

# Effects of Attentional Load on Auditory Scene Analysis

Claude Alain and Aaron Izenberg

## Abstract

■ The effects of attention on the neural processes underlying auditory scene analysis were investigated through the manipulation of auditory task load. Participants were asked to focus their attention on tuned and mistuned stimuli presented to one ear and to ignore similar stimuli presented to the other ear. For both tuned and mistuned sounds, long (standard) and shorter (deviant) duration stimuli were presented in both ears. Auditory task load was manipulated by varying task instructions. In the easier condition, participants were asked to press a button for deviant sounds (target) at the attended location, irrespective of tuning. In the harder condition, participants were further asked to identify whether the targets were tuned

or mistuned. Participants were faster in detecting targets defined by duration only than by both duration and tuning. At the unattended location, deviant stimuli generated a mismatch negativity wave at frontocentral sites whose amplitude decreased with increasing task demand. In comparison, standard mistuned stimuli generated an object-related negativity at central sites whose amplitude was not affected by task difficulty. These results show that the processing of sound sequences is differentially affected by attentional load than is the processing of sounds that occur simultaneously (i.e., sequential vs. simultaneous grouping processes), and that they each recruit distinct neural networks. ■

## INTRODUCTION

Our typical environment is acoustically complex. At any given moment, we might be surrounded by a number of sound-generating elements such as a computer fan, a radio playing music, or a group of people speaking, and several of these elements might be operating simultaneously. In order to make sense of this environment, we must identify, group, and segregate these elements into separate mental representations called *auditory streams* or *objects*. This process is known as *auditory scene analysis* and involves perceptually organizing our environment along two axes: horizontal (time) and vertical (frequency). Organization along the horizontal axis entails sequential grouping of acoustic data over several seconds, whereas along the vertical axis, acoustic elements from simultaneous sound sources are grouped and segregated within hundreds of milliseconds.

Bregman (1990) proposed a comprehensive account of auditory scene analysis that outlines both primitive (bottom-up) and knowledge-based (top-down) modes of processing. Low-level stream segregation involves the grouping together of sounds according to frequency, spatial, and temporal aspects following Gestalt principles (e.g., sounds that start at the same time are more likely to be coming from the same source). This is thought to be an automatic, innate process that can be found in infants as well as in animals such as birds

(MacDougall-Shackleton, Hulse, Gentner, & White, 1998; Hulse, MacDougall-Shackleton, & Wisniewski, 1997) and monkeys (Fishman, Reser, Arezzo, & Steinschneider, 2001). Top-down processes, on the other hand, involve conscious attention and may be based on past experiences with certain classes of sounds—for example, the recognition of a familiar voice or melody as opposed to one that is novel.

A great deal of evidence supports this notion of low-level auditory scene analysis that is automatic, unlearned, and independent of listener attention. For example, performance in auditory tasks can be improved by clustering sounds presented outside the focus of attention (e.g., Alain & Woods, 1993, 1994; Bregman & Rudnicki, 1975). Infants, while still too young to intentionally attend to stimuli, have been shown to be responsive to changes in auditory patterns (e.g., Alho, Sainio, Sajaniemi, Reinikainen, & Naatanen, 1990; Trehub, Bull, & Thorpe, 1984; Chang & Trehub, 1977), suggesting that focused attention is not a prerequisite for extracting and encoding regularities within the auditory stream. There is, however, some evidence from more recent studies that appears to contradict this notion. For example, Carlyon, Cusack, Foxton, and Robertson (2001) showed that attention influences what is thought to be low-level stream segregation. They noticed that selective attention to one spatial location (e.g., right ear) substantially reduced the build-up of stream segregation for sounds presented at the unattended location (i.e., left ear). In addition, they found that patients with unilateral neglect showed

---

University of Toronto

reduced streaming for stimuli presented in the neglected hemifield.

### The Present Study

The aim of the present study is to elucidate the relationship between attention and low-level stream segregation through the recording of human event-related brain potentials (ERPs). ERPs provide a powerful measure for examining the effects of selective attention on auditory scene analysis because they can be recorded for stimuli presented outside the focus of attention. Two electrophysiological events have been associated with the bottom-up aspect of auditory scene analysis: the mismatch negativity wave (MMN) and the object-related negativity (ORN) wave.

The MMN is elicited by deviant stimuli embedded in a sequence of standard stimuli, even when stimuli are presented outside the focus of attention. The deviant may differ from the standard in frequency, intensity, and/or duration (Näätänen, 1992). Sounds that break repeating auditory patterns (e.g., four tones regularly alternating in frequency) also generate an MMN response (for reviews, see Alain & Arnott, 2000; Picton, Alain, Otten, Ritter, & Achim, 2000). The MMN is thought to index an automatic mismatch response between the incoming stimulus and a representation of what is expected based upon the organization of the previously presented stimuli (Alain, Woods, & Ogawa, 1994; Alain, Cortese, & Picton, 1999). In the present study, the MMN is used as an index of sequential integration because its elicitation depends on the extraction of regularities in previously presented stimuli. The fact that the MMN can be recorded outside the focus of attention is consistent with the hypothesis that sequential integration of sounds may occur automatically. However, selective attention experiments that make strong demands on attentional resources have shown that the MMN is smaller for stimuli presented at the unattended location than at the attended location (e.g., Alain & Woods, 1997; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998). Task instructions that draw listeners' attention to the organization of the stimuli also modulate the amplitude of the MMN, indicating that MMN generation is sensitive to top-down controlled processes (Sussman, Winkler, Huotilainen, Ritter, & Naatanen, 2002). Together, these findings suggest that a listener's attention plays an important role in extracting and encoding the acoustic regularities between consecutive acoustic events. It has been suggested that attention affects processes prior to the MMN-generating system rather than the MMN itself (Rinne, Antila, & Winkler, 2001; Ritter, Sussman, Deacon, Cowan, & Vaughan, 1999).

The ORN is a relatively new ERP component thought to index concurrent sound segregation (Alain, Arnott, & Picton, 2001; Alain, Schuler, & McDonald, 2002; Alain,

Theunissen, Chevalier, Batty, & Taylor, 2003). It is present when participants perceive two simultaneous auditory events. Two concurrent sounds are perceived when one harmonic component of a complex sound is mistuned so that it is not an integer multiple of the fundamental. The ORN is present even when participants are asked to ignore the stimuli and read a book of their choice (Alain, Arnott, et al., 2001) or watched a subtitled movie with our sound muted (Alain et al., 2002). The ORN is thought to index concurrent sound segregation because, in contrast with the MMN, its generation does not depend on contextual factors such as the probability of the mistuned harmonic within a sequence of stimuli. Specifically, it likely reflects the discrepancy between the mistuned harmonic and the expected frequency based upon the fundamental. The fact that the ORN was recorded even when participants were not attending to the stimuli is consistent with the proposal that concurrent sound segregation may occur independently of a listener's attention. However, these findings should be interpreted with caution because, in these studies, listener attention was not well controlled, for little effort was made to ensure that they read their book or attended to the subtitled movie. Because there was no objective measure of attention, the possibility that the participant's attention may have wandered to the auditory stimuli cannot be ruled out.

In the present study, we controlled attention by using a selective listening paradigm in which participants were asked to attend to tuned and mistuned stimuli presented in one location (e.g., left ear) while ignoring similar sounds presented in another location (i.e., right ear). In one condition, participants were asked to press a button in response to rare, shorter duration sounds regardless of mistuning. In another condition, participants were presented with the same stimuli, but they were asked to press one key if the shorter sounds were tuned and another key if they were mistuned, thereby increasing the attentional demand of the task. In the present study, the MMN was used as an index of sequential integration, whereas the ORN was used as an index of concurrent sound segregation. It was hypothesized that: (a) if the sequential integration of sounds is independent of listener attention, then the MMN should be present in both listening conditions and should be little affected by auditory

**Table 1.** Group Mean and Standard Error of the Mean for Hits, False Alarms (FAs), and Response Time (RT) to Target Defined by a Single or a Conjunction of Features

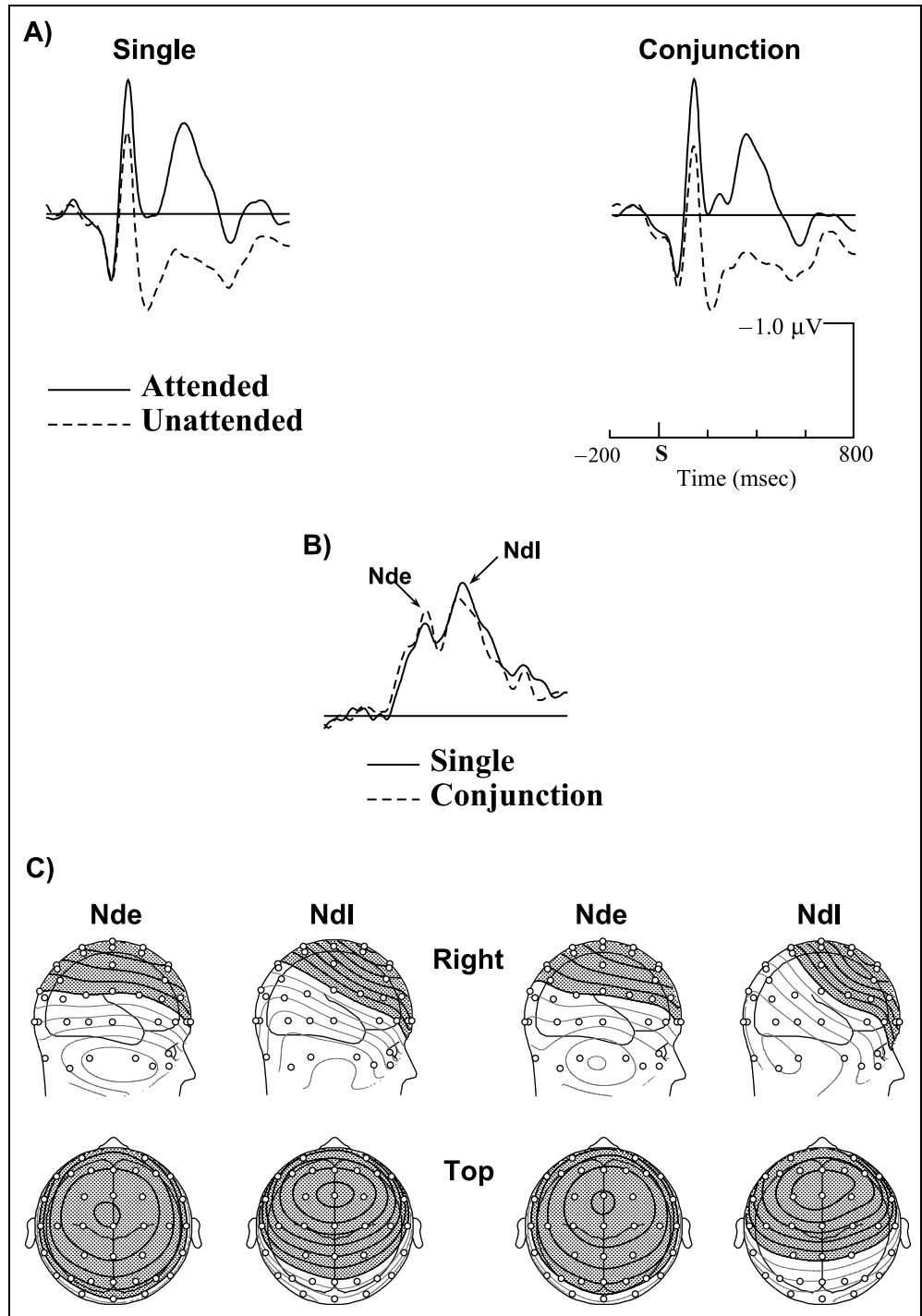
<i>Condition</i>	<i>Hit (%)</i>	<i>FAs (%)</i>	<i>RT (msec)</i>
Single	43 ± 4.7	9.4 ± 2.6	672 ± 16
Conjunction	39 ± 5.5	25.3 ± 3.6	767 ± 19

The data from the left and right ear were lumped together.

attention load; (b) similarly, if concurrent sound segregation is independent of listener attention, the ORN should be present in both listening conditions and should be little affected by task instruction. However,

if the processing of sequentially and/or simultaneously occurring sounds depends on listener attention, then the MMN and/or ORN amplitude should vary with task instructions.

**Figure 1.** (A) Group mean ERPs elicited by tuned stimuli occurring at the attended location (solid) and unattended location (dashed) during the single (left panel) and feature conjunction (right panel) tasks. (B) Group mean negative difference (Nd) waves between the ERPs elicited by standard attended and unattended stimuli when the target was defined by duration only (Single) or by duration and tuning (Conjunction). In this and the subsequent figures, negativity is plotted upwards. S = stimulus onset. (C) Contour maps for the Nd<sub>c</sub> (200 msec) and Nd<sub>i</sub> (375 msec). The data obtained when participants attended to the left and right ear were collapsed together with the electrode position from the attended right ear condition flipped so that electrodes over the right hemisphere are contralateral to the attended location (i.e., left ear). The contour spacing was set at 0.2  $\mu$ V. The negative polarity is illustrated by the shaded area. The open dots indicate electrode positions.



## RESULTS

### Behavioral Data

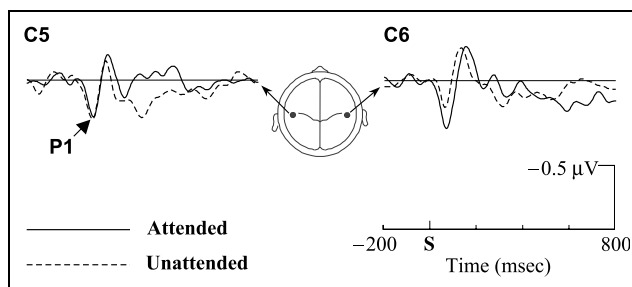
The behavioral results are summarized in Table 1. The single and conjunction tasks were both very demanding with participants accurately identifying only 43% and 39% of the target stimuli, respectively. There was no difference in accuracy between the single and conjunction tasks,  $F(1,9) = 1.46$ . However, participants were faster when targets were defined by duration only than when they were defined by the conjunction of duration and tuning,  $F(1,9) = 33.39$ ,  $p < .001$ . There was no difference in response time when the targets were presented to the left or right ears nor was the interaction between task and ear of presentation significant. Participants also made significantly more errors in the conjunction than in the single detection task,  $F(1,9) = 36.29$ ,  $p < .001$ . This difference in error rate was caused primarily by participants pressing the wrong button in the conjunction condition rather than by participants making responses outside of the hit window (see Methods).

### Electrophysiological Data

The extent to which participants focus their attention at the designated location may have a tremendous impact on how task-irrelevant stimuli are processed. We examined the effects of spatial attention on processing standard stimuli and tested whether the allocation of spatial attention differed between the single and conjunction tasks.

Figure 1 shows ERPs elicited by the standard tuned stimuli when they occurred both at the attended location and at the unattended location. All stimuli generated N1 and P2 waves that were maximum at frontocentral scalp sites. The ERPs elicited by stimuli occurring at the attended location were characterized by a negative displacement at frontocentral sites that began at about 70 msec after sound onset and lasted for several hundred milliseconds. The effects of attention on the ERPs are best isolated as Nd waves between the ERPs elicited by the standard stimuli presented in the attended location and those elicited by the same stimuli when they were unattended. At frontocentral sites, the Nd wave was comprised of early ( $Nd_e$ ) and late ( $Nd_l$ ) portions, peaking at about 200 and 375 msec after sound onset, respectively. Both  $Nd_e$  and  $Nd_l$  inverted in polarity at temporal electrodes, consistent with generators located in the superior temporal gyrus along the Sylvian fissure.

The effects of auditory spatial attention on the ERPs elicited by standard stimuli were confirmed with an ANOVA on the mean amplitude between 90 and 170 msec with Attention (attended vs. unattended), Task (single vs. conjunction), and Electrodes as factors,  $F(1,9) = 17.90$ ,  $p < .01$ . The  $Nd_e$  amplitude was similar



**Figure 2.** Group mean ERPs elicited by tuned stimuli occurring at the attended (solid) and unattended location (dashed) during the single feature conjunction. The solid dots on the schematic of the head indicate electrode positions. The data obtained when participants attended to the left and right ear were collapsed together with the electrode position from the attended right ear condition flipped so that electrodes over the right hemisphere are contralateral to the attended location (i.e., left ear). That is, the P1 recorded at C6 is contralateral to the attended location.

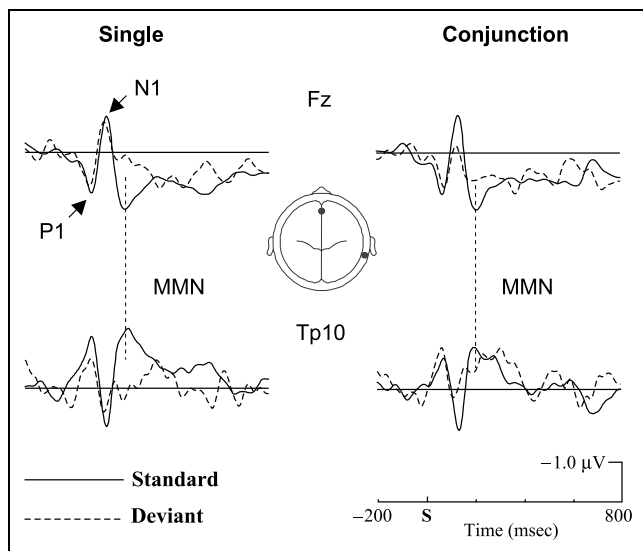
in both single and feature conjunction conditions [Attention  $\times$  Task interaction,  $F(1,9) = .90$ ]. The late effects of attention on ERPs to the standard stimuli were quantified over the 330–410 msec interval. The ANOVA showed a main effect of attention,  $F(1,9) = 54.54$ ,  $p < .001$ . As with the  $Nd_e$ ,  $Nd_l$  was similar in amplitude during the single and feature conjunction conditions [Attention  $\times$  Task interaction,  $F(1,9) = 1.62$ ].

In addition to the effects of selective attention at the frontocentral scalp sites, we also observed a reduced P1 wave for unattended stimuli over the hemisphere contralateral to the attended location (Figure 2). This effect of spatial attention on the P1 was quantified at a subset of electrodes over the hemisphere ipsi- and contralateral to the attended location (i.e., F5/6, F7/8, Fc5/6, C5/6, T7/8, and CP5/6). There was a main effect of attention,  $F(1, 9) = 31.94$ ,  $p < .001$ , and a significant Attention  $\times$  Hemisphere interaction,  $F(1,9) = 13.38$ ,  $p < .005$ . The interaction between Attention and Task was not significant.

### Effects of Task Load on Sequential Integration

The effects of task demand on sequential integration were examined by comparing ERPs elicited by standard tuned and deviant tuned stimuli occurring only at the unattended location (Figure 3). The ERPs elicited by stimuli occurring at the attended location were not included in the analyses because the duration targets generated an MMN with a latency that overlapped with the  $Nd_e$  to the standard stimuli. In other words, it was not possible to dissociate contribution of attention-related enhanced negativity (Nd) from the MMN, making the identification of the MMN response difficult.

Deviant stimuli occurring at the unattended location generated an MMN that peaked at about 220 msec after sound onset [main effect of stimulus type for the 180–260-msec interval,  $F(1,9) = 24.82$ ,  $p < .001$ ]. The



**Figure 3.** Group mean ERPs elicited by standard tuned and deviant tuned stimuli presented at the unattended location as a function of auditory task load. Note that the N1 wave is smaller for the deviant than for the standard because the deviant tones is shorter in duration than the standards and thus has less energy than the standards. The solid dots on the schematic of the head indicate electrode positions.

interaction between Stimulus Type and Task was significant,  $F(1,9) = 7.48, p < .05$ , indicating that the MMN amplitude was significantly smaller during the conjunction than in the single feature discrimination task. A separate ANOVA on the ERPs recorded during the conjunction task indicated that the MMN response failed to reach significance,  $F(1,9) = 4.37, p = .07$ . The reduced MMN amplitude in the conjunction condition was primarily due to an effect of task demand on ERPs elicited by the deviant stimuli,  $F(1,9) = 6.03, p < .05$  (Figure 4). A planned comparison revealed no significant difference in ERP amplitude elicited by the standard stimuli,  $F(1,9) < 1.0$ .

The MMN was largest over the right frontocentral areas and inverted in polarity at inferior temporal-parietal sites (Figure 3). Deviant stimuli generated a significant MMN at inferior parietal and mastoid sites [TP9/10 and T7/8,  $F(1,9) = 23.87, p < .001$ ]. As was the case for the MMN recorded at frontocentral sites, the MMN amplitude at inferior temporal-parietal sites was significantly reduced with increasing auditory task load,  $F(1,9) = 12.38, p < .01$ . A subsequent analysis on the ERPs recorded during the conjunction task revealed no significant MMN response,  $F(1,9) = 1.40, p = 0.27$ .

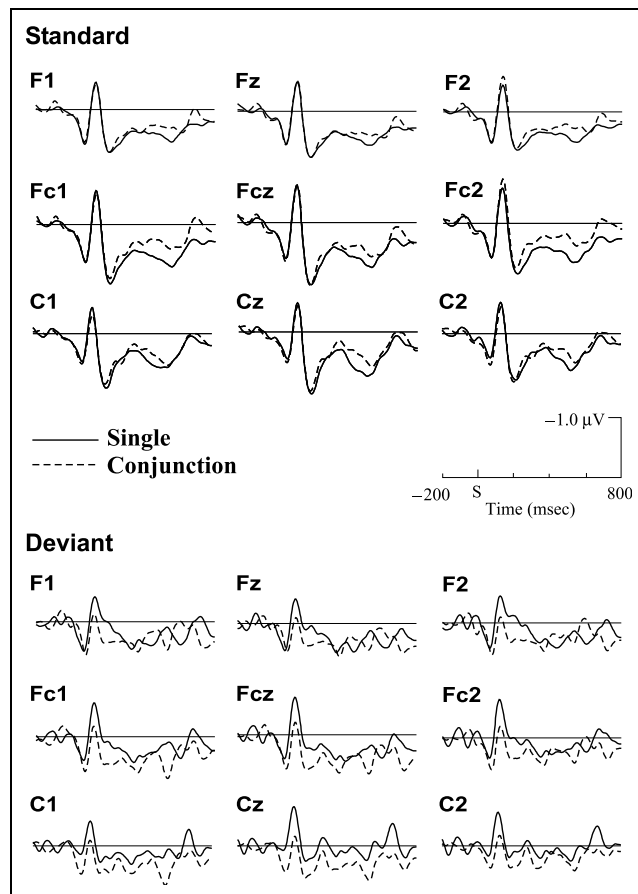
#### *Effects of Task Load on Concurrent Sound Segregation*

The effects of task demand on concurrent sound segregation were examined by contrasting ERPs elicited by standard tuned and standard mistuned stimuli presented at the unattended location (Figure 5). The effects of task load on processing standard tuned and

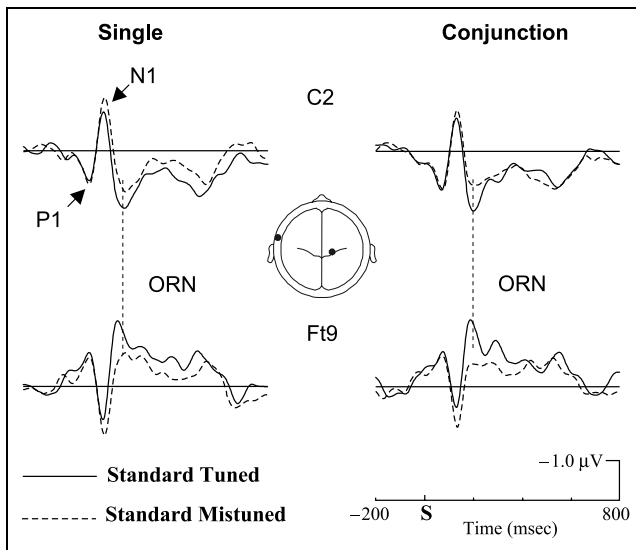
mistuned stimuli are illustrated separately in Figures 4 and 6, respectively. There was a main effect of mistuning on ERPs for the 150–230-msec interval,  $F(1,9) = 15.41, p < .01$ . The interaction between Task and Mistuning was not significant,  $F(1,9) = 0.32$ . The amplitude was maximal over the central scalp region and inverted in polarity at the inferior frontal and temporal-parietal scalp regions, consistent with generators located in auditory cortices along the supratemporal plane. Separate ANOVAs on the ERPs elicited by tuned and mistuned stimuli at the inferior temporal-parietal sites showed a main effect of mistuning,  $F(1,9) = 10.92, p < .01$ . The interaction between Task and Mistuning was not significant,  $F(1,9) = 1.06, p = 0.33$ .

#### *Effects of Task Load: Sequential versus Concurrent Sound Segregation*

To test whether sequential segregation was more sensitive to task demand than concurrent sound segregation, we compared the effects of task load on the MMN and

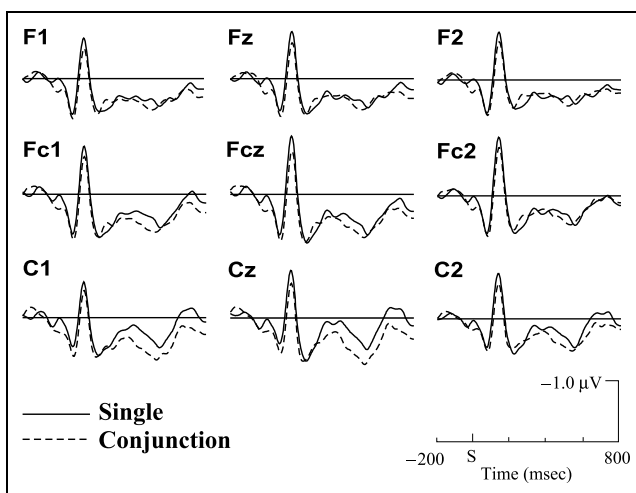


**Figure 4.** Effects of auditory task load on processing standard and deviant tuned stimuli presented outside the focus of attention. Note that the processing of deviant stimuli was modulated by varying task demands, whereas the processing of the standard stimuli was little affected by the attentional resource allocated to the auditory task.

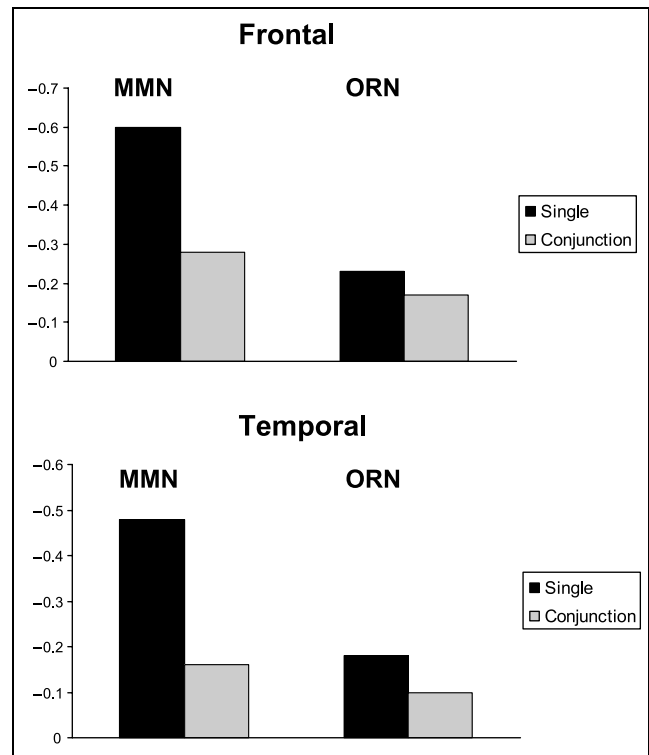


**Figure 5.** Group mean ERPs elicited by standard tuned and standard mistuned stimuli presented at the unattended location as a function of auditory task load. Note that the ORN showed maximum amplitude at different electrode sites than the MMN. The solid dots on the schematic of the head indicate electrode positions.

ORN within the same ANOVA. Overall, the MMN amplitude was significantly larger than the ORN at frontocentral sites,  $F(1,9) = 8.82, p < .05$  (Figure 7). There was also a significant interaction between Task and Component,  $F(1,9) = 6.26, p < .05$ , reflecting greater effects of auditory task load on MMN than on ORN amplitude. At inferior temporal sites, the MMN and ORN amplitude was comparable,  $F(1,9) = 2.98, p = .12$ . However, there was a significant interaction between Task and Component,  $F(1,9) = 10.48, p < .01$ . This finding suggests that sequential integration, as indexed by the MMN, is more



**Figure 6.** Effects of auditory task load on processing standard mistuned stimuli presented outside the focus of attention. Note that the processing of mistuning stimuli between an interval of 100 and 200 msec was little affected by the attentional resource allocated to the auditory task.

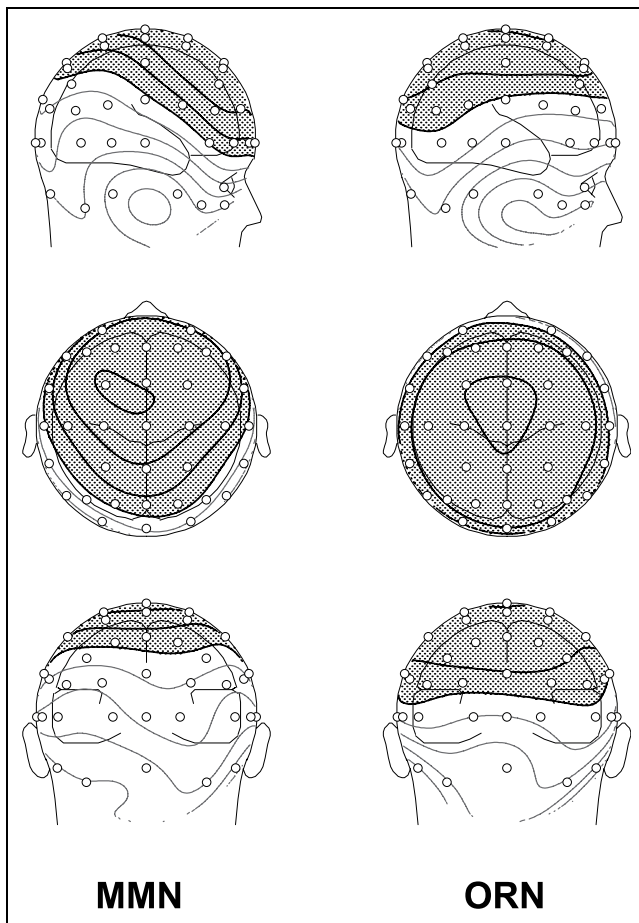


**Figure 7.** Bar graph illustrating the group mean amplitude for the MMN and ORN as a function of the auditory task load. The MMN and ORN amplitude reflect the mean voltage over an 80-msec interval measured at nine frontocentral (top) and four temporal (bottom) sites (see Methods for details). Only the MMN amplitude was significantly reduced with increasing task demand.

sensitive to demand on attentional resources than is concurrent sound segregation.

### Scalp Distribution Analyses

One important assumption of the present study is that the ORN and MMN responses index different aspects of auditory scene analysis, namely concurrent and sequential analysis of sounds, respectively. To test this hypothesis, we compared the scalp distribution of the ORN and MMN obtained from stimuli presented at the unattended location during the single feature detection task. Scalp distributions are an important criterion in distinguishing between ERP components. The postulate is that different scalp distributions indicate different spatial configurations of intracranial sources. In the present study, we analyzed scalp distributions to examine whether the observed ERP component generation (i.e., MMN and ORN) depends on distinct neural networks. Figure 8 shows the amplitude distribution for the MMN and ORN. The ORN was largest at central sites and inverted in polarity at inferior temporal sites. In comparison, the MMN was largest at frontocentral sites and inverted in polarity at inferior parieto-occipital sites. There was a significant difference in topographies between the MMN and ORN,  $F(60, 540) = 5.00, p < .001$ ,



**Figure 8.** Contour maps for the MMN and ORN peak amplitude. The contour spacing was set at 0.2 and 0.1  $\mu\text{V}$  for the MMN and ORN, respectively. The negative polarity is illustrated by the shaded area. The open dots indicate electrode positions.

with the ORN being more centrally distributed than the MMN. The effect of task instruction on the MMN or ORN amplitude distribution was not significant.

## DISCUSSION

The present study was designed to examine the extent to which demand on attentional resources influences auditory scene analysis by manipulating auditory task difficulty. The assumption is that a more difficult auditory task (i.e., conjunction task) puts a greater demand on a fixed pool of resources, leaving less available for processing events occurring outside the “spotlight” of attention.

The observed level of accuracy was relatively low—albeit above chance performance—due largely to high target density. We chose a high rate of target presentation, with one occurring on average every 1500 msec, to help participants maintain their focus of attention at the designated location and to minimize the wandering of attention toward the task-irrelevant location.

Although the hit rate was comparable in both single and conjunction tasks, participants made more FAs and were slower in detecting targets defined by a combination of features than those defined by a single feature. This finding is not surprising in light of previous research on auditory feature conjunction that showed an increased response time when targets were defined by a conjunction of features rather than by a single feature (Woods, Alain, & Ogawa, 1998; Woods, Alain, Diaz, Rhodes, & Ogawa, 2001). In addition, it has been well documented that the threshold for detecting mistuning increases with decreased stimulus duration (Alain, McDonald, Ostroff, & Schneider, 2001; Moore, Peters, & Glasberg, 1985; Moore, Glasberg, & Peters, 1986). Though the degree of mistuning was large, the relatively short target duration used in the present study may have impeded the ability of untrained listeners to detect mistuning. This could also account for the increased response time and error rate observed in the conjunction condition. Most importantly, the manipulation of task instructions, while keeping the stimulus set constant, did indeed affect task difficulty.

The electrophysiological data suggest that task instruction affected the neural correlates of response processes rather than those of early selection processes. The effects of spatial attention on ERPs (i.e., Nd wave) showed comparable onset and amplitude in both single and feature conjunction conditions. This suggests that listeners were equally able to focus and maintain their attention to the designated location, in both conditions. This is not surprising given that the Nd<sub>c</sub> is primarily affected by factors that promote the segregation of the attended and unattended channels such as the frequency and/or spatial separation between the task-relevant and task-irrelevant stimuli (for reviews, see Alain & Woods, 1994; Alain & Arnott, 2000; Alho, Tottola, Reinikainen, Sams, & Naatanen, 1987; e.g., Woods & Alain, 2001; Hansen & Hillyard, 1980). The Nd<sub>i</sub> was also similar in amplitude and latency in both single and feature conjunction conditions. This finding differs from previous work showing an increase in Nd<sub>i</sub> amplitude with increasing task difficulty (Okita, 1989; Maiste & Picton, 1987; Parasuraman, 1980). The discrepancy between our results and those from previous studies is likely due to procedural differences. In previous studies, task difficulty was manipulated by either increasing the number of potential targets or by decreasing the discriminability between the standard and target stimuli. In the present study, both the proportion of target and the discriminability between standards and deviants were kept constant, only the number of features to identify within the target stimulus was varied between the conditions.

In addition to these correlates of selective attention, auditory spatial attention also modulates the amplitude of the P1 wave contralateral to the attended location. This attention effect preceded the onset of the Nd wave at frontocentral sites and may reflect an early gating of

task-irrelevant information (Woldorff et al., 1993; Hillyard, Hink, Schwent, & Picton, 1973). This gating was similar in both single and conjunction tasks. Our results are consistent with previous work showing early attention effects on auditory ERPs in situations that promote auditory streaming (Alain & Woods, 1994) as well as those that place strong attentional demands on the listeners (Woldorff et al., 1993). Together, the P1 and Nd wave data suggest that auditory spatial attention is associated with both the modulation of exogenous sensory responses, and the generation of a slow “processing negativity” that begins at sound onset and remains present through the N1 and P2 waves (Woods, Alho, & Algazi, 1994).

The main objective of the present study was to examine the extent to which sequential and concurrent sound segregation can occur independently of listener attention. We found that the MMN amplitude at the unattended location decreased with increasing task demand within the attended channel. The effects of task demand on the MMN cannot be accounted for by attention-related negativities such as the N2b, because they were present for deviant sounds occurring at the unattended location, which did not require a response from the participants. Our results are consistent with many studies showing an effect of selective attention on the MMN elicited by intensity- (Woldorff et al., 1998; Naatanen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Woldorff et al., 1991), duration- (Alain & Woods, 1994), frequency- (Trejo, Ryan-Jones, & Kramer, 1995), or pattern-deviant stimuli (Alain & Woods, 1997). The fact that a significant MMN was recorded in all conditions, despite a very demanding task, is consistent with the proposal that MMN generation does not depend on listener attention but that attention can modulate the MMN generator (Naatanen et al., 1993). However, we cannot be completely sure that the task, albeit demanding, was optimal in preventing participants from listening to the sounds presented in the unattended location. In a recent study using a more complex auditory scene comprised of three sound sources, we found that the MMN amplitude was abolished for those deviants that occur at furthest from the attended source (Arnott & Alain, 2002). This finding suggests that in auditory attention tasks, the attention “spotlight” may be determined by the spatial distance between the attended and adjacent unattended stream of sounds. This could partly explain the small, but not abolished MMN response in the present study, despite very demanding tasks.

Mistuned stimuli generated an ORN whose latency, amplitude, and scalp distribution were comparable to those obtained in previous studies using similar stimuli (Alain, Arnott, et al., 2001; Alain et al., 2002). In contrast to the MMN data, the ORN did not show significant changes as a function of task demand. This finding is consistent with the proposal that concurrent sound segregation can take place independently of a listener’s attention.

Selective attention may influence sequential integration as indexed by the MMN at various levels of processing. For example, selective attention may strengthen the sensory traces allowing them to remain active for a longer period, thereby easing the extraction of regularities between successive events. Attention may also refine the sensory trace, which, in turn, would facilitate the detection of stimulus deviance. In the present study, increased task demands within the attended channel may impoverish the representation of task-irrelevant stimuli to such an extent that stimulus differences are not as prominent and therefore not easily registered. Another possibility is that the computation of a stimulus difference itself is less efficient in situations that place strong demands on attentional resources. That is, increased perceptual load may have affected the comparison process by making it less precise and less efficient. Assuming a fixed amount of attentional resources (e.g., Lavie, 1995), the smaller MMN in the conjunction condition may indicate less attentional resources allocated to task-irrelevant stimuli because of the increased complexity of the conjunction task.

It has been proposed that selective attention has its effect on processing prior to the MMN-generating system, rather than on the MMN itself (Rinne et al., 2001; Ritter et al., 1999). That is, in tasks that involve strong attention demands, sensory gating may reduce the input to the MMN-generating system. However, this view cannot easily account for the data in the present study because there was no attention effect on the ORN. Presumably, if selective attention affected the input to the MMN generators, then the input to the ORN generators should have been similarly affected, unless the MMN and ORN receive input from distinct auditory pathways. Rather, we propose that selective attention affects the MMN generators.

The results from the present study are consistent with the proposal that the MMN and ORN components index different aspects of auditory scene analysis. Firstly, the effect of task demand was greater on the MMN than on the ORN. This finding suggests that sequential integration, as indexed by the MMN, is more sensitive to attention than is concurrent sound segregation. In addition, the MMN was more frontally distributed than the ORN. This implies different neural generators within the auditory cortices along the Sylvian fissure. Converging evidence from scalp distribution analyses (Giard, Perrin, Pernier, & Bouchet, 1990), neuromagnetic recording (Alho et al., 1996), dipole source modeling (Scherg, Vajsar, & Picton, 1989), and intracerebral recording in nonhuman primates (Javitt, Steinschneider, Schroeder, & Arezzo, 1996) indicate that the scalp-recorded MMN results from neural activity in primary and/or secondary auditory cortices. Furthermore, lesion studies in humans have shown that a unilateral lesion to the superior temporal gyrus reduced the MMN to deviant stimuli presented contralateral to the lesion site, whereas damage to the dorsolateral prefrontal cortex reduced the



MMN amplitude for deviant sounds presented to the ear ipsilateral and contralateral to the lesion site (Alain, Woods, & Knight, 1998). These results suggest that sensory memory, which is critical for the sequential integration of sounds, depends on a fronto-temporal network that includes auditory cortices along the supra-temporal plane and the dorsolateral prefrontal cortex (Alain et al., 1998). The neural generators of the ORN have been less studied. In our previous study, we modeled the scalp-recorded ORN with bilateral dipoles in the medial superior temporal gyrus, which tended to be located more medially than the N1 wave (Alain, Arnott, et al., 2001).

## Conclusions

The role of attention in auditory scene analysis has received considerable attention over the past few years. Although several studies in the areas of developmental psychology (e.g., Trehub et al., 1984; Chang & Trehub, 1977), experimental psychology (e.g., Alain & Woods, 1993; Jones, Macken, & Murray, 1993), and cognitive neuroscience (e.g., Alain, Arnott, et al., 2001) have provided evidence supporting the notion that auditory scene analysis occurs automatically and independently of listener attention, there is also evidence suggesting that attention is critical for the perceptual organization of sounds (e.g., Sussman et al., 2002; Carlyon et al., 2001; Alain & Woods, 1997). Our findings are consistent with the proposal that attention plays an important role in scene analysis. In addition, we found evidence suggesting that different aspects of scene analysis may be more sensitive to attention than others. Sequential integration, as indexed by the MMN, was reduced with increased task demand, whereas the concurrent sound segregation was little affected. This difference in attentional sensitivity may be related to the memory system upon which grouping processes depend on. That is, sequential integration requires processing successive stimuli over several seconds, whereas concurrent sound segregation depends on the integration of acoustic information within hundreds of milliseconds. The difference in attentional sensitivity between sequential and simultaneous grouping processes highlights the intimate link between attention, perception, and memory.

## METHODS

### Participants

Twelve adults provided written, informed consent to participate in the study. The data of two participants were excluded from further analysis because they showed extensive ocular contaminations or had extreme difficulty in distinguishing the different stimuli. Five women and five men form the final sample (aged between 21 and 35 years, mean age =  $26.2 \pm 4.6$  years). All participants had

pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000 Hz (both ears).

### Stimuli and Task

The stimulus sequences consisted of complex sounds with a fundamental frequency of 200 Hz. The sequences were constructed such that half the stimuli were presented to the left ear and half were presented to the right ear in random order. In addition, half of the stimuli were tuned and the other half were mistuned. Tuned stimuli were created by combining 12 pure tones (i.e., 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, and 2400 Hz) with equal intensity. For mistuned stimuli, the third harmonic was shifted either upward or downwards by 16% of its original value (696 or 504 Hz instead of 600 Hz). Within each stimulus type (i.e., tuned and mistuned), there were long duration sounds (standard) and rare, shorter duration sounds (deviant, probability 20% overall, 10% in each ear). Sound duration, including 5 msec rise/fall time, was 100 and 75 msec for the standard and deviant sounds, respectively. Stimuli were generated digitally with a sampling rate of 50-kHz using a 16-bit Tucker Davis System converter and played through Sennheiser HD 265 headphones at 80 dB SPL. Stimuli were presented in random order at variable stimulus onset asynchrony (SOA 200–400 msec with 10-msec steps, rectangular distribution) in blocks of 500 sounds (400 standards, 100 deviants). Participants were presented with 12 blocks of trials.

Each participant took part in a single feature and feature conjunction detection tasks (six blocks of trials in each task). In both tasks, participants were asked to focus their attention to one location (e.g., left ear) in order to detect infrequent target sounds at that site. The stimuli were presented at a high rate to promote the segregation of the sequence into two streams (one in each ear) and to ease the focus of attention at the designated location. The stimulus set was the same in both tasks; task difficulty was manipulated only by varying task instructions. In the single feature detection task, participants were asked to press a button whenever they heard shorter duration sounds (tuned or mistuned). The importance of speed and accuracy were equally emphasized to the participants. In the feature conjunction condition, participants were again asked to press a button for shorter duration sounds, but they were also required to identify whether the sound was tuned or mistuned by pressing button “one” or “two” respectively. The order of listening conditions was counterbalanced across participants.

### *Electrophysiological Recording and Analysis*

The electroencephalogram (EEG) was recorded from an array of 64 electrodes including those from the standard 10–20 placements. Vertical and horizontal

eye movements were recorded with electrodes at the outer canthi and at the superior and inferior orbit. Electrophysiological signals were digitized continuously (bandpass 0.05–50 Hz; 250 Hz sampling rate) via NeuroScan SynAmps and stored for off-line analysis. During the recording, all electrodes were referenced to Cz; for data analysis, they were rereferenced to an average reference, and the electrode Cz was reinstated.

The analysis epoch included 200 msec of prestimulus activity and 800 msec of poststimulus activity. Trials contaminated by excessive peak-to-peak deflection ( $\pm 100 \mu\text{V}$ ) at the channels not adjacent to the eyes were automatically rejected before averaging. The ERPs were then averaged separately for each site, stimulus type, and listening conditions. ERPs were digitally high-pass filtered at 0.1 Hz and low-pass filtered to attenuate frequencies above 15 Hz. For each individual average, the ocular artifacts (e.g., blinks and lateral movements) were corrected by means of ocular source components using the Brain Electrical Source Analysis (BESA) software (Picton, van Roon, et al., 2000; Berg & Scherg, 1994).

The behavioral data (hits, FAs, and RTs) were subjected to a within-subject, repeated-measures analysis of variance (ANOVA) with task and attended ear as factors. A hit was defined as a button press between 150 and 1200 msec after target onset. A FA referred to a wrong button press or a correct button press occurring outside the hit window.

For the ERP data, the independent variables were task type (single, conjunction), attention (attended, unattended), stimulus type (standard deviant) or mistuning (tuned, mistuned), and electrode. Unless otherwise specified, the effects of attention on ERPs were quantified using a subset of electrodes over the frontocentral region (Fz, F1, F2, FCz, FC1, FC2, CZ, C1, and C2) and at left and right temporo-parietal regions (TP9/10, T7/8). For each participant, ERP waveforms were measured by computing mean amplitude values in selected latency regions, relative to the mean amplitude of the 200 msec prestimulus activity. Scalp topographies using the 61 electrodes (omitting the periocular electrodes) were statistically analyzed after scaling the amplitudes to eliminate amplitude differences between stimuli and conditions (McCarthy & Wood, 1985). Whenever appropriate, the degrees of freedom were adjusted with the Greenhouse–Geisser epsilon ( $\epsilon$ ). All reported probability estimates are based on these reduced degrees of freedom.

## Acknowledgments

We are thankful to Yu He and Kelly McDonald for their technical assistance. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes for Health Research. Funding for the equipment used in this project was derived from a Canada Foundation for Innovation and Ontario Innovation Trust (CFI/OIT) to the University of Toronto Functional Imaging Research Network (FIRN).

Reprint requests should be sent to Claude Alain, PhD, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1, or via e-mail: calain@rotman-baycrest.on.ca.

## REFERENCES

- Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, *5*, D202–212.
- Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1072–1089.
- Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *NeuroReport*, *10*, 2429–2434.
- Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2001). Age-related changes in detecting a mistuned harmonic. *Journal of the Acoustical Society of America*, *109*, 2211–2216.
- Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America*, *111*, 990–995.
- Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., & Taylor, M. J. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cognitive Brain Research*, *16*, 210–218.
- Alain, C., Woods, D., & Ogawa, K. (1994). Brain indices of automatic pattern processing. *NeuroReport*, *6*, 140–144.
- Alain, C., & Woods, D. L. (1993). Distractor clustering enhances detection speed and accuracy during selective listening. *Perception and Psychophysics*, *54*, 509–514.
- Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception and Psychophysics*, *56*, 501–516.
- Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, *34*, 534–546.
- Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research*, *812*, 23–37.
- Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Naatanen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, *77*, 151–155.
- Alho, K., Tervaniemi, M., Huottilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., Knuutila, J., & Naatanen, R. (1996). Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology*, *33*, 369–375.
- Alho, K., Tottola, K., Reinikainen, K., Sams, M., & Naatanen, R. (1987). Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *68*, 458–470.
- Arnott, S. E., & Alain, C. (2002). Falling outside the spotlight: Effect of spatial attention on the MMN. *NeuroReport*, *13*, 2209–2212.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*, 229–241.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sounds*. London: MIT Press.
- Bregman, A. S., & Rudnick, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 263–267.

- Carlyon, R. P., Cusack, R., Foxtton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology. Human Perception and Performance*, *27*, 115–127.
- Chang, H. W., & Trehub, S. E. (1977). Infants' perception of temporal grouping in auditory patterns. *Child Development*, *48*, 1666–1670.
- Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, *51*, 167–187.
- Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, *27*, 627–640.
- Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, *49*, 277–290.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*, 177–180.
- Hulse, S. H., MacDougall-Shackleton, S. A., & Wisniewski, A. B. (1997). Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, *111*, 3–13.
- Javitt, D. C., Steinschneider, M., Schroeder, C. E., & Arezzo, J. C. (1996). Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia. *Proceedings of the National Academy Sciences, U.S.A.*, *93*, 11962–11967.
- Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*, 318–328.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology. Human Perception and Performance*, *21*, 451–468.
- MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *Journal of the Acoustical Society of America*, *103*, 3581–3587.
- Maiste, A. C., & Picton, T. W. (1987). Auditory evoked potentials during selective attention. In C. Barber & T. Blum (Eds.), *Evoked potentials: III. The third international evoked potentials symposium* (pp. 385–391). Stoneham, MA: Butterworth.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
- Moore, B. C., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, *80*, 479–483.
- Moore, B. C., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America*, *77*, 1861–1867.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Naatanen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*, 436–450.
- Okita, T. (1989). Within-channel selection and event-related potentials during selective auditory attention. *Psychophysiology*, *26*, 127–139.
- Parasuraman, R. (1980). Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. *Biological Psychology*, *11*, 217–233.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-otology*, *5*, 111–139.
- Picton, T. W., van Roon, P., Armiljo, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, *111*, 53–65.
- Rinne, T., Antila, S., & Winkler, I. (2001). Mismatch negativity is unaffected by top-down predictive information. *NeuroReport*, *12*, 2209–2213.
- Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan, H. G., Jr. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, *36*, 835–838.
- Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, *1*, 326–355.
- Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Naatanen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Brain Research. Cognitive Brain Research*, *13*, 393–405.
- Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*, 821–830.
- Trejo, L. J., Ryan-Jones, D. L., & Kramer, A. F. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology*, *32*, 319–328.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *90*, 8722–8726.
- 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.
- Woods, D. L., & Alain, C. (2001). Conjoining three auditory features: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *13*, 492–509.
- Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology. Human Perception and Performance*, *27*, 65–74.
- Woods, D. L., Alain, C., & Ogawa, K. H. (1998). Conjoining auditory and visual features during high-rate serial presentation: Processing and conjoining two features can be faster than processing one. *Perception and Psychophysics*, *60*, 239–249.
- Woods, D. L., Alho, K., & Algazi, A. (1994). Stages of auditory feature conjunction: An event-related brain potential study. *Journal of Experimental Psychology. Human Perception and Performance*, *20*, 81–94.