Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments

E. S. SUSSMAN
Albert Einstein College of Medicine, New York, New York

A. S. BREGMAN
McGill University, Montreal, Quebec, Canada

W. J. WANG
*Albert Einstein College of Medicine, New York, New York
and City University of New York, New York*

and

F. J. KHAN
Albert Einstein College of Medicine, New York, New York

In three experiments, we addressed the issue of attention effects on unattended sound processing when one auditory stream is selected from three potential streams, creating a simple model of the cocktail party situation. We recorded event-related brain potentials (ERPs) to determine the way in which unattended, task-irrelevant sounds were stored in auditory memory (i.e., as one integrated stream or as two distinct streams). Subjects were instructed to ignore all the sounds and attend to a visual task or to selectively attend to a subset of the sounds and perform a task with the sounds (Experiments 1 and 2). A third (behavioral) experiment was conducted to test whether global pattern violations (used in Experiments 1 and 2) were perceptible when the sounds were segregated. We found that the mismatch negativity ERP component, an index of auditory change detection, was evoked by infrequent pattern violations occurring in the unattended sounds when *all* the sounds were ignored, but not when attention was focused on a subset of the sounds. The results demonstrate that multiple unattended sound streams can segregate by frequency range but that selectively attending to a subset of the sounds can modify the extent to which the unattended sounds are processed. These results are consistent with models in animal and human studies showing that attentional control can limit the processing of unattended input in favor of attended sensory inputs, thereby facilitating the ability to achieve behavioral goals.

In natural situations, the auditory system is confronted with overlapping acoustic input from several simultaneously active sources at any given time. An important aspect of auditory processing is to sort the overlapping inputs into groups, each from a single source, a process called *auditory scene analysis* (ASA; Bregman, 1990). The ASA process facilitates our ability to choose one stream of information within a background of sounds emanating from many different sources, such as when listening to a stream of speech in a crowded room. The *cocktail party* provides an example of a situation in which it is common to attend to one stream of speech (one conversation in the room: the *foreground*) while ignoring the rest of the auditory information (the other conversations and noise in the room: the *background*).

The present study focuses on a central scientific question in ASA theory: Which of the following views is more accurate?

1. Streams are structured without focused attention, by preattentive processes operating over the whole set of sounds that are present. If this is true, more than one stream can be formed at the same time, despite the fact that attention may choose one or another for further processing.

2. Attention is required to form a stream. Under this hypothesis, only a single stream is ever present as a structured perceptual entity—the one created by attention out of a subset of the sounds that are present. The unattended tones are simply an unstructured background. As attention shifts, different streams are formed, enduring only as long as attention remains focused on the stream.

To address the question, we tested the extent to which unattended sounds are processed when one auditory stream is selected from a set of potential ones. Specifically, by using tones spanning three frequency ranges, we were

able to focus on the issue of whether the sounds of an unattended acoustic background, those that are not the immediate focus of our attention, segregate from one another on the basis of frequency separation to form their own individual sound streams. In the cocktail party situation, we may hear the pitch of the clinking of glasses in the background as a sound stream separated from the pitch of the voice stream of the person speaking in front of us. The question was whether the sounds emanating from the clinking glasses will be organized into a coherent stream of sound when attention is directed toward the person speaking in front of us.

The question of the degree of processing of unattended sounds has been focused on by researchers for several decades. According to the information-processing framework, stimulus information enters the sensory system and undergoes a series of transformations before it is perceived (Broadbent, 1958). An enduring view of the role of attention in perception is to limit the processing of sensory information to items that are relevant to current behavior (Broadbent, 1958; Desimone & Duncan, 1995; Neisser, 1967; Treisman, 1969). The function of attention is then to select objects of interest for further processing, so that attention acts as a filter for sensory input. The unattended sensory information is suppressed from further processing at some point prior to response selection, either before the sensory input is fully analyzed (Broadbent, 1958; Treisman, 1969) or after it is (Deutsch & Deutsch, 1963; Duncan & Humphreys, 1989; Norman, 1967). This model of information processing is supported by data obtained from a number of different disciplines (for a review, see Hillyard, Vogel, & Luck, 1998). Single-unit recording studies of the monkey brain demonstrate a bias for attended sensory information that begins before the stimulus is identified (Bushnell, Goldberg, & Robinson, 1981; Desimone, 1998; Moran & Desimone, 1985). Although there is much evidence to support the basic information-processing model of perception, the debate about whether (or where) attention acts as a filter is still unresolved and likely reflects differences both in the paradigms used to assess unattended sound input and in the manner in which attention affects sound processes under different conditions. Moreover, very little is understood about the neural mechanisms associated with stimulus-driven sound organization in the perception of auditory objects. In the present article, we address this issue of stimulus-driven sound organization by using a methodology that does not require a subject to focus on sound input to ascertain how sounds are organized and maintained in auditory memory.

The basic question of whether unattended sounds are organized without focused attention remains unanswered, because there is both behavioral and electrophysiological evidence supporting opposing views.

Recent behavioral data suggest that unattended acoustic input is *not* segregated by source (Botte, Drake, Brochard, & McAdams, 1997; Brochard, Drake, Botte, & McAdams, 1999; Carlyon, Cusack, Foxton, & Robertson,

2001). Carlyon et al. studied the effect of attention on the build-up phase of auditory stream segregation.[1] They presented a pure tone sequence containing high- and low-frequency sounds to one ear (using four levels of Δf) and required subjects to perform a task involving the noise burst input presented to the other ear. After 10 sec, the subjects were told to switch their attention to the tone sequence and to make a judgment about the segregation of the sounds—that is, to tell whether they heard one or two streams of sound. The logic behind their design was that if the build-up process occurred without attention (during the 10 sec when the subjects were performing the noise task in the contralateral ear), when the subjects redirected their attention to the test sounds, they should immediately judge the sounds as being segregated into two sound streams (since the build-up phase had already occurred in the ear receiving the tone sequence). On the other hand, if attention was needed for the build-up phase, the subjects should initially judge the test sounds as a single integrated sound stream, since the build-up phase would not yet have been initiated. Carlyon et al. found that the sounds were heard as a single stream (integrated) in the initial moments at which the subjects switched their attention to the test sequence. The subjects reported hearing two streams only after the sequence had proceeded for at least 5 sec. The authors therefore concluded that attention is essential for auditory stream segregation.

This conclusion is not definitive. The basic finding was that when the listener's attention was switched to the tone task after the white noise had been attended to, there was no evidence of any accumulation of a tendency toward segregation. Carlyon et al. (2001) attributed this to the fact that attention had not been paid to the tones prior to the switch. However, it is possible that the accumulation of a segregation tendency *had* been accumulating while the subject was attending to the noise. The switching of attention to the right ear may have reset the system that accumulates information about the presence of streams. In a recent study by Cusack, Deeks, Aikman, and Carlyon (2004), this hypothesis was tested (Experiment 4). They presented tones to one ear and noise to another ear, requesting subjects to switch attention from the tones (10 sec) to the noise (5 sec) and back to the tones and to judge, each time, whether they heard one or two streams of tones. This was compared with a condition in which the subjects performed the same task with the tones (judged whether they heard one or two streams) but, instead of noise, there was a silent gap (5 sec) between the 10-sec blocks of tonal stimuli. The silent gap has the effect of *resetting* the streaming process (Bregman, 1978). Thus, if switching attention to the noise and back to the tones has an effect similar to that of resetting the stream segregation process by a silent gap, the results between these two conditions should not differ significantly. They found 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 seem to indicate that attention switching can have the same effect as

resetting the system by silence. This hypothesis would have to rest on the assumption that integration is the *default* process—that input is segregated only after enough evidence has accumulated to confirm it. However, the exact nature of this attention effect still remains to be determined.

Brochard et al. (1999) came to a conclusion similar to that in Carlyon et al. (2001), in a study in which listeners responded to temporal irregularities that could be detected only when the listeners perceptually organized the sounds into separate frequency streams. The listeners were presented with sequences that contained one to four potential subsequences based on the frequency differences between the tones. Each frequency stream had a different tempo (300, 400, 500, or 700 msec, onset-to-onset interval). The listeners were instructed to focus on one frequency stream and to detect temporal irregularities within it. The logic was that if the listener perceived the global sequence (i.e., the sounds did not segregate), the tempo of the focused stream would not be detectable, because all of the sounds would be included in the tempo (the perception of rhythm is altered when sounds are perceived as segregated; van Noorden, 1975), and thus, the listeners would not be able to focus on any particular potential subsequence. However, if the sounds segregated by frequency stream (as the result of an increase in the frequency distance between them), the tempo of the target stream could be determined and irregularities could be detected. The authors concluded that because performance did not change as a function of context changes across conditions (changes in the surrounding frequency–tempo combinations), the unattended sounds were not segregated by frequency. However, it is also possible to suggest that once the sounds were segregated on the basis of stimulus-driven factors (i.e., segregation induced by increasing the distance between sounds), attention was easily used to perform the task independently of the context of the unattended sounds. There was no direct evidence that the background unattended sounds had an effect on target detection—only that frequency separation improved detection of irregularity in the attended stream.

In these studies, the state, or organization, of the unattended sounds was assessed in terms of the behavioral measures obtained for the attended subset of sounds. Probing the state of unattended sounds without task distortion is not easy to accomplish with behavioral measures. In Carlyon et al.'s (2001) study, as was mentioned earlier, it cannot be ruled out that the act of attention switching itself may have mediated the perception of stream segregation. And in Brochard et al.'s (1999) study, the consistent performance obtained on the selected channel may have been a function of stimulus-driven factors for stream segregation (i.e., frequency separation), irrespective of attentional focus.

To avoid extraneous effects of attention as much as possible, the use of electrophysiological measures of the brain's response to unattended sounds is highly suitable for investigating this issue. Obtaining an electrophysio-logical index of sound organization when attention is directed to either a visual or an auditory task does not require the listener to make an overt response concerning the unattended sound channel (Ritter, Sussman, & Molholm, 2000; Sussman, Ritter, & Vaughan, 1998, 1999; Winkler et al., 2003).

Because event-related brain potentials (ERPs) provide high temporal resolution, in the order of milliseconds, they have the advantage of providing information about the timing of cognitive processes associated with stimulus events and have been used extensively to index cognitive processes initiated by sound change (Friedman, Cycowicz, & Gaeta, 2001; Picton, Alain, Otten, Ritter, & Achim, 2000; Rugg & Coles, 1995). The mismatch negativity (MMN) component of ERPs is an especially useful tool for assessing the processing of unattended sounds, because it does not require the experimental subject to respond to the sounds or to indicate perception of the sounds for it to be elicited (for a recent review, see Picton et al., 2000). MMN is elicited when a sound is detected as violating some regularity of the preceding auditory sequence (approximately 150 msec from onset of the deviation), whether or not attention is focused on the sounds. It is elicited by infrequent changes in auditory features (such as frequency, intensity, tone duration, or spatial location), as well as by violations of tone patterns (e.g., Sussman, Ritter, & Vaughan, 1999; for a review, see Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). The MMN reflects the outcome of the auditory change detection process, which uses the neural representation of the acoustic regularities extracted from the sound sequence that are stored in auditory memory. Moreover, because detection of change is dependent on how the sound sequence is structured in memory, MMN can be used to investigate how the auditory system organizes sequential patterns extracted from the sound environment (Ritter et al., 2000; Sussman, Čeponienė, Shestakova, Näätänen, & Winkler, 2001; Sussman et al., 1998, 1999; Winkler et al., 2003).

Recent ERP data have suggested that the unattended acoustic input *is* segregated by source (Ritter et al., 2000; Sussman et al., 2001; Sussman et al., 1998, 1999; Winkler et al., 2003). Sussman et al. (1999) used the MMN measure of preattentive[2] sound discrimination to probe stream segregation processes when subjects ignored the sound stimulation and read a book. In this study, alternating high (H) and low (L) tones (three different frequencies in each of these two ranges, labeled H_1, H_2, and H_3 and L_1, L_2, and L_3, respectively) were presented at a rapid pace that was known to induce a streaming effect (Bregman, 1990). The streaming effect is the perception that there are two separate frequency streams occurring concurrently (H_1_H_2_H_3_ and L_1_L_2_L_3_) even though the actual presentation of the stimuli is alternating (e.g., $H_1 L_1 H_2 L_2 H_3 L_3$). This difference between the perception of a set of tones as two concurrent streams of sound and the actual alternating presentation of the high and low tones was exploited to determine the processing stage at

which segregation occurred. The paradigm was set up so that when the H tones segregated from the L tones, the sequential three-tone patterns embedded within the high and low streams would emerge. Once the three-tone sequential patterns emerged (as a result of the stimulus-driven segregation process), the infrequent within-stream pattern violations could be detected, which would elicit the MMN component. In another condition, we showed that if the structure of the sound sequence represented in memory was the six-tone pattern of H and L tones (i.e., the six tones were regarded as belonging to a single sound stream), the same tones that were detected as violations of the three-tone patterns did not elicit MMNs. In other words, an MMN component would be elicited *only* if segregation occurred. Therefore, if the representation of the *alternating* high and low tones were maintained as a single integrated stream in auditory memory in the early processing stages, no MMNs would be elicited. Because the pattern violations within *both* the high and the low tone streams elicited MMNs, it was concluded that the two potential streams of sound were segregated preattentively, without attention being focused on the auditory stimuli. Moreover, this basic finding that stream segregation occurs preattentively and concurrently in two distinct frequency ranges (i.e., from two potential sound streams) has been replicated in different paradigms in which different segregation cues were used, such as frequency proximity (Sussman et al., 2001; Winkler et al., 2003) and a combination of spatial location, frequency, and intensity (Ritter et al., 2000).

These studies indicate that the basic stream segregation process is automatic; the acoustic input is initially sorted into sound streams on the basis of their stimulus characteristics and is maintained in a transient auditory memory. However, even though the initial organizational process may not require attention, we have demonstrated that attention can modify the neural representations of the regularities underlying the MMN process. Because the basis for MMN elicitation is the regularity maintained in memory, modification of the underlying regularity affects what is detected as deviant within the sound sequence and, consequently, affects MMN elicitation (Sussman et al., 1998; Sussman, Winkler, Huoutilainen, Ritter, & Näätänen, 2002).

In Sussman et al. (1998), selective attention was used to segregate the high- from the low-frequency sounds in a sequence that did not segregate automatically (as indexed by MMN) when a subject had no task to perform with the sounds. That is, stream segregation was manipulated by changing the subject's task with the same set of sounds, rather than by manipulating the stimulus characteristics of the sensory input (cf. Sussman et al., 1999, in which changing the stimulus-driven characteristics [slow vs. fast pace] induced a streaming effect). When the subject selectively attended to the high sounds, the unattended low sounds organized into a coherent stream of the three-tone rising pitch pattern (e.g., L_1–L_2–L_3 . . .). The infrequent within-stream pattern violations (e.g.,

L_3–L_2–L_1) elicited MMNs separately in both the attended (high) and the unattended (low) streams. The active segregation of the high sounds, induced by attention, reorganized the memory representation of the sound input from one integrated stream of sounds jumping up and down in pitch in the *ignore* condition (*ignore all sounds*) to two segregated sound streams, one made up of the high sounds and one made up of the low sounds, in the *attend* condition (*attend only to the high tones*). This was shown by the fact that MMN was elicited only in the attend condition; the deviants could be detected within the repeating three-tone patterns that emerged when the high and the low sounds segregated. These results showed that (1) attention could be used to organize the input and (2) attention modified the neural trace representing the organization of the sounds (i.e., one vs. two frequency streams) in the memory used by the deviance detection process.

It is only recently that we have shown in our laboratory that attention can modify processes prior to the MMN-generating process (e.g., Sussman et al., 1998; Sussman, Winkler, Huoutilainen, et al., 2002) but not the MMN process itself (e.g., Ritter, Sussman, Deacon, Cowan, & Vaughan, 1999; Sussman, Winkler, & Wang, 2003). When attention initiates a change in the organization and storage of sound in memory, this is reflected by the absence or presence of specific ERP components (e.g., Sussman et al., 1998). However, simply attending to the deviants (e.g., Sussman et al., 2004) or actively discriminating between them (e.g., Sussman, Winkler, Kreuzer, et al., 2002) without active organization of the sequential sound sequence (i.e., without altering the structure of the sounds held in memory) does not alter MMN elicitation. In other words, the MMN process is highly dependent on context (Sussman, 2004). Thus, the MMN process is subject to attentional control (or top-down processes), in the sense that it uses sound information that can be modified by listener intent. The MMN system uses the structural information maintained in memory and detects deviations within the established regularities. This type of attentional effect on sound processing—in which neurophysiological changes to sound input are induced by manipulating the subject's auditory task (but not the input)—is consistent with the results of a recent study by Fritz, Shamma, Elhilali, and Klein (2003), who demonstrated, using an animal model, that changes in the response properties of neurons in the primary auditory cortex could be induced by changing the task demands on the animal. The results in Fritz et al. suggest a feedback system similar to what we have proposed (e.g., Sussman et al., 1998; Sussman, Winkler, Huouti-lainen, et al., 2002), in which changes in neural representations are initiated by attention and serve to facilitate the behavioral goals of the subject (see also Sussman et al., 2003). None of these results precludes the possibility that the deviance detection process itself does not require attention to operate.

To summarize, these results indicate that MMN can detect changes in the sound organization—the outcome

of attentional processes—and also that sound organization processes occur prior to the deviance detection process. The deviance detection process is operating on the already grouped sound information. Consequently, it is critical to understand the stages of the auditory processes that attention can act on. In the present study, the processing of sound input in a situation in which all the sound input is ignored is compared with one in which only a subset of the sound is ignored. If MMN output reflects changes in attentional focus on the identical set of sounds, we can consider that attention has an effect on the processes that detect deviance.

Sussman et al. (1998) and Sussman, Winkler, Huotilainen, et al. (2002) showed that attention could modify sensory input at the level of sound organization and that the reorganized information would be used by later cognitive processes (such as the deviance detection process). In Sussman et al. (1998), however, there were only two possible frequency streams presented in the sound sequence (high and low). Thus, actively selecting the high sounds left only the low tones as a residual stream in which deviants could be inattentively detected. The results of Sussman et al. (1998) therefore cannot address the issue of how (or whether) unattended sounds can be further organized—for example, whether the unattended sounds could be organized into two frequency-based streams. Therefore, in the present study, three potential frequency streams were used, so that actively attending to the high sounds would leave a mixture of middle- and low-frequency sounds (i.e., two possible unattended frequency streams). In this way, we could determine whether, when the listener selected one sound source (the high sounds) from a mixture of tones consisting of three different frequency ranges, the unattended input would be organized into frequency-based streams or remain as an undifferentiated sequence. Two ERP experiments and one behavioral experiment were conducted to test this. We measured the brain's response to the same set of sounds in a condition in which attention was directed to a subset of the sounds (e.g., the high tones) and compared it with two conditions in which the experimental subjects were instructed to ignore the sounds altogether and direct attention to a visual task. In one condition, the visual task was reading, a commonly used *ignore* paradigm for MMN elicitation. In another experiment, a visual discrimination task was used that matched the processing demands of the attended condition, to control for effects related solely to processing load. We were interested in (1) whether a set of tones spanning three frequency ranges would be segregated into three distinct streams, each formed in one of these ranges, and (2) whether attention to a subset of the sounds would have a different effect on the unattended sounds from that found when attention was focused on a visual task. Would the unattended sounds still organize by frequency range?

Even though earlier experiments had shown that detection of the infrequent tone pattern violations can be made to depend on the segregation of the sounds, we also wanted to test whether the pattern reversals that were thought of as *within-stream* by the experimenters could actually be perceptually detected as violations of the *global* sound sequence. If so, the explanation that stream segregation was responsible for the deviance detection (and hence, for the MMN) would be unnecessary. Therefore, a separate behavioral experiment was conducted to determine whether infrequent pattern reversals would disrupt the global pattern of sounds.

EXPERIMENT 1

Method

Subjects. Nineteen adults (5 males) from 18 to 42 years of age ($M = 30$ years, $SD = 7$ years) were paid for their participation. The subjects were treated in accordance with the ethical guidelines for human subjects research at Albert Einstein College of Medicine, where the study was conducted. All the subjects reported normal hearing and had no history of neurological disorders. Informed consent was obtained after the experimental procedures had been explained to them. The data from 3 subjects were excluded because of excessive eye artifacts in electroencephalogram (EEG) recordings.

Stimuli. The stimuli were pure tones, 50 msec in duration (which included 5-msec of onset and offset linear ramps), equated for intensity at 80 dB SPL (calibrated with a Bruel & Kjaer 2209 sound level meter), the level at which they were presented. They were presented bilaterally through insert earphones (Eartone 3A) at a constant stimulus onset asynchrony (SOA; onset to onset) of 90 msec. Three sets of tones, each occupying a distinct frequency range, were used. Within each range, the tones that were nearest in frequency were separated by one semitone (frequency ratio of about 1.06). The high-frequency range (H) included two tones (2489 and 2637 Hz). The middle-frequency range (M) included three tones (880, 932, and 988 Hz), and the low-frequency range (L) included three tones (311, 330, and 349 Hz).[3] A log frequency scale was used to select the within-range intervals so that the discriminability of the steps would be approximately equal for each of the different frequency ranges. The stimuli were presented in an alternating fashion (L_1, M_1, H_1, L_2, M_2, H_1, L_3, M_3, H_1, and so on). Within the high tones, H_1 (79%) and H_2 (21%) occurred in a random sequence, within which H_2 occasionally occurred twice in succession (7.5%). Within the middle- and low-frequency ranges, a repeating pattern of the three tones occurred (e.g., M_1, M_2, M_3 and L_1, L_2, L_3, respectively).[4] Reversals of the component three-tone patterns (e.g., L_3, L_2, L_1) occurred randomly 7.5% of the time within each frequency range, with the restriction that the infrequent pattern reversals never occurred consecutively either within or across streams. Thus, the target (repeated higher pitched tones) and the pattern reversal *deviants* in the middle- and low-frequency ranges occurred equally often in each stream (7.5%). The stimulus sequence is depicted in Figure 1.

Procedures. The subjects were comfortably seated in a sound-attenuated booth. The stimuli were presented in two conditions (ignore and attend) that differed only in the instructions to the subject. Per condition, 25,200 tones were presented. In the ignore condition, four separately randomized sequences were presented (6,300 stimuli per run), and in the attend condition, they were divided into seven separately randomized sequences (3,600 stimuli per run), reducing the amount of time the subjects had to attend to the sequences and perform the task in one run. Short breaks were given after each run. In the ignore condition, the subjects were instructed to ignore the sounds and watch a silent, subtitled movie. This condition was conducted first for all the subjects. After the recording of the ignore condition, the subjects were given a rest break (to walk around and have a snack), followed by a practice session that preceded the recording of the attend condition. In the attend condition, the subjects were instructed to focus on the two high-frequency tones and ignore the rest of the sounds. The subjects responded to
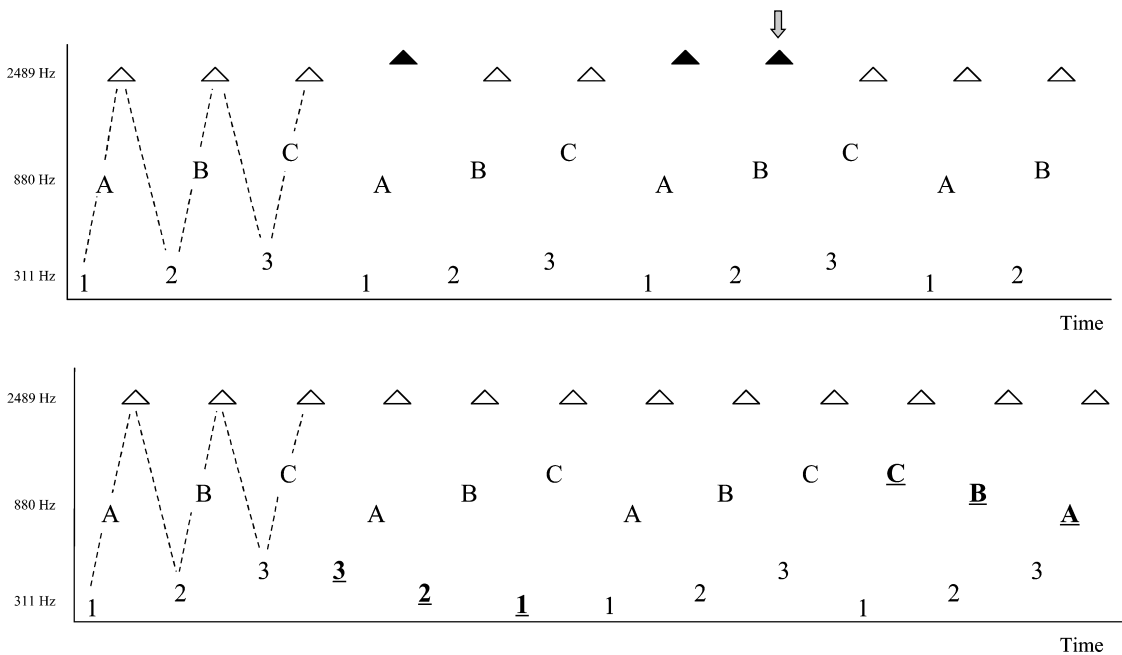
Figure 1. Stimulus paradigm for Experiment 1: schematic diagram showing the presentation of the alternating high- (top row), middle- (middle row), and low- (bottom row) frequency range tones. The top panel shows the frequent high tones (white triangles) and the infrequent higher pitched tones (black triangles), along with the standard three-tone rising patterns occurring in the middle range (denoted with different letters) and in the low one (denoted with different numbers). The target for the attention-based task is designated with an arrow (two successive higher pitched tones). The bottom panel shows the three-tone pattern violations (reversal of order) occurring within the middle- and low-frequency range tones (denoted in boldface and underlined).

the target, the *repetition of the higher pitched tone within the high stream*, by pressing a designated response key. In the practice session for this condition, the subjects were first presented with the high tones only (the high stream as it would occur if presented alone) and were shown a visual diagram of the stimuli, pointing out the difference between the higher pitched tones when they occurred singly and when they were repeated (the target). After the subjects understood and could perform the task with the high tones alone (with about 80% accuracy), they were presented with a 1-min sample of the full sound sequence and were told that they were to select out the high tones with which they had just been familiarized, perform the same task, and ignore all the other sounds. Practice then continued until the subjects could perform the task with at least 70% accuracy (on average, the subjects practiced for three 1-min samples before proceeding).

Data recording and analysis. The EEG was recorded from nine electrode sites (Fz, Cz, Pz, F3, F4, C3, C4, LM, and RM), which by and large, are the primary sites that will provide the scalp distribution needed for assessing MMN elicitation (our dependent measure). The reference electrode was placed on the tip of the nose. An electrooculogram (EOG) was measured using a bipolar electrode configuration between F7 and F8 for horizontal eye movements and between FP1 and an electrode placed below the left eye for vertical eye movements. A horizontal electrooculogram (HEOG) was recorded to monitor eye saccades during reading (when the subjects were watching the video) and to ensure that fixation was maintained (when the subjects were performing the auditory task). The EEG was digitized at a 250-Hz rate (0.05–100 Hz bandpass) and then filtered off line (1–20 Hz). An artifact reject criterion was set at $\pm 75 \, \mu V$, measured after the epochs were baseline corrected on the entire sweep.

ERPs were averaged separately for each stimulus type (standard and deviant), frequency range (high, middle, and low), and condition (ignore and attend). For the low- and middle-frequency ranges,

the standard ERP was an average of the first tone of the standard (rising) three-tone pattern (i.e., L_1 for the low-frequency range and M_1 for the middle-frequency range), and the deviant ERP was an average of the first tone of the deviant (falling) pattern reversals (Sussman et al., 1999). For the high-frequency range, the standard ERP was an average of the frequent high tones (H_1). The deviant (for the ignore condition) or the target (for the attend condition) was averaged from the first of the two successive higher pitched (H_2) tones. Epochs were 600 msec, starting from 100 msec before and ending 500 msec after the onset of the stimuli for the ignore condition, so the ERP response to the successive high-tone deviants (a high tone occurring once every 270 msec) could be seen (MMN was expected to be elicited separately by each one; Sussman, 2005). Because of the rapid stimulus rate (90 msec, onset to onset), ERP responses to successive tones could be seen within the epoch. The stimuli were pseudorandomized so that no two deviants from different frequency ranges overlapped or occurred in succession. Thus, the time-locked responses in the epochs represent the deviant or the standard within the same frequency range, followed by responses to successive standards across all the frequency ranges. For the attend condition, epochs were 1,000 msec, including a prestimulus interval of 100 msec. A longer epoch was used in the attend condition in order to make it possible to observe not only the ERP responses to both of the successive higher pitched tones, but also the expected ERP responses that were elicited by deviants when sounds were in the focus of attention (i.e., N2b/P3b). Averaged waveforms were then baseline corrected: The mean amplitude in the prestimulus period was the reference for measuring the amplitude of the ERP components.

The mean MMN amplitude was measured using a 40-msec window centered on the MMN peak latency obtained in the deviant minus standard difference curves of the grand mean averages, for each condition separately. The ranges were determined at the grand

mean frontal (Fz) amplitude, the electrode site at which MMN shows the best signal-to-noise ratio.

The intervals listed below include the grand mean peak latency of the MMNs. For the high tones in the ignore condition, MMNs were measured from the 164- to 204-msec time interval for the MMN elicited by the first of the repeated pitch changes (peak latency, 184 msec) and from 404 to 444 msec (peak latency, 424 msec) for the second of the pair, calculated from stimulus onset of the first of the two successive deviants. For the middle tones in the ignore condition, the time range was between 168 and 208 msec (peak latency, 188 msec), and for the low tones, it was between 152 and 192 msec (peak latency, 172 msec). For the high tones in the attend condition, successive MMNs elicited by the pitch change pair were measured using 168–208 msec (peak latency, 188 msec) and 384–424 msec (peak latency, 404 msec), calculated from stimulus onset of the epoch. The same time ranges as those in the ignore condition, in which MMNs were elicited by the middle and low sounds (peaks of 188 and 172 msec, respectively), were used to assess the presence of the MMN in the attend condition. All measurements were made separately for each subject and stimulus type. Because the predictions for the MMNs were made a priori, one-sample one-tailed Student's t tests were used to determine whether the mean amplitudes
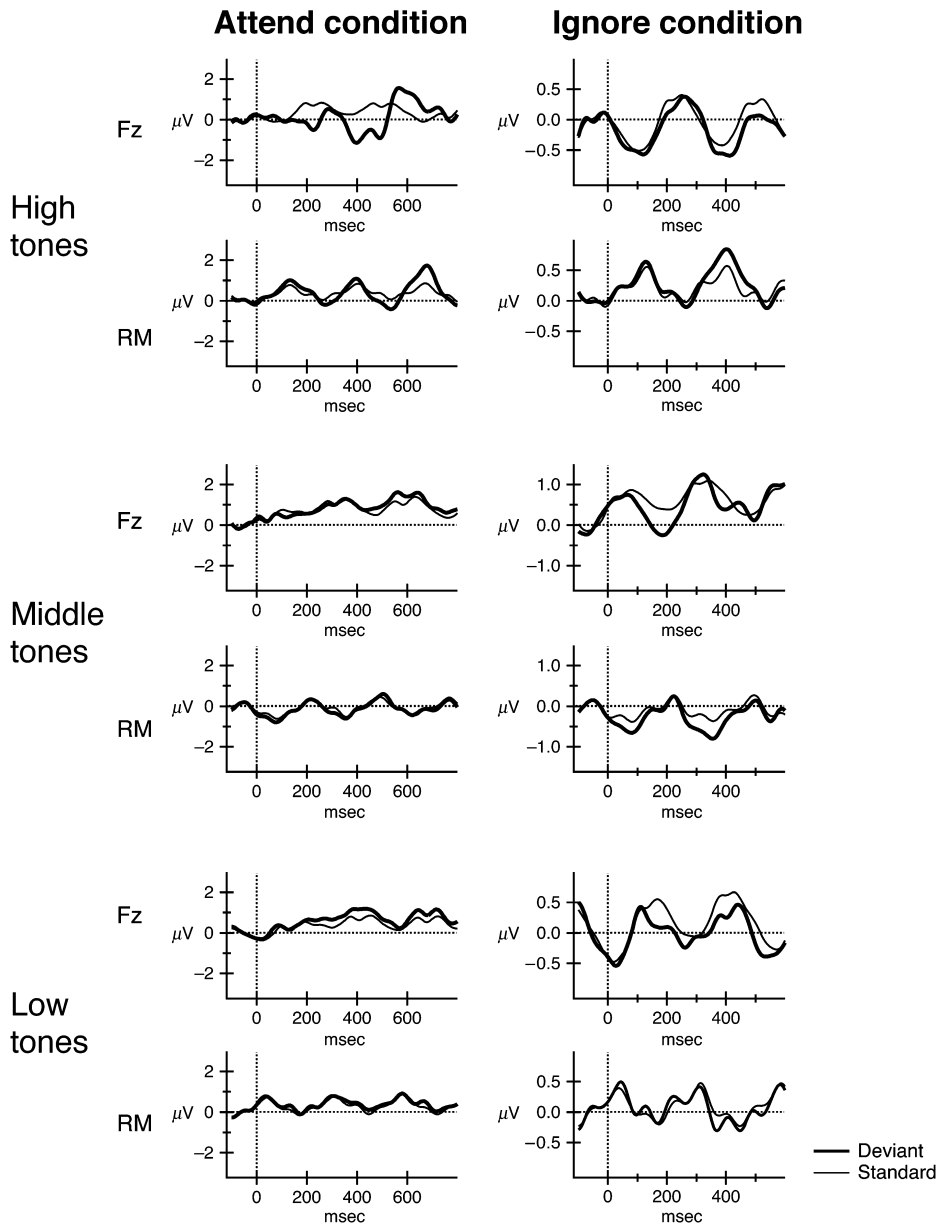

Figure 2. Event-related potential (ERP) data for the ignore and attend conditions in Experiment 1: ERPs elicited at the frontal electrode site (Fz) and the right mastoid (RM) by the standards (thin line) and deviants (thick line) in the attend (left column) and the ignore (right column) conditions for the high (top rows), middle (middle rows), and low (bottom rows) tones separately.

obtained in these time intervals were significantly greater than zero, separately for each condition and frequency range, conducted on the mean voltage of the individual subjects. Paired-sample two-tailed t tests were used to compare mean amplitudes of the MMNs. Attention effects were assessed in two-way repeated measures analyses of variance (ANOVAs) with factors of attentional state (attend or ignore) and frequency range (middle or low).

Behavioral responses were measured by calculating reaction times (RTs), hit rates (HRs), and false alarm rates (FARs) to the targets in the attend condition. Responses were considered correct if they fell between 200 and 1,300 msec from the onset of the second of the two higher pitched tones when they repeated within the high stream.

Results and Discussion

Behavioral results. On average, the subjects responded correctly on 79% of the targets ($SD = 16$). The FAR was less than 1% (0.39%). Mean RT to the target (successive higher pitched tones) was 409 msec ($SD = 78$).

ERP results. In the ignore condition, MMNs were elicited separately by occasional changes occurring randomly within each frequency range (see Figure 2 for ERPs and Figure 3 for corresponding difference waveforms). MMNs were elicited by each of the successive deviant pitched high tones, seen consecutively in the epoch, separated by ~270 msec, which is the onset to onset latency of the successive tones within the high stream [high range peak 1, $t(15) = 1.8$, $p < .05$; high range peak 2, $t(15) = 3.3$, $p < .01$]. The infrequent reversal of the three-tone patterns, which are expected to emerge only when the low and the middle frequency range sounds segregate from each other, gave rise to separate MMNs [middle range, $t(15) = 3.7$, $p < .01$; low range, $t(15) = 3.0$, $p < .01$; see Table 1 for a summary of mean MMN amplitudes]. The absence of attention-related ERP components means that the subjects were not attending to the sounds enough to detect deviants in any of the sound streams. These results demonstrate that the sounds were segregated by frequency range without focused attention to the sound and extend the findings of Sussman et al. (1999) to include preattentive segregation of three frequency streams.

In the attend condition, MMNs were elicited only in the attended stream by each of the successive high-pitched tones [high range peak 1, $t(15) = 3.2$, $p < .05$; high range peak 2, $t(15) = 5.6$, $p < .01$; see Figure 3]. The deviants in the high stream also elicited the N2b/P3b target detection components [$t(15) = 3.9$, $p < .01$, and $t(15) = 5.1$, $p < .01$, respectively]. No MMNs were elicited by the infrequent reversals of the three-tone repeating pattern by either the middle or the low tones [middle range, $t(15) < 1$, $p > .39$; low range, $t(15) < 1$, $p > .47$; see Table 1]. Furthermore, the tones in the middle- and low-frequency ranges elicited no such attention-related ERP components, demonstrating that the subjects were ignoring the sounds sufficiently and did not actively detect any deviants in the middle- and low-frequency ranges. This result suggests that there were attention effects on unattended sound processing.

An ANOVA with attentional state (attend vs. ignore) and frequency range (middle vs. low) as factors revealed a main effect of attention on the MMN [$F(1,15) = 21.7$, $p = .00$] but no main effect of frequency range [$F(1,15) < 1$, $p > .42$] and no interaction [$F(1,15) < 1$, $p > .47$]. This shows that MMNs were elicited when the sounds were ignored, but not when other sounds were being attended. Furthermore, it shows that there was no difference between the amplitude of the MMNs elicited by the low- and the middle-frequency tone pattern deviants when the sounds were ignored and no difference between the amplitudes of the ERPs elicited by the low- and the middle-frequency range tones when the high sounds were attended (no MMNs were elicited).

The MMNs elicited by the two successive high-pitched deviants in the attend condition were larger than those elicited in the ignore condition [peak 1, $t(15) = 2.5$, $p < .025$; peak 2, $t(15) = 3.9$, $p < .01$], which was likely due to overlap with the N2b elicited when the subjects attended to the sounds. The difference in the peak latencies of the MMNs elicited by the low (172 msec) and the middle (188 msec) tones in the ignore condition was not significant ($t < 1$), nor did attention modulate the latency of the MMN elicited by the high sounds in the attend versus ignore conditions, since there was no significant difference in the peak latencies of these MMNs ($F < 1$).

One interpretation of these results is that the stream segregation process was limited by focused attention to another sound stream, so that the unattended sounds did not segregate into low- and middle-range frequency streams. Because detection of the tone pattern violations is dependent on segregation of the sounds by frequency range, attention, which was needed to perform the task with a high degree of accuracy, may have disrupted the formation of streams of the unattended sounds. Attention to one sound stream may have actively suppressed computation of the unattended sound streams. This interpretation may be consistent with Carlyon et al.'s (2001) results, in that highly focused attention to one ear may have preempted the formation of sound streams in the other ear.

An alternative interpretation is that attention did not affect the stream segregation processes at all but, rather,

Table 1
Mean Amplitudes (in μV, with Standard Deviations) at Fz in the MMN Latency Ranges for Each Frequency (High, Middle, and Low) in the Ignore and Attend Conditions in Experiment 1

Ignore (Read a Book)			Attend High Tones		
Frequency (msec)	M	SD	Frequency (msec)	M	SD
		High Tones			
164–204	−0.20	0.46*	168–208	−0.88	1.09**
404–444	−0.30	0.36**	384–424	−1.34	0.95**
		Middle Tones			
168–208	−0.65	0.71**	168–208	−0.06	0.94
		Low Tones			
152–192	−0.39	0.52**	152–192	−0.01	0.48

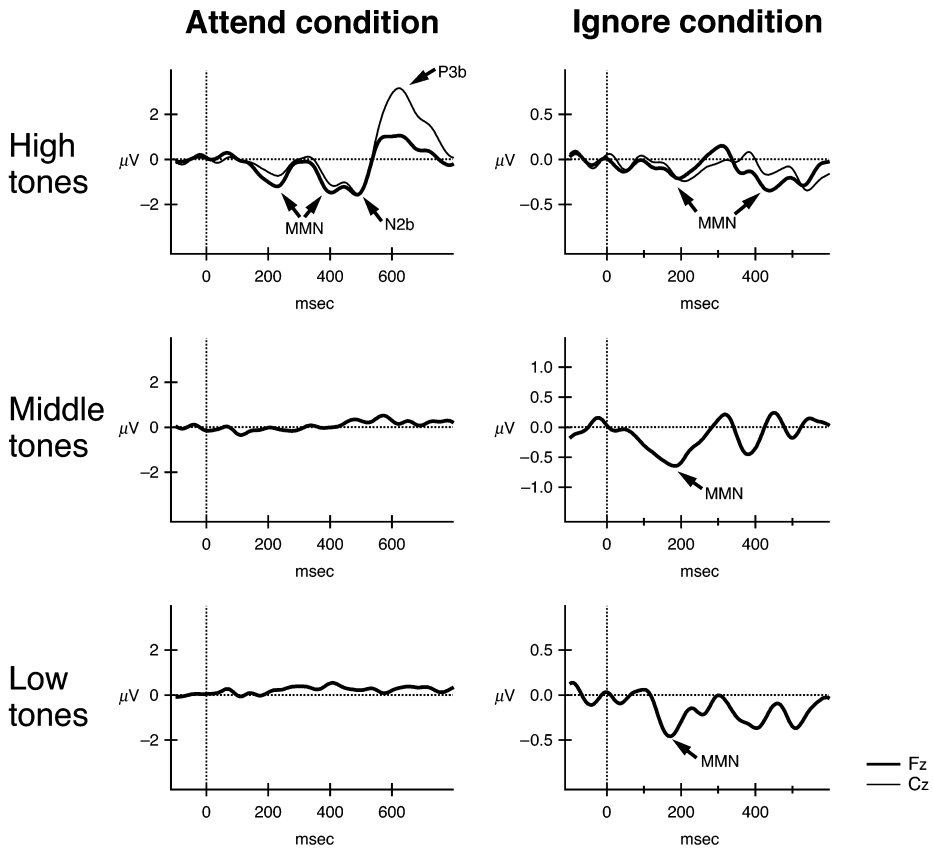Note—Results of the one-sample, one-tailed Student's t tests: *$p < .05$; **$p < .01$.

Figure 3. Difference waveforms for the attend and ignore conditions in Experiment 1: The mismatched negativity (MMN) can be observed as a negative waveform in the difference curves, obtained by subtracting the event-related potential (ERP) responses elicited by the standard from the ERPs elicited by the deviant. In the attend condition (left column), the MMN can be observed for the high tones (top panel). Following the MMN to the target (the second MMN seen in the epoch), the N2b and P3b components can be observed, with larger amplitude at the Cz electrode (overlaid, thin line). MMN, N2b, or P3b components were not observed in the ERPs elicited by the unattended middle tones (middle panel) or the low tones (bottom panel). In the ignore condition (right column), MMNs can be observed in the difference waveforms (right column) as a negative waveform at Fz for the high tones, middle tones, and low tones. No attention-related components (N2b/P3b) could be observed at Cz in the ignore condition. Arrows point to the MMN, N2b, and P3b components where elicited.

affected the MMN process. The unattended sounds may have segregated by frequency range, and attention may have suppressed or limited the MMN deviance detection process. This interpretation suggests that detecting pattern violations within the middle and low streams may have been limited by the highly focused attention needed to perform the task within the subset of high sounds.

If attentional load was a factor in modifying either the stream segregation or the MMN process, the ignore condition task in Experiment 1 (reading a book) would not provide a sufficient control of attention to determine this, because the visual task was not as demanding as the auditory task (detecting repetitions of infrequent high-pitched sounds within a larger set of alternating pitches). Furthermore, the subjects had no task to perform in the ignore condition, and thus, no measure of attentional focus was obtained, such as that obtained in the attend

condition. Thus, when the subjects ignored the sounds in Experiment 1, a useful comparison for measuring the focus of attention away from the sounds (attentional demand) was not obtained.

Although the results in the ignore condition in Experiment 1 do show that focused attention on the sounds was not needed for the stream segregation process to operate (evidenced by the fact that no ERP components associated with any form of attention were elicited in the ignore condition), the task of reading a book and ignoring the sounds in Experiment 1 did not test whether *highly* focused attention to a nonauditory task with an attentional demand similar to that in the attend condition could also preempt the elicitation of MMN. Therefore, another experiment was conducted, in which the subjects performed a visual discrimination task (detecting repetitions of infrequent visual stimuli) that was equated in

style to the auditory task and with a difficulty level comparable to that task, to test whether highly focused attention away from the sounds would modify their processing. We should expect that if highly focused attention was a critical factor influencing unattended sound processing, MMN elicitation should be similarly affected by highly focused attention onto a visual discrimination task, as occurred when highly focused attention was used to perform the auditory task (i.e., no MMNs elicited by some or all of the auditory deviants). If, however, deviants in all the frequency ranges elicited MMNs, this result would support the notion that highly focused attention, in and of itself, does not modify the stream segregation process and may then point toward an explanation involving competition for auditory resources when attention selects a subset of sounds (from many).

EXPERIMENT 2

The purpose of Experiment 2 was to determine whether highly focused attention, in and of itself, might have contributed to the attention effect observed in Experiment 1. To test this, we presented the same auditory input as that in Experiment 1, except that we controlled the focus of the subjects' attention by (1) requiring the subjects to perform a task with the visual input and (2) equating task difficulty with that of the auditory task performed in the attend condition in Experiment 1 (~80% HR).

Method

Subjects. Ten adults (5 males) ranging from 22 to 44 years in age ($M = 28$ years, $SD = 7$ years) were paid for their participation. The subjects were treated in accordance with the ethical guidelines for human subjects research at Albert Einstein College of Medicine, where the study was conducted. All the subjects reported normal hearing and had no history of neurological disorders. Informed consent was obtained after the experimental procedures were explained to them. None of the subjects had participated in Experiment 1.

Auditory Stimuli. The auditory stimulus parameters and the protocol were exactly the same as those in Experiment 1 (see Figure 1).

Visual Stimuli. Visual images were created using the Neuroscan ImageWord program to obtain two sizes of 10 letters (A, C, E, F, G, J, K, M, P, and R). The font used for the 10 larger stimuli and for the 10 smaller stimuli was Times New Roman. The target event was the repetition of the same letter of the same size (e.g., AA or AA; see Figure 4). A total of 1,863 trials were presented in a pseudorandom order (restricted to prevent targets from being presented succes-

sively) divided equally into nine separate runs of 207 stimuli each. Targets occurred on 10% of the trials. A total of 90 targets and approximately 65 distractor events (two of the same letter of different sizes) were presented overall. A pilot test was conducted (without simultaneous sound input) with three individuals (who did not participate in the main study) to match task difficulty level with the auditory task in the attend condition in Experiment 1 (approximately 79% hit rate and <1% false alarm rate). The visual stimuli were presented on the screen of a computer monitor (via Neuroscan STIM software), individually for 150 msec, with an offset-to-onset interval of 1.4 sec. The subjects were instructed to fixate their gaze at a cross that was maintained in the center of the screen and to press the computer mouse key whenever they detected a letter of the same size that was repeated (see Figure 4).

Procedures. The testing environment was exactly the same as that in Experiment 1. A computer monitor was in front of the subject at a distance of 1.3 m. The large letters subtended 3.12° of visual angle, and the small letters subtended 2.77° of visual angle.

Prior to the EEG recording session, a brief practice session was given to acquaint the subjects with the visual task and to obtain a baseline measure of subject performance on the visual task without sound input. First, each of the 10 letters was displayed (large, then small) on the screen. Then, after the task instructions were explained, two runs of visual stimuli were presented, in which the experimenter also checked that the subjects could perform the task with at least 50% accuracy. These data then served as a reference to ensure that task performance was maintained similarly with and without sound input. Thus, if attention were diverted away from the main visual task (when sounds were simultaneously presented), we might expect a decrease in performance level on the visual task during the recording session.

During the EEG recording session, sounds were presented bilaterally, with the same sounds to both ears, via insert earphones, and the subjects were instructed to ignore the sounds and perform the same visual task as the one in the practice session in which no sounds were presented. Nine separately randomized sequences (3,600 auditory stimuli per run; 207 visual stimuli per run) were presented. Short breaks were given as needed and a larger break was provided about halfway through the session to allow the subjects to have a snack and to walk around.

EEG recording. EEG recording was exactly the same as that in Experiment 1, with the following exceptions. The EEG was digitized at a rate of 500 Hz (0.05–100 Hz bandpass) and then filtered off line (1–15 Hz). Epochs were 600 msec, starting from 100 msec before and ending 500 msec after the onset of the auditory stimuli.

Data analysis. Mean MMN amplitude was measured using a 40-msec window (time interval) centered on the MMN peak latency obtained in the deviant minus standard difference curves of the grand mean averages, for each condition separately, and was referred to the average amplitude in the prestimulus period. The time intervals were determined at the grand mean frontal (Fz) amplitude, the electrode site at which MMN shows the best signal-to-noise ratio. The following intervals included the grand mean peak latency
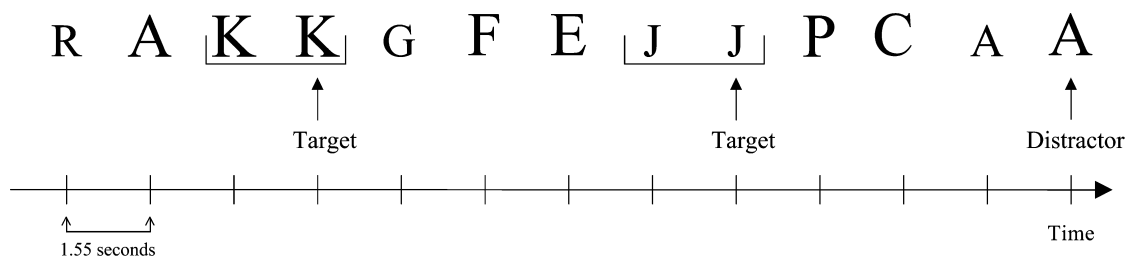

Figure 4. Visual discrimination paradigm for Experiment 2: schematic diagram showing an example of the target (repetition of the same-sized letter) and the distractor stimuli (repetition of different-sized letters). The abscissa denotes time. One visual stimulus was presented on the screen every 1.55 sec.

of the MMNs. For the high tones, MMNs were measured using the 154- to 194-msec time range for the MMN elicited by the first tone of the target pair (a repeated H_2 tone; peak latency, 174 msec) and, for the second of the pair, the range from 380 to 420 msec (peak latency, 400 msec), calculated from stimulus onset of the first of the successive deviants that could be seen within the epoch—for the middle tones, between 136 and 176 msec (peak latency, 156 msec), and for the low tones, between 206 and 246 msec (peak latency, 226 msec). Because the predictions for the MMNs were made a priori, one-sample one-tailed Student's t tests were used to determine whether the mean amplitudes obtained in these time ranges were significantly greater than zero, separately for each frequency range, conducted on the mean voltage of the individual subjects. MMN amplitudes were compared using ANOVAs with repeated measures. Greenhouse–Geisser corrections were reported. Paired two-sample t tests were used to compare behavioral responses for the visual task.

Behavioral responses to visual targets were measured by calculating RTs, HRs, and FARs. Responses were considered correct if they fell within 1,200 msec from the onset of the visual target stimulus.

Results and Discussion

Behavioral results. In the visual-alone condition (prerecording practice session), the subjects responded correctly on an average of 79% of the trials ($SD = 0.15$) with a 1.4% FAR. The majority of false alarms were responses to the distractor stimuli (1.3%). In the ignore condition (the EEG recording session), when the subjects ignored the auditory stimuli and performed the visual task, they responded correctly on 81% of the trials ($SD = 0.15$), with a 1.8% FAR. The majority of false
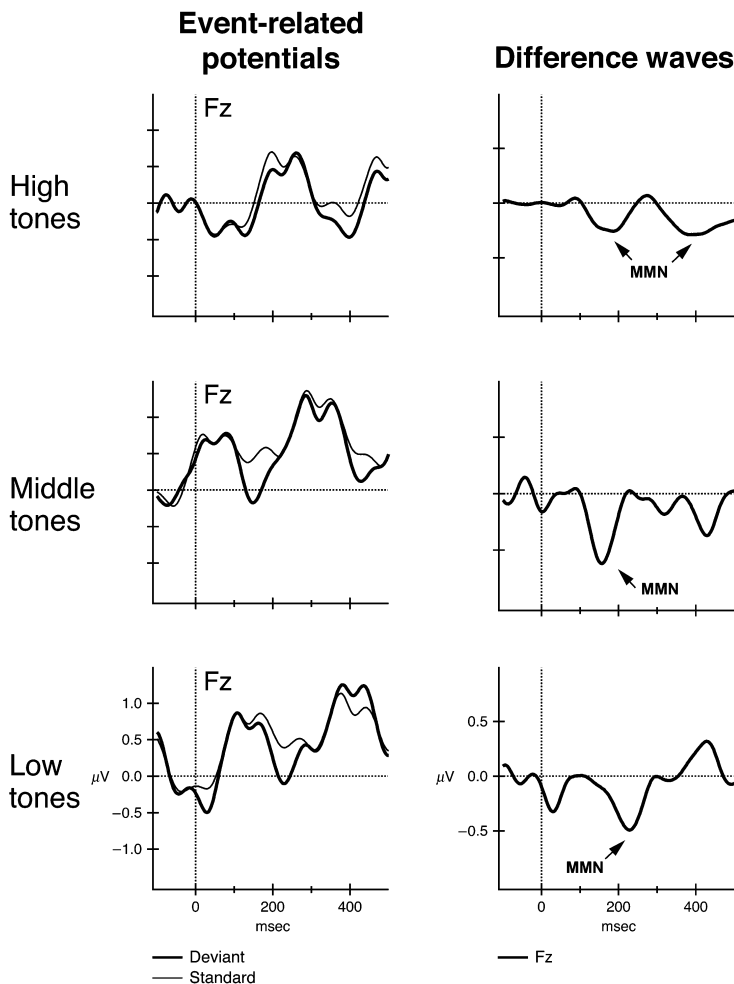

Figure 5. Event-related potential (ERP) and difference waveforms for the ignore condition in Experiment 2: ERPs (left column) elicited at the frontal electrode site (Fz) by the standards (thin line) and deviants (thick line) for the high (top row), middle (middle row), and low (bottom row) tones, displayed separately. Mismatched negatives (MMNs) can be observed in the difference waveforms, separately by frequency range (right column), as a negative waveform at Fz for the high tones, middle tones, and low tones. No attention-related components (N2b/P3b) can be observed. Arrows point to the MMNs.

alarms were responses to the distractor stimuli (1.7%). There was no significant difference between performance on the visual task when the sounds were presented simultaneously [$t(9) < 1, p > .70$]. RTs also did not differ significantly between these conditions [$t(9) < 1, p > .77$]—710 msec ($SD = 155$ msec) for the visual-alone condition and 705 msec ($SD = 142$ msec) for the ignore condition.

ERP results. MMNs were elicited separately by the infrequent sound changes occurring randomly within each frequency range (see Figure 5 for ERPs and corresponding difference waveforms). MMNs were elicited by each of the successive deviant pitched high tones, seen consecutively in the epoch [High Range Peak 1, $t(9) = 2, p < .05$; High Range Peak 2, $t(9) = 2.2, p < .05$]. The response to the infrequent reversal of the three-tone patterns, which emerge when the low and the middle frequency range sounds segregate from each other, gave rise to separate MMNs [middle range, $t(9) = 4.5, p < .01$; low range, $t(9) = 3.2, p < .01$; see Table 2 for summary of mean MMN amplitudes]. There were no significant effects of frequency range on MMN amplitude [$F(3,27) = 1.7, p > .33$]. The absence of any attention-related ERP components (e.g., P3a, N2b, or P3b) shows that the subjects were not attending to the sounds enough to detect deviants in any of the sound streams. These results are very similar to those for the ignore condition in Experiment 1 (see Tables 1 and 2), showing no attention effect on unattended sound processing. Even with highly focused attention on a primary visual task, stream segregation and MMN processes were not modified for the unattended sound input.

The same sound input was used in Experiments 1 and 2; thus, we show that MMNs were elicited by unattended deviants when a video was watched (the ignore condition in Experiment 1) or when a demanding visual discrimination task was performed (the ignore condition of Experiment 2), but not by the same deviants when another auditory stream was attended to (the attend condition in Experiment 1). This indicates that attentional demand by itself is not the main factor contributing to the attention effect. Segregation was induced by the frequency separation of the sound input, so that deviants were detected and MMNs elicited, separately in each of the streams, even when attentional demand was increased in

the ignore condition to match that in the attend condition. Thus, taken together, the results of Experiments 1 and 2 point toward a different explanation for the attention effect than simply attentional demand (see the General Discussion section). This may be due in part to some independence of visual and auditory attentional resources (Duncan, Martens, & Ward, 1997). Further testing is needed to assess this interpretation for sound organization processes.

EXPERIMENT 3

This was a behavioral experiment whose purpose was to validate the assumptions made in Experiments 1 and 2 and in other MMN research in which auditory streams have been used. There have been a number of experiments in which global sequences segregate into subsequences (auditory streams) that occupy different frequency ranges (Ritter et al., 2000; Sussman et al., 2001; Sussman et al., 1998, 1999; Winkler et al., 2003; see Bregman, 1990, for a review). When a change is introduced into one of these subsequences and an MMN response occurs, it has been assumed that elicitation of the MMN is related to the detection of a change *within that subsequence*. However, any changes in a subsequence will, of necessity, introduce a change in the overall (global) sequence. Could the MMN simply be a response to a change in the global sequence? This experiment was an attempt to rule out this interpretation by determining whether deviants eliciting MMN in Experiment 1 could have been part of a global perception of deviance of those same sounds. Many studies have shown that the auditory information encoded in the memory trace that generates MMN strongly corresponds with behavioral perception of the same sounds (e.g., Kisley et al., 2004; Stekelenburg, Vroomen, & de Gelder, 2004; Sussman et al., 2001; Tiitinen, May, Reinikainen, & Näätänen, 1994). To determine how the sounds were perceived, we obtained behavioral performance measures on the unattended sounds in the attend condition in Experiment 1 (i.e., the middle- and low-frequency sounds) to assess correspondence between perception of the sounds (no EEG recording) and neurophysiological evidence (i.e., MMN elicitation) concerning the same unattended sounds (the ignore conditions in Experiments 1 and 2). Thus, in Experiment 3, the middle- and low-frequency sounds were presented without the high tones to behaviorally test whether the change in the global pattern was perceptible even when the sounds were segregated. Support for the notion that MMN elicitation in the ignore conditions in Experiments 1 and 2 indexed stream segregation would be found if the global patterns were not perceptible when the sounds were perceptually segregated.

Method

Subjects. Twenty adults (14 females) ranging in age from 18 to 42 years ($M = 31$ years, $SD = 7$ years) with reported normal hearing were paid to participate in the study. Informed consent was obtained after the experimental procedures had been explained to them. One subject's data were excluded due to experimenter error.

Table 2
Mean Amplitudes in μV (and Standard Deviations) at Fz in the MMN Latency Ranges (in Milliseconds) for Each Frequency Range (High, Middle, Low) in the Ignore Condition (Visual Discrimination Task) in Experiment 2

Tones	Range (msec)	Amplitude	
		M	SD
High	142–182	−0.24	0.42*
	380–420	−0.29	0.40*
Middle	136–176	−0.57	0.40**
Low	152–192	−0.46	0.46**

Note—Results of the one-sample, one-tailed Student's t tests: *$p < .05$; **$p < .01$.

All the subjects in Experiment 1 participated in Experiment 3 at the Albert Einstein College of Medicine.

Stimuli. Sounds and sound sequences were created using the CoolEdit Pro software and were presented bilaterally through insert earphones (Eartone 3A), using the same testing environment as that in Experiments 1 and 2. Two sets of three different tones (represented by A, B, and C and 1, 2, and 3) were arranged in two different orders to create two different repeating global tone patterns. In Global Tone Pattern 1 (G1), the order of the six tones was A–1–B–2–C–3–A–1–B–2–C–3 . . . In Global Tone Pattern 2 (G2), the order of the six tones was A–2–B–3–C–1–A–2–B–3– C–1 . . . (the stimulus sequences are depicted in Figure 6). For either G1 or G2, when all the tones occupy a similar frequency range, the subject should perceptually integrate it as a single six-tone perceptual stream and should, therefore, be able to distinguish G1 from G2 by the difference in the order of the six tones. When the A, B, C tones

occupy a different frequency range from the 1, 2, 3 tones, the tones should segregate by frequency range into two streams. When this happens, each of the within-stream patterns, the higher (A, B, C, A, B, C, . . .) and the lower (1, 2, 3, 1, 2, 3, . . .), is the same in both the G1 and the G2 patterns; the only thing that differs in G1 and G2 is the starting point of the lower cycle (i.e., 1, 2, 3, 1, 2, 3 . . . and 2, 3, 1, 2, 3, 1 . . . are the same cycle that simply starts on a different tone). A 1-sec fade-in and 1-sec fade-out were used at the beginning and end of the tone cycles to avoid the possibility that the subjects could use the cues of the beginning or ending of the sequences to determine the difference between the lower streams of G1 and G2 (see Figure 4). Without these cues, any subject who segregates the upper and the lower streams should not be able to distinguish G1 from G2.

Two conditions were presented, near (N) and far (F), in which the value of the A, B, and C tones, separated by one-semitone steps,

A Far Condition

Near Condition

B

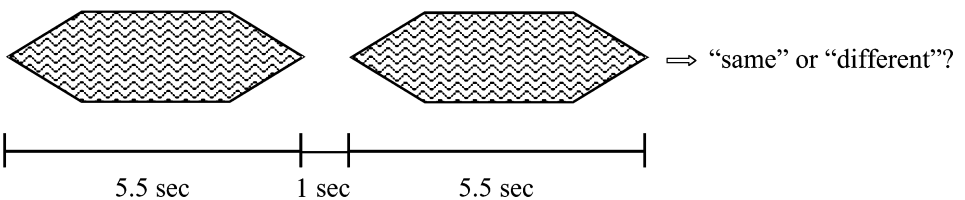⟹ "same" or "different"?

5.5 sec   1 sec   5.5 sec
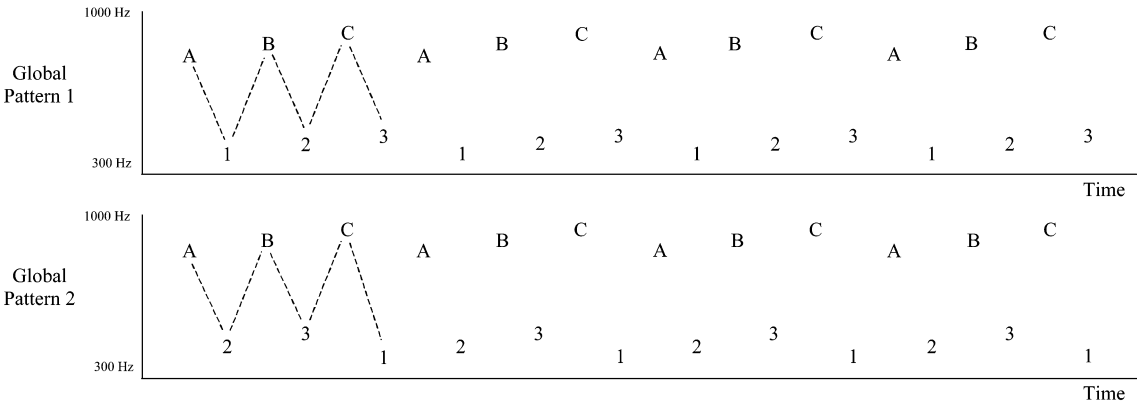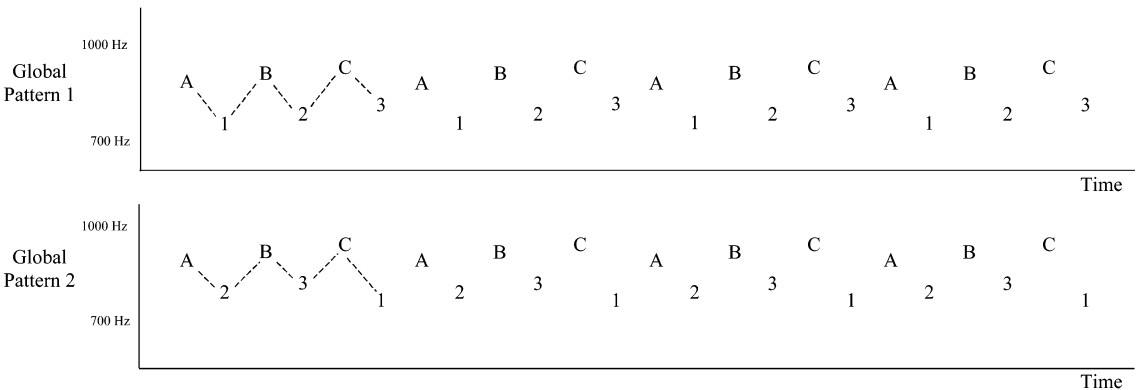
Figure 6. Stimulus paradigm for Experiment 3. (A) Schematic diagram showing the global presentations of the alternating tones presented in the far (top two rows) and near (bottom two rows) conditions. The *y*-axis represents sound frequency, and the *x*-axis represents time. (B) Schematic representation of a single trial. The fade-in, steady-intensity, and fade-out parts of the amplitude envelopes are shown for Global Patterns 1 and 2.

were identical in both conditions. The A tone was 880 Hz, the B tone was 932 Hz, and the C tone was 988 Hz. As in Experiment 1, these are referred to as the *middle* frequency range. The value of the 1, 2, 3 tones differed to create the two conditions. In the N condition, the three tones (A, B, and C) making up the middle-frequency range were *near* in frequency to the three tones making up the *low*-frequency range (1, 2, 3), which were separated, among themselves, by one-semitone steps. The tone represented by the number 1 was 740 Hz, the number 2 was 784 Hz, and the number 3 was 831 Hz, being near in frequency to the middle tones (the separation between A, the lowest tone in the middle range, and 3, the highest tone in the low range, being one semitone).

In contrast, in the F condition, the A, B, C tones making up the middle-frequency range and the 1, 2, 3 tones making up the low-frequency range were *far* in frequency from each other. The 1, 2, and 3 tones were separated, among themselves, by only one-semitone steps: Tone number 1 was 312 Hz, tone number 2 was 330 Hz, and tone 3 was 349 Hz, but they were far in frequency from the A, B, and C tones (the separation between tones A and 3 being six semitones). The F condition tones in Experiment 3 thus were identical in frequency value to the six tones used for the middle and low tones in Experiment 1 (those that were unattended in the attend condition in Experiment 1). In addition, in both the N and the F conditions, the tones were presented using a 135-msec SOA between successive tones (yielding a *within*-stream SOA of 270 msec), to match the SOA of the middle and low tones in Experiment 1.

Each trial consisted of two sequences separated by 1 sec of silence (see Figure 6). Each sequence consisted of one cycle of six tones repeated for 5.5 sec. One such sequence was created for each of the two global sequences each of the two conditions (creating four different global sequences: NG1, NG2, FG1, and FG2). There were eight combinations of these four sequences of stimuli: FG1–FG1, FG1–FG2, FG2–FG1, FG2–FG2, NG1–NG1, NG1–NG2, NG2–NG1, and NG2–NG2.

In each condition, there were either two sequences that were the same (FG1–FG1 and FG2–FG2, or NG1–NG1 and NG2–NG2) or two sequences that were different (FG1–FG2 and FG2–FG1, or NG1–NG2 and NG2–NG1). If the order of the sounds could be detected in the global sequences whether or not the sounds segregated, it was expected that the *different* sequences should be significantly discriminated from the *same* sequences in both the N and the F conditions. On the other hand, if the order of the sounds in the global sequences could not be detected when the sounds segregated (in the F condition), *different* should be discriminated from the *same* sequences only in the N condition.

Procedures. EEG was not recorded during the testing session in Experiment 3. A total of 48 trials, 6 trials of each of the eight possible sequence combinations, made up the experiment. The order of trials was counterbalanced across subjects. The subjects were told that they would hear a short sequence of sounds, followed by a short silence, and then another short sequence of sounds. They were instructed to listen to the two sequences and then indicate whether the two sequences sounded the *same* or *different* by marking their answer next to the trial number (1–48) on a prepared sheet given in advance. A response was required after each trial. There was no time limitation for recording responses; the next trial was started after the subject had indicated that he or she had completed his or her response.

Data analysis. For each subject, the HRs, the FARs, and d' were calculated in each condition separately. The d' measure 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 subject's decision is based. In the present experiment, we used a *same–different* model for independent observations to obtain d' (Macmillan & Creelman, 1991), a measure of the subject's ability to discriminate *same* from *different* sequences. The following procedure was used to calculate d'. Responses were counted as hits if the subject indicated *same*

Near Condition			Far Condition		
	Stimulus Sequence			Stimulus Sequence	
Response	Same	Different	Response	Same	Different
Same	74	26	Same	79	72
Different	33	67	Different	21	28

when the two sequences were the same or if subjects indicated *different* when the two sequences were different. Responses were counted as false alarms if the subjects indicated *different* when the sequences were the same or if the subjects indicated *same* when the sequences were different. We calculated one d' score for each subject separately in each condition.

Results and Discussion

If the tones were segregated, the within-stream patterns (1–2–3 and A–B–C) were identical in both G1 and G2 (apart from their starting or ending at different points in the within-stream cycle). Therefore, the measure of whether the global pattern was available to the subjects was whether or not they could tell the difference between G1 and G2. Table 3 provides a summary of the subjects' responses.

The mean d' score was significantly greater than zero in the N condition [mean $d' = 1.9$, $t(18) = 7.5$, $p < .001$], but not in the F condition [mean $d' = 0.43$, $t(18) = 1.4$, $p > .18$]. Furthermore, there was a significantly greater mean d' in the N than in the F condition [$t(18) = 4.8$, $p < .001$]. There also appears to be a response bias in favor of responding *same* in the F condition ($c = -0.10$ and -0.96, in the N and F conditions, respectively). We believe that the bias for responding *same* occurred because the global sequences (G1 and G2) were not detected as being different from each other in the F condition. This pattern of results indicates that segregation of the high and the low sounds occurred in the F condition, so that the segregation of the two streams made it impossible to distinguish the global patterns of the six-tone cycles from each other. Only in the N condition were the subjects able to discriminate the *same* (G1–G1 and G2–G2) from the *different* (G1–G2 and G2–G1) sequences, which is consistent with the hypothesis that they heard each of them as a single global sequence.

Taken together, these results demonstrate that when the sounds were integrated to form a single frequency stream (N condition), the global patterns were detectable to listeners but that when the sounds were segregated to form different frequency streams (F condition), the global patterns were not detectable. This suggests that global violations were not responsible for eliciting MMN in Experiments 1 and 2. The MMN process seems to involve two stages; one is the extraction and maintenance of the regularities in the ongoing input, and the other is the deviance detection process, which uses the information maintained in the regularities. The process that extracts and maintains sound regularities is likely shared with other auditory processes, perhaps subserving many pro-

cesses in parallel or at various stages. It seems unreasonable to assume that the formation of records of regularities in the sounds of the environment would exist solely for the purpose of detecting deviants, considering that identification of the sound sources (and finding regularities in the acoustic signal) is so central to auditory functioning. Deviance detection is not needed to establish regularities, even though establishment of the regularities is necessary to detect deviants. The established close correspondence between perception and MMN elicitation suggests that the deviance detection part of the process, eliciting MMN, occurs at a relatively *late* processing stage, just prior to or coincident with perception. This is also consistent with our findings that attention can alter auditory processes preceding the deviance detection stage (see the introduction for further details). Thus, we can conclude, in this and in our other studies in which streams have been used, that MMN elicitation by pattern violations indexes the detection of violations of the within-stream patterns, and not of disruptions of the global patterns. These data provide further support for the notion that MMN elicitation can index stream segregation (Sussman et al., 1999).

GENERAL DISCUSSION

Three experiments tested whether attention could modify the manner in which unattended sounds were processed when one frequency range was selected from three possible frequency streams. When none of the sounds was attended, evidence for stream segregation was obtained; MMNs were elicited within each stream. Stream segregation appeared to be unaffected by attention when the focus of attention was on a visual task, whether the visual task load was easy (reading a book in Experiment 1) or difficult (a visual discrimination task in Experiment 2). In contrast, when one auditory stream among the three possible streams was selected in order to perform a task with the sounds (Experiment 1, attend condition), evidence of segregation of the unattended sounds was not obtained (no MMNs and no attention-related ERP components [e.g., N2b and P3b] were elicited by them). Although the ERP responses associated with sound change detection were elicited by the subset of high sounds within the listener's focus of attention (the MMN, N2b, and P3b components were elicited by targets), there was no neurophysiological indication that the low and the middle sounds were segregated from each other when the high sounds were selected out and a task performed with them. These results indicate a modality-specific attention effect, suggesting that focusing attention on an acoustic stream (but not on a visually based task) can modify the manner, or extent, to which the unattended acoustic information is processed.

Experiment 3 demonstrated that discrimination between global patterns formed by different temporal relationships of the same component sounds could be achieved only when the all the sounds were heard as belonging to the same stream (N condition). When the distance between the

component sounds was increased (F condition), the sounds formed separate streams based on the frequency separation among the component sounds, and the global pattern differences were no longer heard. This result indicates that the MMNs elicited by pattern violations in the ignore conditions (Experiments 1 and 2) reflected *within-stream* deviance detection processes and not detection of global violations of the sequence. These results are also concordant with other behavioral studies that have suggested that primitive processes are responsible for stream segregation (e.g., Bey & McAdams, 2003). Bey and McAdams investigated the recognition of interleaved melodies by varying the frequency separation between the target melody and the distractor sounds on five levels of Δf (0, 6, 12, and 24 semitones). They found that listeners were able to accurately recognize the target melody interleaved within distractor sounds as the Δf between melody and distractor increased. The authors concluded that the segregation of tones was induced by the frequency separation between distractor and melody sounds, influencing the ability to detect the melody. Their results would predict that in the present study, in which two three-tone rising pitch patterns were alternated to form six-tone global patterns, that the three-tone "melodies" would be perceived separately from each other only when the component sounds of the global patterns were presented with greater than 10-semitone distance from each other (as was the case in the F condition in Experiment 3). And we concluded from the results of Experiment 3 that the observed bias (for the subjects to hear the patterns as the same) occurred because, when the sounds segregated, there was no difference between the corresponding within-stream three-tone "melodies" in G1 and G2. These data provide further evidence that MMN elicitation can serve as an index of stream segregation.

Taken together, the present results point toward an explanation of modality-specific processes mediating an effect of attention on unattended sound processing. The results in Experiment 2 demonstrated that there was no effect of attention on the segregation or the MMN elicitation processes for unattended sounds even when a visual task was performed that was equal in processing load to the auditory task in Experiment 1. Thus, processing load itself cannot be considered the critical factor contributing to the attention effect. The results could be attributable to a modality-specific interference (Duncan et al., 1997). This interpretation suggests that interference will be more likely to occur if attention is focused on a sound source, rather than on a visual source, because attentional capacities for visual and auditory processing may be largely independent from each other. When subjects attend a visual channel, the visual-processing resources needed to perform the task will not interfere with simultaneous auditory processes. This means that when attention is assigned to an auditory channel, processing of other auditory channels may be restricted. The difference in processing unattended sounds between the attend and the ignore conditions in the present study may be attributable to a difference in the quantity of resources used in the

modality performing the task. This may have been set up by focusing attention to one of many sound streams (when frequency was the only cue for stream segregation).

It should be noted, that the locus of this attentional biasing for sound processing remains to be determined. We cannot determine whether attention modified the stream segregation process or the MMN process (or both). Previous studies have indicated that the two processes (auditory stream segregation and MMN change detection process) are distinct, by showing that the stream segregation process precedes the MMN deviance detection process in time (Sussman, 2005; Sussman et al., 1998, 1999; Sussman et al., 2001; Yabe et al., 2001). This suggests that the two processes might have operated independently of one another and attention could have modified one and not the other. But because the MMN process is the later of the two, the absence of MMN does not indicate which process was modified by attention. MMN would not be elicited in either case.

It is possible that attention could have modified the level of the MMN deviance detection process, even if the sounds had already been organized via stimulus-driven factors. The pace of the sounds in the present study was fast (90 msec offset to onset), which automatically induces stream segregation when the sounds are completely ignored. This was seen in the ignore conditions, in which deviants in all three streams elicited MMNs and no ERP components associated with any forms of attention (e.g., no P3a, N2b, or P3b; and also in Sussman et al., 1999, in which the alternating high and low sounds were presented at a rapid rate). In contrast, in our previous studies, in which attention modified earlier organizational processes that influenced the MMN deviance detection process (Sussman et al., 1998; Sussman, Winkler, Huotilainen, et al., 2002), the pace of the sounds was slower (e.g., 1,250 and 1,000 msec offset to onset, respectively). In both of these studies, segregation and grouping of the sounds were *not* automatically induced by stimulus-driven factors, likely due to the slow pace, and no MMNs were elicited (when the auditory signal was completely ignored). We concluded in these studies that the difference in MMN results was due to a difference in sound organization between ignore and attend conditions. That is, attention modified the organization of the sounds, not the MMN process. Thus, it is possible in the present study, that the unattended sounds segregated by frequency range but the highly focused attention on the high stream may have suppressed or limited the MMN deviance detection process.

In other studies, the effects of attention on the MMN process have been examined. Some studies have shown, in selective listening paradigms, that highly selected attention to one auditory channel attenuated the amplitude of the MMN elicited by unattended deviants in another auditory channel (e.g., Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998). An unsolved issue among these studies was that only the amplitude of the intensity-based (not the frequency-based) MMN was reduced by attention. Recently, Suss-

man et al. (2003) resolved this conflict by demonstrating that highly focused selective attention, in and of itself, does not alter deviant detection on the unattended channel. Rather, the reduced MMN amplitude was the result of a competition for MMN deviance detection that was set up when identical deviants (e.g., frequency-based deviants) occurring on both attended and unattended channels vied for MMN elicitation. In Sussman et al.'s (2003) study, when the subjects ignored the stimulation altogether (in a separate condition), MMNs were elicited by deviants presented to both ears. When frequency deviants were targets in the attended ear, frequency deviants in the unattended ear elicited no MMNs. We concluded that competition for deviance detection was biased by attention so that the attended channel "won out." The presence of deviants without attentional bias did not alter the MMN result. This competition hypothesis may support an MMN limitation hypothesis of the present results. Competition for MMN deviance detection may have been set up between the middle- and the low-frequency streams in which similar (falling pitch) pattern violations were vying for the same processes. Highly focused attention to the sounds for detecting the successive pitch deviants in the high stream may have biased the MMN system to the attended stream (attention facilitating the deviance detection process in the attended stream) without affecting the stream segregation process at all. Moreover, similar to what was found by Sussman et al. (2003), in the present study when a visual task was the primary focus (ignore conditions), all the deviants in all the frequency ranges elicited MMN. In other words, attention to sounds may be needed to bias processing of the unattended sounds, because MMN was not biased simply by the presence of the deviants occurring in both ears.

Alternatively, it is possible that it was not the MMN system that was limited or biased by attention but that attention limited the organizational processes associated with stream segregation. Attention to one sound stream may have disrupted the formation of streams in the unattended sounds. If this were the case, the highly focused attention needed to perform the auditory task may have modified processes preceding the MMN system—those associated with representation and storage of sequential sounds. If attention modified the organizational process, so that the unattended sounds did not segregate, the within-stream patterns would not be detected (and consequently, violations of the within-stream patterns not detected), and thus, no MMNs would be elicited. Attention may heighten power for the task at hand but preempt certain auditory resources that would otherwise be used by *preattentive* auditory processes.

In summary, the results of the three experiments showed that even though more than one stream can be formed at the same time, attention can modify sound processing when one sound stream is selected over others. MMNs were elicited by unattended deviants while the subjects performed a visual task (the ignore conditions in Experiments 1 and 2), but not by the same deviants when the subjects attended to another auditory stream (the attend con-

dition in Experiment 1). When experimental subjects had no task to perform with any of the sounds and focused, instead, on simultaneous visual sensory input, auditory stream segregation was induced by the frequency separation of the sound input, and MMNs were thus elicited separately in each of the streams. This may be explained, at least in part, by some independence of visual and auditory attentional resources (Duncan et al., 1997). In contrast, when attention was focused on one frequency range of the total sound input, attention appears to have biased the sensory input to facilitate the task at hand, at the possible expense of the unattended sound input. This result is consistent with previous animal (Desimone, 1998; Desimone & Duncan, 1995; Fritz et al., 2003; Moran & Desimone, 1985) and human (Sussman et al., 2003; Woldorff et al., 1991; Woldorff et al., 1998) studies showing that attentional control can limit the unattended channel in favor of the attended sensory input. The active selection of a subset of sensory information modifies the input at an early processing stage, which is generally observed as an enhancement of the neural activity for the relevant features of the input. This type of attentional biasing facilitates the ability to achieve behavioral goals.

REFERENCES

Bey, C., & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 267-279.

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*. Cambridge, MA: MIT Press.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.

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.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex: I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology*, **46**, 755-772.

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.

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.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London*, **353**, 1245-1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193-222.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, **70**, 80-90.

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, **387**, 808-810.

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.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, **6**, 1216-1223.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

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.

Kisley, M. A., Davalos, D. B., Layton, H. S., Pratt, D., Ellis, J. K., & Seger, C. A. (2004). Small changes in temporal deviance modulate mismatch negativity amplitude in humans. *Neuroscience Letters*, **358**, 197-200.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, **229**, 782-784.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, **30**, 436-450.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, **24**, 283-288.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Norman, D. A. (1967). Temporal confusion and limited capacity processors. *Acta Psychologica*, **27**, 293-297.

Picton, D. 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., Deacon, D., Cowan, N., & Vaughan, H. G., Jr. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, **36**, 835-839.

Ritter, W., Sussman, E., & Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, **11**, 61-63.

Rugg, M. D., & Coles, G. H. (1995). *Electrophysiology of mind. Event-related brain potentials and cognition* (Oxford Psychology Series No. 25). Oxford: Oxford University Press.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, **357**, 163-166.

Sussman, E. (2004). Auditory scene analysis: Examining the role of nonlinguistic auditory processing in speech perception. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 5-12). New York: Kluwer.

Sussman, E. (2005). Integration and segregation in auditory scene analysis. *Journal of the Acoustical Society of America*, **117**, 1285-1298.

Sussman E., Čeponienè, R., Shestakova, A., Näätänen, R., & Winkler, I. (2001). Auditory stream segregation processes operate similarly in school-aged children as adults. *Hearing Research*, **153**, 108-114.

Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research*, **190**, 128-140.

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

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

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

Sussman, E., Winkler, I., Kreuzer, J., Saher, M., Näätänen, R., & Ritter, W. (2002). Temporal integration: Intentional sound discrimination does not modify stimulus-driven processes in auditory event synthesis. *Clinical Neurophysiology*, **113**, 909-920.

Sussman, E., Winkler, I., & Wang, W. J. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, **40**, 430-435.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, **372**, 90-92.

Treisman, A. (1969). Strategies and models in selective attention. *Psychological Review*, **76**, 242-299.

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

Winkler, I., Sussman, E., Tervaniemi, M., Ritter, W., Horvath, J., & Näätänen, R. (2003). Pre-attentive auditory context effects. *Cognitive, Affective, & Behavioral Neuroscience*, **3**, 57-77.

Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, **28**, 30-42.

Woldorff, 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., Hiruma, T., & Kaneko, S. (2001). Organizing sound sequences in the human brain: The interplay of auditory streaming and temporal integration. *Brain Research*, **897**, 222-227.

NOTES

1. Bregman (1978) determined that stream segregation takes time to build up. Initially, when a mixture of high and low sounds is presented, subjects perceive a single integrated stream and, after some time (<7 sec), the sounds are perceived as two segregated streams. The build-up phase varies as a function of Δf, with the judgment for two streams occurring more quickly the larger the Δf.

2. The term *preattentive* is meant to describe a process that does not require attention to operate. It is a misnomer, in the sense that processes occurring prior to the deviance detection process can be affected by attention, and a process that can be affected by attention is not strictly preattentive. However, the MMN process (the deviance detection aspect of it) itself does not appear to be altered by attention (see the introduction and the Results and Discussion section for Experiment 3).

3. The tones are numbered within each frequency range (H, M, and L), with the lowest tone represented by the lowest number (e.g., H_1 is 2489 Hz, and H_2 is 2637 Hz).

4. It should be noted that because of the random distribution of single and successive high deviants (H_2), a repeating pattern of the alternating L–M–H tones could not be established within the global sequences. Therefore, the occurrence of a deviant did not violate a global pattern of the alternating sequence. This would make an explanation in which a global violation in the sequence was responsible for eliciting MMN less plausible.