Published in final edited form as:

Psychophysiology. 2020 February ; 57(2): e13487. doi:10.1111/psyp.13487.

Cognitive resources are distributed amongst the entire auditory landscape in auditory scene analysis.

Renee M. Symonds¹, Juin W. Zhou^{1,2}, Sally L. Cole³, Kelin M. Brace¹, Elyse S. Sussman¹

¹Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY

²Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY

³Teachers College, Columbia University, New York, NY

Abstract

Although attention has been shown to enhance neural representations of selected inputs, the fate of unselected, background sounds is still debated. The goal of the current study was to understand how processing resources are distributed amongst attended and unattended sounds during auditory scene analysis. We used a three-stream paradigm with four acoustic features uniquely defining each sound stream (frequency, envelope shape, spatial location, and tone quality). We manipulated task load by having participants perform a difficult auditory task and an easy movie-viewing task with the same set of sounds in separate conditions. The mismatch negativity (MMN) component of event-related brain potentials (ERPs) was measured to evaluate sound processing in both conditions. We found no effect of task demands on unattended sound processing: MMNs were elicited by unattended deviants during both low- and high-load task conditions. A key factor of this result was the use of unique tone feature combinations to distinguish each of the three sound streams, strengthening the segregation of streams. In the auditory task, the P3b component demonstrates a two-stage process of target evaluation. Thus, these results, in conjunction with results of previous studies, suggest that stimulus-driven factors that strengthen stream segregation can free up processing capacity for higher-level analyses. The results illustrate the interactive nature of top-down and stimulus-driven processes in stream formation, supporting a distributive theory of attention that balances the strength of the bottom-up input with perceptual goals in analyzing the auditory scene.

Keywords

event-related potentials (ERPs); cognitive load; Attention; auditory scene analysis; mismatch negativity (MMN); P3b component

1. Introduction

To listen successfully in noisy environments, individuals must selectively attend to relevant sounds amidst a myriad of irrelevant background sounds. Although it has been demonstrated

that selective attention can enhance neural representations of the attended sounds (Desimone & Duncan, 1995; Hillyard, Hink, Schwent, & Picton, 1973; Mesgarani & Chang, 2012; Moran & Desimone, 1985), there is ongoing debate about the fate of the unattended sounds and the role attention has in influencing working memory representations during selective listening and task performance (Konstantinou et al., 2013; Luck & Vogel, 2013; Murphy, Spence, & Dalton, 2017; Pannese, Herrmann, & Sussman, 2015; Shamma, Elhilali, & Michey, 2011). It is unclear, for example, what selective attention does to modify processing of unattended sound inputs, especially how the unattended auditory space competes for perceptive resources used to process the entire auditory landscape. The goal of the current study was to understand how processing resources are distributed amongst attended and unattended sounds during auditory scene analysis.

Theories of attention have been largely formulated through mechanisms of visual perception. Attention is often described in a binary manner, with attention (to an object, a feature, or a location in space) serving to enhance neural activity for attended inputs and suppress or inhibit activity of distracting or irrelevant information (Chelazzi, Duncan, Miller, & Desimone, 1998; Treue & Maunsell, 1996; Desimone & Duncan, 1995; Maunsell & Treue, 2006; Störmer & Alvarez, 2014). A distinguishing factor among theoretical perspectives is where the attentional filter acts to inhibit processing of the unattended stimuli. Attentional gain control theories propose early stage filters. Competition for representation of items in working memory posed by crowded scenes increases the ‘gain’ of attended, task-relevant stimuli, and suppresses unattended, competing inputs (Cusack, Deeks, Aikman, & Carlyon, 2004; Desimone & Duncan, 1995; Hillyard, Vogel, & Luck, 1998; Mesgarani, David, Fritz, & Shamma, 2014; Störmer & Alvarez, 2014). Thus, the role of attention is to accentuate attended information and suppress irrelevant, ‘distractors’ from further processing, leading to improved behavioral performance (Desimone & Duncan, 1995; Herrmann, Heeger, & Carrasco, 2012; Hillyard et al., 1998; Maunsell & Treue, 2006; Mesgarani & Chang, 2012; Moran & Desimone, 1985; Puvvada & Simon, 2017; Störmer & Alvarez, 2014; Woldorff, Hackley, & Hillyard, 1991).

One of these competition theories is feature-based, proposing that when there is competition for perception, attention actively inhibits processing of similar features in the visual field that would be easily confusable with the attended target feature (Störmer & Alvarez, 2014). Thus, for example, neural signals for an attended color would be enhanced while colors similar to the target would be suppressed, enabling more accurate target selection by reducing confusing input information (Störmer & Alvarez, 2014). In contrast, cognitive load theories propose later stage filters, suggesting that unattended stimuli are fully processed in parallel with attended stimuli and it is the cognitive load of the task that then interacts with working memory. High task demands restrict availability of resources to process irrelevant or distracting information, enhancing task performance by reducing interference from distracting stimuli (Lavie, 2005; Lavie & Tsai, 1994; Lavie, Hirst, de Fockert, & Viding, 2004; Raveh & Lavie, 2015).

The auditory system relies more heavily on time than space for object identity; sound source-identity builds up over time (Anstis & Saida, 1985; Bregman, 1978). If unattended sounds were inhibited from further processing, their identity would be lost without the

possibility of resampling the input once it's gone. Accordingly, there is evidence that unattended sounds are represented distinctly in working memory, and that sound streams are processed in parallel, competing for resources in working memory (de Fockert, Rees, Frith, & Lavie, 2001; Horvath, Czigler, Sussman, & Winkler, 2001; Hughes, Vachon, & Jones, 2007; Lavie et al., 2004; Miller, Chen, Lee, & Sussman, 2015; Pannese et al., 2015; Raveh & Lavie, 2015; Sussman, Bregman, Wang, & Khan, 2005). This suggests a degree of processing for unattended sounds that allows access through memory.

In auditory scene analysis, the ability to segregate sound input has been attributed to the distance between sounds, with larger frequency separations prompting easier perceptual segregation of streams (e.g., Carlyon, Cusack, Foxton, & Robertson, 2001; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Sussman, Ritter, & Vaughan, 1999; Weise, Ritter, & Schröger, 2011; Winkler, Denham, Mill, Bohm, & Bendixen, 2012). Most previous studies on attention and scene analysis have used 'two-stream' paradigms to investigate how sounds are maintained in memory. In this way, one set of sounds forms an attended sound stream, within which a task is performed, and the rest of the sounds are unattended, forming a distractor stream or broadband noise (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Shamma et al., 2011; Zion-Golumbic, Ding, Bickel, Lakatos, Schevon, McKhann, & Schroeder, 2013; Woldorff et al., 1991). Easier segregation and perception of two streams when tone frequencies are far apart is easily explained by peripheral channeling (Beauvois & Meddis, 1991; Hartmann & Johnson, 1991; Rose and Moore, 2000). The peripheral channeling theory of streaming is an analogue of the feature-based theory of attention. It is based on spatial mapping of hair cells along the basilar membrane. Tones that are distant in frequency have corresponding hair cells that are spatially distant from each other along the basilar membrane. Thus, this theory predicts that it is easier to distinguish tones that are far in frequency from each other because tones near in frequency are more confusable due to overlapping excitation of hair cells within the cochlea. However, several studies have shown that the stream segregation process involves more than simple frequency resolution of peripheral mechanisms (Bey & McAdams, 2003; Cusack & Roberts, 2000; Iverson, 1995; Moore & Glockel, 2012; Roberts et al., 2002; Sussman, 2005; Vliegen & Oxenham, 1999) and does not explain modulation by attention (e.g., Sussman, Ritter, & Vaughan, 1998).

While the two-stream paradigm is a parsimonious way of comparing responses within the focus of attention with those outside attentional focus, the simplicity of the paradigm cannot account for the meaningful complexity of soundscapes encountered in everyday life that generally contain multiple unattended streams. Moreover, in more naturalistic auditory scenes, combined auditory features characterize a sound stream. For example, the voice of your conversational partner at a cocktail party is distinct in more than one feature from the sounds of the other guests (e.g., pitch, relative loudness, and location to you) and from the background music. It is likely that these multiple features increase stream distinctiveness and facilitate the ability to segregate one speaker's voice from the background din. Thus, in two stream paradigms, the use of a single tone feature for segregation may be less effective for distinguishing sound streams since all the other sound features are shared between the two streams (e.g., intensity, spatial location, tone duration). Additionally, there may be greater confusability of sound events when there is only one cue for segregation and more chance

for stimulus uncertainty due to overlap of the other features within the global sound sequence (Störmer & Alvarez, 2014). When sounds are only weakly segregated and overlap in spectral-temporal characteristics, there may be a greater demand of attentional resources (Dinca & Sussman, 2017). If greater attentional resources are required to segregate sounds due to stimulus confusability there may be limited resources for processing higher-level within-stream sound events. This is also indicated in results of previous studies on informational masking (Dykstra, Cariani, & Gutschalk, 2017; Kidd, Richards, Mason, Gallun, & Huang, 2008) and would explain the results of numerous previous studies (Dykstra & Gutschalk, 2015; Kidd et al., 2008; Pannese et al., 2015; Sussman et al., 2005; Woldorff et al., 1991). In the current study, we used a three-stream paradigm similar to ones we have used previously but with multiple tone features converging in each frequency range, lessening the influence of frequency itself as the distinguishing factor for perceiving auditory streams.

To evaluate the brain's response to both attended and unattended sounds, event-related brain potentials (ERPs) were recorded. ERPs provide a direct measure of unattended sound processing when they can be elicited without the participant focusing attention toward the sounds or making an overt response. One such component is the mismatch negativity (MMN). The MMN is an auditory-specific response characterized by a larger negative response to infrequently occurring tones detected as deviants with respect to frequently repeating standard tones in the sequence (Näätänen, Gaillard, & Mantysalo, 1978). The MMN indexes ongoing processing of attended and unattended sounds during a selective attention task because it can be elicited regardless of the direction of attention (Winkler, Czigler, Sussman, Horvath, & Balazs, 2005). When MMN is elicited it indicates that the repeating standard was detected and held in auditory memory to allow deviance detection. We have previously used MMN to assess segregation of sounds in working memory (Sussman et al., 1999). MMNs elicited by unattended within-stream deviants served as an index of stream segregation, as well as an index of detection of within-stream standards and deviants (Dykstra & Gutschalk, 2015; Horvath et al., 2001; Macken, Tremblay, Houghton, Nicholls, & Jones, 2003; Pannese et al., 2015; Sussman et al., 2005; Sussman et al., 1998; Sussman et al., 1999; Sussman et al., 2005; Sussman, Horvath, Winkler, & Orr, 2007; Weise et al., 2011). In contrast, elicitation of the P3b component of ERPs requires attention to be focused on the sounds. The P3b component is a nonmodality specific index of target processing (Sutton, Braren, Zubin, & John, 1965), elicited by detected novel auditory or visual events. In the current study, the P3b component provided a measure (in latency and amplitude) of attention-based, sound task processing (Kok, 2001; Polich, 2007).

In our previous three-stream studies that used one tone feature (frequency) to distinguish streams, results indicated that attention limited processing of unattended sounds during auditory selective attention (Pannese et al., 2015; Sussman et al., 2005). When attention segregated out the high-frequency stream tones to identify high tone target deviants, MMNs were elicited by the attended target deviants but not by the unattended middle- and low-frequency stream deviants. However, when participants ignored all of the sounds and watched a silent movie, MMNs were elicited by within-stream deviants in all of the three frequency streams. This suggested that highly focused attention to one stream limited processing of unattended sounds, but it was not clear at what processing level. The absence

of MMN could have meant either that attention suppressed processing of the unattended sounds and preempted stream segregation or that stream segregation occurred but task demands restricted higher-level processes of within-stream pattern and deviance detection.

In the current study, we manipulated task load in a three-stream paradigm, defining streams by four tone features (frequency, envelope, spatial location, and tone quality) to enhance stimulus-driven factors in segregation. The key question was whether or not MMNs would be elicited by deviants embedded within the two unattended streams with enhanced stream distinctiveness. If stimulus-driven factors strengthen automatic streaming for unattended sounds and thus lower processing demands for unattended sounds, MMNs would be elicited by within-stream deviants. This finding would contrast those of our two previous experiments (Sussman et al., 2005 and Pannese et al., 2015) and would be consistent with the hypothesis that there is a distribution of cognitive resources amongst the entire auditory landscape, which is modulated by task demands and supported by stimulus factors. In contrast, no MMN elicitation by unattended deviants during active segregation would suggest that processing of unattended sounds is more globally dampened when selectively attending to a part of the total auditory scene.

2. Method

2.1 Participants

Ten normal-hearing adults were paid for their participation (5 men; $M = 30$ years, $SD = 5$). We derived our power calculation based on the MMN component, which is smallest and most variable. Based on the amplitude estimation of the MMN obtained from our previous study with a similar paradigm (Sussman et al., 2005), there is ample power in an auditory-ignore condition ($1-\beta=0.77$) and in an auditory-attend condition ($1-\beta=0.99$) to detect an MMN with an alpha level of .05 and 10 adults. All participants passed a hearing screening (20 dB HL or better bilaterally at 500, 1000, 2000, and 4000 Hz) and provided written informed consent in accordance with the Declaration of Helsinki after the experiment was explained to them, and prior to the test session. The protocol was approved by the Institutional Review Board of the Albert Einstein College of Medicine where the study was conducted.

2.2 Stimuli

Stimuli used in this experiment were six tones created using Adobe Audition® software. Tones were equated in duration (80 ms) and intensity (74 dBA, calibrated with a Brüel & Kjær® sound level meter with an artificial ear). Table 1 summarizes the tone characteristics for each frequency stream (high, middle, and low). For descriptive purposes, the fundamental frequency of the tones will heretofore refer to the streams as high, middle, and low.

Eighteen semitones separated frequency streams (329.6–923.3–2367 Hz for low, middle, and high, respectively) so that sounds would segregate automatically, even without attention focused on the sounds (Sussman et al., 1999; Winkler et al., 2005) (Figure 1A). Tones were alternated (L, M, H, L, M, H, and so on; Figure 1B) with a 90 ms stimulus onset asynchrony

(SOA, onset-to-onset) and presented with NeuroStim hardware and software for PC (Compumedics Inc., Texas, USA). Tones in each stream had a unique convergence of tone characteristics. The high stream tones were octaves -- the fundamental and second harmonic -- with a slow-rise envelope shape (78 ms rise/2 ms fall), presented to both ears (center location). The high tone stream included standard tones (90% occurrence) and deviant frequency tones (2793 Hz) that randomly replaced 10% of the high tones, and occurred either once ('single', 5%) or twice successively ('double', 5%) (Figure 1C). The middle stream tones were pure tones with a slow fall sound envelope (2 ms rise/78 ms fall time), presented to the right ear (right location). Low stream tones were complex (four harmonics above the fundamental frequency), with a diamond envelope shape (40 ms rise/40 ms fall time) presented to the left ear (left location). Middle deviants were envelope deviants (40 ms rise/40 ms fall time) that randomly replaced 5% of middle standard tones and low deviants were envelope deviants (2ms rise/78ms fall time) that replaced 5% of low standard tones (Figure 1C). That is, deviants for the middle stream had the envelope shape of the low stream standards, and deviants for the low stream had the envelope shape of the middle stream standards. Thus, detection of "deviant" envelope shape would only be detected if the tones segregated to frequency streams. Each condition consisted of 21,600 tones (6480 standards (H), 180 single frequency-deviants (HD) and 180 double frequency-deviants (HDHD); 6840 M and L standards, and 360 M- and L-envelope deviants, respectively) presented in 12 separately randomized blocks of 1800 tones each (2.7 minutes per presentation block, 32.4 minutes total per condition).

2.3 Electroencephalographic (EEG) recording and data reduction

Electroencephalogram (EEG) was recorded with a 32-channel electrode cap with the International modified 10–20 system. Additional electrodes were placed at the left (LM) and right mastoids (RM). Horizontal electro-oculogram (EOG) was recorded using a bipolar configuration between F7 and F8 electrodes. Vertical EOG was recorded using the FP1 electrode in a bipolar configuration with an external electrode placed below the left eye. The reference electrode was placed at the tip of the nose, and the P09 electrode served as the ground. Impedances were maintained below 5 k Ω . The EEG and EOG were digitized (Neuroscan Synamps amplifier, Compumedics Corp., El Paso, Texas) with a 500 Hz sampling rate and bandpass of 0.05–100 Hz. The continuous EEG was then filtered offline (0.1–30 Hz) using a finite impulse response filter with zero phase shift and a roll-off slope of 24 dB/octave using Neuroscan SCAN software 4.3 for PC. The filtered EEG was then segmented into 1 s epochs with a 100 ms prestimulus period. Epochs were baseline corrected before artifact rejection was applied with a criterion set at $\pm 75\mu\text{V}$ on all electrodes (EOG and EEG). For four participants who had excessive eye-blink activity, Independent Component Analysis was performed (EEGLAB, Delorme & Makeig, 2004) and the signal reconstructed without those components highly correlated to the EOG. The corrected data were then baseline corrected across the whole epoch and artifact reject criterion was applied. Overall, 13% of epochs were rejected due to artifact. The remaining EEG epochs were averaged separately by stimulus type in each stream and in each condition, and then baseline corrected to the prestimulus period.

2.4 Procedures

Participants sat comfortably in a sound attenuated booth (IAC Acoustics, Bronx, NY). There were two task conditions: *attend-visual* and *attend-auditory*. In the *attend-visual condition*, participants were instructed to ignore the sounds presented to their ears via insert earphones (E-a--rtones® 3A, Indianapolis, IN) and watch a self-selected subtitled movie with no audio. In the *attend-auditory condition*, participants performed a go/no-go task. Participants were instructed to selectively attend to the high-frequency tones and press a response key when they detected the high-frequency double deviants and refrain from pressing the response key for the high frequency single deviants (Figure 1C). Participants were given a practice sequence prior to recording for the *attend-auditory condition*. The practice sequence was not used in the main experiment. Correct identification of at least 50% of the targets in the practice trial was required to proceed. All participants met this criterion.

All participants performed the *attend-visual condition* first followed by the *attend-auditory condition*. This was done to obtain a baseline measure of the sounds prior to any top-down knowledge about the structure or content of the sequence. A within-participants design was used so that individuals served as their own control to compare responses when attention was directed to the movie (*attend-visual*) or to the high tones in the sound sequence (*attend-auditory*). Total session time, including electrode cap placement, practice, and breaks, was approximately two hours.

2.5 Data analysis

2.5.1 Behavioral data.—Reaction time (RT), hit rate (HR), and false alarm rate (FAR) were calculated for button-press responses in the *attend-auditory condition*. Responses were considered correct when they occurred 100–900 ms from deviant onset. ‘Hits’ were correct button presses to the double deviants in the high tone stream. Reaction time was calculated as the difference between the sound onset of the second deviant of the double deviants and the button press response. Correct rejections were no-go responses to any other tones. False alarms were considered as button presses to any other tone that was not the second deviant of the double deviant. Due to the rapid pace of the stimuli, calculating FAR based on the total number of non-target stimuli possibly yields an underestimate of the FAR because there can be no expectation that a response can occur to every tone (Bendixen & Andersen, 2013). To provide a more conservative estimate, false alarm rate was adjusted for rapid presentation rate and calculated with the following formula:

$$f = \frac{F}{\left(\frac{T_s}{T_R}\right) - N_T} \text{ where } F \text{ is the number of false alarms, } T_s \text{ duration of the entire stimulus}$$

sequence (ms), T_R is length of the response window (ms) and N_T is the number of target events (Bendixen & Andersen, 2013).

2.5.2 Event-related potentials.—The MMN and P3b components were separately delineated in each stream and in each task condition in the grand-mean difference waveforms by subtracting the ERP response to the standard from the ERP response to the deviant, separately for each deviant type. High stream standards were subtracted from high

stream deviants (single and double separately), middle stream standards from middle stream deviants, and low stream standards from low stream deviants.

The mean amplitude of the MMN and P3b components were calculated by first identifying the peak latency of the component in the grand-mean difference waveform and then calculating the mean over an interval centered on the peak (50ms for the MMN and 60ms for the P3b) for each individual and stimulus type (standard and deviant). The peak was chosen from the right mastoid (RM) for the MMN due to potential overlap with the attention-related components at the frontal electrodes and from the Pz electrode for the P3b component. The intervals used to calculate the mean amplitudes are shown in Table 2.

2.6 Statistical verification and quantification of event-related potentials.

Separate omnibus repeated measures analyses of variance (rmANOVA) were calculated for MMN and P3b components. To verify the presence of the MMN component, determine its scalp distribution, and assess attention effects, a four-way rmANOVA with factors of task (auditory, visual), stream (H, M, L), stimulus type (standard, deviant), and electrode (Fz, Cz, Pz) was calculated on the mean amplitude referenced to the right mastoid. To verify the presence and topography of the P3b component, a three-way rmANOVA was conducted with factors of order (first, second), stimulus type (standard, deviant), and electrode (Fz, Cz, Pz) on the mean amplitude of the single (non-target) and double (target) deviants, separately.

Where data violated the assumption of sphericity, Greenhouse-Geisser corrections were applied. Corrected df and p values are reported, as well as epsilon values. For post hoc analyses, Tukey HSD for repeated measures was conducted on pairwise contrasts only when main effects or interactions of the omnibus ANOVA were significant. Contrasts were reported as significantly different at $p < .05$. All statistical analyses were performed using Statistica 12 software (Statsoft, Inc., Tulsa, OK).

3. Results

3.1 Behavioral performance

Overall, participants could perform the task well. Mean HR to target double deviants was 82% ($SD = 8$), with FAR less than one percent (0.001, $SD = 0.001$). Mean reaction time to the double deviant in the high stream was 690 ms ($SD = 46$).

3.2 Event-related potentials

Figure 2 displays the ERPs evoked by deviants in high, middle, and low streams overlain with the ERPs evoked by comparison standards in *attend-visual* and *attend-auditory conditions*. Difference waveforms showing the subtraction (deviant-minus-standard) are displayed in Figure 3. Table 2 summarizes the mean amplitudes of the deviants and standards for the interval measured to assess MMN and P3b components, and their respective difference values.

3.2.1 MMN amplitude.—MMN was elicited by all deviants (main effect of stimulus type, $F_{1,9} = 85.99$, $p < .001$, $\eta_p^2 = .91$). Overall, the amplitude of responses to deviant tones

was significantly more negative than the amplitude of the ERP responses to standards. The topographic distribution of responses along the midline electrodes (Fz, Cz, Pz) was characteristic of the MMN component (main effect: electrode, $F_{1.09, 9.84} = 24.01$, $\epsilon = .55$, $p < .001$, $\eta_p^2 = .73$), with more negative amplitude observed at Fz and Cz electrodes (Giard, Perrin, Pernier, & Bouchet, 1990). Post hoc calculations showed that Fz and Cz: -0.63 ; -0.76 μV were larger than Pz: -0.37 μV .

There was a significant interaction between stream, stimulus type, and electrode ($F_{1.81, 16.3} = 13.03$, $\epsilon = .45$, $p < .001$, $\eta_p^2 = .60$). Post hoc analysis revealed this was due to the ERP response to the middle envelope deviant having a smaller amplitude (less negative) than that elicited by the low and high stream deviants at Fz. There was no difference among standards elicited within low, middle, and high stream at any electrode. This is likely due to the rapid stimulus rate attenuating the overall amplitude of the obligatory components (Budd, Barry, Gordon, Rennie, & Michie, 1998; Näätänen & Picton, 1987).

There was no main effect of task on MMN amplitude ($F_{1, 9} = 0.42$, $p = .53$). Post hoc calculation showed no amplitude difference between MMNs elicited by middle and low deviants when high tones were attended (*attend-auditory*) compared to when all tones were ignored and attention was on watching a movie (*attend-visual condition*). There was a task effect on MMNs elicited by high stream deviants, which was shown by a significant interaction between attention, stream, stimulus type, and electrode ($F_{2.36, 21.26} = 4.56$, $\epsilon = .59$, $p = .018$, $\eta_p^2 = .34$). Post hoc calculations showed a larger (more negative) ERP response elicited by the high-frequency double deviants when attended in the *attend-auditory condition* (Fz: -3.0 μV) than when ignored in the *attend-visual condition* (Fz: -1.74 μV). There was with no difference in the response to high stream standards whether attended or ignored, which is likely due to the rapid stimulus rate attenuating the overall amplitude of the obligatory components (Budd et al., 1998; Näätänen & Picton, 1987). Thus, the task effect was due to a difference in the deviant response. However, there is overlap of the N2 target response with MMN in the *attend-auditory condition*. Thus, the larger response to attended vs. unattended high stream targets cannot be solely attributed to the MMN amplitude (Sussman, 2007).

In summary, MMNs were elicited by high, middle, and low streams deviants when all the sounds were ignored (*attend-visual condition*) as well as when the high stream was segregated out to perform a task (*attend-auditory condition*). MMNs were larger for high stream deviants when attended than when ignored, but the task performed (attend-visual vs. attend-auditory) did not affect the MMN amplitude of the unattended middle and low stream envelope deviants.

3.2.3 P3b amplitude.—Figure 4 displays the deviant-minus-standard difference waveforms and the voltage distribution maps of the P3b component elicited by the single and double deviants. Neither P3a (which is largest at Fz) nor P3b (which is largest at Pz) were elicited by envelope deviants in unattended low or middle streams in either *attend-visual* or *attend-auditory* conditions (Figure 2 for Fz electrode and Figure 4 for Pz electrode). ERPs elicited by target deviants in the *attend-auditory condition* were larger (more positive) than ERPs elicited by standard tones (main effect of stimulus type, $F_{1, 9} =$

9.55, $p < .013$, $\eta_p^2 = .77$), and were largest (most positive) at the midline Pz electrode, consistent with P3b scalp topography (Polich, 2007; Sutton et al., 1965). P3b amplitude was smaller in response to the first of the high stream double deviants than to the second of the high stream double deviants (main effect of order, $F_{1,9} = 30.78$, $p < .001$, $\eta_p^2 = .55$). Thus, the larger P3b amplitude was elicited by the second deviant that designated it as a target. There was a significant interaction between order, stimulus type, and electrode ($F_{1,40,12.55} = 26.70$, $\epsilon = .70$, $p < .001$, $\eta_p^2 = .75$). Post hoc calculations showed that the amplitude of the deviants was overall larger than the amplitude of the standards, and that the response to the second of the deviants had a broader scalp distribution than that elicited by the first. Specifically, standard and deviant differed at Fz, Cz, and Pz, whereas for the first of the deviants the distribution was focused over centro-parietal electrodes. Standard and deviant differed at Cz and Pz electrodes, and not at Fz (Figure 4).

P3b was elicited by the high stream single (non-target) deviants in the *attend-auditory condition* and had typical P3b scalp distribution (main effect of electrode, $F_{1,04,9.38} = 5.79$, $\epsilon = .52$, $p = .037$, $\eta_p^2 = .39$). The amplitude of the ERP response to the standard following the single deviant was larger (more positive) than the amplitude of the response to standards not following a single deviant (interaction between order, stimulus type, and electrode, $F_{1,35,12.18} = 5.40$, $\epsilon = .68$, $p = .03$, $\eta_p^2 = .38$). That is, the amplitude to the standard was larger when it was possible that a second deviant could occur than when there was no expectation that a deviant could occur. There was, however, no defined peak response to the standard in that latency. Instead, there was a sustained response that resolved (went back to baseline) after the expected time of occurrence of the second deviant of the double deviants (Figure 4A, pink trace, 600–800 ms range). This was during the waiting period in determining whether or not the second deviant would occur to signal a target response. In contrast, when participants ignored the sounds and watched a movie, and when the double deviant was not designated as a target for a button press response, P3b was not elicited by single or double deviants (*attend-visual condition*, Figure 4B).

In sum, P3bs were elicited by the first and by the second of the target double high frequency deviants, showing a temporal progression of events time-locked to the deviants. First, there was an evaluation period at the occurrence of the first deviant, and then there was detection of the target at the second deviant that then required a button press response. P3b amplitude was smaller to the first deviant (target evaluation) than to the second deviant (target detection). P3b was also elicited by single high-frequency deviants (target evaluation), and during the time of the high tone standard that followed the high tone single deviant, there was a sustained positive-going waveform where the second of the double deviants could be expected to occur (Figure 4A, 600–800 ms range, pink trace).

4. Discussion

The overarching goal of this study was to understand how attentional resources are distributed to attended and unattended sounds within an entire auditory soundscape. A key factor of this study was the unique combination of multiple tone features (frequency, envelope, spatial location, and tone quality) in the identity of each of three auditory streams, providing strong stimulus-driven cues for cohering distinct streams. The MMN ERP brain

component was elicited by deviants within unattended streams during visual and auditory selective attention tasks demonstrating allocation of processes across multiple streams regardless of the direction of attention. Our results do not fully support any one theory of attention but rather support a distributive theory of attention that balances stimulus-driven factors and task goals in analyzing the entire soundscape.

4.1 Theories of attention and stream segregation

We used four tone features to enhance stream distinctiveness by stimulus-driven factors and manipulated what the participant attended to (i.e., *attend-visual* vs. *attend-auditory* conditions) to determine if the differences in task demands modulated attentional resources available to process irrelevant, unattended information. MMNs were elicited by deviants within all three streams when attention was focused on movie-viewing (*attend-visual* condition). When attention actively segregated the attended stream from the other sounds (*attend-auditory* condition), MMNs were elicited by attended deviants and by deviants in the two unattended streams. This result contrasts our previous results using a similar paradigm as no MMNs were elicited by unattended deviants during selective auditory attention (Pannese et al., 2015; Sussman et al., 2005).

For MMN elicitation to occur by unattended within-stream deviants, several processes had to occur. The sounds that were irrelevant to performing the task had to be segregated to distinct streams. The tone patterns within each of those unattended streams had to be identified, and deviants within each stream had to be detected. These processes all occurred in parallel with those processes that segregate out one stream and perform a task with the attended, task-relevant sounds. This result thus demonstrates that the irrelevant sounds were not globally inhibited by selective attention. Because gain theories suggest bottom-up interference, this result would be more closely aligned with load theories of attention.

However, we found no effect of task load on processing unattended sounds. Similar MMN amplitudes were elicited by within-stream unattended deviants -- when watching the movie with all three streams unattended, and when attending to one of the three streams with two of the streams unattended. This suggests that task demands were not a key factor in modulating the degree of unattended sound processing. Arguably, task load is lower for watching a movie and ignoring all sound input (*attend-visual* condition) than it is for selectively attending to a subset of tones within an alternating L-M-H tone sequence, and detecting double frequency deviants within the attended stream (*attend-auditory* condition) while ignoring task-irrelevant sounds. In our previous three-stream studies, MMNs were not elicited by ignored sounds when this same task was performed with the high stream tones. Within-stream deviants elicited MMN in each of the frequency streams only when participants watched a movie -- ignoring all of the sounds (Pannese et al., 2015; Sussman et al., 2005). One might speculate that while watching a movie, attention was covertly available to segregate and detect within-stream pattern deviants, and that in the more demanding condition of performing a task with a subset of tones, resources were unavailable for higher-level processing of the unattended sounds. Thus, load theory seems a reasonable explanation for the results of our *previous* three-stream studies. However, task load cannot explain the results of the current study because MMNs of equal amplitude were elicited by unattended

deviants in the *attend-auditory* (high load) and *attend-visual* (low load) task conditions (Figure 3). Working memory load explanations also do not explain the results because these experiments all involved processing three sound streams, imposing a similar load on working memory.

Feature-based theories of attention may explain the discrepancy in results because in our previous studies frequency separation was the only cue for stream segregation. The other tone features (intensity, duration, and spatial location of the sounds) were common to all three streams. This may have resulted in greater confusability among the common features across sounds, destabilizing distinct streams. In the current study, unique features characterizing each stream could have reduced confusability and strengthened stream segregation.

Feature-based theories may also explain why attentional demand is greater when only a single feature distinguishes streams and processing is limited by it. In previous three-stream studies, a single tone feature for segregation may have been less effective for distinguishing sound streams. There would be more chance for stimulus uncertainty due to overlap of the other features within the global sound sequence (Störmer & Alvarez, 2014). If the streams were more confusable because most tone features overlapped across streams, then processing resources had to be expended to differentiate the streams, to segregate the global auditory scene. Thus, the difference in results between *attend-auditory* and *attend-visual* conditions in previous studies may have been due to greater processing resources required to segregate background sounds thus limiting resources for processing the higher-level within-stream structure of unattended streams. However, feature-based theories of attention only partially explain these results because if greater processing resources are required to segregate sounds due to stimulus confusability, the amount of resources available for processing higher-level within-stream sound events is affected. Thus, it is not simply a feature-based issue of increased contrast across the streams to reduce confusability; resources are still needed to process unattended within-stream patterns and deviants. Results thus demonstrate that stimulus-driven factors were in play, interacting with overall processing capacity in scene analysis (Murphy et al., 2017).

Our results are also compatible with the results of Dykstra and Gutschalk (2015) that used an informational masking paradigm to test whether MMNs would be elicited by deviants when the standards were not detected. In the Dykstra and Gutschalk study, using a demanding segregation task, MMNs were elicited by deviants only when the standards could be perceived; they were not masked. Thus, we can extrapolate that if processing demands are such to obscure the distinctiveness of the unattended standard (tones or patterns as in the current study), this could explain why in previous studies MMNs are not elicited when stimulus-driven factors only weakly segregate sounds. Unattended sounds would have been more susceptible to perceptual masking effects, which may have prevented the standards from emerging even if segregation of the sounds had occurred on the basis of frequency separation. Thus, with weak stimulus-driven factors in combination with a demanding auditory task expending attentional resources, unattended within-stream pattern processing may have been limited and MMNs not elicited, in previous studies. If we consider that task processing takes first priority for overall resources, then processing unattended sounds will

be limited according to both stimulus-driven and task demands. In this way, stimulus-driven segregation is a factor that can decrease the ‘cognitive load’ for processing the entire auditory soundscape. The current results along with others indicate that priority for processing is dependent on a number of factors, not just the direction or the load of attention (Dykstra et al., 2017; Dykstra & Gutschalk, 2015; Kidd et al., 2008; Pannese et al., 2015; Sussman et al., 2005; Störmer & Alverez, 2014).

4.2 Chronometry of task-specific responses

We used two types of deviants in the high tone stream (single and double deviants). This was done to increase difficulty of the task so that the listener could not correctly press the response key any time a high frequency deviant was heard. Participants had to withhold a response to the single deviants and press the response key for the double deviants. Thus, every time a high frequency deviant occurred, participants had to wait to find out if the next tone in the high stream would be another frequency deviant or a frequency standard. This evaluation period, initiated by the first deviant of the double deviant or by the single deviant, elicited a P3b component (Figure 4A, black and pink traces, 400 ms range). The second deviant, when the target could be identified, elicited another P3b component (Figure 4A, black trace, 600–800 ms range). The standard following the single deviant, the time that identity of a “nontarget” (no-go) response could be confirmed, elicited a sustained positive waveform (Figure 4A, pink trace 600–800 ms range). This positive potential was likely indicative of the continued processing until the identity of a “non-target” was confirmed by the occurrence of the next high stream standard tone.

The two distinct P3b responses elicited by the target double deviant show a dissociation in time between detection of a potential deviant and identification of the target requiring a motor response. The mean P3b amplitude was smaller during the evaluation phase (‘will it be a target?’) than the mean amplitude of the P3b elicited by the second deviant when participants resolved the identity of the target. However, the topography of the P3b elicited by the single deviants, and by the first tone of the double deviants, had similar and typical P3b topography (Figure 4C). The second deviant of the double deviant was larger in magnitude but also with typical P3b scalp distribution, having its maxima at the Pz electrode. This suggests that the two stages of target evaluation and detection represent disassociated aspects of the P3b target processing response rather than representing two different neural processes altogether.

These results are consistent with Sutton, Tueting, Zubin, & John (1967) who used a one-click vs. two-click target paradigm. P3b responses were evoked by the single clicks and by the two successive clicks. The response to the second of the two clicks had a larger amplitude, as was observed in the current experiment with single and double tone deviants. Sutton et al. attributed the larger amplitude response of the second click as a resolution of uncertainty. When the first click occurred, it was not yet certain whether a second click would be absent or present. Similarly, in the current experiment, when the second successive high-frequency deviant occurred, it provided participants with information to resolve the identity of the target with certainty. Thus, the first deviant tone initiated the evaluation

period and the second deviant tone initiated a resolution of uncertainty with respect to the identity of the target.

When participants watched a movie and deviants were irrelevant to the task, no such target-processing components (evaluation or detection) were elicited (Figure 4B). The absence of P3b demonstrates task-irrelevance of deviants while watching the movie, even though attention may have been sporadically diverted to the unattended sounds.

4.3 Conclusions

The current study tested how task demands modulate processing of unattended, background sounds. We used a three-stream paradigm with four unique acoustic features defining each sound stream (frequency, envelope shape, spatial location, and tone quality) to enhance stream distinctiveness, and compared effects of task load by having participants perform a difficult auditory task and an easy movie-viewing task. We found no effect of task load on unattended sound processing. MMNs were elicited by unattended deviants during both low- and high-load task conditions. In comparison with our previous studies, in which MMNs were not elicited by unattended deviants during high task load conditions, the results of this study demonstrate that deviance detection within unattended streams can occur in parallel with task demands for the attended stream. Further, the attention-based P3b response demonstrated a two-stage process of target evaluation. Thus, results reveal that stimulus-driven factors (e.g., stream distinctiveness) play a contributing role in the overall processing capacity of the auditory scene. Overall, results indicate that processing resources are distributed amongst attended and unattended sounds. In sum, results show that attention is preferentially distributed to the task, and the limitation placed on processing resources for irrelevant sounds can be altered by stimulus-driven factors, such as the strength of streaming. Task load is not a necessarily defining factor in limiting attentional resources (Murphy et al., 2017). Results of this study support a distributive theory of attention that balances the strength of the bottom-up input with perceptual goals in analyzing the auditory scene.

Acknowledgments

This research was funded by the NIDCD of the NIH (Grant # R01DC004263, E.S.S.). All of the content is the sole responsibility of the authors and does not necessarily reflect the opinions of the NIH. This work was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Graduate Division of Medical Sciences, Albert Einstein College of Medicine, New York (R.S.). J.W.Z. was a student in the Summer Undergraduate Research Program (SURP) of the Albert Einstein College of Medicine, Graduate Division of Biomedical Sciences. We thank G. Sadia, W. Lee, and J. DeMarco for assistance with data collection and analysis. The authors declare no conflict of interest.

References

- Anstis SM, & Saida S (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, 11(3), 257–271. 10.1037/0096-1523.11.3.257
- Beauvois MW, & Meddis R (1991). A computer model of auditory stream segregation. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 43(3), 517–541. 10.1080/14640749108400985 [PubMed: 1775655]
- Bendixen A, & Andersen SK (2013). Measuring target detection performance in paradigms with high event rates. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 124(5), 928–940. 10.1016/j.clinph.2012.11.012

- Bey C, & McAdams S (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology. Human Perception and Performance*, 29(2), 267–279. 10.1037/0096-1523.29.2.267 [PubMed: 12760614]
- Bregman AS (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology Human Perception and Performance*, 4(3), 380–387. 10.1037/0096-1523.4.3.380 [PubMed: 681887]
- Budd TW, Barry RJ, Gordon E, Rennie C, & Michie PT (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 31(1), 51–68. 10.1016/S0167-8760(98)00040-3
- Carlyon RP, Cusack R, Foxton JM, & Robertson IH (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception & Performance*, 27, 115–127. 10.1037/00961523.27.1.115
- Chelazzi L, Duncan J, Miller EK, & Desimone R (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, 80(6), 2918–2940. 10.1152/jn.1998.80.6.2918 [PubMed: 9862896]
- Cusack R, Deeks J, Aikman G, & Carlyon RP (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance*, 30(4), 643–656. 10.1037/0096-1523.30.4.643 [PubMed: 15301615]
- Cusack R, & Roberts B (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, 62(5), 1112–1120. 10.3758/bf03212092 [PubMed: 10997053]
- de Fockert JW, Rees G, Frith CD, & Lavie N (2001). The Role of Working Memory in Visual Selective Attention. *Science*, 291(5509), 1803. 10.1126/science.1056496 [PubMed: 11230699]
- Delorme A, & Makeig S (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. 10.1016/j.jneumeth.2003.10.009 [PubMed: 15102499]
- Desimone R, & Duncan J (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. 10.1146/annurev.ne.18.030195.001205
- Dinces E, & Sussman ES (2017). Attentional Resources Are Needed for Auditory Stream Segregation in Aging. *Frontiers in Aging Neuroscience*, 9, 414. 10.3389/fnagi.2017.00414. [PubMed: 29311902]
- Dykstra AR, Cariani PA, & Gutschalk A (2017). A roadmap for the study of conscious audition and its neural basis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372(1714). 10.1098/rstb.2016.0103
- Dykstra AR, & Gutschalk A (2015). Does the mismatch negativity operate on a consciously accessible memory trace? *Science Advances*, 1(10), e1500677. 10.1126/sciadv.1500677 [PubMed: 26702432]
- Giard MH, 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. 10.1111/j.1469-8986.1990.tb03184.x [PubMed: 2100348]
- Hartmann WM, & Johnson D (1991). Stream segregation and peripheral channeling. *Music Perception*, 9, 155–184. 10.2307/40285527
- Herrmann K, Heeger DJ, & Carrasco M (2012). Feature-based attention enhances performance by increasing response gain. *Vision Research*, 74, 10–20. 10.1016/j.visres.2012.04.016 [PubMed: 22580017]
- Hillyard SA, Hink RF, Schwent VL, & Picton TW (1973). Electrical signs of selective attention in the human brain. *Science*, 182(4108), 177–180. 10.1126/science.182.4108.177 [PubMed: 4730062]
- Hillyard SA, Vogel EK, & Luck SJ (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 353(1373), 1257–1270. 10.1098/rstb.1998.0281 [PubMed: 9770220]
- Horvath J, Czigler I, Sussman E, & Winkler I (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Brain Research. Cognitive Brain Research*, 12(1), 131–144. 10.1016/S0926-6410(01)00038-6 [PubMed: 11489616]

- Hughes RW, Vachon F, & Jones DM (2007). Disruption of short-term memory by changing and deviant sounds: support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 33(6), 1050–1061. 10.1037/0278-7393.33.6.1050
- Iverson P (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology. Human Perception and Performance*, 21(4), 751–763. 10.1037//0096-1523.21.4.751 [PubMed: 7643047]
- Kidd G Jr., Richards VM, Mason CR, Gallun FJ, & Huang R (2008). Informational masking increases the costs of monitoring multiple channels. *The Journal of the Acoustical Society of America*, 124(4), E1223–229. 10.1121/1.2968302 [PubMed: 19062790]
- Kok A (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3), 557–577. 10.1017/s0048577201990559 [PubMed: 11352145]
- Konstantinou N, Beal E, King JR, & Lavie N (2014). Working memory load and distraction: dissociable effects of visual maintenance and cognitive control. *The Journal of the Acoustical Society of America*, 136(7), 4085–4097. 10.3758/s13414-014-0742-z
- Lavie N (2005). Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82. 10.1016/j.tics.2004.12.004 [PubMed: 15668100]
- Lavie N, Hirst A, de Fockert JW, & Viding E (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology*, 133(3), 339–354. 10.1037/0096-3445.133.3.339 [PubMed: 15355143]
- Lavie N, & Tsai Y (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, 56(2), 183–197. 10.3758/bf03213897 [PubMed: 7971119]
- Luck SJ, & Vogel EK (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cognitive Science*, 17(8), 391–400. 10.1016/j.tics.2013.06.006
- Macken WJ, Tremblay S, Houghton RJ, Nicholls AP, & Jones DM (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology. Human Perception and Performance*, 29(1), 43–51. 10.1037/0096-1523.29.1.43 [PubMed: 12669746]
- Maunsell JH, & Treue S (2006). Feature-based attention in visual cortex. *Trends Neurosciences*, 29(6), 317–322. 10.1016/j.tins.2006.04.001
- Mesgarani N, & Chang EF (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397), 233–236. 10.1038/nature11020 [PubMed: 22522927]
- Mesgarani N, David SV, Fritz JB, & Shamma SA (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18), 6792–6797. 10.1073/pnas.1318017111 [PubMed: 24753585]
- Micheyl C, Tian B, Carlyon RP, & Rauschecker JP (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48(1), 139–148. 10.1016/j.neuron.2005.08.039 [PubMed: 16202714]
- Miller T, Chen S, Lee WW, & Sussman ES (2015). Multitasking: Effects of processing multiple auditory feature patterns. *Psychophysiology*, 52(9), 1140–1148. 10.1111/psyp.12446 [PubMed: 25939456]
- Moore BC, & Gockel HE (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1591), 919–931. 10.1098/rstb.2011.0355 [PubMed: 22371614]
- Moran J, & Desimone R (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715), 782–784. 10.1126/science.4023713 [PubMed: 4023713]
- Murphy S, Spence C, & Dalton P (2017). Auditory perceptual load: A review. *Hearing Research*, 352, 40–48. 10.1016/j.heares.2017.02.005 [PubMed: 28189838]
- Näätänen R, Gaillard AW, & Mantysalo S (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica (Amst)*, 42(4), 313–329. 10.1016/0001-6918(78)90006-9
- Näätänen R, & Picton T (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425. 10.1111/j.1469-8986.1987.tb00311.x [PubMed: 3615753]

- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, & Winkler I (2001). "Primitive intelligence" in the auditory cortex. *Trends Neurosciences*, 24(5), 283–288. 10.1016/S0166-2236(00)01790-2
- Pannese A, Herrmann CS, & 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. 10.1007/s10548-014-0368-4 [PubMed: 24771006]
- Polich J (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 118(10), 2128–2148. 10.1016/j.clinph.2007.04.019 [PubMed: 17573239]
- Puvvada KC, & Simon JZ (2017). Cortical Representations of Speech in a Multitalker Auditory Scene. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 37(38), 9189–9196. 10.1523/jneurosci.0938-17.2017
- Raveh D, & Lavie N (2015). Load-induced inattentional deafness. *Attention, Perception & Psychophysics*, 77(2), 483–492. 10.3758/s13414-014-0776-2
- Roberts B, Glasberg BR, & Moore BC (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *The Journal of the Acoustical Society of America*, 112(5 Pt 1), 2074–2085. 10.1121/1.1508784 [PubMed: 12430819]
- Rose MM, & Moore BC (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America*, 108(3 Pt 1), 1209–1214. 10.1121/1.1287708 [PubMed: 11008821]
- Shamma SA, Elhilali M, & Micheyl C (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosciences*, 34(3), 114–123. 10.1016/j.tins.2010.11.002
- Störmer VS, & Alvarez GA (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology: CB*, 24(17), 1985–1998. 10.1016/j.cub.2014.07.030 [PubMed: 25155510]
- Sussman ES (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, 117(3 Pt 1), 1285–1298. 10.1121/1.1854312 [PubMed: 15807017]
- Sussman E (2007). A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology*, 21(3), 164–175. 10.1027/0269-8803.21.34.164
- Sussman E, Ritter W, & Vaughan HG Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, 789(1), 130–138. 10.1016/S0006-8993(97)01443-1 [PubMed: 9602095]
- Sussman E, Ritter W, & Vaughan HG Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, 36(1), 22–34. 10.1016/S0006-8993(97)01443-1 [PubMed: 10098377]
- Sussman ES, Bregman AS, Wang WJ, & Khan FJ (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cognitive, Affective & Behavioral Neuroscience*, 5(1), 93–110. 10.3758/CABN.5.1.93
- Sussman ES, Horvath J, Winkler I, & Orr M (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, 69(1), 136–152. 10.3758/bf03194460 [PubMed: 17515223]
- Sutton S, Braren M, Zubin J, & John ER (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700), 1187–1188. 10.1126/science.150.3700.1187 [PubMed: 5852977]
- Sutton S, Tueting P, Zubin J, & John ER (1967). Information delivery and the sensory evoked potential. *Science*, 155(3768), 1436–1439. 10.1126/science.155.3768.1436 [PubMed: 6018511]
- Treue S, & Maunsell JH (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382(6591), 539–541. 10.1038/382539a0 [PubMed: 8700227]
- Vliegen J, & Oxenham AJ (1999). Sequential stream segregation in the absence of spectral cues. *The Journal of the Acoustical Society of America*, 105(1), 339–346. 10.1121/1.424503 [PubMed: 9921660]
- Winkler I, Czigler I, Sussman E, Horvath J, & Balazs L (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, 17(2), 320–339. 10.1162/0898929053124866 [PubMed: 15811243]
- Winkler I, Denham S, Mill R, Bohm TM, & Bendixen A (2012). Multistability in auditory stream segregation: a predictive coding view. *Philosophical Transactions of the Royal Society of London*.

Series B, Biological Sciences, 367(1591), 1001–1012. 10.1098/rstb.2011.0359 [PubMed: 22371621]

Woldorff MG, Hackley SA, & Hillyard SA (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28(1), 30–42. 10.1111/j.1469-8986.1991.tb03384.x [PubMed: 1886962]

Zion-Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, & Schroeder CE (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, 77(5), 980–991. 10.1016/j.neuron.2012.12.037 [PubMed: 23473326]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

A. *Stimulus features.* Sound streams were distinct from each other in four acoustic features: fundamental frequency, envelope shape, spatial location, and tone quality. Fundamental frequency for each tone stream is shown on the ordinate, with low tones at 329.6 Hz, middle tones at 923.3 Hz, and high tones at 2637 Hz. The stimulus quality (sound type) varied for tones within each stream, with low tones having four harmonics above the fundamental (complex tone, blue diamond), middle tones being sine wave tones (pure tone, green triangle), and high tones having one harmonic above the fundamental (octave tone, magenta triangle). The sound envelope (envelope shape) is depicted as slow rise and slow fall for the low tones (blue diamond), sharp rise and slow fall for middle tones (green triangle), and slow rise and sharp fall for the high tones (magenta triangle). Spatial location of the tones is labeled with high frequency tones presented to both ears (center), middle frequency tones presented to the right ear (right), and low frequency tones presented to the left ear (left). B. *Stimulus presentation.* Tones were presented in an alternating sequence (LMHLMH...). C. *Deviant types.* Within-stream deviants were embedded in all sound streams. The high stream deviants were randomly and seldomly occurring single (non-target) or double (target) frequency deviants (top row, black triangles). The middle and low stream deviants were the envelope shapes from the low and middle streams respectively (right two panels). Thus, the envelope shape itself was not deviant to the global sequence. Its detection depended upon stream segregation in which the deviant envelope shape in one stream was the standard envelope shape of the other stream.

Figure 2: Event-related potentials (ERPs).

Grand-mean ERPs elicited by standards (dashed lines) and deviants (solid lines) are displayed separately for the high tone stream (top row), middle tone stream (middle row), and low tone stream (bottom row) at the Fz electrode in the *attend auditory* (left column) and *attend visual* (right column) conditions. Amplitude is displayed in microvolts (μV) on the ordinate. The timing of the three tones following the deviant (HD1 and HD2, top row; MD, middle row, and LD, bottom row) is indicated along the abscissa, with time shown in milliseconds (ms).

Figure 3: Mismatch negativity (MMN) component.

Grand-mean difference waveforms, derived by subtracting the ERP response of the standard from the ERP response of the deviant (those illustrated in Figure 2), are displayed separately for the high (top row), middle (middle row), and low (bottom row) frequency streams in the *attend auditory* (left column) and *attend visual* (right column) conditions. The focus of attention onto or away from the tones is denoted for each deviant type in each frequency stream (e.g., attended or unattended). Fz (thick, solid black line) and RM (thin, solid black line) demarcate the MMN component, and significant MMN components are labeled. Amplitude is displayed in microvolts (μV) on the ordinate. The timing of the three tones following the deviant (HD1 and HD2, top row; MD, middle row, and LD, bottom row) is indicated along the abscissa, with time shown in milliseconds (ms).

Figure 4: P3b component.

A. Attend auditory condition. Grand-mean difference waveforms at Pz are displayed to show high stream responses to single, non-target deviant (pink lines) and the double deviant target (black lines). The first deviant response is consistent with target evaluation and the second deviant response is consistent with target detection. **B. Attend visual condition.** Grand-mean difference waveforms at Pz are displayed to show high stream responses to single, non-target deviant (pink lines) and the double deviant target (black lines). Tones were irrelevant to the task of watching a movie and deviants did not elicit P3b components. **C. Scalp voltage topography.** P3bs elicited by the single and double deviant in the *attend auditory* condition are displayed. Similar topography elicited by single and double deviants suggests different chronological phases of the target detection process.

Table 1:

Stimulus characteristics.

Stream	Fundamental (in Hz)		Sound Envelope (rise/fall in ms)		Tone Quality		Spatial Location	
	<i>STD</i>	<i>DEV</i>	<i>STD</i>	<i>DEV</i>	<i>STD</i>	<i>DEV</i>	<i>STD</i>	<i>DEV</i>
High	2637	2793	78/2	78/2	Octave	Octave	Center	Center
Middle	932.3	932.3	2/78	40/40	Pure	Pure	Right	Right
Low	329.6	329.6	40/40	2/78	Harmonics	Harmonics	Left	Left

Table 2:

MMN and P3b grand-mean amplitude (in μV) for the standard (STD) and deviants (DEV) with standard deviation (in parentheses), and differences (DIFF, deviant-minus-standard).

Attend-visual condition					
ERP Component	Interval measured (ms)	Stream	STD	DEV	DIFF
MMN (Fz)	119–169	High	0.01 (–0.22)	–0.91 (–1.02)	–0.92
	97–147	Med	–0.01 (0.29)	–0.57 (0.68)	–0.56
	101–151	Low	0.32 (0.30)	–1.44 (0.71)	–1.76
Attend-auditory condition					
MMN (Fz)	123–173	High	0.29 (0.42)	–1.73 (1.37)	–2.02
	105–155	Med	0.01 (0.43)	–0.54 (0.71)	–0.55
	101–151	Low	0.20 (0.29)	–0.98 (0.85)	–1.18
Attend-auditory condition Double deviant					
P3b (Pz)	368–428	High	–0.11 (0.73)	4.00 (5.14)	4.11
P3b (Pz)	632–692	High	–0.08 (0.72)	11.33 (6.19)	11.41
Attend-auditory condition Single deviant					
P3b (Pz)	374–434	High	–0.11 (0.73)	3.61 (5.14)	3.72
P3b (Pz)	632–692	High	–0.08 (0.72)	2.56 (3.62)	2.64