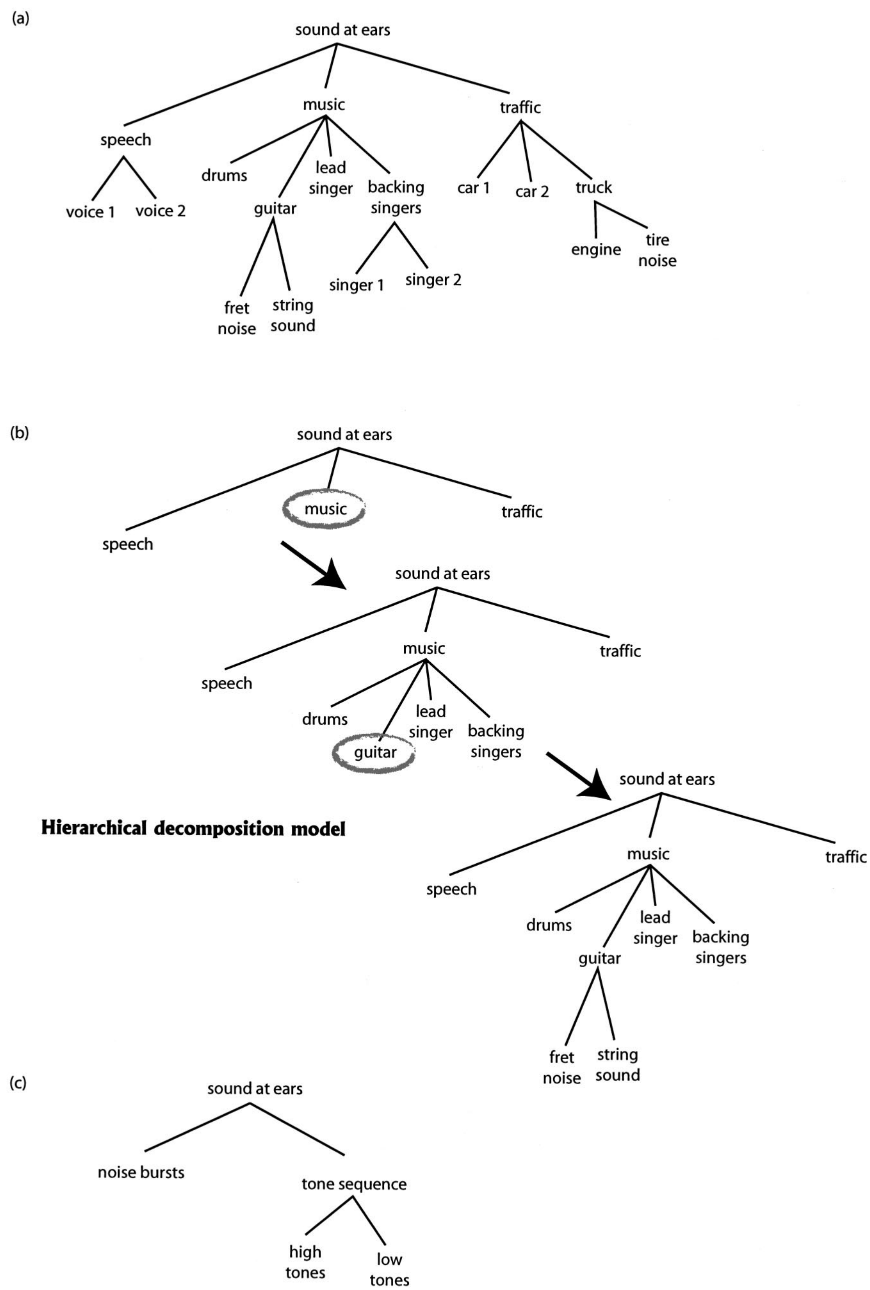

Figure 12. (a) A schematic example of a hierarchy in perceptual grouping in the real world. (b) In the hierarchical decomposition model, it is proposed that this hierarchy is only partially formed, with only attended branches fully elaborated. The circled items correspond to those branches that are attended to, and the arrows indicate the resulting changes in perceptual grouping. (c) The hierarchy in Experiments 1, 2, and 4.

from each other—but that they are not fully elaborated, so the components within each, such as different musical instruments, do not form separate streams. It might be interesting to conduct further experiments to investigate what determines the degree of automatic fragmentation. In Experiments 1 and 2, there was some buildup of streaming, and in particular, in both experiments, there was greater segregation in the absence of attention when the frequency separation of the tones was wide. This might reflect automatic fragmentation. Future experiments could investigate this further by varying the frequency separation of the tones over a wider range or manipulating the context by varying the number of other sequences in the auditory scene. The results from Experiment 4, in which we manipulated attention over time, are consistent with the hierarchical decomposition model. When attention is switched to a different object for a brief period, the streaming of the unattended streams seems to be reset. This is equivalent to subbranches in a branch of the hierarchy collapsing when attention is withdrawn from that branch.

In Experiment 3, we found that a gap in the sequence also reset the perceptual grouping. Here we consider two slightly different explanations for what might be happening. It might be that there is a basic automatic perceptual grouping rule along the lines of *sounds in a similar frequency region that start at similar times are likely to come from the same object and should be allocated to the same stream.* Alternatively, the effect might actually be mediated by attention. If, as suggested above, starting to apply attention to some set of sounds forces the sounds to form a single perceptual group, then perhaps what is happening after a gap is that attention is being exogenously drawn to the new sounds, and the resultant application of attention biases them toward a single perceptual group.

A Flexible or Fixed Hierarchy?

We have established that attention affects perceptual grouping. However, it is not clear whether (a) the top-down signal that is a result of selective attention merely permits further segregation of a stream or (b) it can actually bias the organization that takes place, so that a different kind of grouping can result. Biasing the organization corresponds to a change in the form of the grouping hierarchy rather than just its degree of elaboration. Rather than just a choice between being able to hear the music from Figure 12a as a single perceptual group or being able to break it down into separate instruments, can top-down influences modify the grouping so that, for example, the drums and a rhythmic truck sound form a single perceptual group? Such effects might have interesting implications for experiments investigating streaming. In vision, it has been established that it is harder to attend to two features of separate objects than two features of the same object (Duncan, 1984). An auditory analogue of this would suggest that it should be easier to selectively attend to all of the tones if they form a single perceptual group. In streaming experiments (including our own), listeners are instructed to "attend to the tones," perhaps encouraging the tones' grouping into a single stream at the start of their presentation. To our knowledge, the effect of task demands on grouping has not been investigated. It might be investigated by experiments in which the task set is changed, so that one grouping structure is more helpful than another, and then an electrophysiological (e.g., MMN) or behavioral measure of streaming is taken.

An Early Selection Template?

A further addition to our general account of the interaction between selective attention and perceptual grouping can be made by including a greater degree of early selection. Such a model is illustrated in Figure 1d. It might be that, initially, sources are segregated on a general level. The target source is identified and selectively attended to. Perhaps then some kind of early selection before perceptual grouping is initiated by setting up a template for the target source using its basic features. These basic features might be the specification of a particular set of peripheral frequency channels: It is well established that a substantial amount of selection can performed in this way (e.g., Beauvois & Meddis, 1996; Hartmann & Johnson, 1991). After the selection is initiated, perceptual grouping would be free to form new, more fragmented groups within the parts selected by this template. The advantage of such a system would be that the handing down of some of the selection to a lower level would reduce demand on the subsequent perceptual grouping and selective attention stages. A disadvantage would be that it might not handle sources that vary rapidly in the peripheral channels that they excite. However, this model could easily explain the data presented here. Perceptual grouping at a general level would initially form two streams: the tones and the noises. When the tones start to be attended, a low-level template would be set up that selects them, and perceptual grouping would gradually act upon the output of this process of selection. When attention is switched to the noises, the template would be re-formed, the tones would no longer enter the perceptual grouping level, and streaming would be reset.

Is Perceptual Grouping Capacity Limited?

In earlier sections, we considered the possibility that perceptual grouping mechanisms have a limited capacity. There are good reasons for suspecting that there might be a limitation. Perhaps the most likely form of neural coding for perceptual grouping is that of synchronous firing, an idea that has become popular in vision (Singer & Gray, 1995) and has been suggested in audition (Wang, 1996). There might be limitations to how many simultaneous different firing patterns (or perhaps oscillatory phases) can be reliably coded without interference between them. There are perhaps even more complications in audition than in vision because of the intrinsically temporal nature of sounds. Temporal coding plays a very important role in early auditory processing centers, carrying pitch and envelope information. If the temporal pattern of firing at higher levels is to also carry information about perceptual grouping, either these two types of information must coexist in the temporal code, or the temporal pattern must be recoded as a firing rate (or *place* code) before the level of perceptual grouping.

Summary

The experiments presented here have further elucidated the effects of attention on perceptual grouping. Unattended streams show substantially reduced fragmentation, even if they are in the same ear and frequency region. Short gaps, or short switches in attention, are sufficient to reset the perceptual grouping of a sequence. We propose a hierarchical model of stream formation in which basic streaming is performed and a single stream attended, and then this stream can fragment further, but unattended streams do not.

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. (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.

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.* Cambridge, MA: MIT Press.

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.

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

Culling, J. F., & Darwin, C. J. (1993). The role of timbre in the segregation of simultaneous voices with intersecting F0 contours. *Perception & Psychophysics, 54,* 303–309.

Cusack, R., & Roberts, B. (1999). Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception, 28,* 1281–1289.

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

Dannenbring, G. L., & Bregman, A. S. (1976). Effect of silence between tones on auditory stream segregation. *Journal of the Acoustical Society of America, 59,* 987–989.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. Moore (Ed.), *Hearing* (pp. 387–424). San Diego, CA: Academic Press.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology, 113,* 501–517.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *Journal of the Acoustical Society of America, 108,* 263–271.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–183.

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & Macken, W. J. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 464–473.

Jones, D. M., & Macken, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition, 23,* 192–200.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83,* 323–355.

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.

Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics, 60,* 1216–1227.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18,* 555–586.

Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America, 102,* 1943–1952.

Smith, J., Hausfeld, S., Power, R. P., & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics, 32,* 454–464.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research, 789,* 130–138.

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

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Naatanen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research, 13,* 393–405.

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

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.

Wang, D. (1996). Primitive auditory segregation based on oscillatory correlation. *Cognitive Science, 20,* 409–456.

Received November 5, 2002
Revision received August 6, 2003
Accepted September 29, 2003 ■