

# The role of attention in the formation of auditory streams

**ELYSE S. SUSSMAN**

*Albert Einstein College of Medicine, Bronx, New York*

**JÁNOS HORVÁTH**

*Hungarian Academy of Sciences, Budapest, Hungary*

**ISTVÁN WINKLER**

*Hungarian Academy of Sciences, Budapest, Hungary  
and Helsinki University, Helsinki, Finland*

AND

**MARK ORR**

*Albert Einstein College of Medicine, Bronx, New York*

There is controversy over whether stream segregation is an attention-dependent process. Part of the argument is related to the initial formation of auditory streams. It has been suggested that attention is needed only to form the streams, but not to maintain them once they have been segregated. The question of whether covert attention at the beginning of a to-be-ignored set of sounds will be enough to initiate the segregation process remains open. Here, we investigate this question by (1) using a methodology that does not require the participant to make an overt response to assess how the unattended sounds are organized and (2) structuring the test sound sequence to account for the *covert attention* explanation. The results of four experiments provide evidence to support the view that attention is not always required for the formation of auditory streams.

From infancy, we experience a complex auditory environment made up of several simultaneously active sound sources that often overlap in many acoustic parameters. Despite the confluence of sound, we are able to hear distinct auditory objects and to experience a coherent environment consisting of identifiable auditory events. Because the acoustic information entering one's ears is a mixture of all the sounds in the environment, a key function of the auditory system is to disentangle the mixture of sound and to construct, from the concurrent inputs, neural representations of the sound events that maintain the integrity of the original sources (i.e., auditory scene analysis [ASA]; Bregman, 1990). This is a crucial step in auditory information processing that allows us to hear a single voice in a crowd or to distinguish a voice coming from the left or right in a noisy room. ASA, therefore, plays a critical role in how we experience the auditory environment.

The purpose of the present study was to determine the role of attention in the formation of auditory streams. Bregman (1978) demonstrated that the perception of a sequential sound sequence as segregated into two streams does not occur immediately; it takes several seconds for evidence about the acoustic characteristics of the input to

accumulate (the *buildup* phase). Initially, listeners judge all the sounds as coming from the same source (integration), and only after sufficient information about the characteristics of the acoustic input has been gathered do two (or more) streams emerge in perception (segregation; Bregman, 1978; Carlyon, Cusack, Foxtan, & Robertson, 2001). Functionally, this could mean that in complex auditory environments, there is a bias toward integration (or toward maintaining the organization that is current) that could regulate the system in such a way as, for example, to prevent switching of the sound organization on the basis of spurious transient sounds.

There has been considerable debate about whether the mechanisms responsible for the formation of auditory streams (for the buildup phase, but not necessarily for the maintenance of the streams) require attention. Bregman (1990) suggested that certain forms of stream segregation occur outside the focus of attention, on the basis of the stimulus-driven characteristics of the auditory input, which he termed *primitive processes*. Other forms of stream segregation occur as a function of attentional control, on the basis of top-down or *schema-driven processes*. Among the processes considered to be primitive is the phenomenon

of auditory streaming, in which the hearing of streams of sounds is governed by the reciprocal relationship between presentation rate and frequency separation ( $\Delta f$ ). This was demonstrated by van Noorden (1975), in a study in which a sequence of high- (A) and low-frequency (B) sounds were presented in triplets, with a silent interval between the triplets (ABA-ABA-ABA- . . .). Sounds that were presented at a fast rate with a sufficient frequency separation segregated into two streams, whereas sounds that were presented at a slow rate or with low levels of  $\Delta f$  were perceived as belonging to the same stream. Furthermore, the perceived rate of presentation was slower *within* each stream when the sounds were heard as segregated (A-A-A-A and B-B-B), in comparison with the faster rhythm of *galloping* (ABA-ABA-ABA) that was heard when the sounds were perceived as belonging to the same stream. The perception of the overall rhythm of a sound sequence changes according to the sound organization that is perceived (Bregman, 1990).

Some researchers have argued that attention is crucial even for stream formation that is governed by primitive processes (Alain & Woods, 1997; Botte, Drake, Brochard, & McAdams, 1997; Brochard, Drake, Botte, & McAdams, 1999; Carlyon et al., 2001; M. R. Jones, 1976), whereas other researchers have argued that the primitive stream formation processes can occur outside the focus of attention (Bregman, 1990; Macken, Tremblay, Houghton, Nicholls, & Jones, 2003; Ritter, Sussman, & Molholm, 2000; Sussman, Čeponienė, Shestakova, Nääänen, & Winkler, 2001; Sussman, Ritter, & Vaughan, 1998a, 1999; Winkler, Sussman, et al., 2003).

Supporting the view that attention is required for the initial formation of auditory streams, Carlyon et al.'s (2001) study tested the role of attention in the buildup phase of auditory stream segregation. This was accomplished by assessing whether listeners judged a (previously unattended) sequence of sounds as containing one or two streams. Carlyon et al. presented, to one ear of the participants, a 21-sec sequence containing high- and low-frequency pure tones in an ABA triplet pattern (ABA-ABA-ABA- . . .). When these ABA triplets are perceived as a single integrated stream, participants hear them in a galloping rhythm, whereas when the ABA triplets are perceived as two separate streams, participants hear them as two isochronous rhythms. Carlyon et al. presented four levels of  $\Delta f$  between the high and the low tones in a blocked design (keeping the presentation rate constant across blocks). In the other ear, 400-msec-duration noise bursts were presented once every second, and the participants were instructed to identify each noise burst as either continuously increasing or continuously decreasing in amplitude. Ten seconds into the 21-sec-long train, the participants were cued to switch their attention to the ABA tone sequence in the opposite ear. At the switch, they made a judgment about whether they heard the ABA tones in one stream (in a galloping rhythm) or in two separate streams. The logic of the experiment was that if the buildup process occurred without the participants' attention being focused on the ABA test sounds during the 10-sec time while they were performing the noise task in the contralateral ear, the participants should then immediately judge the ABA tone

sequence as two streams when they directed attention to it. The buildup phase would have already been exceeded during the first 10 sec of the train (while the tone sequence was unattended). However, if attention was needed to surpass the buildup phase, the unattended ABA tones would not yet have segregated at the time of the attention switch, and the participants should initially judge the tone sequence as a single stream in the galloping rhythm. The latter result would also presume that there is an initial bias toward the integrated organization of the auditory input. Carlyon et al. found that after switching their attention, the participants initially judged the tone sequences as a single stream and, later, as two streams, once the buildup phase had been exceeded.<sup>1</sup> Therefore, the authors concluded that attention is essential for the formation of auditory streams.

However, it should be considered that the act of attention switching itself could influence the stream segregation process (Carlyon, 2004; Cusack, Deeks, Aikman, & Carlyon, 2004; Moore & Gockel, 2002). The act of switching attention from one ear to the other may have *reset* the stream segregation process, so that the buildup process began again when attention was directed to the tone sequence. Evidence supporting this interpretation was provided in a recent experiment by Cusack et al. (2004, Experiment 4). They tested whether switching attention has an effect similar to that of resetting the stream segregation process by means of a silent period. The participants switched attention from tones (10 sec) to noise (5 sec) and back to the tones, judging each time whether they heard one or two streams of tones. These results were compared with those from a condition in which the 5 sec of tones were replaced with 5 sec of silence and the participants judged whether they heard one or two streams in the 10-sec blocks of tonal stimuli. The results indicated that switching attention from noise to tones had the same effect on stream judgments as placing a gap between the target tones. Thus, their results indicate that attention switching can reset the stream segregation process.

If attention switching reset the stream segregation process, it is possible, in Carlyon et al.'s (2001) study, that the buildup process and segregation of the ABA sounds occurred without attention in the first 10 sec, while the participant was performing the noise task in the contralateral ear. However, the method of assessing the organization of the unattended sounds (judging the number of sound streams heard) may have been limited by the task's involving attention switching. Carlyon et al.'s results, therefore, cannot definitively answer the question of whether attention is required for the buildup process.

Macken et al. (2003) addressed the difficulty arising from asking participants to make a judgment about streams for a set of test sounds by cleverly making use of a different paradigm in which the participants were not required to perform a task with the test sounds. They made use of the *irrelevant sound effect* (ISE), in which presentation of a sequence of sounds that are irrelevant to the listener's task (i.e., occurring in an unattended stream) disrupts performance on a serial recall task of visually presented items (D. M. Jones, Alford, Bridges, Tremblay, & Macken, 1999). The level of disruption in performance on the visual task

increases as the *changing state* of the irrelevant background sounds increases (D. M. Jones, Alford, Macken, Banbury, & Tremblay, 2000). *Changing state* refers to the alternation of different tokens that belong to the same stream, such as three different syllables presented in a continuous loop. The stipulation is that there must be more than one type of stimulus token in the irrelevant stream for disruption of performance to occur (Tremblay & Jones, 1998). A single repeating sound does not cause a change in performance, in comparison with task performance of the serial recall task in quiet. Thus, alternating different tones belonging to a single stream has a more disruptive effect on serial recall performance than does presenting a sequence of alternating tones that belong to two separate monotonic streams, each stream consisting of one repeating tone (D. M. Jones et al., 1999). Macken et al. presented irrelevant sounds at a fixed seven-semitone difference between them, in a sequence using the ABA triplet pattern (van Noorden, 1975), and varied the rate of presentation of the sounds in three main test conditions (low, medium, and high rates). The participants were instructed to perform a visual serial recall test and to ignore the irrelevant sounds. Because there is a reciprocal relationship between frequency separation and presentation rate (the faster the presentation rate, the smaller the frequency separation before streaming is perceived and vice versa), Macken et al. reasoned that increasing the rate of presentation would change the ISE. When the presentation rate was speeded up, this would lead to streaming at the high rate, effectively changing the functional rhythm of the sounds, and thus, performance should improve on the visual serial recall task. The changing state, the necessary precondition of the ISE, would be eliminated if the two different frequency tones were part of two single monotonic streams. Without a changing state in the background sound, the level of disruption on performance would decrease. The authors reasoned that as the rate of presentation increased, the degree of interference from the irrelevant sounds would increase until the two sounds split into two monotonic streams. In this way, the performance level for the simultaneously presented serial visual recall task would indicate whether the irrelevant unattended ABA sounds were integrated into one frequency stream of alternating tones of two different pitches (i.e., a changing state) or were segregated into two monotonic streams. If attention to irrelevant sounds is required for stream segregation to occur, increasing the stimulus rate should lead to a *linear* increase in the amount of interference with performance on the recall task. With no segregation of the sounds (and maintenance of the changing state), as rate increases, interference with performance should also increase. However, this was not the case; increasing the rate of stimulation interfered with performance on the visual serial recall task up to the point at which the two tones split into separate streams, showing a *nonmonotonic* effect of presentation rate on performance. Therefore, Macken et al. concluded that stream segregation processes do not necessarily require attention.

Macken et al.'s (2003) results provide support for the view that auditory stream segregation is part of a primitive process that does not always require focused attention. Even

so, their method is open to the criticism that attention may be needed only during the initial *formation* of the auditory streams and may not be required for the maintenance of the streams. It is possible that the listeners *covertly* attended the irrelevant sounds for enough time to initiate the segregation process. The term *covert attention* has most often been used to refer to visual processes—specifically, the ability to attend to something in the visual field without moving one's eyes. However, the term has sometimes been used to refer to auditory attention, in the case in which participants are asked to ignore a sound source but may have the opportunity to “sneak a peek” toward the to-be-ignored sound source when simultaneously performing a visual task or a task involving a different set of sounds. It has been conjectured that this type of covert attention, the “sneak a peek” toward the ignored sounds, may be enough to initiate the formation of the streams, even if continual or focused attention is not directed toward the sounds. The covert attention explanation may be considered a possibility in Macken et al.'s study, especially since the primary task could be considered a *low-load* task (Lavie, 2005).

In the present study, we addressed this *covert attention hypothesis* in four experiments, testing the question of whether the *buildup phase* of stream segregation requires attention. Electrophysiological measures of the brain's response to sounds are highly suitable for addressing this question. The mismatch negativity (MMN) component of event-related brain potentials (ERPs) has been an especially useful tool, because MMN elicitation does not require the experimental participant to respond to the sounds or to indicate perception of the sounds. MMN can be used to index the organization of one set of sounds when attention is directed to another set of sounds to perform an auditory task. The MMN component (occurring about 150 msec from the onset of deviance) reflects the outcome of a deviance detection process that is based on the memory of the regularities in the auditory input (Näätänen & Winkler, 1999; Picton, Alain, Otten, Ritter, & Achim, 2000; Sussman, Ritter, & Vaughan, 1998a, 1998b; Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Winkler, Karmos, & Näätänen, 1996). Sound input detected as being different from the stored regularities elicits the MMN. Infrequent changes in such auditory features as frequency, intensity, tone duration, or spatial location, as well as violations of sequential tone patterns, elicit MMN (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Sussman et al., 1999). The main generators of the MMN component are located in the auditory cortex, which accounts for its maximally negative amplitude over the frontocentral scalp region (Giard, Perrin, Pernier, & Bouchet, 1990; Scherg, Vajsar, & Picton, 1989).

When unattended sound processes indexed by MMN are assessed, the issue of whether highly focused attention can affect MMN generation is brought into question. A number of studies have shown that the MMN amplitude elicited by intensity deviants presented to the unattended ear was attenuated, as in a dichotic-listening paradigm. Focusing one's attention on sounds presented to one ear and detecting infrequent deviants among them attenuates the MMN amplitude to similar deviants presented simultaneously to the ignored ear (Näätänen, Paavilainen, Tiitinen,

Jiang, & Alho, 1993; Trejo, Ryan-Jones, & Kramer, 1995; Woldorff & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998). However, this effect appears to be specific to the situation in which the same feature deviant occurs in both ears and is biased by selectively attending to deviants in one ear and ignoring the same deviants in the other ear (Sussman, Winkler, & Wang, 2003). If a feature deviant is presented to the ignored ear that is different from that which is being detected in the attended ear (e.g., intensity deviation when frequency is attended to), the MMN amplitude is unaffected (Sussman et al., 2003). Sussman et al.'s (2003) finding directly relates to the present study, in that we used a selective-listening paradigm that, by itself, would not bias the MMN measure (see also Winkler, Czigler, Sussman, Horváth, & Balázs, 2005).

Another ERP component, the P3a (or *novelty P3*), is often elicited by infrequent or novel sounds. P3a is generally associated with involuntary attention switching to salient sounds that are task irrelevant or unattended (Escera, Alho, Schröger, & Winkler, 2000; Friedman, Cycowicz, & Gaeta, 2001; Knight & Scabini, 1998). The P3a component usually peaks around 200–300 msec from stimulus onset and has a maximal frontocentral scalp distribution. When elicited together with MMN, the P3a follows the MMN in time. P3a and MMN have in common that they can be elicited by deviations in a sound sequence. However, an important distinction between them is that the MMN is thought to reflect the detection of the deviance, whereas the P3a component is thought to reflect the action of attention switching (Friedman et al., 2001). P3a can also be distinguished from the P3b component (which requires attention to be focused on the sounds to be elicited) by its scalp topography. P3a is maximal at frontocentral scalp sites, and P3b is maximal at parietal sites (Friedman et al., 2001).

The neural mechanisms that allow us to perceive separate sound sources in the environment have recently been studied in humans, using auditory ERPs (Ritter et al., 2000; Sussman, Bregman, Wang, & Kahn, 2005; Sussman et al., 2001; Sussman et al., 1998a, 1999; Winkler, Horváth, Teder-Sälejärvi, Näätänen, & Sussman, 2003; Winkler, Kushnerenko, et al., 2003; Winkler, Sussman, et al., 2003; Yabe et al., 2001). The results of these studies indicate that stream segregation can occur in the absence of focused attention. Evidence from both animal and human studies suggests that the basic stream segregation mechanisms are part of vertebrates' (Bee & Klump, 2004; Fay, 2000; Fishman, Reser, Arezzo, & Steinschneider, 2001; Hulse, MacDougall-Shackleton, & Wisniewski, 1997; Izumi, 2002; Kanwal, Medvedev, & Micheyl, 2003; Winkler, Kushnerenko, et al., 2003) and invertebrates' (Schul & Sheridan, 2006) auditory systems, consistent with a role for primitive processes in governing the basic segregation of auditory information.

Sussman et al. (1999) used the MMN measure to probe the stream segregation processes while participants ignored the sound stimulation and read a book. Alternating high (H) and low (L) tones were presented at a rapid pace that was known to induce a streaming effect (Bregman, 1990). The paradigm was set up so that when the tones segregated into two streams, repeating three-tone sequen-

tial patterns would emerge separately within each stream. When the sequential tone patterns emerged as a result of segregation, the infrequent within-stream pattern violations could be detected. Detection of the pattern violations would then elicit the MMN component. Therefore, if the representation of *alternating* high and low tones was maintained as a single integrated stream in auditory memory, no MMN would be elicited. The unattended pattern violations elicited MMN within both the high- and low-tone streams; thus, it was concluded that the high and low tones segregated into separate streams without attention being focused on the input. A control condition was used to demonstrate that no MMN was elicited by the deviants when the sounds alternated in frequency within a single stream of high and low tones.

Although the results of this and other previous studies are compatible with the view that auditory stream segregation processes can occur outside the focus of attention, these data have not ruled out the possibility that focused attention is needed for the *initial* buildup phase (or the formation of the streams), even if attention is not needed to maintain the segregated organization after the sounds have been sorted into streams. In these previous studies, continuous sequences of sounds were presented in separate blocks of approximately 5 min each, thus leaving open the possibility that the participants could covertly attend the sounds during the initial seconds in which the stimulus blocks were presented.<sup>2</sup> On this basis, it could be argued that using *ignore* conditions did not sufficiently control attention; covert attention to the irrelevant sounds at the beginning of the sound sequence might have been enough to initiate the segregation process. Furthermore, one could argue that auditory attentional resources might have been available to process the unattended sounds to some degree, especially when the primary task was visual, since there is evidence to suggest that attentional capacities for visual and auditory processing may, at least in part, operate independently of each other (e.g., Duncan, Martens, & Ward, 1997).

The covert attention hypothesis suggests that attention (enough to surpass the buildup phase) will be needed only for a short time to initiate segregation of unattended sounds. Once the segregation process has been initiated, attention will no longer be needed to maintain segregation of the unattended sounds. Thus, in experimental paradigms in which a continuous sequence of sounds is presented, attention will be needed only at or near the beginning of the continuous sequence.

Therefore, in the present study, to determine whether the buildup phase of stream segregation requires attention, we used the following methodological strategies: (1) a demanding primary task with a set of sounds different from the irrelevant (to-be-ignored) sounds, (2) no task with the irrelevant sounds, and (3) a paradigm designed to test the buildup phase of the stream segregation process, in which short trains, rather than long, continuous sequences of task-irrelevant sounds, were used.

We presented short sound sequences (*trains*), separated by silence, while the participants performed a noise change detection task. The purpose of the silence between trains was to *reset* the stream segregation process (Bregman,



1978; Cusack et al., 2004) so that the buildup to stream segregation (called the *buildup phase*) would have to be renewed at the start of every train. For the two segregated streams to emerge, this phase must be completed as the train unfolds (i.e., the buildup phase must be surpassed). By presenting short sequences, we eliminated confounding factors between the buildup phase of stream segregation and the onset of a sound sequence in the more commonly used presentation of long continuous sequences of sound. We reasoned that if attention were needed to initiate the stream segregation process, covert attention to the beginning of the first short sequence of the long continuous sequence (train–silence–train–silence and so on) would not be enough to account for the results, because at least some attention would have to be directed toward *all* the irrelevant trains presented. This would likely prevent the participants from effectively performing the primary noise change detection task.

A second feature of the paradigm was that an MMN component would be elicited only if the sounds of the train sequences segregated into two streams. If the buildup phase required attention, no MMNs would be elicited by deviants within the unattended stimuli, because the segregation would not occur. On the other hand, if the deviants did elicit MMNs, we could conclude that attention is not always required for the buildup of stream segregation.

Experiments 1–3 assessed electrophysiological measures of sound organization and features of MMN generation in the train sound sequences when participants have no task with the sounds. Experiment 4 was conducted to obtain a measure of the perception of the tone sequences.

## EXPERIMENT 1

### Method

#### Participants

Twelve young adults (18–26 years of age; average age, 22 years; 5 males) with normal hearing (screened for a hearing threshold of 20 dB HL or better from 250 through 4000 Hz) were paid for their participation in the study. Written informed consent was obtained after the procedures had been explained to them. Two participants' data were excluded due to extensive electrical artifacts.

#### Stimuli and Procedure

**Procedure.** The stimuli were presented in a free-field listening environment in a sound-attenuated chamber to participants who sat in a comfortable chair with two loudspeakers placed behind them (positioned on an arc of 1-m radius around the participants) bilaterally presenting the sounds that were irrelevant to the listeners' task and to be ignored. A third loudspeaker was placed 1.9 m directly in front of them, presenting acoustic input that was relevant and to be attended.

**Task-irrelevant sound sequences.** We present 3.55-sec sequences (trains) of X and O tones (OOXOOXOOX . . . ; 12 X and 24 O tones; see Figure 1) to the participants. The tones were 50 msec in duration (5-msec rise/fall times), with an interstimulus interval (ISI; offset to onset) of 50 msec. The X tones had the same frequency and intensity value in both conditions (1500 Hz; 74 dB). The 10th X tone of every train was presented with 83-dB intensity (frequency remained 1500 Hz; this we call the *deviant*; the other X tones we call the *standard*). Two O tones always intervened between the X tones. There were two types of trains, *near* and *far*, which denoted whether the intervening O tones were near to or far from the X tones in frequency. For the near trains (*near condition*), the O tones had a frequency of 1400 Hz (1 semitone [ST] away from the X tone),

whereas for the far trains (*far condition*) the O tones had a frequency of 500 Hz (19 STs away from the X tone). The intensity of the intervening O tones randomly varied in 3-dB increments from 65 to 92 dB (excluding the 74-dB value of the standard X tone and the 83-dB value of the deviant X tone). Near and far trains were presented randomly, using an intertrain interval of 4.05 sec (offset to onset). In each condition, 208 trains were delivered. There were 10 stimulus blocks of near- and far-condition trains (A blocks), and 10 blocks of two types of control condition trains, one for the near-condition trains and one for the far-condition trains (B blocks; see the Data Recording and Analysis section for a description). The blocks were presented in a counterbalanced order.

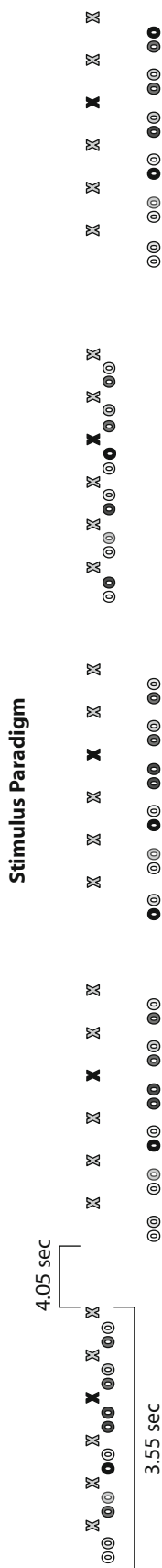
The deviant (83-dB) X tone was the *probe tone* used to determine how the sounds were organized for the near and far trains in the memory underlying MMN generation. From our previous studies, we knew that the probe stimulus elicits MMN only when the X and the O tones are segregated into two separate streams, but not when they are integrated into a single frequency stream (Sussman et al., 2001; Winkler, Kushnarenko, et al., 2003; Winkler, Sussman, et al., 2003). The paradigm was set up so that segregation was cued by frequency separation between the X and the O tones, whereas intensity was used to elicit MMN. When the X and the O tones are part of the same frequency stream, the intensity variation of the O tones prevents the emergence of a regularity in intensity (the prerequisite for MMN elicitation), and thus the probe tones do not violate any regularity. In contrast, when the X and O tones are segregated into separate streams, a regularity of intensity emerges in the X-tone stream (they all have the same intensity value), upon which the probe tones (which differ in intensity from the standard X tones) violate the intensity regularity and elicit the MMN component.

**Primary task.** Continuous band-filtered (100–1200 Hz) white noise was presented through a loudspeaker placed in front of the participant. Infrequently, the noise intensity changed. The base level of noise, measured at the participant's head, was 65 dB SPL. Intensity changes (1–2 dB), moving up or down from the base level, were 5-msec-long ramps that occurred approximately once every 15.25 sec (with even distribution between 0.5 and 30 sec). After the change (up or down), the noise intensity stayed on the new level until the next change (up or down). The participant was instructed to listen to the noise and indicate when he or she detected these slight changes in noise intensity by pressing a designated button on a response keypad.

Practice was given to the participant prior to the test session, in which the amount of intensity change was determined separately for each participant (between 1 and 3 dB). This was done to obtain performance close to 80% accuracy for each participant during the initial practice session. Starting from 3 dB, the amount of intensity change decreased as long as the participant still performed above 80% and increased when detection performance did not reach 80% in two consecutive blocks of the same intensity change. The training period ended when a stable above-80% level of performance was established—typically, after 10–12 training blocks. During this practice session, the intensity changes occurred about once every 5 sec. The participant received feedback on his or her task performance after each stimulus block (both during training and in the main experiment) and was motivated to perform well by a bonus payment schedule that depended on task performance during the main experiment.

#### Data Recording and Analysis

An electroencephalogram (EEG) was recorded with Ag/AgCl electrodes from Fz, Cz, Pz, F3, C3, P3, F4, C4, and P4 (10–20 system) and from the left and right mastoids (Lm and Rm, respectively). The reference electrode was placed on the tip of nose. Horizontal electrooculogram (EOG) was monitored with electrodes placed lateral to the outer canthi of the two eyes, and vertical EOG was monitored with electrodes placed above and below the right eye. The EEG and EOG were amplified (Neuroscan SynAmps), digitized with a 250-Hz sampling rate, and filtered offline (1.5–30 Hz). Epochs of 300 msec prestimulus and 400 msec poststimulus periods were averaged separately for the standard and the deviant tones (the 300-msec



**Figure 1.** Experiment 1: Schematic diagram of the unattended sound sequence showing the near and far conditions. In the far condition, the O tones were far in frequency from the X tones (19-semitone [ST] difference), in comparison with the near condition (1-ST difference). This is denoted by distance from the X tones along the ordinate axis. The shades of gray represent the variations in intensity of the intervening O tones. The dark X tone represents the probe tone that could elicit mismatch negativity with respect to the intensity of the previous X tones, occurring only when the X and O sounds were segregated. The trains were 3.55 sec in duration, and the silence between the trains was 4.05 sec.

prestimulus baseline is the length of one full stimulation cycle—i.e., an “XOO” subsequence). Epochs in which the signal difference between two consecutive sampling points exceeded  $14 \mu\text{V}$  or the overall signal range exceeded  $75 \mu\text{V}$  on any EOG or EEG channel were excluded from averaging. We also excluded epochs that overlapped with a noise change in the task-relevant sounds (2-sec intervals following a noise change and 1-sec intervals around buttonpresses).

**Control comparison for the MMN.** To control for stimulus-specific ERP effects, another condition was presented to the participants in which all the X tones took the intensity value of the deviant tone in the main conditions (83 dB), while all other parameters were kept identical to those in the main experimental conditions. The X tone delivered in the same position as that for the deviant tone in the main conditions (the 10th X tone of the trains) served as the comparison for the deviants in the main experiment. In this way, a tone that was a deviant in the train could be compared with a tone with exactly the same tone intensity (83 dB) that was a standard in the train, thereby eliminating any stimulus-specific ERP effects that may have occurred by comparing tones of two different intensities.

The ERPs elicited by the control and deviant tones were averaged separately. The elicitation of MMN was assessed by comparing the responses elicited by control and deviant tones separately for the near and the far conditions. A 40-msec latency measurement window was centered on the negative peak of the grand mean deviant-minus-control difference curve in the expected MMN latency range at Fz for the far condition, in which MMN could be observed (116–156 msec). The average voltage in this time interval was used to test for the presence of MMN in both the near and the far conditions. Statistical testing was conducted using a one-sample, one-tailed *t* test at the Fz electrode (the site of maximal signal-to-noise ratio for MMN) to determine whether the mean probe-minus-control difference voltage was significantly greater than zero. To test for the presence of P3a, a 40-msec measurement latency window was centered on the positive peak of the grand mean deviant-minus-control difference curve in the expected P3a latency range at Cz for the far condition, in which P3a could be observed (232–272 msec). This time interval was used to assess the presence of P3a in both the near and the far conditions. One-sample, one-tailed *t* tests were conducted at the Cz electrode (the site of maximal signal-to-noise ratio for P3a) to determine whether the mean probe-minus-control difference voltage was significantly greater than zero. Responses in the noise change detection task were considered correct if they fell within 100–2,000 msec from the onset of the noise change. Student *t* tests were used to compare reaction times (RTs) and hit rates (HRs) across conditions.

## Results and Discussion

### Performance on the Primary Task

The mean intensity change for the participants was 1.47 dB ( $SD = 0.14$ ). There was no difference in performance on the noise change detection task between the near and the far conditions for either HR [ $t(9) < 1, p > .3$ ] or RT [ $t(9) = 2.1, p > .06$ ], although there was a trend for the RTs in the near condition to be shorter; thus, the overall mean performance will be reported. The mean HR on the noise change detection task was 68% ( $SD = 6\%$ ), with a mean RT of 774 msec ( $SD = 68$  msec). The HR was lower for the experimental blocks than for the training phase.

### Event-Related Potentials

The grand mean ERPs elicited by the probe (deviant) and control stimuli at all scalp-recorded electrode sites are presented for the near (Figure 2) and the far (Figure 3) conditions separately. In the far condition, a negative difference between the control and the deviant ERPs can be seen, followed by a positive difference. These differences

denote MMN and P3a, respectively. These components can be clearly seen by subtracting the control from the probe (deviant) waveforms (see Figure 4). In the near condition, no such difference can be seen. The response to the control tone appears more negative than the response to the probe tone in the expected latency range for MMN (Figure 2).

MMN was elicited by probe tones in the far condition [ $t(9) = 3.9, p < .01$ ], whereas no MMN was elicited in the near condition. The control waveform was more negative than the deviant waveform in the expected latency of the MMN of the near condition. This produced a positive-going waveform in the difference curves, which was significantly greater than zero [ $t(9) = 2.6, p < .05$ ]. These results show that the sounds were segregated with a 1000-Hz (approximately 19 STs) difference, but not with a 100-Hz (approximately 1 ST) difference between the X and the O tones. In addition, a P3a component was elicited by the probe tones in the far condition [ $t(9) = 5.0, p < .01$ ]. In contrast, there was no evidence for a P3a in the near condition [ $t(9) < 1, p > .42$ ]. This is further evidence suggesting that the sounds were segregated only in the far condition, in which the probe tone was detected as a salient infrequent sound in the X-tone stream.

MMN was elicited in the far condition, in which there was about a 19-ST difference between the X and the O tones, but not in the near condition, in which there was about a 1-ST difference between them. This demonstrates that there was a distinction in how the probe sounds were processed in these two conditions, which we interpret as a difference in the *context* of the sounds set up by the frequency separation between the sounds. In the near condition, the intensity of the probe tones could not be distinguished as being deviant with respect to the intensity of the other tones in the sequence; thus, they elicited no MMN. In the far condition, the X tones segregated from

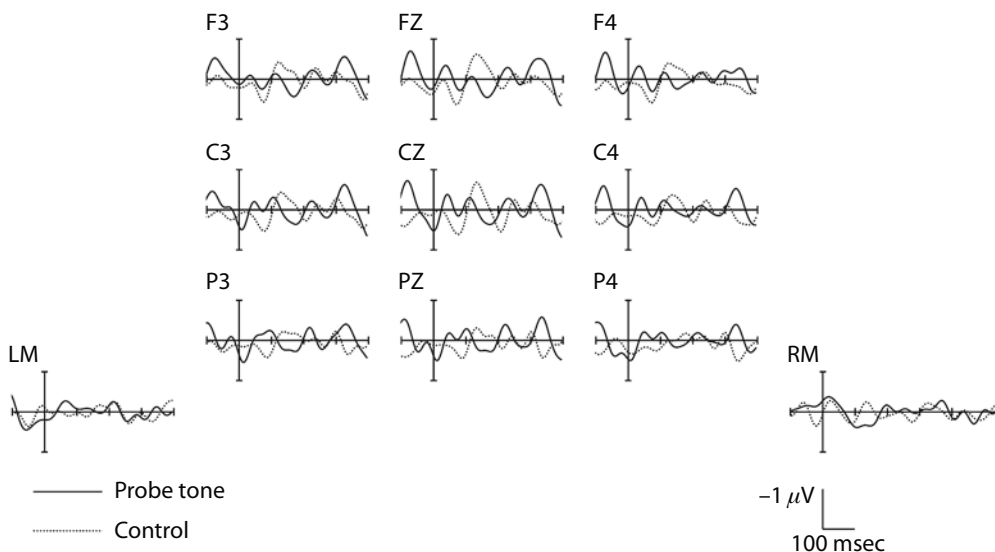
the O tones. This is demonstrated by the X tones being detected as deviant in the X-tone stream (elicitation of the MMN); moreover, they were salient “oddballs” in this stream, because they subsequently elicited the P3a (which follows MMN in time). These results provide evidence supporting the notion that attention is not always required to initiate the stream segregation process.

The underlying assumption in the interpretation of the results of Experiment 1 was that the buildup phase occurred at the beginning of every train, which was precipitated by the resetting of the stream segregation process by the silence between the trains. If this assumption is true, then, in addition to an MMN being elicited at the end of the trains, no MMN should be elicited by probe tones occurring at the beginning of the trains, because the buildup phase would not yet be completed. Thus, if the ending deviants elicited MMN, but not the beginning deviants, in the far-condition trains, this would provide strong evidence that the buildup phase does not require attention. Experiment 2 was run to test this.

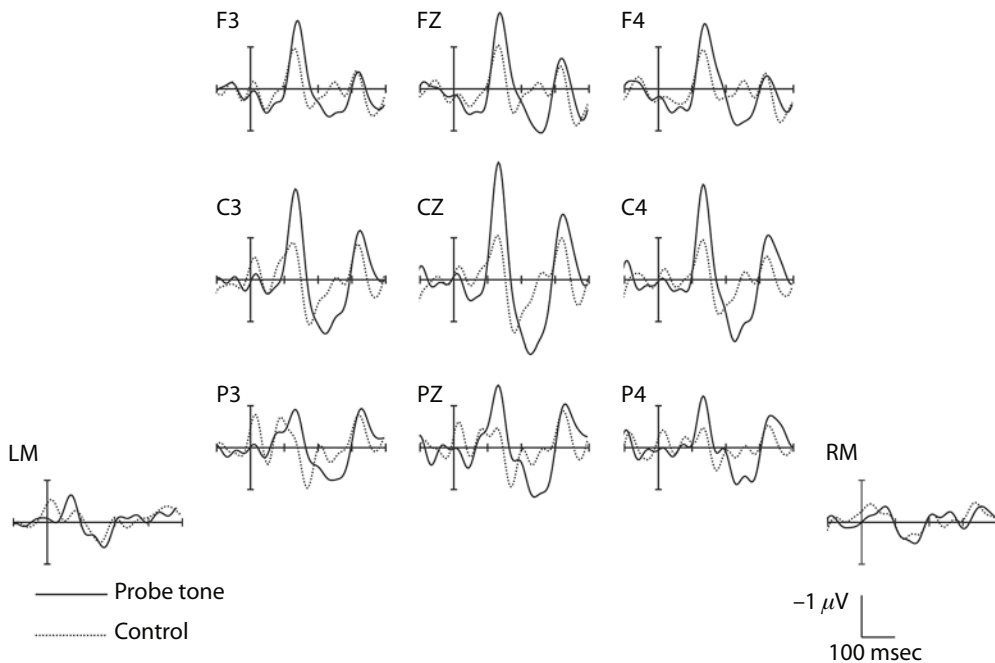
It should also be considered that with the large 19-ST difference, these data do not conclusively rule out a peripheral channel explanation for the MMN results. This alternative explanation suggests that because there was no overlap along the basilar membrane for tones with a 19-ST difference, detecting the probe tones was easier for these tones than for those with a 1-ST difference, where there was overlap of excitation along the basilar membrane. This issue was addressed further in Experiment 2.

## EXPERIMENT 2

The main purpose for Experiment 2 was to test the assumption of Experiment 1 that the silence between trains reset the stream segregation process, so that the buildup



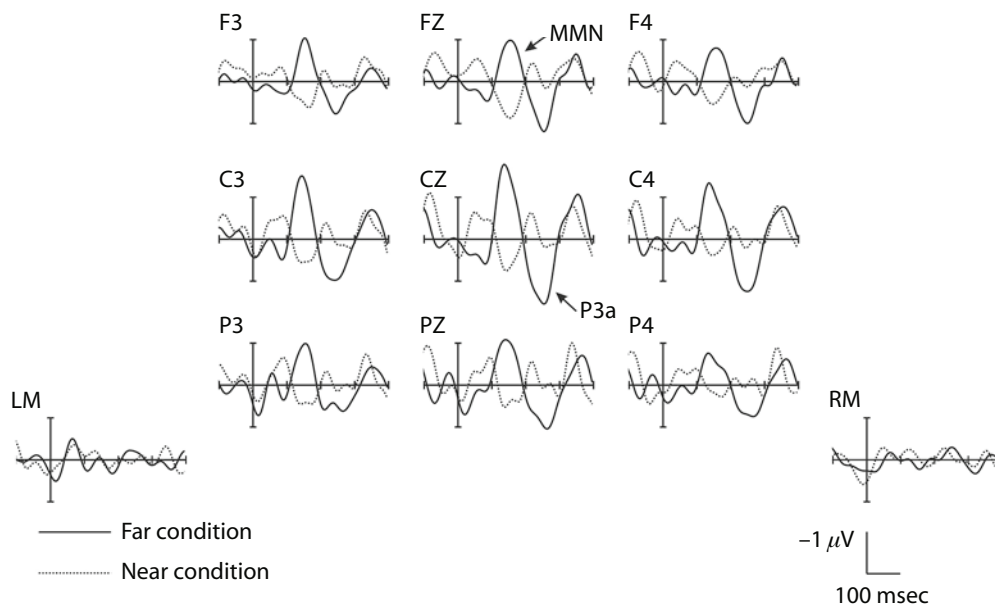
**Figure 2.** Experiment 1: Grand-averaged event-related potentials (ERPs) for the near condition. The ERP response elicited by the probe tones (solid line) is superimposed on the ERP response elicited by the control tones (dotted line; see the Method section) at all scalp-recorded sites. Note that for display purposes only, a 100-msec prestimulus baseline is shown (as in Figures 3 and 4).



**Figure 3. Experiment 1: Grand-averaged event-related potentials (ERPs) for the far condition. The ERP response elicited by the probe tones (solid line) is superimposed on the ERP response elicited by the control tones (dotted line) at all scalp-recorded sites.**

phase would have to be surpassed at the start of every train before the segregation of sounds within trains could occur. Thus, probe tones were placed in both beginning and ending positions of the far-condition trains. If MMNs were elicited by the probe tones at the ending position of the far-condition

trains, but *not* by those at the beginning position, this would show that the buildup phase had not yet been completed at the beginning of the trains but had been surpassed by the end of the short train sequences. That is, the absence of MMN to probe tones at the beginning of the trains would



**Figure 4. Experiment 1: Difference waveforms for the near (dotted line) and far (solid line) conditions at all recording sites. The difference waveforms were obtained by subtracting the event-related potential (ERP) response elicited by the control tones from the ERP response elicited by the probe tones separately for each condition. The mismatch negativity (MMN) and P3a components (shown with arrows) can be clearly seen in the far condition, whereas neither of these components can be observed in the near condition.**



indicate that the buildup phase had not yet been completed. Experiment 3 provided additional confirmation, showing that MMN was elicited for stimuli at the beginning of a train when there were no intervening tones.

## Method

### Participants

Sixteen young adults (18–25 years of age; mean age, 21 years; 5 males) with normal hearing (screened for a hearing threshold of 20 dB HL or better from 250 through 4000 Hz) were paid for their participation in the experiment. Written informed consent was obtained after the procedures had been explained to them.

### Stimuli and Procedure

**Procedure.** The procedure was exactly the same as that in Experiment 1. The stimuli were presented in a free-field listening environment with two loudspeakers placed behind the participants. These loudspeakers presented acoustic input that was irrelevant to the listeners' task and were to be ignored. A third loudspeaker was placed directly in front of the participants. The participants performed the noise change detection task with the input from the loudspeaker placed in front of them.

**Task-irrelevant sound sequences.** The stimuli were presented in 2.52-sec tone sequences (trains), whose composition was similar to those used in Experiment 1. Thirty-six tones made up each train (XOOXOOXOO . . . ; 12 X tones and 24 O tones). Tone duration was 30 msec (5-msec rise/fall times). The ISI (offset-to-onset) was 40 msec. The trains were separated by 4.07 sec of silence. Five different types of trains were randomly presented in blocks of 50 trains. Two hundred trains of each type were presented in 20 blocks. Two types of trains made up the far condition (one in which the deviant occurred in the beginning of the train and one in which the deviant occurred at the end of the train) and one type for the near condition (Figure 5 provides a schematic diagram of the different trains). The other two types of trains served as controls for delineating the MMN component (see the Control comparison for the MMN section below), one for the near-condition trains and one for the far-condition trains. The  $\Delta f$  between the X and the O tones in the far condition was 8 ST, and it was 1 ST in the near condition. The frequency separation between the X and the O tones was decreased, and the presentation rate was increased, in comparison with Experiment 1, since there is an inverse relationship between frequency separation and presentation rate for the perceiving of stream segregation (Bregman, 1990). Increasing the presentation rate allowed us to place the probe tone in the fourth position of the train without surpassing the buildup phase at this frequency separation (Carlyon et al., 2001). Placing the probe tone in the fourth position allowed sufficient time for the standard representation to be established, so that a deviant in the fourth position of the train could elicit MMN (Cowan, Winkler, Teder, & Näätänen, 1993; Horváth, Czigler, Sussman, & Winkler, 2001; Schröger, 1997). Cowan et al. presented trains of tones and showed that MMN could be elicited after three standards (i.e., a fourth-position deviant), but not after one standard (i.e., a second-position deviant) in their *roving standard* condition (similar to our paradigm, with trains that roved in frequency being used; see the next paragraph). Thus, we ensured that a sufficient number of standards were presented for a standard representation to be established before the occurrence of the fourth-position probe tone.

To avoid possible carryover effects from one train to the next, the following steps were taken: (1) The presentation of all of the train types was randomized within each block, (2) the frequency of the X tone roved across eight levels of frequency (1396.9–3135.9 Hz in 2-ST steps), and (3) the base X tone intensity roved across five levels (in 4-dB steps, from 62 to 78 dB). Within each train, the *standard* X tone maintained the same base frequency and intensity level. Either the 4th-position or the 10th-position X tone was 9 dB higher in intensity than the other X tones (this we call the *deviant* or *probe*

tone). The O tones varied randomly among four other intensity levels (–3, +3, +6, and +12 dB, as compared with that of the base standard intensity in the given train). For example, for the far condition, if the base frequency/intensity of a train was 1567.98 Hz/62 dB, all of the O tones had a frequency of 987.77 Hz (8-ST  $\Delta f$ ) and varied randomly, having one of the following intensity levels: 59, 65, 68, or 74 dB. The deviant X tone was 9 dB higher than the standard X tone (e.g., 71 dB). Thus, the standard X tone was 1567.98 Hz/62 dB, and the deviant X tone was 1567.98 Hz/71 dB. In this way, as also in Experiment 1,  $\Delta f$  cued segregation between the X and the O tones, whereas intensity was used to elicit MMN (see the Stimuli and Procedure section in Experiment 1 for a more detailed description of the rationale for the protocol).

**Primary task.** The primary task in Experiment 2 was exactly the same as that used in Experiment 1.

### Data Recording and Analysis

EEG recordings and data analysis procedures were exactly the same as those used in Experiment 1, with the following exceptions.

The prestimulus baseline period was 210 msec, to accommodate exactly one full cycle of stimulation (i.e., one XOO subsequence).

**Control comparison for the MMN.** Five intensity levels were equiprobably distributed randomly across all the X tone positions of the trains, with the +9-dB value occurring in the 4th or 10th position of the far-condition trains and in the 10th position of the near-condition trains. Thus, the same position tone from the control trains served as the comparison for the deviants in the main experiment. In this way, a tone that was a deviant in the train could be compared with a tone with exactly the same tone intensity that was a standard in the train, in the same position. This comparison procedure would eliminate any stimulus-specific ERP effects from the deviant-minus-control difference waveforms.

A 40-msec latency measurement window was centered on the negative peak of the grand mean deviant-minus-control difference curve in the expected MMN latency range at Fz for Position 10 in the far condition, in which MMN could be observed (116–156 msec). The average voltage in this time interval was used to test for the presence of MMN in both the near and the far conditions (for both probe positions). Statistical testing was conducted using a one-sample, one-tailed *t* test at the Fz electrode to determine whether the mean probe-minus-control difference voltage was significantly greater than zero.

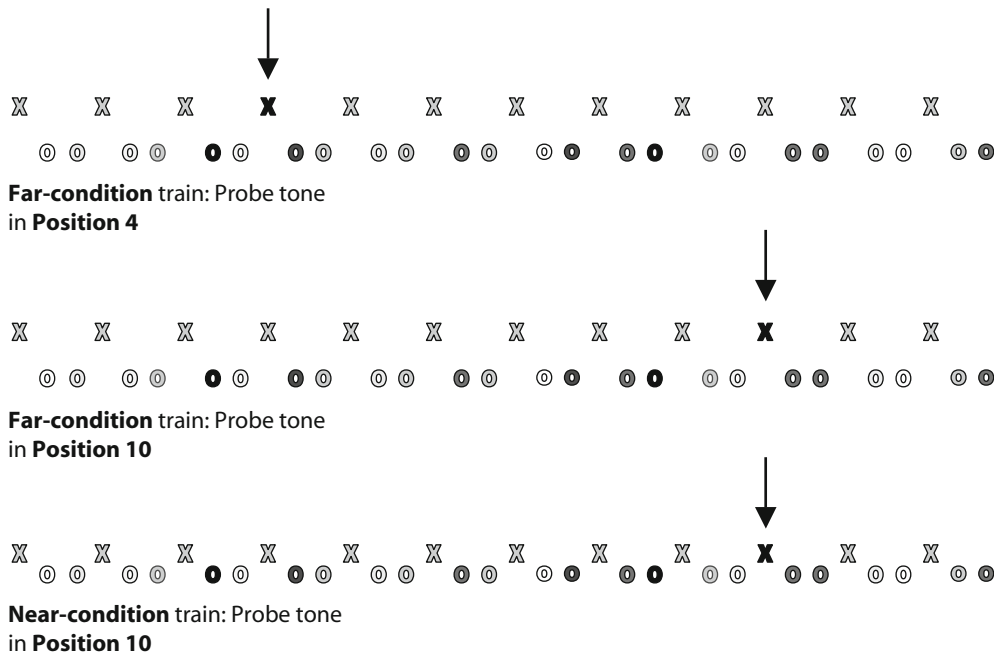
To test for the presence of P3a, a 40-msec measurement latency window was centered on the positive peak of the grand mean deviant-minus-control difference curve in the expected P3a latency range at Cz for position 10 in the far condition, in which P3a could be observed (204–244 msec). This time interval was used to measure for the presence of P3a in both the near and the far conditions (for both probe positions). Statistical testing was conducted using a one-sample, one-tailed *t* test at the Cz electrode to determine whether the mean probe-minus-control difference voltage was significantly greater than zero.

Responses in the noise change detection task were considered correct if they fell within 100–2,000 msec from the onset of the noise change. Paired *t* tests were used to compare RTs and HRs across conditions.

## Results and Discussion

### Performance on the Primary Task

The mean intensity change for the participants was 1.7 dB ( $SD = 0.25$ ). There was no difference in performance on the noise change detection task between the near and the far conditions for either HR [ $t(15) = 1.3, p > .2$ ] or RT [ $t(15) = 0.1, p > .9$ ]; thus, the overall mean performance will be reported. The mean HR was 61% ( $SD = 14\%$ ), with a mean RT of 888 msec ( $SD = 145$  msec). The HR was lower for the experimental blocks than for the training phase.



**Figure 5.** Experiment 2: Schematic diagram of the train types for the task-irrelevant tones. The three experimental train types are depicted. In the far condition, the O tones were 8 semitones (STs) higher in frequency than the X tones. In the near condition, the O tones were 1 ST higher in frequency than the X tones. This is denoted by distance from the X tones along the ordinate axis. The shades of gray represent the variations in intensity of the intervening O tones. The dark X tone represents the probe tone that could elicit mismatch negativity with respect to the intensity of the previous X tones only when the X and the O sounds were segregated. Arrows point to the positions of the probe tones.

### Event-Related Potentials

The grand mean ERPs elicited by the probe tone and the control stimuli and the corresponding probe tone-minus-control difference waveforms are presented in Figures 6 and 7, respectively. In the far condition, the 10th position probe tone elicited a negative difference, in comparison with the control, followed by a positive difference between the two. These differences denote MMN [ $t(15) = 3.2, p < .01$ ] and P3a [ $t(15) = 7.2, p < .01$ ], respectively. In contrast, the mean voltage in the MMN or P3a range was not significantly different from zero for the 4th-position probe tones in the far condition [MMN range,  $t(15) < 1, p > .47$ ; P3a range,  $t(15) = 1.1, p > .14$ ] or the 10th-position probe tones in the near condition [MMN range,  $t(15) < 1, p > .34$ ; P3a range,  $t(15) = 1.1, p > .14$ ]. The presence of the P3a elicited by the probe tones in the 10th-position far-condition trains indicates that the probe tone was salient, as compared with the other X tones within the same stream. This same probe tone was not detectable as different from the other tones in its stream; it did not elicit P3a or MMN when it occurred within the near-condition trains or at the beginning (4th) position of the far-condition trains. The absence of MMN in the fourth position of the far-condition trains shows that the buildup phase was not yet completed at that point in time. By the 10th position, the buildup phase was surpassed, and MMN was elicited by the probe tone. The results of Experiment 2 thus replicate and extend the findings of Experiment 1, indicating that focused attention is not always required for the formation of auditory streams.

### EXPERIMENT 3

The conclusions of Experiment 2 were based on the result that MMN was not elicited by the 4th-position probe tone of the trains but was elicited by the 10th-position probe tone of the trains. MMN, we conclude, was not elicited by 4th-position deviant intensity tones because the buildup phase had not yet finished, so the regularity of intensity in that stream could not yet be passively detected. However, if three repetitions of the X tones were not enough to establish the regularity from which MMN could be elicited, MMN would also not have been elicited. Although previous studies have shown that three repetitions of a standard tone are sufficient to elicit MMN for other features, such as duration (Winkler et al., 1996) and frequency (Cowan et al., 1993), no study has specifically tested intensity. Therefore, Experiment 3 was conducted to test whether MMN could be elicited by an intensity deviant after three repetitions of a standard tone without intervening (O) tones when no buildup of stream segregation would be required before the regularity eliciting MMN could be passively detected.

### METHOD

#### Participants

Twelve young adults (19–24 years of age; mean age, 21 years; 5 females) who reported having normal hearing were paid for their participation in the experiment. None had participated in Experiment 1 or 2. Written informed consent was obtained after the procedures had been explained to them.

### Stimuli and Procedure

**Task-irrelevant sound sequences.** The stimuli were presented in 1.68-sec tone sequences (trains) consisting of eight X tones and no intervening O tones. Tone duration was 30 msec (5-msec rise/fall times). The ISI (offset to onset) was 180 msec, making the onset-to-onset pace of the X tones 210 msec (tone duration plus ISI), the same pace as that of the X tones occurring within the train sequences in Experiment 2. The trains were separated by 4.07 sec of silence. Two different types of trains were randomly presented in blocks of 64 trains. One type of train was the main experimental sequence, and the other type was the control needed to delineate MMN (see the Control Comparison for the MMN section below). Six hundred and forty trains were presented (320 per train type) in 10 blocks.

**Procedure.** The procedure was exactly the same as those in Experiments 1 and 2. The stimuli were presented in a free-field listening environment with two loudspeakers placed behind the participants. These loudspeakers presented acoustic input that was irrelevant to the listeners' task and were to be ignored. A third loudspeaker was placed directly in front of the participants. The participants performed the noise change detection task with the input from the loudspeaker placed in front of them.

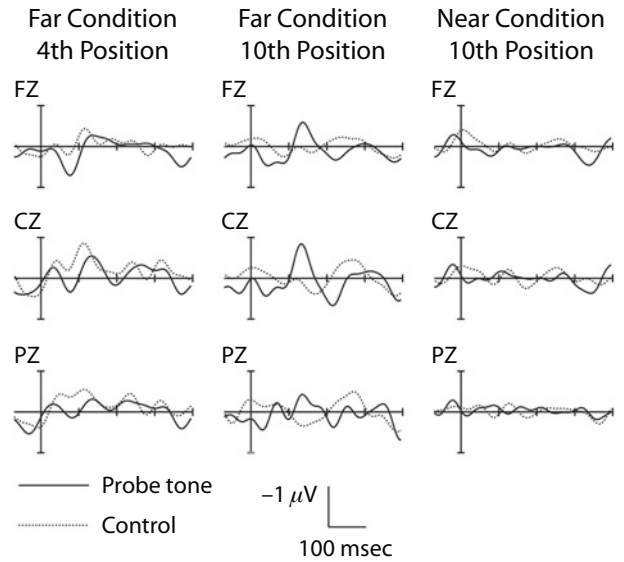
**Primary task.** The primary task in Experiment 3 was exactly the same as that used in Experiments 1 and 2.

### Data Recording and Analysis

EEG recordings and data analysis procedures were exactly the same as those used in Experiment 2.

**Control comparison for the MMN.** To compare exactly the same tone intensity when the tone was a deviant in the train with that when the tone was a standard in the train (occurring in the same position), control trains were presented, randomly intermixed with the main experimental trains. Five intensity levels in 3-dB steps were equiprobably distributed randomly across the tone positions of the control trains, with the +9-dB value (equal to the deviant in the experimental trains) occurring in the fourth position of these trains.

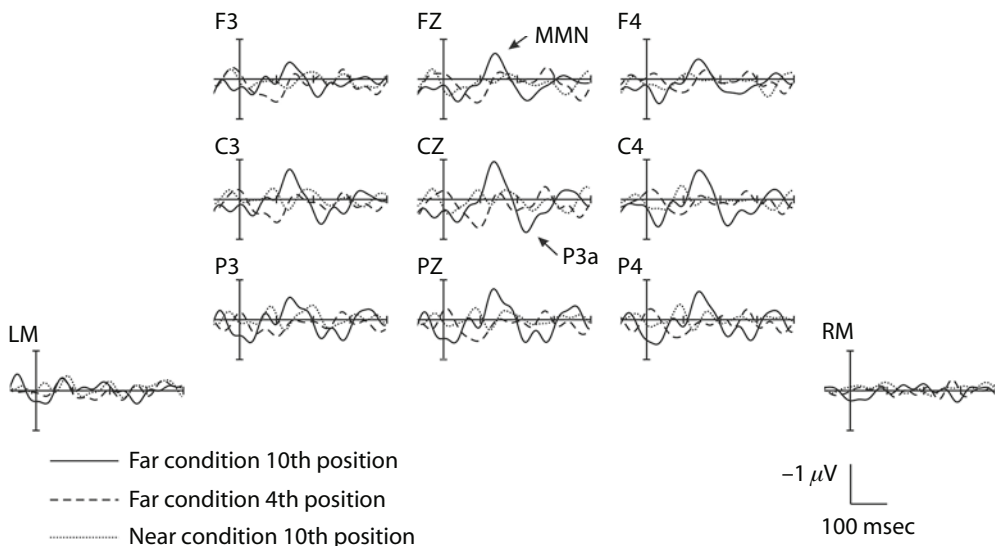
The average voltage of 116–156 msec, measured in each participant, was used to test for the presence of MMN. The 40-msec latency measurement window was obtained by visual inspection of the grand mean deviant-minus-control difference curve at Fz where



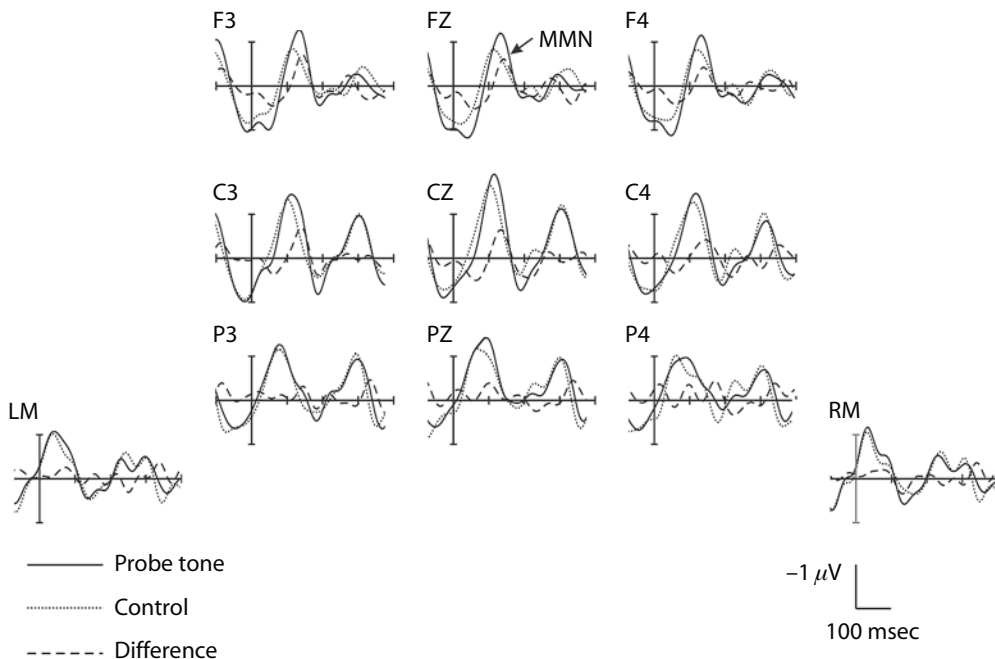
**Figure 6. Experiment 2: Event-related brain potentials elicited by the probe tones (solid lines) and control tones (dotted lines) for the three train types. Note that for display purposes only, a 70-msec prestimulus baseline is shown.**

MMN was observed (peak latency was 136 msec). Statistical testing was conducted using a one-sample, one-tailed *t* test at the Fz electrode to determine whether the mean probe-minus-control difference voltage was significantly greater than zero.

No P3a was visually observed in the deviant-minus-control waveform. Therefore, to statistically test for the presence of P3a, a 40-msec measurement latency window was chosen, which was the same time interval as that used to measure for the presence of P3a in Experiment 2, where P3a was observed (204–244 msec). Statistical testing was conducted using a one-sample, one-tailed *t* test at the Cz electrode



**Figure 7. Experiment 2: Probe-minus-control difference waveforms showing the responses elicited by the probe tones occurring in the 4th position (dashed line), overlaid with the responses elicited by the probe tones in the 10th position (solid line) for the far condition and those elicited by the 10th position (dotted line) for the near condition. Mismatch negativity (MMN) and P3a were elicited only by the probe tones in the 10th position of the far condition (shown with arrows). Note that for display purposes only, a 70-msec prestimulus baseline is shown.**



**Figure 8. Experiment 3: Event-related brain potentials elicited by the probe (solid lines) and control (dotted lines) tones are displayed together with the probe-minus-control difference waveforms (dashed line). Mismatch negativity (MMN) was elicited by the probe tones (shown with an arrow). Note that only 70 msec of the prestimulus baseline is displayed.**

to determine whether the mean probe-minus-control difference voltage was significantly greater than zero.

Responses in the noise change detection task were considered correct if they fell within 100–2,000 msec from the onset of the noise change. Paired *t* tests were used to compare RTs and HRs across conditions.

## Results and Discussion

### Performance on the Primary Task

The mean intensity change for the participants was 1.62 dB ( $SD = 0.20$ ). The mean HR was 80% ( $SD = 0.08\%$ ), with a mean RT of 787 msec ( $SD = 114$  msec).

### Event-Related Potentials

The grand mean ERPs elicited by the deviant and control stimuli and the corresponding deviant-minus-control difference waveforms are presented in Figure 8. The fourth-position probe tone elicited a negative displacement, in comparison with the control. This difference denotes the MMN [ $t(11) = 2.42, p < .02$ ]. No P3a was elicited by fourth-position deviants [ $t(11) = 1.76, p > .05$ ]. Thus, we can conclude that MMN would have been elicited by fourth-position probe tones in Experiment 2, had the two streams been segregated at the time the fourth-position probe tones occurred. These data suggest, therefore, that the buildup phase had not yet been completed by the fourth-position tone of the train in Experiment 2.

## EXPERIMENT 4

The purpose of Experiment 4 was to obtain a behavioral measure of the participants' perception of the sound

sequences, to determine whether they heard them as integrated or segregated.

## Method

### Participants

Fifteen young adults (22–34 years of age; mean age, 27 years; 4 males) who reported having normal hearing were paid for their participation in the experiment. Written informed consent had been obtained after the procedures were explained to them. None had participated in Experiments 1–3. This precluded the possibility that active listening (Experiment 4), if it were conducted first, or passive listening (Experiments 1–3), if it were conducted first, would influence responses to subsequent testing on the same sounds.

### Stimuli and Procedure

**Procedure.** The stimuli were presented in a free-field listening environment in a sound-attenuated IAC chamber. The participants sat in a comfortable chair with two loudspeakers placed at a distance of 1 m from the center of their heads. Speaker placement was to the left and right of the azimuth (approximately 45° angles).

**Sound sequences.** One hundred twenty trials (one trial is equal to one train) were presented (a randomly selected subset of those from Experiment 2 that included the full range of intensity and frequency variation). The control trains in Experiment 2 were not used. Two blocks of 60 trials were presented, randomized in order, both containing 30 near trains (1 ST) and 30 far trains (8 ST), 120 trials in total (60 near and 60 far). Although the deviant tones were task irrelevant in Experiment 4, they were included in the sequences so that perception would be based on the same sequences as those presented when ERPs were recorded. The deviant position (beginning or end position) was randomly determined for each trial (see the Method section in Experiment 2).

**Task.** The participants were instructed that they would hear a short sequence of sounds (~2.5 sec), followed by a silent pause, and that they were to tell whether they heard the sequence of sounds as one integrated stream of sounds or as two segregated streams of sounds.



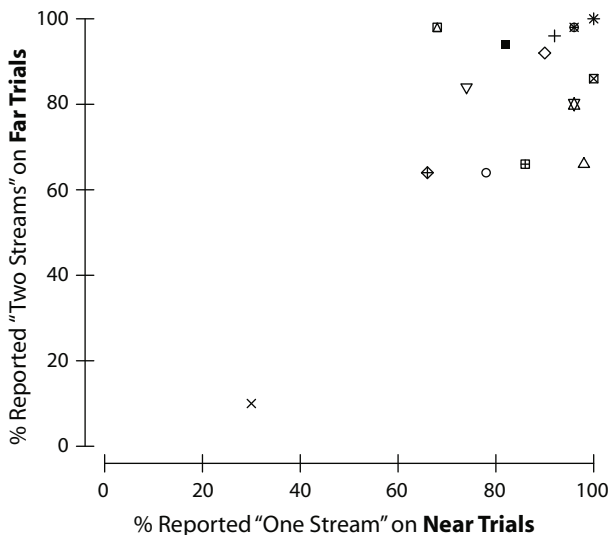
During the instruction period and prior to presentation of the sounds, a visual illustration of one versus two streams was presented. The first 10 trials were counted as practice and were not included in the statistical analysis. The participants judged whether each sound sequence was perceived as having one or two perceptual streams by giving a verbal response to the experimenter, who wrote their responses down (i.e., “one stream” or “two streams”). There was no time limit for providing a response. As soon as the response was given, the next sequence was started. Feedback was not given during the task, because there was no *incorrect* perception of the sounds.

### Data Analysis

For each participant, HRs, false alarm rates, and  $d'$  were calculated. The measure  $d'$  is derived from signal detection theory (Green & Swets, 1966) and measures the separation between two classes of stimuli on a hypothetical inner perceptual dimension upon which the participant's decision is based. In the present experiment, we used a *yes-no* model for independent observations to obtain  $d'$  (Macmillan & Creelman, 2005), a measure of the participant's ability to discriminate *near* from *far* sequences. The  $d'$  was calculated using the following procedure. Responses were counted as *hits* if the participants said “one stream” when the near trials were presented; responses were counted as *misses* if they said “two streams” when near trials were presented; responses were counted as *false alarms* if they said “two streams” when the near trials were presented and as *correct rejections* if they said “one stream” when the far conditions were presented. Thus, we calculated one  $d'$  score for each participant that would reveal how sensitive he or she was to the different organizations. A  $d'$  score of 0 would indicate an inability to discriminate the near from the far trials (i.e., the participant responds “one stream” and “two streams” equally on the far and near trials). A one-sample two-tailed Student's  $t$  test was used to determine whether the  $d'$  score was significantly greater than zero.

## Results and Discussion

The participants were sensitive to the near versus far manipulation. Table 1 displays a summary of the par-



**Figure 9. Experiment 4: Behavioral performance per trial type (near vs. far trials).** The abscissa represents the percentage of trials on which the participants responded “one stream” for the near trials, and the ordinate represents the percentage of trials on which the participants responded “two streams” for the far trials. Each data point represents 1 participant. There was a high level of sensitivity for judging the organization of the sequences as one or two streams.

**Table 1**  
**Summary of the Mean Responses (in Percentages)**  
**for the Near and Far Trials in Experiment 4**

Stimulus Type	Participant Response	
	“One Stream”	“Two Streams”
Near	83	17
Far	20	80

ticipants' responses. On average, the participants judged the near trials as integrated 82% of the time ( $SD = 19\%$ ) and the far trials as segregated 76% of the time ( $SD = 21\%$ ). The mean  $d'$  score was 2.3, which was significantly greater than zero [ $t(14) = 5.77, p < .01$ ], with no bias for responding toward one organization or the other ( $c = -0.06$ ). This shows that the participants distinguished between the two organizations. When they said “one stream,” they heard the sounds as integrated, and when they said “two streams,” they heard the sounds as segregated. Figure 9 displays the data for each participant.

## GENERAL DISCUSSION

In this study, four experiments were conducted to investigate the role of attention in the formation of auditory streams. The main goal was to determine whether attention was required during the initial buildup phase of auditory stream segregation. Our results provide evidence that focused attention is not always required for the initial formation of auditory streams.

A difficulty in addressing this issue has been the control of attention. The results of studies in which experimental paradigms were used in which the participants did not perform a task with test sounds could be explained in terms of covert attention's being directed momentarily toward task-irrelevant sounds (particularly at the beginning of the sound sequence), even though the participants in such experiments were instructed not to do so. Therefore, despite previous evidence supporting the view that auditory stream segregation is largely an attention-independent process, the question of whether listeners covertly attend to the to-be-ignored auditory information enough to initiate the stream segregation process was still open.

The covert attention hypothesis suggests that it is possible that some attention to the irrelevant and to-be-ignored sound information may be enough to initiate the stream segregation process, even if attention is not needed to maintain the organization once segregation has been initiated by attention. Thus, it could be reasoned that in experimental paradigms in which long continuous sequences of sound are presented, attention will be needed only for a short time at or near the beginning of the continuous sequence for the formation of the auditory streams.

For that reason, in the present study, we presented short trains of sounds with sufficient silence between them to reset the segregation process prior to the onset of every train. The absence of MMN at the beginning of the far-condition trains in Experiment 2 and the presence of MMN at the end of those trains provide compelling evidence that the resetting of the stream segregation process by silence

precipitated the buildup phase for each far-condition train. The results of Experiment 3 confirmed that if the buildup phase was surpassed, MMN could be elicited by fourth-position probe tones within trains. Therefore, if the participants focused only at the start of the whole sequence presented (i.e., train–silence–train–silence and so on), it would not be enough attention to initiate the buildup process for every train in the whole sequence. The participants would have had to attend to each of the irrelevant trains in the sequence—in particular, to the start of all of the trains—for a short time in order for the buildup process to be surpassed and the segregation to occur. Attention to the middle of the train would not allow enough time for the buildup process to be surpassed before the next silent segment came on. Attention would have to be focused continuously on the to-be-ignored sounds at the start of every train in the long sequence of trains. It is unlikely that the participants could attend to all of the trains while simultaneously performing the noise change detection task. Therefore, covert attention likely cannot explain the present results, because covertly attending the test sounds once in a while would not have been sufficient to initiate the buildup process.

To control the participants' focus of attention, they performed a difficult noise change detection task simultaneously with presentation of the to-be-ignored train stimuli. It should be noted that the HR during the training phase for the noise change detection task, when no competing train stimuli were presented, was higher than that during simultaneous presentation of the noise and the train stimuli. This lower HR was likely due to masking effects from overlapping tone frequencies that interfered with the change detection task. Winkler et al. (2005) used a similar noise change detection task with concurrent pure tone stimuli presented in the background. They found a decrease in HR that could be attributed to masking, by calculating HR as a function of the time between the noise change and the temporally closest tone. The HR decreased when noise-change-to-tone intervals were within 200 msec, but there was no significant effect on HR when they were outside the effective range of masking (greater than 200 msec). In the present study, due to the very fast tone presentation rate, every noise change fell within the temporal range of masking. Therefore, it is likely that the decrease in HR was primarily due to auditory masking. Since we cannot separate the masking effect in the present study, it should be acknowledged that the lower HR may also reflect shifts of attention to the to-be-ignored train stimuli. Although we do not think that the present results were due to attentional shifts, this possibility needs to be recognized and explored in future studies.

In addition to the main finding that attention is not always required for the buildup phase of stream segregation, the present results also demonstrate that the acoustic characteristics of the sound input govern whether or not segregation occurs for irrelevant sounds. The characteristics of the sound input in the near condition resulted in the formation of a single, integrated, frequency-varying stream, whereas the characteristics of the sound input in the far condition resulted in two separate streams. Fur-

thermore, the presence of the MMN generated by 10th-position probe tones and its absence for 4th-position probe tones in the far condition provide evidence to support the view that attention is not necessary to induce the stream segregation process. The acoustic characteristics of the input can initiate the process, determining whether sound input is represented as one or two sound streams in auditory memory.

An alternative explanation is that overlap of excitation along the basilar membrane was greater for the near than for the far trains. Thus, the probe tone may have been easier to detect when the frequency separation was greater (far trains), because there was less interference from the other tones. On this view, it does not matter whether or not the sounds were segregated; the MMN effects could be explained by overlap at the level of the cochlea. However, this explanation cannot account for the discrepancy between no MMN elicitation by probe tones in the 4th position and MMN elicitation by probe tones in the 10th position of the far trains in Experiment 2. The mixture of frequencies exciting the basilar membrane was the same throughout these trains. The difference between elicitation of the MMN and no elicitation of the MMN was determined by whether or not the sounds were segregated at the point at which the probe tone occurred (no MMN means that the sounds were *not* segregated, and the presence of MMN means the sounds *were* segregated). The differential elicitation of MMN by position within the train for the far condition strengthens the argument that the results were due to stream segregation involving central auditory processes, and not to processes related solely to peripheral mechanisms.

A more complicated issue is the definitive role of attention in stream formation. Even though we find evidence that attention is not required for the formation of auditory streams, this does not mean that there are no circumstances in which attention is needed to strengthen segregation or to resolve an ambiguous auditory scene. Cusack et al. (2004) found evidence that attention is needed for stream segregation and, therefore, postulated the *hierarchical decomposition* model. In this model, attention may be needed to *fine tune* segmentation of a complex auditory scene. Thus, consideration should be given to the fact that attention may be required for stream segregation under certain circumstances, but not under others.

In another example of attentional effects on stream segregation, Sussman et al. (2005) presented three frequency ranges of pure tones and manipulated the direction of attention. The question raised in the study was whether the background sounds would be organized into streams when attention was focused on a subset of the sounds. They found that when all the sounds were ignored and attention was focused on easy or difficult visual tasks, MMN was elicited by deviants in all frequency ranges. This suggested that the irrelevant sounds were segregated into three streams without focused attention on them. However, when attention was focused onto one of the three frequency ranges to perform a difficult auditory task, no indication of streaming was evident for the unattended background sounds. The same deviants that elicited MMN in the ignore-all-

sounds conditions elicited no MMNs when attention was focused on one of the streams. These results show an effect of attention on ignored sounds that appears to preempt the segregation process for unattended frequencies. However, it should be noted that it was not evident whether attention preempted the deviance detection process or the stream segregation process. The absence of MMN could not distinguish between these two alternatives, because no MMN would be elicited either way. Nonetheless, the focus of attention modified the neurophysiological response to the sounds. Further studies are underway.

Shedding light on the neural basis of auditory stream segregation is necessary to further our understanding of the role that attention plays in modifying associated processes. The neural mechanisms responsible for the buildup phase of auditory stream segregation have only recently been studied and are not well understood. An elegant study by Micheyl, Tian, Carlyon, and Rauschecker (2005) has addressed this difficult issue by comparing human behavioral responses with monkey single-unit recordings in response to the same sounds. They found good correspondence, measured over a 10-sec time, between the probability that a human would say that they heard "two streams" and the neural responses in the primary auditory cortex of awake monkeys listening to the same  $\Delta f$  conditions. The rapid adaptation of the single-unit responses is fully consistent with forward masking (Fishman et al., 2001) but does not necessarily distinguish the underlying mechanism responsible for the formation of streams. Nevertheless, the data do suggest that adaptation may play a role in the buildup response and, as such, indicate that neural mechanisms for the perception of the buildup phase are intrinsic to the auditory system.

The present results are consistent with the view that intrinsic neural mechanisms are responsible for the buildup phase, indicating that it can operate in the absence of focused attention. This does not rule out a role for attention in the formation of auditory streams under certain circumstances, as has been indicated by other studies. A recent study by Snyder, Alain, and Picton (2006) provides converging neurophysiological evidence that the buildup phase can occur without attention but also shows an attention effect on the buildup phase. Snyder et al. found larger amplitude responses associated with the buildup phase in the FCz and T8 electrodes for attend than for ignore conditions and concluded that there is a strong influence of attention on the stream formation process. Although their results suggest that attention modulates the buildup phase, the study leaves the question of the mechanism responsible for the buildup phase still largely unanswered, since it is difficult to dissociate the neural activity associated with the overall gain of the evoked responses resulting from attention to the signal (Hillyard, Hink, Schwent, & Picton, 1973; Hillyard, Vogel, & Luck, 1998) and that related specifically to the buildup process. The need for further studies is evident by the range of results found when neural activity has been measured at different levels of the auditory system.

In the present study, attention may not have been needed to initiate the buildup phase because the cues for streaming were relatively strong (e.g., an 8-ST frequency dif-

ference between sounds) and there were few competing sources. In other situations in which the cues for streaming are less strong or are ambiguous (e.g., input occurring in a noisy environment, with more overlapping properties of the sound input), attention may be needed to strengthen or maintain segregation. In Carlyon et al.'s (2001) study, it is possible that the stimuli fell into this ambiguous domain in which attention is necessary for segregation. However, Carlyon et al.'s results cannot fully address whether the buildup phase was affected by attention, because attention switching has been shown to reset the stream segregation process (Cusack et al., 2004). If the act of switching attention to test sounds resets the stream segregation process, the results cannot confirm whether the sounds were segregated when attention was focused away from them. A situation in which streaming can occur automatically under certain circumstances and can be influenced by attention in others may indicate the flexibility of the human auditory system for facilitating behavioral performance, even if at the "expense" of processing depth for unattended information. Further studies are needed to specify the role of attention in forming auditory streams.

Functionally, a purpose for the automatic organization of auditory input would be to free up attentional resources, which have limited capacity. If attention is not needed for basic segregation mechanisms, attentional resources would be free to focus on one or another sound stream occurring within the environment. Attention could thus be used to extract meaning from the attended sound stream, as when one listens to a speaker's message. Thus, automatic segregation processes would be advantageous for preserving attentional capacity. This model implies a balance between processes resulting from attended and unattended inputs.

#### AUTHOR NOTE

This research was supported by the National Institutes of Health (R01 DC004263) and the Hungarian Science Research Fund (OTKA T034112). We thank Wenjung Wang and Kinga Gyimesi for their assistance with data collection. M.O. is now affiliated with the Department of Population and Family Health, Columbia University. Correspondence concerning this article should be addressed to E. S. Sussman, Department of Neuroscience, Albert Einstein College of Medicine, Kennedy Center, Room 925, 1410 Pelham Parkway South, Bronx, NY 10461 (e-mail: esussman@aecom.yu.edu).

#### REFERENCES

- ALAIN, C., & WOODS, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, *34*, 534-546.
- BEE, M. A., & KLUMP, G. M. (2004). Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *Journal of Neurophysiology*, *92*, 1088-1104.
- BOTTE, M.-C., DRAKE, C., BROCHARD, R., & MCADAMS, S. (1997). Perceptual attenuation of nonfocused auditory streams. *Perception & Psychophysics*, *59*, 419-425.
- BREGMAN, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception & Performance*, *4*, 380-387.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- 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 & Performance*, **25**, 1742-1759.
- CARLYON, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, **8**, 465-471.
- 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 & Performance*, **27**, 115-127.
- COWAN, N., WINKLER, I., TEDER, W., & NÄÄTÄNEN, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 909-921.
- 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 & Performance*, **30**, 643-656.
- DUNCAN, J., MARTENS, S., & WARD, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, **387**, 808-810.
- ESCERA, C., ALHO, K., SCHRÖGER, E., & WINKLER, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology & Neuro-Otology*, **5**, 151-166.
- FAY, R. R. (2000). Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *Journal of Associative Research in Otolaryngology*, **1**, 120-128.
- 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. *Hearing Research*, **151**, 167-187.
- FRIEDMAN, D., CYCOWICZ, Y. M., & GAETA, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, **25**, 355-373.
- GIARD, M. H., PERRIN, F., PERNIER, J., & BOUCHET, P. (1990). Brain generators implicated in processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, **27**, 627-640.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HILLYARD, S. A., HINK, R. F., SCHWENT, V. L., & PICTON, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, **182**, 177-180.
- HILLYARD, S. A., VOGEL, E. K., & LUCK, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London: Series B*, **353**, 1257-1270.
- HORVÁTH, J., CZIGLER, I., SUSSMAN, E. [S.], & WINKLER, I. (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Cognitive Brain Research*, **12**, 131-144.
- HULSE, S. H., MACDOUGALL-SHACKLETON, S. A., & WISNIEWSKI, A. B. (1997). Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, **111**, 3-13.
- IZUMI, A. (2002). Auditory stream segregation in Japanese monkeys. *Cognition*, **82**, B113-B122.
- JONES, D. M., ALFORD, D., BRIDGES, A., TREMBLAY, S., & MACKEN, B. (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, & Cognition*, **25**, 464-473.
- JONES, D. M., ALFORD, D., MACKEN, W. J., BANBURY, S. P., & TREMBLAY, S. (2000). Interference from degraded auditory stimuli: Linear effects of changing-state in the irrelevant sequence. *Journal of the Acoustical Society of America*, **108**, 1082-1088.
- JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, **83**, 323-355.
- KANWAL, J. S., MEDVEDEV, A. V., & MICHEYL, C. (2003). Neurodynamics for auditory stream segregation: Tracking sounds in the mustached bat's natural environment. *Network*, **14**, 413-435.
- KNIGHT, R. T., & SCABINI, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology*, **15**, 3-13.
- LAVIE, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, **9**, 75-82.
- LOVELESS, N. E. (1986). Potentials evoked by temporal deviance. *Biological Psychology*, **22**, 149-167.
- 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 & Performance*, **29**, 43-51.
- MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- MICHEYL, C., TIAN, B., CARLYON, R. P., & RAUSCHHECKER, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, **48**, 139-148.
- MOORE, B. C. J., & GOCKEL, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica united with Acustica*, **88**, 320-333.
- NÄÄTÄNEN, R., PAAVILAINEN, P., TIITINEN, H., JIANG, D., & ALHO, K. (1993). Attention and mismatch negativity. *Psychophysiology*, **30**, 436-450.
- NÄÄTÄNEN, R., SIMPSON, R., & LOVELESS, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology*, **14**, 53-98.
- NÄÄTÄNEN, R., TERVANIEMI, M., SUSSMAN, E., PAAVILAINEN, P., & WINKLER, I. (2001). Pre-attentive cognitive processing ("primitive intelligence") in the auditory cortex as revealed by the mismatch negativity (MMN). *Trends in Neurosciences*, **24**, 283-288.
- NÄÄTÄNEN, R., & WINKLER, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, **125**, 826-859.
- PICTON, T. W., ALAIN, C., OTTEN, L., RITTER, W., & ACHIM, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, **5**, 111-139.
- RITTER, W., SUSSMAN, E., & MOLHOLM, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, **11**, 61-63.
- SCHERG, M., VAJSAR, J., & PICTON, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, **1**, 336-355.
- SCHRÖGER, E. (1997). On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology*, **34**, 245-257.
- SCHUL, J., & SHERIDAN, R. A. (2006). Auditory stream segregation in an insect. *Neuroscience*, **138**, 1-4.
- 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.
- SUSSMAN, E. S., BREGMAN, A. S., WANG, W. J., & KAHN, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds in multistream auditory environments. *Cognitive, Affective, & Behavioral Neuroscience*, **5**, 93-110.
- SUSSMAN, E. [S.], ČEPONIENĖ, R., SHESTAKOVA, A., NÄÄTÄNEN, R., & WINKLER, I. (2001). Auditory stream segregation processes operate similarly in school-aged children and adults. *Hearing Research*, **153**, 108-114.
- SUSSMAN, E. [S.], RITTER, W., & VAUGHAN, H. G., JR. (1998a). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, **789**, 130-138.
- SUSSMAN, E. [S.], RITTER, W., & VAUGHAN, H. G., JR. (1998b). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport*, **9**, 4167-4170.
- SUSSMAN, E. [S.], 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. [S.], WINKLER, I., HUOTILAINEN, M., RITTER, W., & NÄÄTÄNEN, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, **13**, 393-405.
- SUSSMAN, E. [S.], WINKLER, I., & WANG, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, **40**, 430-435.
- TREJO, L. J., RYAN-JONES, D. L., & KRAMER, A. F. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology*, **32**, 319-328.
- TREMBLAY, S., & JONES, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and



- rate of transition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 659-671.
- VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology.
- WINKLER, I., CZIGLER, I., SUSSMAN, E. [S.], HORVÁTH, J., & BALÁZS, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, **17**, 320-339.
- WINKLER, I., HORVÁTH, J., TEDER-SÄLEJÄRVI, W. A., NÄÄTÄNEN, R., & SUSSMAN, E. (2003). Neural model of sound sources in auditory cortex. *NeuroReport*, **14**, 2053-2056.
- WINKLER, I., KARMOS, G., & NÄÄTÄNEN, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, **742**, 239-252.
- WINKLER, I., KUSHNERENKO, E., HORVÁTH, J., ČEPONIENĖ, R., FELLMAN, V., HUOTILAINEN, M., ET AL. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, **100**, 11812-11815.
- WINKLER, I., SUSSMAN, E., TERVANIEMI, M., HORVÁTH, J., RITTER, W., & NÄÄTÄNEN, R. (2003). Preattentive auditory context effects. *Cognitive, Affective, & Behavioral Neuroscience*, **3**, 57-77.
- WÖLDORFF, M. G., & HILLYARD, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography & Clinical Neurophysiology*, **79**, 170-191.
- WÖLDORFF, M. G., HILLYARD, S. A., GALLEN, C. C., HAMPSON, S. R., & BLOOM, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, **35**, 283-292.
- YABE, H., WINKLER, I., CZIGLER, I., KOYAMA, S., KAKIGI, R., SUTOH, T., ET AL. (2001). Organizing sound sequences in the human brain: The interplay of auditory streaming and temporal integration. *Brain Research*, **897**, 222-227.

## NOTES

1. The length of the buildup period, when the participants shifted from judging the ABA sequence as one stream to two streams, was dependent upon the level of  $\Delta f$ , in that the larger the  $\Delta f$ , the shorter the buildup phase (Carlyon et al., 2001; cf. Bregman, 1978).

2. Although performing a primary visual task (for an *ignore* condition) does not conclusively rule out the possibility that participants could have attended the auditory stimuli, it should be noted that in the above-mentioned studies, there were no electrophysiological indications that attention was directed toward the to-be-ignored sounds, since no attention-related ERP components (e.g., N2b-P3b) were elicited by them. N2b can be elicited if sounds are detected as being deviant in a sequence even without an overt response, provided that attention is directed toward the sounds (Loveless, 1986; Näätänen, Simpson, & Loveless, 1982).

(Manuscript received October 1, 2004;  
revision accepted for publication March 3, 2006.)