SELECTIVELY ATTENDING TO AUDITORY OBJECTS

Claude Alain and Stephen R Arnott

Rotman Research Institute, Baycrest Centre for Geriatric Care, and Department of Psychology, University of Toronto, Canada

TABLE OF CONTENTS

1. ABSTRACT

The ability to maintain a conversation with one person while at a noisy cocktail party has often been used to illustrate a general characteristic of auditory selective attention, namely that perceivers' attention is usually directed to a particular set of sounds and not to others. Part of the cocktail party problem involves parsing co-occurring speech sounds and simultaneously integrating these various speech tokens into meaningful units ("auditory scene analysis"). Here, we review auditory perception and selective attention studies in an attempt to determine the role of perceptual organization in selective attention. Results from several behavioral and electrophysiological studies indicate that the ability to focus attention selectively on a particular sound source depends on a preliminary analysis that partitions the auditory input into distinct perceptual objects. Most findings can be accounted for by an object-based hypothesis in which auditory attention is allocated to perceptual objects derived from the auditory scene according to perceptual grouping principles.

2. INTRODUCTION

In most everyday situations, there is often more than one audible sound source at any given moment. For example, during a quiet walk around the neighborhood you may hear birds singing, a dog barking and a car's engine. You may decide to listen carefully to the birds or you may pay attention to the car passing. Selective attention is this essential component of our day-to-day life that enables us to preferentially process a particular set of sounds (e.g., the birds singing) at the expense of other sounds. Although this is an ability that we take for granted, the psychological and neural mechanisms underlying selective attention are not well understood. Part of the difficulty in understanding selective attention resides in defining the "object"

of attention and in clarifying the role of selective attention in forming and localizing auditory objects.

Although we can selectively focus our attention to a particular frequency or location, in this paper we will argue that auditory attention operates on perceptual objects (1-5). The term auditory object refers to a mental description of a sound source in the environment rather than the source itself or the sounds it emits (6). Here, a sound source is defined as a physical entity that generates an acoustic wave. An auditory object, on the other hand, is the percept of a group of sounds as a coherent whole seeming to emanate from a single source. The perception of a physical sound source in the environment has also been referred to as an auditory event, auditory entity, and auditory stream (6-8). In the present review, we make a distinction between an auditory object and an auditory event. While the former refers to a perception of a sound source and its behavior over time, the latter is used when referring to the perceptual dimension of hearing a sound that is occurring at a particular time, in a particular space and having particular attributes (e.g., intensity, duration, timbre). The event can be part of a larger entity, i.e., the auditory object (9). For example, in dichotic experiments the co-occurring conversation would be defined as the auditory objects (i.e., each is perceived to originate from a separate auditory source) whereas the elements (e.g., listener's own name) within the conversation would be defined as events.

One possible role of selective attention may be to link together auditory input at the focus of attention in order to create a perceptual object (10). In the above example, selectively attending to a particular frequency region (e.g., high frequency sounds) may allow the integration of sounds within this region, so that a separate, and a potentially meaningful, auditory object (i.e., the birds singing) can be accurately perceived and identified

from other sound sources (e.g., the dog barking). Given that perception of auditory objects often involves the integration of sounds over time, selective attention may be required to link together auditory stimuli into meaningful units. This view is akin to the feature integration theory in which object features are initially represented independently in different feature maps and then are bound together through attention to the object's location in space (10, 11). If the system is operating optimally, only features that occupy the same location in the attended space are bound together (i.e., are perceived to be features of the same object). However, if attention is not adequately focused on an object's location, then the features of that object are more likely to remain unbound, or perhaps be miscombined with those of another, producing feature conjunction errors, also called illusory conjunctions.

Another view suggests that auditory attention may be brought to bear after the preliminary analysis of objects has occurred, and that one object at a time is then selected for a complete conscious analysis (6, 12, 13). From this perspective, object formation occurs without attention or awareness, the role of attention being to "bring" one of those objects to conscious experience. Taken to the one extreme, it implies that our perception is imposed on us by the incoming acoustic data. The attended object stands out perceptually while the rest of the sounds become less prominent as in the figure-ground phenomenon in visual perception. In the cocktail party example, the conversation that we attend to stands out as the "figure" while the other conversations recede to the background. With these two views in mind, we will review the results from behavioral and electrophysiological studies of selective attention with an emphasis on the interaction between selection, attention and perceptual organization.

3. DICHOTIC LISTENING EXPERIMENTS

Research on auditory selective attention considers how listeners can process simultaneous sources of information and how attention affects the processing of task-relevant and task-irrelevant stimuli. This is evaluated primarily by requiring listeners to attend to a particular set of sounds in the presence of other unwanted sounds. In early and now classic behavioral studies, auditory selective attention was examined during dichotic listening tasks in which different auditory messages (e.g., prose) were presented simultaneously to each ear via headphones. The participants' task was to shadow one message presented at a designated ear and simultaneously ignore the message presented in the opposite ear. Typically, shadowing performance improved with increasing physical difference between the wanted and unwanted speech sound. For example, an increase in spatial separation between two concurrent messages improves performance in identifying the task-relevant message (14). Similarly, increasing the distance between the frequency bands of the two messages also improved individual's ability to attend to either of the messages selectively (14). It is also easier to ignore two irrelevant messages when they come from the same spatial location than when they occur in two separate locations (15). Importantly, Treisman found that a decrease in spatial discriminability between the two task-irrelevant messages does not cause any increase in interference on shadowing

performance. On the contrary, performance improves slightly as the irrelevant messages are brought closer together in location.

In shadowing experiments, individuals are also often unable to report the contents of the unattended ear after the shadowing task has ended, although they are usually aware if the voice presented to the unattended ear switched gender (i.e., pitch), or was replaced with a 400-Hz tone (16, 17). However, in some cases information arriving at the so-called "unattended" channel (i.e., the ear to be ignored) can be processed at a semantic level of representation. For example, listeners are usually able to respond to the occurrence of their own name, or the meaning of a significant item presented in the unattended ear (18-20). This suggests that stimuli that are highly pertinent to one's self interest, such as one's name, can capture a listener's attention when presented in the unattended location (18, 21). Individuals who recalled hearing their own name also showed an increase in shadowing errors and a decrease in shadowing speed, following the presentation of their own name (21). Although the same individuals also reported that their attention occasionally wandered to the irrelevant message, the pattern of errors could not be easily accounted for in terms of a lapse of attention because the changes in shadowing speed occurred only after the listener had heard their own name (21). The results appear more consistent with the idea that, in some individuals, hearing their own name momentarily captured their attention thereby reducing shadowing performance. Although the content of the task-irrelevant message may cause some interference, it is important to keep in mind that its effect on performance is usually small and occurs only for a subset of individuals who also report that their attention wanders to the irrelevant message.

Findings from dichotic listening studies demonstrate that the ability to focus attention selectively on a particular message depends on the acoustical factors that promote the segregation of co-occurring speech stimuli into distinct groups of sounds thereby improving the speech discriminability. They can also be taken to suggest that the analysis of a simple physical attribute, which usually defines the information channel, may occur automatically and that the bottleneck in processing auditory information arises in a subsequent stage, after the co-occurring sounds have been perceptually segregated into distinct sources.

4. AUDITORY SCENE ANALYSIS

The results from dichotic listening experiments are consistent with a stage model of attention in which selective attention operates on the output from a pre-attentive stage of analysis involved in forming and localizing auditory objects ("auditory scene analysis"). Bregman proposed two groups of processes involved in auditory scene analysis (6). One is a pre-attentive process that partitions the auditory input according to gestalt principles, such as grouping by physical similarity, temporal proximity, and good continuity. Sounds are more likely to be assigned to separate sources if they differ widely in frequency, intensity, and spatial location. The other is a schema-driven attention process that "searches" for patterns in the acoustic data. While the preattentive processes group sounds based on physical similarity, the schema-driven
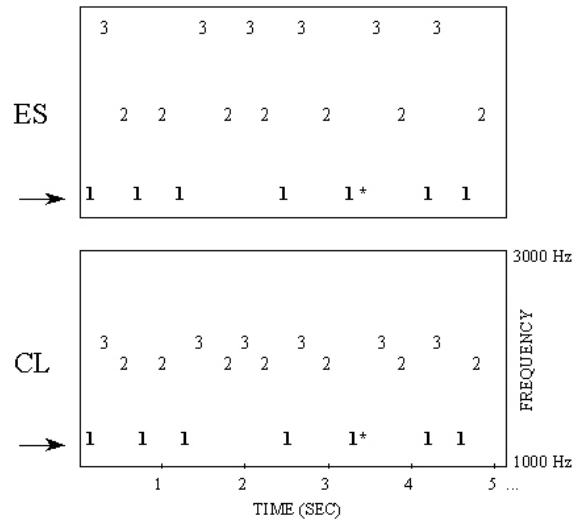
Figure 1. Schemata of the stimuli presented in the two different clustering conditions. Arrows indicate the frequency to be attended in each condition. A target is shown with an asterisk. (Adapted from Alain and Woods, 1993). ES = Evenly Spaced condition. CL = Clustered condition.

processes use prior knowledge to extract meaning from the acoustic data. As such, the schema-driven process depends on representations of previous experiences that have been acquired through learning and a matching process between those representations and the complex acoustic wave that reach our ears. The use of context and/or knowledge is particularly evident in an unfavorable signal-to-noise listening situation such as in the cocktail party example. In the laboratory, individuals correctly perceive more words when the sentence final words are predictable (i.e., fit with the context) than unpredictable (i.e., do not fit with the context) (22). Schema-driven processes provide a way to resolve perceptual ambiguity in complex listening situations.

4.1. Grouping of task-irrelevant material

Central to the auditory scene analysis account is the notion that object formation can occur independently of a listener's attention. One way to examine whether object formation can occur outside the focus of attention is to manipulate the physical similarity between task-irrelevant stimuli to promote their perceptual grouping. Presumably, if the clustering of task-relevant material influences performance, then one can assume that the unattended sounds were perceptually organized in some way. To our knowledge, Bregman and Rudnicky were the first to examine whether the ability to attend and process elements within a particular stream of sounds could be influenced by the perceptual relations of task-irrelevant materials (23). Participants were asked to judge the sequential order of two sounds differing in pitch (targets), each flanked by an identical lower frequency sound referred to here as "flanker tones." The target and flanker tones were kept constant while additional "captor" tones of still lower or identical pitch preceded and followed the block of four tones.

Bregman and Rudnicky noted that the ability to judge the sequential order of the target sounds could be improved significantly by increasing the frequency similarity between the captor and flanker tones. The magnitude of the captor effect has been shown to vary with temporal predictability (24). Bregman and Rudnicky proposed that captor and flanker tones that are similar or identical in pitch form a perceptual stream that separates them from the target tones because listeners tend to perceptually group together tones that are nearest in frequency and share a similar temporal structure. This causes the target tones to be perceived as a separate auditory object and makes it much easier for the listener to make a judgment about their sequential order.

However, factors other than perceptual grouping could have contributed to the observed captor effect in Bregman and Rudnicky's study. Many studies have shown that the ability to detect or judge the duration of a particular target stimulus can be improved by providing listeners with advance information indicating the frequency or the location of the upcoming target (1-3, 25). Usually, cues that provide accurate information about the frequency or the spatial location of the target lead to greater accuracy and faster target detection. Similarly, in selective attention tasks using rapid serial auditory presentation, targets preceded by distractor tones that share a similar location or frequency can reduce target reaction time (26, 27). Thus, in Bregman and Rudnicky's study the captor tones may have acted as a cue or a prime, improving performance when the captors were similar or identical in frequency to flanker tones than when they differed widely in frequency with the flankers, and consequently with the target sounds.

The idea that auditory objects can be formed outside the focus of attention was further examined in a series of experiments by Alain and Woods (28). Participants were presented with a rapid sequence of binaural stimuli varying randomly along three different frequencies (see Figure 1). The participants' task was to focus their attention on one of the extreme frequencies (the middle tones were never to be attended) in order to detect infrequent deviant (target) sounds differing slightly from the standard stimuli. The discriminability between standard and deviant sounds was adjusted so that participants had to first pay attention to the most salient dimension (i.e., frequency) in order to be able to detect rare target stimuli differing from the standard in duration (Experiment 1) or intensity (Experiment 2). In the evenly spaced condition, the tones composing the sequence were equally spaced in frequency. In the clustered condition, the high or low frequency tones were increased or decreased in frequency (depending on the condition) so that they were one semitone apart from the middle frequency. This manipulation promotes the perceptual grouping of the middle tones and the extreme tones, based on frequency similarity. If selective attention is necessary for perceptual grouping, then the clustering of task-irrelevant stimuli should not affect performance. On the other hand, if object formation can occur outside the focus of attention then one would expect improvements in performance with distractor clustering. Participants were faster and more accurate in detecting target sounds when the distractor stimuli were

clustered together, despite increases in frequency similarity between one of the distractors and the target stimuli. To dissociate the effects of clustering from the effects of priming or cuing, the effect of the preceding stimulus on response time was examined in both clustering conditions. A sequential analysis of targets preceded by middle distractors revealed faster RTs for the clustered condition, even if the middle distractor was identical in both conditions. This shows that cuing or priming alone cannot account for the difference in performance between the two conditions. Rather, the results are consistent with the proposal that the increase in similarity between the two irrelevant frequencies allowed them to be perceptually grouped into a separate object from the relevant frequency. The effects of clustering task-irrelevant material on performance during selective attention tasks are robust, being present in both synchronous and isochronous sequences of stimuli as well as in both young and older adults (28, 29).

The preceding studies indicate that clustering of task-irrelevant material modulates performance during selective listening. Although these findings are consistent with the proposal that object formation may occur outside the focus of attention, a possible effect of attention cannot be ruled out. Evidence from scalp recordings of cortical evoked brain activity and functional neuroimaging studies has shown that selective attention not only modulates the processing of task-relevant stimuli but also modifies neural activity elicited by task-irrelevant sounds (30-33). Such findings are consistent with a dual-process of attention in which the processing of task-relevant stimuli may be facilitated while the processing of task-irrelevant stimuli may be suppressed during selective listening (34). In the studies described above, the clustering effects might therefore have been mediated by an active inhibition of the unwanted sounds. Clustering may make this inhibition easier to maintain because both nearby distractor frequencies could be inhibited simultaneously (34). To determine whether object formation can occur outside the focus of attention, one must examine auditory pattern processing in situations that entail neither active selection nor active rejection of the auditory stimuli.

4.2. Electrophysiological studies of auditory organization

The extent to which auditory patterns are automatically processed by the human brain has been recently investigated using human event-related brain potentials (ERPs). ERPs are particularly well-suited for studying automatic processes because they provide a real-time measure of information-processing throughout the auditory system, even for those sounds that are outside the focus of attention. When individuals are presented with a sequence of standard and deviant stimuli while they perform another task such as reading, rare deviant sounds elicit a negative wave (mismatch negativity or MMN) that superimposes the N1-P2 deflection. The MMN amplitude peaks between 120 and 220 ms following deviant onset and is maximum over the frontocentral region. Scalp topography analysis, dipole source modeling, animal models and lesion studies in humans all provide converging evidence consistent with generators in auditory cortices with contribution from the dorsolateral prefrontal cortex (35-40).

It is now well documented that deviant stimuli differing from repetitive standard stimuli along a physical dimension such as frequency, intensity, or spatial location elicit an MMN response. More relevant to the present review are the studies showing an MMN to changes in auditory patterns. For example, deviation from simple auditory patterns such as sequences of tones that alternated or decreased regularly in pitch elicited an MMN wave with maximum amplitude over the frontocentral region (41-43). The MMN is also elicited by changes in the sequence of four tones varying regularly in frequency or in frequency and spatial location (44, 45). The scalp topography elicited by pattern-deviant stimuli is usually more centrally distributed than the one elicited by deviant stimuli differing from the standard along a physical dimension such as frequency (46). This indicates that auditory pattern processing engages neural circuits that are distinct from those involved in simple detection of changes in physical attributes.

The MMN to pattern deviant stimuli depends on stimulus-related factors, such as the rate of presentation and the frequency separation of the elements that compose the pattern (41, 47). That is, the MMN to pattern deviant stimuli is larger when the stimuli are presented at short ISIs and when the tones composing the pattern are clearly distinguishable in pitch. The amplitude and latency of the MMN also vary with the magnitude of deviation in the spectral and temporal transition between consecutive elements within the pattern. An enhanced MMN amplitude and decreased MMN latency are associated with a greater magnitude of deviation (44). Similarly, performance in detecting target sounds that are inconsistent with pattern structure depends on the magnitude of violation (44, 48).

The MMN wave to pattern deviant stimuli is thought to reflect a neural mismatch between the incoming stimulus and the expected stimulus based upon the organization of the previous stimuli. In other words, the MMN to pattern deviant sounds reflects a violation of expectancy established by the preceding stimuli. Evidence from scalp topography analysis suggests that an auditory pattern is encoded as a Gestalt including both frequency and temporal transition among the elements composing the pattern (44).

The fact that an MMN to pattern deviant stimuli can be recorded when participants are not actively attending to auditory stimuli suggests that some pattern recognition can occur during passive listening. These findings provide some support for the auditory scene analysis account in which object formation is thought to occur at an early stage of processing, independently of listeners' attention. However, the MMN results should be interpreted with caution because in most studies, listeners' attention was not well controlled. For example, participants were often required to read a book of their

choice and little effort was made to verify whether they carried out this task. When participants' attention is monitored, such as during an auditory selective attention task, the MMN to pattern deviant sounds presented in the unattended location is markedly reduced in amplitude compared to the MMN elicited by deviant sounds occurring at the attended location or during reading (45). The MMN amplitude differences between the visual and auditory tasks may reflect differences in the amount of attention required in the two situations or differences related to the within vs. cross-modal nature of the attentional demand. It may be easier to automatically detect deviant auditory stimuli when attention is allocated to visual stimuli than auditory stimuli because in the former case attention may tap into a different "pool" of resources.

The MMN is also larger in amplitude when participants have extended experience with the auditory material. Koelsch, Schroger, and Tervaniemi, compared the MMNs elicited by major chords and single tones in professional musicians and non-musicians (49). Slightly impure chords presented among perfect major chords elicited a distinct MMN in professional musicians, but not in non-musicians. This finding may indicate a greater sensitivity to tonal difference in musicians, which most likely results from extensive musical training. Similar results were found for speech sounds. The MMN is larger for acoustical changes made in listeners' own language than for similar acoustic changes made in a different language (50).

To sum up, the MMN to pattern-deviant stimuli is sensitive to stimulus-related factors that promote the formation of auditory objects such as the rate of presentation and frequency separation among the elements composing the object. The MMN is also sensitive to subject-related factors such as listeners' knowledge and prior experience with the auditory material. The results are consistent with the notion that object formation can occur outside the focus of attention and suggest that both preattentive and schema-driven processes interact early during auditory scene analysis.

5. ELECTROPHYSIOLOGICAL STUDIES OF AUDITORY SELECTIVE ATTENTION

The paper will now turn to electrophysiological studies of auditory selective attention. The recording of ERPs has been one of the most widely used neuroimaging techniques in studying the neural basis of auditory selective attention (31, 34, 51, 52). The typical paradigm involves a rapid serial presentation of auditory stimuli varying along two or three physical dimensions. For example, participants may be presented with low or high tones in the left or right ear in a random order and required to detect infrequent longer duration tones at a designated frequency and location (targets).

In auditory selective attention tasks, ERPs to attended tones show a negative displacement that overlaps with the N1 wave relative to the ERPs elicited by the same tones when they are unattended (53). The attended target elicits an additional longer latency positive deflection peaking between 300-500 ms referred to as the P300 or P3b. The effects of selective attention on auditory evoked potentials are often illustrated by subtracting the ERPs to unattended stimuli from the ERPs to the same stimuli when they are attended. The resulting negative difference (Nd) wave has at least two partially overlapping components, an early component (Nde) and a late component (Ndl) that peak at 200 and 400 ms post-stimulus, respectively (54).

The Nde is more closely related to the discriminability of attended and unattended sequences than to the discriminability of the standard and target stimuli within the attended sequence. For example, when listeners are presented with two concurrent sequences of tone bursts and are required to respond to occasional targets in one of them, the latency and amplitude of the Nde varies with the distinctiveness of the sequences -- not with the distinctiveness of standard and target tone bursts within the attended sequence (55). When the task-relevant and task-irrelevant tones are highly distinctive and presented at short ISIs (e.g., 200-400 msec), the selective attention effects on ERPs can occur at latencies as short as 30 msec following stimulus onset (56, 57). The attention-related changes in ERPs occur when participants attend to a particular stream of sounds in the presence of one or more different streams of distracting stimuli and when stimulus sequences are easily discriminated, whether they are distinguished by spatial position, frequency, or both spatial position and frequency (58-60). Although there is one report of the habituation of the Nde (61), performance and Nde amplitude are usually maintained over long sessions (59, 62, 63). The Nde amplitudes and latencies are also little affected by repeated testing (34, 64).

5.1. Stimulus set account

Several models have been proposed for this attention-related negativity. Hillyard and colleagues suggested that the Nde reflects an early stimulus selection based upon easily discriminable information that defines the information channel (65,66), i.e., stimulus set (67). From this perspective, stimuli are accepted or rejected based on their dominant relevant physical attributes, and further perceptual analysis is contingent on this early selection. This selection process would be possible when attended and unattended stimuli differ in some highly distinguishable physical attributes, such as frequency and/or location. Hillyard and colleagues also proposed that the Ndl reflects a late selection of targets from non-targets within the attended channel based on less discriminable cues. Such an account (referred to here as the feature-based account) assumes that the physical dimensions of the target sounds can be used by the executive or control system to set up an appropriate attentional "filter". According to this model, distractors physically similar to a target produce interference because they fall within the attentional filter, and therefore are selected along

with targets for further processing and a potential response. By contrast, distractors highly distinguishable from the targets fall outside the attentional filter, and are easily dismissed (4, 68).

5.2. Attentional trace account

Naatanen proposed a more specific interpretation in which the Nd_e reflects an early selection of stimuli that is based upon a gradual comparison process between the sensory input and an attentional trace (51). The attentional trace refers to a temporary neuronal representation of the distinctive features of the task-relevant stimuli that is actively formed and maintained during selective listening and supports identification of the stimuli that must be further processed for a potential response. According to Naatanen, all incoming stimuli are compared to the attentional trace. The comparison process generates a surface negative potential (the processing negativity or PN), with the duration of the comparison process depending on the similarity between the stimulus and the attentional trace. The Nd reflects the difference between the PN elicited by the comparison process executed when confronted with matching (task-relevant) and mismatching (task-irrelevant) stimuli. The onset of the Nd directly reflects the time needed to stop comparing the task-irrelevant stimuli to the trace, and is earlier for distractors differing substantially from the targets than for distractors similar to the target. In addition, Naatanen proposed that the Nd_l might reflect either a late selection of the target within the attended channel or a rehearsal of the attentional trace.

5.3. Object-based account

The attentional trace model implies that each transient stimulus is treated as an isolated event independent of the sequential context. In most experimental situations involving one attended and one unattended stimulus sequence (e.g., dichotic listening tasks), the obvious perceptual information available to orient processing is that provided by the physical features of individual stimuli. In such experimental situations, all other potential information is confounded with the difference between the physical features of the stimuli. Realistic listening situations contain information on continuity and other interstimulus relationships that promote the formation of auditory objects.

An alternative hypothesis, that we call the object-based hypothesis, is introduced here to account for the selective attention effects on auditory evoked potentials. According to this hypothesis, the observer's attention is allocated to an auditory object, as opposed to a feature per se. Although the basis of such an object-based account is theoretically different from a feature-based one, an auditory object may also be defined by its physical features, such as pitch and location. Thus, selectively attending to a particular auditory object should also result in attention-related changes in those brain areas involved in processing the properties of the object, such as frequency and location. According to an object-based view, only those parts and properties defining the object should receive preferential processing over other competing stimuli. These two alternative models can be distinguished by contrasting the effects of perceptual context with the effects of physical similarities on the Nd wave.

In a series of experiments, we have begun to examine the role of perceptual grouping on the ERP attention effects. In one set of experiments, participants were presented with a sequence of tones varying randomly between four different frequencies (69). They were asked to attend to the lowest or highest tone frequency and to detect occasional longer duration tones at that particular frequency. The middle tones were never to be attended. Two clustering conditions were used. An evenly spaced condition in which the four tones were equally spaced along the musical scale and a clustered condition in which the two middle frequencies were closer to the extreme frequencies. In the first experiment, the two extreme frequencies were kept constant whereas the two middle frequencies were manipulated between conditions. An earlier attention effect was found when the middle frequencies were clustered with the extreme frequencies. A closer examination of the ERPs revealed that clustering effects on ERPs were mediated primarily by a decrease in PN elicited by the task-irrelevant stimuli, consistent with the proposal that distractor clustering reduces interference on target detection (28). The results are also consistent with the idea that only the parts and properties of the attended object received extended processing and that the processing allocated to a particular sound feature does not depend solely on the physical similarity between the attended and distractor stimuli.

The role of perceptual organization on the ERP effects of attention was examined further by contrasting the effect of grouping against the effects of physical similarity. As mentioned earlier, the clustering of distractors during selective listening can improve performance in detecting infrequent targets at a designated frequency (28). Clustering effects are thought to reduce the interference of task-irrelevant material so that maintaining their focus of attention on the relevant stream of sounds is easier for listeners. This implies that perceptual organization of sounds would affect the processing of both task-relevant and task-irrelevant material. To test this hypothesis, Alain and Woods used a paradigm similar to the one illustrated in Figure 1 (70). Participants were presented with a rapid sequence of binaural stimuli varying along three different frequencies. In the evenly spaced condition, the tones composing the sequence were equally spaced in frequency. In the clustered condition, one extreme frequency (either the high or low, depending on the condition) was moved closer to the middle frequency. This manipulation was thought to promote the grouping of middle tones with the extreme tones that were moved closer. The participants' task was to focus their attention to the extreme low frequency tones for half the session and to the high frequency tones for the other half of the session. Again, the middle tones were never to be attended. According to the object-based view, one should expect both an improvement in performance and enhanced Nd amplitude in the clustered condition because the clustering of distrators should ease their segregation into a separate auditory object. This
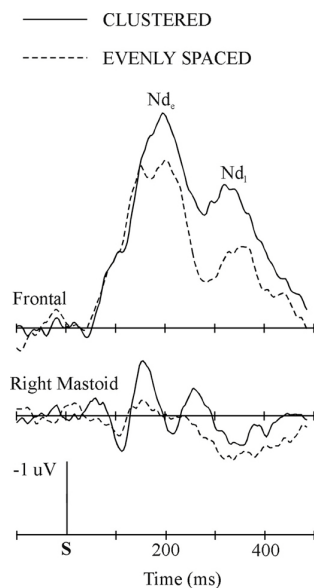
Figure 2. Negative difference (Nd) wave obtained in both evenly spaced (ES) and clustering (CL) conditions at the midline frontal electrode and the right mastoid electrodes. Adapted from Alain and Woods (1994). The vertical bar indicates stimulus onset. In this and the subsequent figures, negativity is plotted upward.
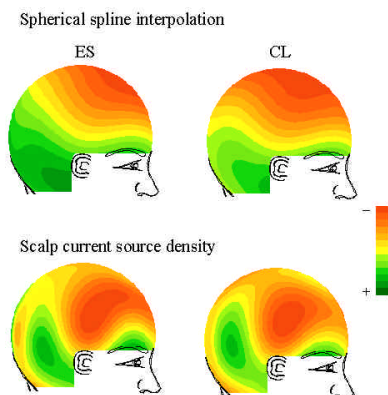
Figure 3. Isopotential color maps (top) and scalp current density mapping (bottom) of the normalized distribution of the negative difference (Nd_e) wave as a function of the clustering condition.

should increase the figure-ground separation, thereby, easing the allocation of attention to the relevant stream of sounds.

As predicted, Alain and Woods found a significant effect of frequency clustering on both performance and Nd wave. That is, participants were faster and more accurate in detecting infrequent targets in the attended stream when the distractor tones were grouped together. Figure 2 shows the effects of clustering on the Nd wave obtained at the midline frontal electrode (i.e., Fz) and the right mastoid. For the evenly spaced and the clustering conditions, an early attention effect was

found beginning at 50 ms post-stimulus. In both conditions, the Nde onset was similar but the ERP attention effect was larger when the two distractor frequencies were clustered together than when the attended and the distractor frequencies were evenly spaced. Both isopotential and scalp density of the Nde peak amplitude mappings were consistent with generators in auditory cortices along the supra temporal plane (Figure 3). The Nde in the clustered condition was associated with stronger current sources over the temporal regions than the Nde in the evenly spaced condition. These findings suggest that auditory cortices play an important role in grouping sounds that are similar in frequency.

In the experiments discussed so far, the effects of perceptual grouping on the Nd were examined only by varying the frequency similarity between the elements composing the sequence. We have recently begun to examine whether the clustering of auditory elements based on spatial location would also modulate the amplitude of the Nd wave. In a preliminary study, six participants were presented with broadband noise bursts at three possible azimuth locations (see Figure 4). In the evenly spaced condition, the stimuli were presented either at 60o left, center, or 60o right relative to the listeners' head. In the clustered condition, either the middle location was moved closer to the extreme location or the extreme location was moved closer to the middle location, so that locations were separated by 30o. The clustering of distractors based on spatial location generated a similar pattern of results as those observed with clustering frequency. That is, we found a larger Nd wave when the distractors were clustered together based on spatial location (Figure 5). In fact, this is consistent with an early observation made by Treisman who noticed that performance in shadowing improves slightly when the task-irrelevant messages were brought closer together in location (15). The effect of selective attention was larger over the hemisphere contralateral to the attended location (Figure 6). As in the previous study, the Nd amplitude was more centrally distributed in the clustered than in the evenly spaced condition, indicating that different generators may be active in situations that promote the formation of auditory objects.

We have reviewed studies that show that perceptual context (as manipulated by the frequency or spatial separation among the stimuli composing a sequence) modulates the amplitude of potentials (ERPs) recorded during selective attention tasks. These results imply that the selection of auditory stimuli is affected by the perceptual context in which they are embedded and that selection does not depend solely on the physical attributes of the stimuli. Moreover, they are consistent with an object-based account of attentional filtering in which attention is first allocated to an auditory object. This object-based account can address some basic aspects of selective attention that otherwise remain puzzling. For instance, individuals can easily monitor, or pay attention to, a source that is varying in spatial location and frequency. Theories assuming attentional filtering based on the physical attributes, i.e., a feature-based one, cannot easily describe how individuals can perform in this way.
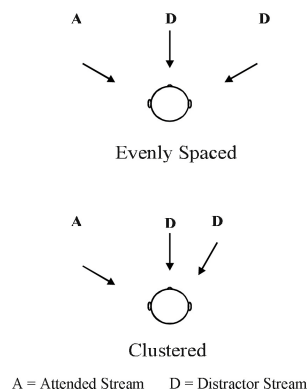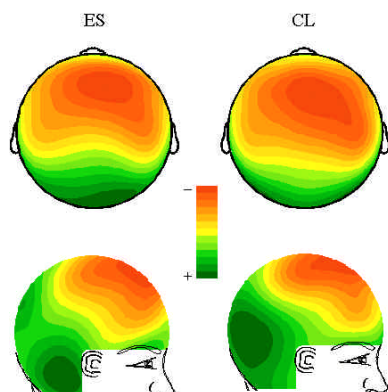
Figure 4. Schemata of the stimuli presented in the two different clustering conditions.
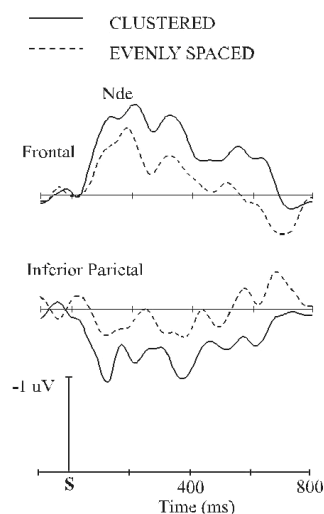
Figure 5. Negative difference (Nd) wave obtained in both evenly spaced (ES) and clustering (CL) conditions at the frontal and right inferior parietal sites. The vertical bar indicates stimulus onset.

Figure 6. Isopotential color maps of the normalized distribution of the negative difference (Nde) wave as a function of the clustering condition. The original data (27 scalp sites) were interpolated with a spherical spline algorithm [77].

Evidence from several previous ERP studies of selective attention may be considered consistent with an object-based account. First, the Nd_e onset latency increases with decreasing frequency separation between the task-relevant and task-irrelevant stimuli (30, 54, 71, 72). The Nd_e onset latency also increases with decreasing spatial separation between the task-relevant and task-irrelevant stimuli (73). Psychophysical studies have shown that individuals are more likely to report hearing two distinct streams of sounds when the physical separation (frequency and/or location) between the task-relevant and task-irrelevant streams of sounds is large. Second, the Nd_e onset latency increases with decreasing rate of stimulus presentation (55, 74, 75). Similarly, decreasing the rate of stimulus presentation makes the perception of distinct perceptual objects more difficult. The critical variable may be the overall rate of stimulus delivery, not the rate of repetition of attended tone bursts. Hink et al. obtained short Nd_e onset latencies in an experiment in which tones were presented in 5 locations, with tone bursts in the attended sequence repeating at mean ISIs of 1.5 sec (59). Lastly, as for the perception of distinct perceptual auditory objects, the Nd_e may require some time to develop after a task begins (76).

5. CONCLUDING COMMENTS

Both behavioral and electrophysiological data indicate that the ability to selectively attend to a particular auditory object depends on the stimulation context, with the ability improving in situations that promote the organization of sounds into distinct groups. Most of the data can be accounted for by object-based hypothesis in which parts and properties of the attended object receive extended processing. Most of the results are consistent with Bregman's model of auditory scene analysis in which an initial pre-attentive process partitions the auditory input into distinct groups of sounds according to Gestalt's principles. This pre-attentive analysis of stimuli may assist the attentional processes by easing the allocation and the maintenance of the attentional focus to a particular subset of stimuli. In other words, the attention to a subset of stimuli would depend on the outcome of the pre-attentive analysis. When the outcome reveals only one sound source, then the elements within the perceptual object must undergo a serial self-terminating process. In contrast, when the outcome of the pre-attentive system reveals more than one sound source then attention can be efficiently allocated to only one of these sources, allowing us to automatically exclude those elements that do not belong to the attended object. However, the implication of attention in forming and localizing an auditory object cannot be ruled out entirely, emphasizing the intimate link between attention and perception.

6. ACKNOWLEDGMENT

8. REFERENCES

1. Mondor, T. A. & S. A. Bregman: Allocating attention to frequency regions. *Percept Psychophys* 56, 268-76 (1994)

2. Hubner, R. & E. R. Hafter: Cuing mechanisms in auditory signal detection. *Percept Psychophys* 57, 197-202 (1995)

3. Mondor, T. A. & R. J., Zatorre: Shifting and focusing auditory spatial attention. *J Exp Psychol Hum Percept Perform* 21, 387-409 (1995)

4. Teder-Salejarvi, W. A. & S. A., Hillyard: The gradient of spatial auditory attention in free field: an event- related potential study. *Percept Psychophys* 60, 1228-1242 (1998)

5. Duncan, J. Selective attention and the organization of visual information. *J Exp Psychol Hum Percept Perform* 113, 510-517 (1984)

6. A. S. Bregman: Auditory Scene Analysis: The Perceptual Organization of Sounds. The MIT Press, London, England (1990)

7. J. Blauert: Spatial Hearing: The Psychophysics of Human Sound Localization. The MIT Press, London (1997)

8. W. M. Hartmann: Pitch, perception and the segregation and integration of auditory entities. In: Auditory function: Neurobiological bases of hearing. Eds: Edelman, G.M., Gall, W.E. & Cowan, W.M., John Wiley & Sons, NY (1988)

9. M. R. Jones & W. Yee: Attending to auditory events: The role of temporal organization. In: Thinking in Sounds: The Cognitive Psychology og Human Audition. Eds: McAdams, S. & Bigand, E., Oxford University Press, Oxford (1993)

10. Treisman, A. M. & G. Gelade: A feature-integration theory of attention. *Cognit Psychol* 12, 97-136 (1980)

11. Treisman, A. M. & S. Sato: Conjunction search revisited. *J Exp Psychol Hum Percept Perform* 16, 459-478 (1990)

12. U. Neisser: Cognitive Psychology. Appleton-Century-Croft, NY (1967)

13. B. C. J. Moore: An Introduction to the Psychology of Hearing. Academic Press, San Diego (1988)

14. Spieth, W., J. F. Curtis & J. C. Webster: Responding to one of two simultaneous messages. *J Acoust Soc Am* 26, 391-396 (1954)

15. Treisman, A. The effect of irrelevant material on the efficiency of selective listening. *Am J Psychol* 77, 533-546 (1964)

16. Cherry, E. C. Some experiments on the recognition of speech, with one and two ears. *J Acoust Soc Am* 25, 975-979 (1953)

17. Wood, N. L. & N. Cowan: The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953) *J Exp Psychol Gen* 124, 243-262 (1995)

18. Moray, N. Attention in dichotic listening: Affective cues and the influence of instruction. *Q J Exp Psychol* 9, 56-60 (1959)

19. N. Moray: Attention: Selective Processes in Vision and Hearing. Academic Press, NY (1970)

20. Treisman, A. M. Contextual cues in selective listening. *Q J Exp Psychol* 12, 242-248 (1960)

21. Wood, N. L. & N. Cowan: The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? *J Exp Psychol Learn, Mem Cog* 21, 255-260 (1995)

22. Pichora-Fuller, M. K., B. A. Schneider & M. Daneman: How young and old adults listen to and remember speech in noise. *J Acoust Soc Am* 97, 593-608 (1995)

23. Bregman, A. S. & A. I. Rudnicky: Auditory segregation: stream or streams? *J Exp Psychol Hum Percept Perform* 1, 263-267 (1975)

24. Jones, M. R., G. Kidd & R. Wetzel: Evidence for rhythmic attention. *J Exp Psychol Hum Percept Perform* 7, 1059-1073 (1981)

25. Ward, L. M. Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Can J Exp Psychol* 48, 242-259 (1994)

26. Woods, D. L. & C. Alain: Feature processing during high-rate auditory selective attention. *Percept Psychophys* 53, 391-402 (1993)

27. Woods, D. L., K. Alho & A. Algazi: A. Stages of auditory feature conjunction: an event-related brain potential study. *J Exp Psychol Hum Percept Perform* 20, 81-94 (1994)

28. Alain, C. & D. L. Woods: Distractor clustering enhances detection speed and accuracy during selective listening. *Percept Psychophys* 54, 509-514 (1993)

29. Alain, C., K. H. Ogawa & D. L. Woods: Aging and the segregation of auditory stimulus sequences. *J Gerontol Psychol Sci* 51B, 91-93 (1996)

30. Michie, P. T., N. Solowij, J. M. Crawford & L. C. Glue: The effects of between-source discriminability on attended and unattended auditory ERPs. *Psychophysiol* 30, 205-220 (1993)

31. Alho, K. Selective attention in auditory processing as reflected by event- related brain potentials. *Psychophysiol* 29, 247-263 (1992)

32. Alho, K., S. V. Medvedev, S. V. Pakhomov, M. S. Roudas, M. Tervaniemi, K. Reinikainen, T. Zeffiro & R. Naatanen: Selective tuning of the left and right auditory cortices during spatially directed attention. *Brain Res Cogn Brain Res* 7, 335-341 (1999)

33. O'Leary, D.S., N. C. Andreasen, R. R. Hurtig, I. J. Torres, L. A. Flashman, M. L. Kesler, S. V. Arndt, T. J., Cizadlo, L. L. B. Ponto, G. L. Watkins & R. D. Hichwa: Auditory and visual attention assessed with PET. *Hum Brain Mapp* 5, 422-436 (1997)

34. D. L. Woods: The physiological basis of selective attention: Implications of event-related potential studies. In: Event-related brain potentials: Basic issues and applications. Eds: Rohrbaugh, J.W., Parasuraman, R. & Johnson, R.J., Oxford University Press, NY (1990)

35. Giard, M. H., F. Perrin, J. Pernier & P. Bouchet: Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiol* 27, 627-640 (1990)

36. Scherg, M., J. Vajsar & T. W. Picton: A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience* 1, 326-355 (1989)

37. Javitt, D. C., M. Steinschneider, C. E. Schroeder, H.G. Jr. Vaughan & J. C. Arezzo: Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Res* 667, 192-200 (1994)

38. Csepe, V., G. Karmos & M. Molnar: Evoked potential correlates of stimulus deviance during wakefulness and sleep in cat -- animal model of the mismatch negativity. *Electroencephalogr Clin Neurophysiol* 66, 571-578 (1987)

39. Alain, C., D. L. Woods & R. T. Knight: A distributed cortical network for auditory sensory memory in humans. *Brain Res* 812, 23-37 (1998)

40. Alho, K., D. L. Woods, A. Algazi, R. T. Knight & R. Naatanen: Lesions of frontal cortex diminish the auditory mismatch negativity. *Electroencephalogr Clin Neurophysiol* 91, 353-362 (1994)

41. Alain, C., D. L. Woods & K. Ogawa: Brain indices of automatic pattern processing. *Neuroreport* 6, 140-144 (1994)

42. Nordby, H., W. T. Roth & A. Pfefferbaum: Event-related potentials to time-deviant and pitch-deviant tones. *Psychophysiol* 25, 249-261 (1988)

43. Tervaniemi, M., S. Maury & R. Naatanen: Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* 5, 844-846 (1994)

44. Alain, C., F. Cortese & T. W. Picton: Event-related brain activity associated with auditory pattern processing. *Neuroreport* 10, 2429-2434 (1999)

45. Alain, C. & D. L. Woods: Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiol* 34, 534-546 (1997)

46. Alain, C., A. Achim & D. L. Woods: Separate memory related processing for auditory frequency and patterns. *Psychophysiol* 36, 737-744 (1999)

47. Sussman, E., W. Ritter & H. J. Jr. Vaughan: Predictability of stimulus deviance and the mismatch negativity. *Neuroreport* 9, 4167-4170 (1998)

48. Mondor, T. A. & N. A. Terrio: Mechanisms of perceptual organization and auditory selective attention: the role of pattern structure. *J Exp Psychol Hum Percept Perform* 24, 1628-1641 (1998)

49. Koelsch, S., E. Schroger & M. Tervaniemi: Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309-1313 (1999)

50. Näätänen, R. A. Lehtokoski, M. Lennes, M. Cheour-Luhtanen, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R. J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen & K. Alho: Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432-434 (1997)

51. R. Näätänen: Attention and brain function. Erlbaum, Hillsdale (1992)

52. Mangun, G. R. Neural mechanisms of visual selective attention. *Psychophysiol* 32, 4-18 (1995)

53. Hillyard, S. A., R. F. Hink, V. L. Schwent & T. W. Picton: Electrical signs of selective attention in the human brain. *Science* 182, 177-180 (1973)

54. Hansen, J. C. & S. A. Hillyard: Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr Clin Neurophysiol* 49, 277-290 (1980)

55. Parasuraman, R. Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. *Biol Psychol* 11, 217-233 (1980)

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

57. Woldorff, M. G., C. C. Gallen, S. A. Hampson, S. A. Hillyard, C. Pantev, D. Sobel & F. Bloom: Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc Natl Acad Sci USA* 90, 8722-8726 (1993)

58. Schwent, V. L. & S. A. Hillyard: Evoked potential correlates of selective attention with multi-channel auditory inputs. *Electroencephalogr Clin Neurophysiol* 38, 131-138 (1975)

59. Hink, R. F., W. H. Jr. Fenton, A. Pfefferbaum, J. R. Tinklenberg & B. S. Kopell: The distribution of attention across auditory input channels: an assessment using the human evoked potential. *Psychophysiol* 15, 466-473 (1978)

60. Schwent, V., E. Snyder & S. A. Hillyard, Auditory evoked potentials during multichannel selective listening: Role of pitch and localization cues. *J Exp Psychol* 2, 313-325 (1976)

61. Donald, M. W. & R. Little: The analysis of stimulus probability inside and outside the focus of attention, as reflected by the auditory N1 and P3 components. *Can J Psychol* 35, 175-187 (1981)

62. Woods, D. L., S. A. Hillyard & J. C. Hansen: Event-related brain potentials reveal similar attentional mechanisms during selective listening and shadowing. *J Exp Psychol Hum Percept Perform* 10, 761-777 (1984)

63. Woods, D. L. & C. C. Clayworth: Scalp topographies dissociate N1 and Nd components during auditory selective attention. *Electroencephalogr Clin Neurophysiol Suppl* 40, 155-160 (1987)

64. Shelley, A. M., P. B. Ward, P. T. Michie, S. Andrews, P F. Mitchell, S. V. Catts & N. McConaghy: The effect of repeated testing on ERP components during auditory selective attention. *Psychophysiol* 28, 496-510 (1991)

65. Hillyard, S. A., M. Woldorff, G. R. Mangun & J. C. Hansen: Mechanisms of early selective attention in auditory and visual modalities. *Electroencephalogr Clin Neurophysiol Suppl* 39, 317-324 (1987)

66. Hillyard, S. A. Electrical and magnetic brain recordings: contributions to cognitive neuroscience. *Current Opinion in Neurobiology* 3, 217-224 (1993)

67. D. E. Broadbent: Decision and Stress. Academic Press, NY (1971)

68. LaBerge, D. Spatial extent of attention to letters and words. *J Exp Psychol Hum Percept Perform* 9, 371-379 (1983)

69. Alain, C., A. Achim & F. Richer: Perceptual context and the selective attention effect on auditory event-related brain potentials. *Psychophysiol* 30, 572-580 (1993)

70. Alain, C. & D. L. Woods: Signal clustering modulates auditory cortical activity in humans. *Percept Psychophys* 56, 501-516 (1994)

71. Alho, K., P. Paavilainen, K. Reinikainen, M. Sams & R. Naatanen: Separability of different negative components of the event-related potential associated with auditory stimulus processing. *Psychophysiol* 23, 613-623 (1986)

72. Alho, K., K. Tottola, K. Reinikainen, M. Sams & R. Naatanen: Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalogr Clin Neurophysiol* 68, 458-470 (1987)

73. Alho, K., N Donauer, P Paavilainen, K. Reinikainen, M. Sams & R. Naatanen: Stimulus selection during auditory

spatial attention as expressed by event-related potentials. *Biol Psychol* 24, 153-162 (1987)

74. Hansen, J. C. & S. A. Hillyard: Effects of stimulation rate and attribute cuing on event-related potentials during selective auditory attention. *Psychophysiol* 21, 394-405 (1984)

75. Teder, W., K. Alho, K. Reinikainen & R. Naatanen: Interstimulus interval and the selective-attention effect on auditory ERPs: "N1 enhancement" versus processing negativity. *Psychophysiol* 30, 71-81 (1993)

76. Donald, M. W. & M. J. Young: A time-course analysis of attentional tuning of the auditory evoked response. *Exp Brain Res* 46, 357-367 (1982)

Send correspondence to: Dr Claude Alain, Rotman Research Institute, Baycrest Centre for Geriatric Care, and Department of Psychology, University of Toronto, Canada, Tel:416-785-2500, ext. 3523, Fax:416-785-2862, E-mail: calain@rotman-baycrest.on.ca

This manuscript is available on line at:

http://www.bioscience.org/2000/v5/d/alain/fulltext.htm

Footnote:

[1] In this paper, the terms "clustered" and "clustering" refer to the physical manipulation of the stimuli whereas the term "grouping" is reserved for the psychological dimension.