Review Article

Auditory Scene Analysis: An Attention Perspective

Elyse S. Sussman^a

Purpose: This review article provides a new perspective on the role of attention in auditory scene analysis.

Method: A framework for understanding how attention interacts with stimulus-driven processes to facilitate task goals is presented. Previously reported data obtained through behavioral and electrophysiological measures in adults with normal hearing are summarized to demonstrate attention effects on auditory perception—from passive processes that organize unattended input to attention effects that act at different levels of the system. Data will show that attention can sharpen stream organization toward behavioral goals, identify auditory events obscured by noise, and limit passive processing capacity.

Conclusions: A model of attention is provided that illustrates how the auditory system performs multilevel

analyses that involve interactions between stimulus-driven input and top-down processes. Overall, these studies show that (a) stream segregation occurs automatically and sets the basis for auditory event formation; (b) attention interacts with automatic processing to facilitate task goals; and (c) information about unattended sounds is not lost when selecting one organization over another. Our results support a neural model that allows multiple sound organizations to be held in memory and accessed simultaneously through a balance of automatic and task-specific processes, allowing flexibility for navigating noisy environments with competing sound sources.

Presentation Video: <http://cred.pubs.asha.org/article.aspx?articleid=2601618>

This research forum contains papers from the 2016 Research Symposium at the ASHA Convention held in Philadelphia, PA.

Attention plays an important role in how we understand and perceive the environment. This review article provides a summary of a talk given at the 2016 American Speech-Language-Hearing Association (ASHA) Research Symposium on the role of attention in auditory scene analysis in adults with normal hearing. The goal was to provide a new framework for understanding how task goals interact with stimulus-driven processes to facilitate our ability to navigate in noisy environments.

Auditory Scene Analysis

Auditory scene analysis is a fundamental skill of the auditory system that allows us to perceive and identify sound events¹ in the environment. Imagine yourself walking

on a busy city street. Noises from the environment converge and digress: A car horn honks; a jet flies by; a jackhammer blasts; people talk as they walk past you. The sound waves reach your ears in a mixture of all the sources, overlapping in time. The ability to listen to your friend talking while you walk down a noisy city street requires brain mechanisms that disentangle this sound mixture, separating your friend's voice from the sounds of the cars and other passing conversations and providing neural representations that maintain the integrity of the individual and distinct sound sources.

We do not fully understand what the brain does to facilitate the ability to select and listen to one voice in the midst of the din of environmental sounds. It is not clear what happens automatically by stimulus-driven processes or how attention modifies neural activity to support scene analysis. This review article discusses the contributions of passive and active listening processes in navigating noisy environments with multiple competing sound sources, providing a framework of how attention interacts with stimulus-driven

^aDepartments of Neuroscience and Otorhinolaryngology–Head and Neck Surgery, Albert Einstein College of Medicine, Bronx, NY
Correspondence to Elyse S. Sussman: elyse.sussman@einstein.yu.edu
Editor-in-Chief: Frederick (Erick) Gallun
Editor: Karen Helfer

Received February 1, 2017

Revision received April 25, 2017

Accepted May 12, 2017

https://doi.org/10.1044/2017_JSLHR-H-17-0041

¹The word “event” is being used to describe an object of auditory perception—any perceptually defined unit of sound dictated by task demands. This definition is not bound to a particular concept of sound and thus may refer to a feature of a tone, a single tone, or a pattern of sequential tones, and differentiates an event from distinct sound streams.

Disclosure: The author has declared that no competing interests existed at the time of publication.

processes to facilitate the formation and maintenance of sound streams.

William James famously suggested that we can only attend to one thing at a time (James, 1890), a precursor to the limited-capacity models of attention advanced by 20th century psychologists (Kahneman, 1973). When selecting one of many competing sound sources, what then is the fate of the unattended? There has been much controversy regarding the degree to which unattended sensory input is processed. Broadbent (1958) originally proposed an early selection filter, in which the unattended inputs are subject to limited processing (e.g., only the features of the sound). Others later proposed that all inputs are fully processed, but the information is forgotten if not used (Deutsch & Deutsch, 1963). Kahneman's approach of a limited-capacity model, in contrast, suggested that, because attention is a limited resource, the complexity of the input influences the degree of processing of the unattended input. However, it is still not well understood when, where in the processing

hierarchy, or how complexity affects the processing of unattended sensory inputs. One reason is that it is difficult to directly measure responses to sensory stimulation that is not being attended. Behavioral measures can be used to deduce the fate of the unattended, such as by quantifying the influence of unattended information on behavioral performance, but do not provide a direct measure of processing. Event-related brain potentials (ERPs) give a direct quantifiable measure of brain activity to both attended and unattended stimuli while the stimuli are being presented. In combination with behavioral measures, ERPs provide a powerful tool to assess the influence of human attention on perception of the auditory scene.

Measuring Brain Processes Associated With Auditory Scene Analysis

One of the challenges in understanding how sounds are processed in noisy environments when there are competing

Figure 1. Schematic model of attention effects on auditory scene analysis. (A) Input is analyzed from spectrotemporal characteristics, driving initial stream formation. Event formation then occurs on the segregated streams. Deviance detection (which leads to mismatch negativity [MMN] elicitation) is a “higher level” process occurring on the already formulated streams. (B) Attention can modulate the stream segregation process, which in turn influences event formation and deviance detection. (C) Attention can be used to perceive within-stream events that may have been otherwise obscured in noisy environments. Attention-based detection of repetitive events can then form the basis for deviance detection. (D) Attention can select a subset of information from the mixture of input that highlights events in the attended stream. However, there is a cost of this attention. The unattended sounds are subject to resource limitations affecting one or more levels of passive processes when selecting a subset of information from the mixture of sounds that enter the ears.

Attention effects at different stages of auditory processing

MMN as a tool for probing the sound trace has many advantages. It is (a) modality specific—generated within auditory cortices (Alho, 1995; Giard, Perrin, Pernier, & Bouchet, 1990; Opitz, Mecklinger, Von Cramon, & Krugge, 1999); (b) elicited whether or not attention is focused on the sounds (Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Sussman, 2007; Sussman, Bregman, Wang, & Khan, 2005; Winkler, Czigler, Sussman, Horváth, & Balázs, 2005); (c) an index of how sounds are held in memory (Javitt, Steinschneider, Schroeder, & Arezzo, 1996; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001); (d) distinguishable from non-modality-specific responses, those associated with attention and target detection (e.g., P3b component; Novak, Ritter, Vaughan, & Wiznitzer, 1990; Sussman, 2007); and (e) highly context dependent (Sussman, 2007; Sussman, Chen, Sussman-Fort, & Dinces, 2014). Thus, with this brain measure, we can learn how background, unattended sounds are processed, represented, and organized in auditory memory and compare that with the effects of attention to the same sounds.

Auditory processes that are driven by the stimulus characteristics of the input, independent of attentional manipulation, are generally called *stimulus-driven* or “bottom-up” (Figure 1a). For example, when you initially walk into a cocktail party it is the degree of processing that occurs before

Figure 2. Context effects on event formation. Stimulus paradigm figure adapted from Sussman (2005). Frequency of the tones is depicted on the y-axis in hertz, and time is depicted on the x-axis in milliseconds. Two tones, one low (L; 440 Hz) and one high (H; 1568 Hz), alternated at a rapid pace of 75 ms onset to onset (LHLHLH...). The within-stream pace (e.g., L-L-L...) was 150-ms stimulus onset to onset. In the blocked condition (top row), every time a frequency deviant occurred in the low-tone stream (LD), a second deviant immediately followed it in the low stream (pink squares denote the low-deviant stimuli). In the mixed condition (bottom row), every time a frequency deviant occurred, it was not fully predicted that a second deviant would follow: Single and double deviants were intermixed randomly. If stream segregation occurs first and event formation occurs on the already segregated streams, then one mismatch negativity (MMN) should be elicited in the blocked condition and two MMNs in the mixed condition based on results of previous studies (see text). In contrast, if there was a global context effect on event formation, then two MMNs would be elicited by double deviants in both conditions; the alternation would preclude specific context effects from being exerted on the low-tone stream.

Alho, & Naatanen, 1999), event formation was determined by whether one or two MMNs were elicited by a “double-deviant” stimulus (i.e., two deviant stimuli presented successively). The same double-deviant stimulus was presented in different sound contexts. The sound context influenced within-stream event formation and affected whether one or two MMNs were elicited by the successive deviant stimuli (Sussman & Winkler, 2001). This manipulation of contextual cues was used in a streaming paradigm in which an alternation of the tones would preclude the within-stream context effects (see Figure 2). Only when the streams were physiologically segregated would the within-stream context exert influence and have an effect on MMNs’ elicitation by the double deviants. Results demonstrated context effects as were found for the single-stream paradigm: one MMN elicited by double deviants in the blocked condition and two MMNs elicited by double deviants in the mixed condition (see Figure 3). Thus, the sound context influenced within-stream event formation, indicating that stream segregation occurs first and within-stream events are formed on the already segregated streams. These results suggest that background sounds are monitored to a greater degree than may have been previously thought, with multiple processes (segregation and integration) acting on unattended sounds when the sounds are irrelevant to the main task being performed.

This automatic level of sound organization thus plays an important role in auditory scene analysis (see Figure 1a).

Figure 3. Segregation occurs before event formation. Data figure adapted from Sussman (2005). Context effects were demonstrated on the low-tone stream (LD) similarly as if the low-tone stream had been presented alone (without alternating high tones). Arrows point to significant mismatch negativity components (MMNs) elicited by the double-deviant stimuli in the blocked and mixed conditions (pink squares denote the low deviants). One MMN was elicited by double deviants in the blocked condition, and two MMNs were elicited by double deviants in the mixed condition. These results demonstrate multiple levels of bottom-up processing: Stream segregation occurs first and then event formation occurs on the already formed streams. L = low tone; H = high tone.

The real-world implication is that, when you walk into a noisy room, sounds are sorted on the basis of stimulus characteristics of the input and represented in memory as distinct sound streams. Stream segregation occurs first, and then sound events are detected and identified on the already sorted streams. Attention, which is a limited resource, can then be used to focus on and process the within-stream events of the already formed streams (e.g., to comprehend the speech stream). That is, attentional resources are conserved when some level of sorting occurs by automatic processes. These results provide evidence for multiple stages of processing on unattended sounds—both the segregation of sounds to streams and the integration of within-stream events to perceptual units. Passive levels of processing therefore play an important role in how we perceive the environment and can facilitate goal-directed behavior.

In addition to stimulus-driven processes, attention is needed to refine scene analysis to highlight what we perceive. Attention interacts with passive processes and plays

Figure 4. Frequency separation as a cue for stream segregation. Stimulus paradigm figure adapted from Sussman and Steinschneider (2009). Frequency of the tones is depicted on the y-axis in hertz, and time is depicted on the x-axis in milliseconds. Rectangles marked “X” represent the low-tone stream (440 Hz). Gray shading of the rectangles indicates the intensity value of the tones. In the intensity oddball condition (top row), tones are presented at 300 ms onset to onset. The standard tone has an intensity value of 71 dB and the deviant 12 dB higher, randomly occurring among 10% of the low tones. In the semitone (ST) conditions (bottom row), two higher frequency tones intervene between each of the “X” tones, with randomly varying intensity values. The presentation rate is 100 ms onset to onset. Intervening tones were presented in separate conditions at 1, 5, 7, and 11 ST higher than the “X” tones. Thus, frequency distance from the “X” tones cued for sound segregation and tone intensity were used to elicit mismatch negativity (MMN). Note that the standard (Std) and deviant (Dev) were neither the lowest or highest intensity values in the global sequence. Thus, detection of the standard-to-deviant relationship of the intensity values in the ST conditions was dependent upon the sounds segregating to two streams, such that MMNs would be elicited only when the sounds were segregated.

multiple roles in auditory scene analysis to facilitate task goals (Sussman, 2006). Discussed next are three effects of focused attention in auditory scene analysis, each influencing a different level of the system.

Attention Influences the Stream Formation Process

When you walk into a lively cocktail party, a level of sound organization occurs—brain mechanisms disentangle the sound input to form identifiable sound streams based on the mixture of sound input that enters the ears (e.g., a person talking, glasses clinking, music playing). We found that attention interacts with the stimulus-driven processes to sharpen the stream segregation process (Figure 1b). A recent study demonstrated that attention could effectively

segregate sounds that were not segregated automatically when the same sounds were in the background and irrelevant to the task (Sussman & Steinschneider, 2009). Participants were presented with several conditions of alternating tones that differed in the frequency difference (Δf) between the lower (440 Hz) and higher frequency tones. Two conditions of attention were compared: active and passive. In the active condition, the task was to listen to the lower frequency tones (440 Hz), ignore the higher frequency sounds, and press the response key whenever a louder intensity tone occurred randomly among the lower frequency tones (Figure 4). When participants selected the low set of sounds to perform the task (active listening), they could segregate the sounds at a smaller frequency separation than what

Figure 5. Attention modulates stream segregation. Data figure adapted from Sussman and Steinschneider (2009). Difference waveforms (deviant – control standard) for the Fz and the left mastoid (LM) are displayed for the passive (left panel) and active (right panel) conditions. Significant mismatch negativity (MMN) elicitation in the oddball and semitone (ST) conditions is shown with single and double asterisks and labeled with an arrow pointing toward the mastoid electrode (dashed line), where the MMN can be distinguished from overlapping N2b components. When subjects watched a movie (passive listening), tones were segregated at 7 ST and larger frequency separations. No MMNs were elicited at 5 and 1 ST (marked with an “X”), indicating that sounds were not segregated in auditory memory. In contrast, when subjects selected the high tones to perform a task (active listening), segregation of the sounds occurred at 5 ST, a smaller frequency separation than what occurred automatically during the passive listening condition. The black dashed circles highlight the different results obtained for the 5 ST conditions during passive and active listening tasks. The distance of 1 ST was too small to segregate high from low tones, at this pace, passively or actively (no MMNs were elicited, marked with an “X”).

occurred automatically during the passive listening condition (see Figure 5, dashed circles). These results demonstrate that attention can refine the stream segregation process by modulating the stream formation process (see Figure 1b). Actively selecting a set of sounds can prompt stream formation when it does not occur automatically by the stimulus characteristics of the input.

Attention Can Influence the Event Formation Process

Once sounds are sorted to distinct streams, attention focuses to specific streams to identify the sound events within them (e.g., words or musical phrases). You listen to a speaker for the content of the speech stream, or you listen to music to hear a familiar melody playing. When the room is noisy, there may be common spectral components in the music and speech streams occurring simultaneously. The within-stream events from background (unattended) sources may thus be obscured. Without attention focused on any particular event, perceptual failures of within-stream events may occur due to overlapping characteristics of the sounds contributing to the ambient noise. Attention can overcome limitations of the automatic system to facilitate task goals (Figure 1c).

In a previous study, we found that when two across-stream deviants occurred closely in time, the second deviant did not elicit MMN (Sussman, Ritter, et al., 1999). The second deviant was seemingly not detected as a separate deviant event when the sounds were unattended and the subjects read a book. We used this paradigm to assess whether attention to the sounds could modulate the event formation process and pull out perceptual sound units that were obscured by inattention to the sounds. To do this, we compared a passive condition with an active condition in which subjects attended to the high tones to detect the within-stream pattern events (see Figure 6). When subjects watched a movie and ignored the sounds, no MMN was elicited by the second of the successive across-stream deviants (see Figure 7, left panel), replicating findings of Sussman, Ritter, et al. (1999). However, when subjects actively detected high-tone pattern reversals, indicated by a response-key button press, separate MMNs were elicited by each of the successive deviants (i.e., attended and unattended sounds; see Figure 7, right panel). Thus, neural activity associated with discrete events was modulated to match task goals. The target event elicited its own MMN. This shows that attention acted on the event formation process, overriding passive listening processes (see Figure 1c). These data indicate that attention can sharpen event formation processes to highlight perceptual units that may have been missed in the cacophony of background sounds.

Interaction Between Bottom-Up and Top-Down Processes

There are limits to how much of the auditory scene we can process at once (Cowan, Blume, & Saults, 2013; Molloy, Griffiths, Chait, & Lavie, 2015; Saults & Cowan, 2007). Passive and active processes interact in scene analysis to facilitate task goals, but the nature of the interaction may limit other processes (Figure 1d). In many noisy

Figure 6. Across-stream pattern deviants occur successively. Stimulus paradigm figure adapted from Sussman, Ritter, et al. (1999). Frequency of the tones is depicted on the y-axis in hertz, and time is depicted on the x-axis in milliseconds. Six pure tones, three in the high-frequency range (denoted by A-B-C) and three in the low-frequency range (denoted by 1-2-3), were presented in alternation, such that an ascending pattern was repeated within each frequency stream (the standard patterns). Tones were 50 ms in duration, with a 50-ms silence between them (i.e., 100-ms onset-to-onset pace). Deviants were pattern reversals (descending patterns) such that no new tonal elements were added to the global sound sequence. Within-stream sound patterns emerged when sounds segregated in memory. Mismatch negativities would then be elicited by pattern reversals if they were detected. Note that, every time the low-deviant pattern occurred, the high-deviant pattern followed successively.

situations, there are multiple sound streams, from which we attend to and process one of them. However, in most laboratory studies, selecting one sound stream leaves only one other sound stream in the background (e.g., Sussman, Ritter, & Vaughan, 1998a). What, then, happens to the unattended background when there are more than one sound stream unattended and we select one of them? To assess the fate of the unattended and determine whether background sounds would be represented in memory as structured frequency streams or as a background of unstructured sounds, Sussman et al. (2005) presented alternating sounds, including three frequency ranges (high, middle, and low). Thus, when subjects were instructed to select one sound stream (high-frequency tones), two potential streams would remain in the background. This condition was compared to when no streams were selected and all of the sounds were in the background (while subjects watched a movie). The selected high-tone stream was a modified oddball paradigm in which single and double (two successive) frequency deviants occurred randomly and infrequently. The subjects' task was to press the response key when they detected the double deviants (see Figure 8). This was done so that subjects could not press the response key whenever a frequency deviant occurred within the high-tone stream (i.e., not to rely on a pop-out effect); subjects had to identify the high-frequency deviants that occurred two times in a row. It was a complex task. Listening to the high tones as a cohesive stream meant ignoring the task-irrelevant middle and low tones that intervened between them (see Figure 8).

When all of the sounds were unattended, MMNs were elicited by the single and double high-frequency deviants and, separately, by the middle and low pattern deviants.

Figure 7. Attention overrides limitations of passive processes. Grand-averaged difference waveforms are displayed for the passive (left panel) and active (right panel) conditions. In the passive condition, mismatch negativity (MMN) was significantly elicited by the low pattern deviants (the first of the two successive deviants, blue line), and no MMN was elicited by high pattern deviants (black line). In contrast, when the task was to detect the high pattern standards and press a response key when the reversals were detected, MMNs were now elicited by the high pattern deviants (black line) in the active condition. Accordingly, N2b and P3b target-detection event-related brain potential components (labeled with an arrow) were also elicited by high-tone pattern deviant targets. Intervening task-irrelevant low deviants did not evoke target responses, thus demonstrating task performance accuracy. Also, MMNs were elicited by low pattern deviants (blue line), which preceded the high pattern deviants, when performing the high-tone task. Thus, deviance detection was initiated by passive processes even when the low sounds were irrelevant to the task. Note the similarity in MMN amplitude and latency elicited by the low deviants in passive and active conditions.

Figure 8. Three-stream paradigm. Stimulus figure adapted from Sussman et al. (2005). Three frequency ranges of pure tones were presented in an alternating pattern (high, middle, and low). Triangles represent the high tones, letters for the middle tones, and numbers for the low tones (the sequential input presented to the ears is depicted below the main panel). Tone duration was 30 ms. Tone onset-to-onset pace was 90 ms. The main panel depicts the segregation of the input to three frequency ranges. The low and middle tones include three-tone ascending standard patterns and descending deviant patterns (similar to that shown in Figure 6). In the active condition, subjects attended to the high tones and ignored the low and middle tones (shaded in gray) and pressed a key to identify the two frequency deviants occurring successively within the high-tone stream (red arrow points to the target). In the passive condition (not shown), all of the sounds were task irrelevant, and subjects watched a movie. The question was whether the unattended background would be structured or unstructured, as demonstrated by whether the sounds segregated by frequency. Thus, if the middle- and low-tone pattern deviants elicited mismatch negativities, then it would indicate that the sounds were segregated by passive processes.

This suggests that the streams were segregated and that respective within-stream deviants were detected (see Figure 9, passive condition). In contrast, when attention was focused onto the high-tone stream to perform the target detection task, no MMNs were elicited by the task-irrelevant pattern deviants in the middle and low streams (see Figure 9, active condition). This result suggests that focused attention onto a subset of sounds preempted the stream segregation process for the background sounds. It is notable that there were differential results depending upon where attention was directed, to visual or auditory input. When attention was focused on the visual input (watching a movie), segregation of the three streams occurred, whereas when attention was focused on a subset of the auditory input, to segregate the high-frequency sounds and perform the auditory deviance detection task, there was no MMN indication that segregation occurred. The results taken together can thus lead to

another interpretation, in which highly focused attention to a subset of sounds preempted within-stream event formation (not segregation) on the unattended set of sounds. That is, a primitive level of stream segregation still occurred for the unattended sounds, but there were not enough resources to process the complex within-stream patterns that would be indexed by MMN (see Figure 1d). We tested this alternative in a follow-up study and found that within-stream event formation, and not stream segregation, was precluded by highly selective attention (Pannese, Herrmann, & Sussman, 2015). Thus, when attention parses out a subset of information, it can impact the degree of processing of the irrelevant, unattended streams, such that not all of the information is processed by passive systems while performing a task (see Figure 1d). In the cocktail party setting, this may be akin to knowing that there are other speakers in the background, with the pitch of the voices indicating male

Figure 9. Focused attention can limit passive processes. Data figure adapted from Sussman et al. (2005). Difference waveforms for the passive condition (left panel) and active condition (right panel) are displayed for the Fz (thick solid line) and Cz (thin solid line) electrodes. Gray shading is for the unattended sounds in passive and active conditions. When all of the sounds were in the background and were task irrelevant (passive condition), mismatch negativity components (MMNs) were elicited by all deviants (labeled with arrows), indicating that segregation of the sounds to frequency streams occurred automatically. In contrast, when subjects actively segregated the high tones (active condition), MMNs were elicited by the two successive deviants in the attended stream (labeled with an arrow), but no MMNs were elicited by the unattended background deviants (marked with an "X"). The P3b target detection response was elicited to the second of the two successive deviants (labeled with an arrow), showing that subjects were accurately performing the task. Thus, active segregation of a subset of the sounds modulated the passive processing capacity.

Figure 10. Task-switching paradigm for biperceptual stimuli. Stimulus paradigm figure adapted from Sussman, Bregman, & Lee (2014). (A) At a 5 semitone distance between sounds, the same set of stimuli can be perceived as either integrated (left panel) or segregated (right panel). (B) For the *integrated percept*, subjects were presented with a block of sounds and were instructed to identify one of the three patterns. Patterns were randomly presented, with Patterns 1 and 2 occurring 45% each and Pattern 3 occurring 10% within a block. Thus, Pattern 3 was the deviant pattern (though subjects were not told anything about probability of occurrence). For the *segregated percept*, subjects segregated out, and focused on, the low tones to perform a loudness detection task, pressing a response key when the louder sound among the low tones was detected. Mismatch negativity (MMN) would be elicited by loudness deviants only when the tones segregated, similarly as described in Figure 4. (C) The instruction of what to focus on (patterns or stream of low tones) and what task to do (pattern identification or loudness detection) was randomized across stimulus blocks and cued visually prior to the onset of the sound blocks. Thus, subjects switched attention back and forth between integrating and segregating the same set of stimuli to perform the task as visually cued to do so.

and female, but not being able to perceive the content of any unattended speech stream among background speakers. There is a level of processing that occurs passively, which segregates the sounds to sources, but the processing is limited by attentional resources needed to make meaning of the within-stream events.

To summarize this section, we demonstrate that attention acts on different levels of the auditory system. Passive processes organize the unattended input based on stimulus characteristics. Attention can then sharpen stream organization toward behavioral goals, overcome limitations of automatic processes to identify events obscured by noise, and select a subset of information, but with a cost to the extent of processing for unattended sound events (see Figure 1).

Perceptual Ambiguity and Auditory Scene Analysis

This section deals with the question of how the auditory system resolves perceptual ambiguity, which can arise when there are overlapping sound sources. When sound input can be perceived in multiple ways, how are the sounds stored in memory? Facilitation and suppression models suggest that attended input should dominate over unattended input, in that attended information is enhanced and unattended information is suppressed (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Desimone & Duncan, 1995; Mesgarani & Chang, 2012). Studies that have measured responses to different simultaneous perceptual organizations derived from the same set of stimuli have indicated that

Figure 11. Integrated and segregated organizations both held in memory. Data figure adapted from Sussman et al. (2005). Grand-averaged difference waveforms are displayed from the Fz electrode (left panel) and the Pz electrode (right panel) to show deviance detection (mismatch negativity [MMN]) and target detection (P3b) responses to both potential organizations. The heads show the voltage distribution maps (black dots represent the electrodes) at the peak latency of the corresponding components (labeled on the x-axis in milliseconds). Blue designates negative polarity, and red designates positive polarity. When subjects segregated sounds to perform the loudness detection task (attend intensity, target, top row, left column), MMN was elicited by the intensity deviants (thick solid line) and by the task-irrelevant pattern deviants (thin solid line), with similar amplitude and latency. Thus, even when the sounds were segregated, the integrated organization was passively monitored and deviants were detected. Likewise, when the sounds were integrated to identify the patterns (attend pattern, target, bottom row, left column), MMNs were elicited by them (thick solid line) and by the task-irrelevant intensity deviants (non-target, thin solid line, bottom row, left column), also with similar amplitude and latency. Consistent with task goals, the target detection responses (P3b, right column) were elicited by the targets associated with the organization needed to perform the task and not by deviants of the task-irrelevant organization. Thus, we show that neural traces for both organizations were simultaneously held in memory, and a distinction between the levels of passive and active processes.

one organization (integration or segregation) is perceptually held at a time and switches back and forth spontaneously (e.g., Denham et al., 2012; Pressnitzer & Hupé, 2006). However, such studies did not establish whether both organizations were held in memory when one was perceived. Sussman, Bregman, and Lee (2014) designed such a paradigm in which integrated and segregated percepts had different deviants

associated with them. Either or both deviants could elicit MMNs, which would reflect what was currently held in memory. The subject's task was to switch attention between integrating all of the stimuli to detect patterns and segregating out the low tones to detect loudness deviants, in separately cued trials (see Figure 10). The question of the study was whether multiple sound organizations would be

represented in memory even when only one appeared in perception. That is, the hypothesis was that attention would “resolve” the ambiguity toward the sound organization that was used to perform the task, consistent with a facilitative-suppression model. Thus, it was predicted that performing one task would preclude neural representation of the alternative organization and MMNs would be elicited only by deviants of the attended task. However, that was not the case. Even when one organization dominated in perception (integration or segregation), MMNs were elicited by deviants in the alternative organizations. This suggests that neural traces for both organizations are simultaneously held in memory (see Figure 11) and that attention has access to more than one organization from the same input, held in memory simultaneously. Further, the results indicate that information about the auditory scene is not lost when selecting a subset of information. This flexibility allows the capacity to retrieve information, for example, if we did not initially hear what we wanted to listen to and had to retrieve from memory the trace of the original source. Having access to multiple representations allows for rapid and flexible attention switching. This is especially needed in noisy environments to facilitate the ability to listen to various different sound events in the room.

Conclusions

The auditory system performs multilevel analyses that involve interactions between stimulus-driven input and top-down processes to facilitate task goals and allow flexibility for processing in noisy environments. Together, the studies discussed show that (a) stream segregation occurs automatically and sets the basis for event formation; (b) attention interacts with automatic processing to facilitate task goals; and (c) information about unattended sounds is not lost when selecting one organization over another. Our results thus support a neural model that allows multiple sound organizations to be held in memory and accessed simultaneously through a balance of automatic and task-specific processes.

To interpret the results within a larger, real-world framework, we can think of the orchestra as an example of how these multilevel processes are engaged. During a Beethoven symphony, we can sit back and listen to the global picture of the orchestra, the harmonies and passing melodies that trade off among the strings, reeds, and brass. Or we can take a closer listen to the melody playing in the violin. As we listen, we have access to multiple organizations held in neural memory simultaneously, which allows us to hear both the harmonic and the melodic aspects of the music and switch readily back and forth between them. Listening to an orchestra uses global harmony without losing access to the individual melodies of the various instruments, and while listening to a single melody we still hear the harmonic structure of the orchestra playing together. We have access to both global and local organizations. However, the way in which attention is directed to the sounds may preclude processing all aspects of the complex

global scene. Listening to the flute, for example, may enhance processing of that melodic stream and by virtue of limited attentional resources may limit the ability to capture specific events occurring in other instruments.

Overall, the results of the studies described here suggest that, from a busy auditory scene, when there are multiple sound sources, we have access to all of the sounds. Attention does not simply enhance the attended and gate out the unattended inputs at an early processing stage. Our results indicate that the input is processed and maintained in memory for access by attentive systems and can be modified by attention at different levels of the system (see Figure 1). Attention is a limited resource; thus, having transient access to the sound memory, rather than simply filtering out the unattended, promotes a great deal of flexibility. The multiple organizations represented in the brain enhance the ability to switch attention from one sound object in the environment to another and detect specific sound events when navigating noisy environments.

Acknowledgments

Research reported in this publication was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Awards R13DC003383 and R01DC004263 to E.S.S.. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, 36(6), 737–744.
- Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*, 16(1), 38–51.
- Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound of sound*. Cambridge, MA: MIT Press.
- Broadbent, D. E. (1958). Selective listening to speech. In *Perception and communication* (pp. 11–35). Elmsford, NY: Pergamon Press. <https://doi.org/10.1037/10037-002>
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248(4962), 1556–1559.
- Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 731–747. <https://doi.org/10.1037/a0029687>
- Denham, S., Bendixen, A., Mill, R., Toth, D., Wennekers, T., Coath, M., ... Winkler, I. (2012). Characterising switching behaviour in perceptual multi-stability. *Journal of Neuroscience Methods*, 210(1), 79–92. <https://doi.org/10.1016/j.jneumeth.2012.04.004>
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- Deutsch, J. A., & Deutsch, D. (1963). Some theoretical considerations. *Psychological Review*, 70, 80–90.

- 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(6), 627–640.
- James, W. (1890). *The principles of psychology* (Vol. 1). New York, NY: Dover Publications.
- 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 of Sciences*, 93(21), 11962–11967. <https://doi.org/10.1073/pnas.93.21.11962>
- Kahneman, D. (1973). Attention and effort. *The American Journal of Psychology*, 88, 339–340. <https://doi.org/10.2307/1421603>
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397), 233–236. <https://doi.org/10.1038/nature11020>
- Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattention deafness: Visual load leads to time-specific suppression of auditory evoked responses. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(49), 16046–16054. <https://doi.org/10.1523/JNEUROSCI.2931-15.2015>
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329.
- Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, 30(5), 436–450.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). “Primitive intelligence” in the auditory cortex. *Trends in Neurosciences*, 24(5), 283–288.
- Novak, G. P., Ritter, W., Vaughan, H. G. J., & Witznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalography and Clinical Neurophysiology*, 75(4), 255–275.
- Opitz, B., Mecklinger, A., Von Cramon, D. Y., & Kruggel, F. (1999). Combining electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology*, 36(1), 142–147.
- Pannese, A., Herrmann, C. S., & Sussman, E. (2015). Analyzing the auditory scene: Neurophysiologic evidence of a dissociation between detection of regularity and detection of change. *Brain Topography*, 28(3), 411–422. <https://doi.org/10.1007/s10548-014-0368-4>
- Pressnitzer, D., & Hupé, J.-M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16(13), 1351–1357. <https://doi.org/10.1016/j.cub.2006.05.054>
- Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 136(4), 663–684. <https://doi.org/10.1037/0096-3445.136.4.663>
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38(4), 387–401.
- Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, 117(3, Pt. 1), 1285–1298.
- Sussman, E. S. (2006). Multiple mechanisms of auditory attention. *The Journal of the Acoustical Society of America*, 119(5), 3416. <https://doi.org/10.1121/1.4786818>
- Sussman, E. (2007). A new view on the MMN and attention debate: The role of context in processing auditory events. *Journal of Psychophysiology*, 21(3), 164–175. <https://doi.org/10.1027/0269-8803.21.3.xxx>
- Sussman, E. S., Bregman, A. S., & Lee, W. W. (2014). Effect of task-switching on neural representation of ambiguous sound input. *Neuropsychologia*, 64, 218–229.
- Sussman, E. S., Bregman, A. S., Wang, W. J., & Khan, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multi-stream auditory environments. *Cognitive, Affective, & Behavioral Neuroscience*, 5(1), 93–110. <https://doi.org/10.3758/CABN.5.1.93>
- Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The five myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography*, 27(4), 553–564. <https://doi.org/10.1007/s10548-013-0326-6>
- Sussman, E. S., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *NeuroReport*, 16(13), 1519–1523. <https://doi.org/10.1097/01.wnr.0000177002.35193.4c>
- Sussman, E., Ritter, W., & Vaughan, H. G. (1998a). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, 789(1), 130–138.
- Sussman, E., Ritter, W., & Vaughan, H. G. (1998b). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport*, 9(18), 4167–4170. <https://doi.org/10.1097/00001756-199812210-00031>
- Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, 36(1), 22–34.
- Sussman, E., & Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia*, 47(3), 771–785. <https://doi.org/10.1016/j.neuropsychologia.2008.12.007>
- Sussman, E., & Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Brain Research. Cognitive Brain Research*, 12(3), 431–439.
- Sussman, E., Winkler, I., Huottilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, 13(3), 393–405. [https://doi.org/10.1016/S0926-6410\(01\)00131-8](https://doi.org/10.1016/S0926-6410(01)00131-8)
- Sussman, E., Winkler, I., Kreuzer, J., Saher, M., Näätänen, R., & Ritter, W. (2002). Temporal integration: Intentional sound discrimination does not modulate stimulus-driven processes in auditory event synthesis. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 113(12), 1909–1920. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12464328>
- Sussman, E., Winkler, I., Ritter, W., Alho, K., & Näätänen, R. (1999). Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. *Neuroscience Letters*, 264(1–3), 161–164. [https://doi.org/10.1016/S0304-3940\(99\)00214-1](https://doi.org/10.1016/S0304-3940(99)00214-1)
- Winkler, I., Czigler, I., Sussman, E. S., Horváth, J., & Balázs, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, 17(2), 320–339. <https://doi.org/10.1162/0898929053124866>