Auditory Object Segregation: Investigation Using Computer Modelling and Empirical Event-Related Potential Measures

Laurence Morissette, MSc.

Thesis submitted to the School of Psychology

in partial fulfillment of the requirements

for the PhD degree in Psychology

School of Psychology

Faculty of Social Sciences

University of Ottawa

Remerciements

Je tiens à remercier mon directeur de thèse, Dr. Sylvain Chartier, pour son travail de supervision. Je voudrais remercier les membres de mon comité, Dr. Kenneth Campbell, Dr. Denis Cousineau, et Dr. Charles Collins pour leur temps et leurs commentaires. Je me dois de remercier l'Université d'Ottawa et le Fond de recherche québécois de recherche nature et technologies pour le support monétaire qui a rendu possible la recherche présentée dans cette thèse. Je voudrais remercier mes parents pour leur support tout au cours de mes études universitaires. Thank you, Andrew Abel, for trying to cure me of my undue love for run on sentences. Thank you also to Melissa Johnson, Eric Kuebler and Idu Azogu for the interesting discussions about psychology research, and even more so for the interesting discussions not about psychology research.

Abstract

There are multiple factors that influence auditory steaming. Some, like frequency separation or rate of presentation, have effects that are well understood while others remain contentious. Human behavioural studies and event-related potential (ERP) studies have shown dissociation between a pre-attentive sound segregation process and an attention-dependent process in forming perceptual objects and streams. This thesis first presents a model that synthetises the processes involved in auditory object creation. It includes sensory feature extraction based on research by Bregman (1990), sensory feature binding through an oscillatory neural network based on work by Wang (1995; 1996; 1999; 2005; 2008), work by Itti and Koch (2001a) for the saliency map, and finally, work by Wrigley and Brown (2004) for the architecture of single feature processing streams, the inhibition of return of the activation and the attentional leaky integrate and fire neuron. The model was tested using stimuli and an experimental paradigm used by Carlyon, Cusack, Foxton and Robertson (2001). Several modifications were then implemented to the initial model to bring it closer to psychological and cognitive validity. The second part of the thesis furthers the knowledge available concerning the influence of the time spent attending to a task on streaming. Two deviant detection experiments using triplet stimuli are presented. The first experiment is a follow-up of Thompson, Carlyon and Cusack (2011) and replicated their behavioural findings, showing that the time spent attending to a task enhances streaming, and that deviant detection is easier when one stream is perceived. The ERP results showed double decisions markers indicating that subjects may have made their deviant detection based on the absence of the time delayed deviant and confirmed their decision with its later presence. The second experiment investigated the effect of the time spent attending to the task in presence of a continuity illusion on streaming. It was found that the

presence of this illusion prevented streaming in such a way that the pattern of the triplet was strengthened through time instead of separated into two streams, and that the deviant detection was easier the longer the subjects attended to the sound sequence.

Table of contents

List of Figures

List of Tables

List of Appendices

Introduction

This thesis investigates the influence of attention on auditory streaming and the formation of auditory objects. Chapter 1 presents a review of the current theories of auditory object creation and streaming, followed by various empirical and electroencephalographic studies of the variables that influence streaming. This includes the time spent attending to the task, the physical characteristics of the stimuli, and attention. Chapter 1 also presents existing computational models of auditory object creation, including oscillatory models and saliency maps, as they are the building blocks of the neural network model of auditory object creation presented in this thesis. Using information obtained from experimental studies on saliency, and the modulatory influence of attention on this perceptual process, the next part of this thesis, presented in Chapter 2, focuses on the creation of a model of the auditory system using artificial neurons. These artificial neurons are used to represent the structural and functional workings of the creation of auditory objects. The initial model was presented at the IJCNN conference (Morissette & Chartier, 2015), and the model was subsequently refined.

The final part of the thesis, presented in Chapter 3, gathers more experimental data on the influence of the time spent attending to an auditory streaming task. The level of influence that attention has on auditory object formation and stream separation is a controversial subject, with some research showing that streaming is pre-attentive (Bregman, 1990; Sussman, Ritter & Vaughan, 1999) and that auditory object creation is based on the intrinsic characteristics of the stimuli. However, other studies have shown that attention has a modulatory effect on the speed of creation of those auditory objects (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Shamma, Elhilali & Micheyl, 2011; Sussman, Winkler, Huotilainen, Ritter & Näätänen, 2002; Sussman, 2007;

Thompson, Carlyon & Cusack, 2011). Two deviant detection experiments using triplet stimuli, based on an existing methodology by Thompson, Carlyon and Cusack (2011), are presented. Event-related potentials (ERPs) measures were added to the behavioural measures to gain insight into the cognitive processes behind the deviant detection. The first experiment is a direct follow-up of the auditory streaming task carried out by Thompson et al. and the second experiment investigated the effect of the time spent attending to the task in the presence of a continuity illusion on streaming.

Chapter 1

Literature Review

Neuroanatomy of audition

Audition is a sensory modality in which the flow of information never ceases, even during sleep, and in which the information received comes from all directions. By contrast, our visual field is much more limited as we cannot see when our eyes are closed. Light does not bend around corners, yet sound will rebound off objects. Because we have 360 degrees of hearing, cannot close our ears during sleep, and can hear what we cannot see, the auditory modality is well-suited to alert us of impending danger and for this reason is often used in alarm systems.

The auditory system converts sound waves into neural influx in the cochlea in the inner ear, more specifically through the basilar membrane that runs along its length (see Figure 1). When sounds reach the ear, the sound waves travel along the basilar membrane and different locations along the basilar membrane vibrate at different frequencies. Frequency discrimination of the incoming sound is done by the localization upon the basilar membrane of the peak vibration, and by the frequency of the vibration of the basilar membrane, which matches the stimulus. The sound wave is transformed into a neural signal by the mechanical movement of the inner hair cells on the basilar membrane when the basilar membrane moves. Since the hair can only move in one direction, the hair cells encode a half-wave rectified version of the stimulus, representing only one half of the incoming sound waves. The signal from the hair cell then goes to the spiral ganglion cells that form the auditory nerve. Auditory cells in the nerve respond preferentially to certain frequencies and fire at a rate that follows a sigmoid function (s shaped function where $S(x) = \frac{1}{1+e^{-x}}$) of the stimulus rate.

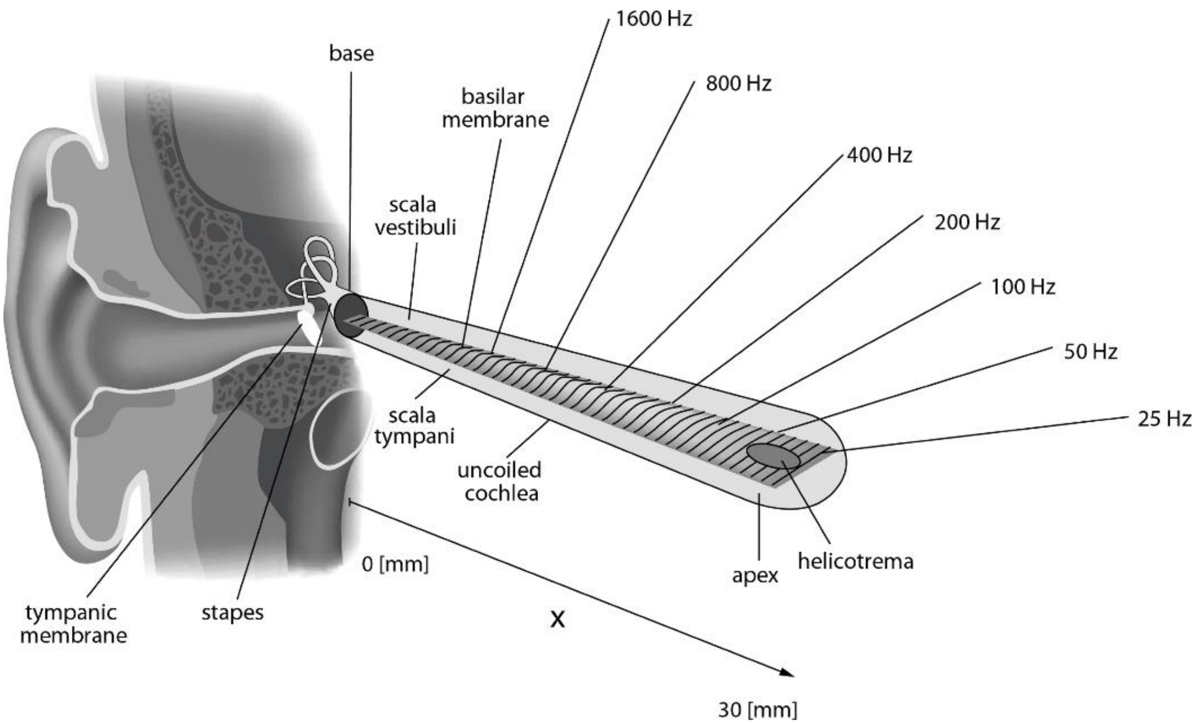


Figure 1. Uncoiled cochlea with basilar membrane (Diagram retrieved from Wikimedia)

In audition, spatial representation refers to the position on the cochlea of each frequency. The auditory nerve encodes the fine temporal structure of basilar membrane vibrations by synchronizing (termed ‘phase-locking’) them up to high frequencies; 4 kHz in mammals. Within this range, the frequency of a tone is encoded in the auditory nerve both spatially, by its characteristic frequency location, and temporally, by the periodicity of the activation in the auditory cells that respond preferentially to this central frequency. Complex sounds (sounds composed of several pure tones) are encoded as the superposition of the activation elicited by each component (Shamma, 2001). This process will be included in the sensory feature extraction of the model that will be presented in Chapter 2.

The activation from the cochlea then proceeds to the cochlear nucleus. The cochlear nucleus is organised tonotopically (creating a spatial tonotopic axis), so nerve cells responsive to low frequencies are at one end and nerve cells that are responsive to high frequencies are at the opposite end. The activation is then transmitted to the superior olivary complex, which is involved in the detection of interaural level and time differences. The neural information then continues to the inferior colliculus and the medial geniculate body (thalamus), where the ventral subnucleus is responsible for relaying frequency, intensity (amplitude) and binaural information to the cortex and the dorsal subnucleus is responsible for the detection of the location of a sound. The information is then passed to the primary auditory cortex of the brain, situated in the temporal lobe. The information treatment by the auditory information pathway just presented is implemented in the auditory process model that will be presented in Chapter 2, which models the treatment of the location, frequency and amplitude information.

Pitch/Frequency

The most studied attribute in auditory analysis is pitch. Pitch refers to the percept evoked by the objective stimulus frequency of a single tone or a harmonic tone complex. When a mixture of pure tones reaches the ear, pitch normally corresponds to the fundamental (lowest) frequency of the harmonics (multiples of the fundamental frequency) of the tones contained in the mixture. Frenquency is usually measured in Hertz (Hz). The human hearing range is between 20 Hz to 20 kHz, with a peak sensitivity between 2 and 4 kHz. There are two main theories of how the auditory system encodes frequencies (Shamma, 2001). The temporal theory, based on frequency and pioneered by Seebeck (1841), hypothesises that the frequency of vibrations in the basilar membrane is encoded by the number (or frequency) of axon potentials in the auditory nerve.

Shamma (2001) added the periodicity of the response in each auditory nerve fiber is evaluated the autocorrelation of the response with itself, which she implemented in her model of frequency estimation that will be used in the next chapter. Periodicity is the measure of the interval necessary for a sound wave to complete a cycle while autocorrelation is the comparison of a signal with itself over a time period. The results are combined from across all fibers to get the final estimate without reference to a spatial tonotopic axis (Langner, 1992; Licklider, 1951; Meddis & Hewitt, 1991; Slaney & Lyon, 1993).

The spectral (or spatial) hypothesis (Helmholtz, 1863), also called place theory, hypothesises that it is the place that is stimulated along the basilar membrane of the cochlea that encodes frequency. Each point along the basilar membrane has a frequency to which it is most responsive, creating a topographic mapping of the frequencies called the tonotopic axis. Low frequencies are encoded by a general movement of the basilar membrane.

There is some psychoacoustical evidence for both place and temporal codes. Pitch discrimination deteriorate at high frequencies, making temporal code necessary for accurate pitch discrimination (Oxenham, Micheyl, Keebler, Loper, Santurette, 2011). On the other hand, another study found that frequency discrimination is based on temporal coding at low frequencies but beyond 8 kHz the temporal information is inferior to place information (Moore & Ernst, 2012). Place information is important for frequency discrimination even at lower frequencies. A study (van de Par & Kohlrausch, 1997) used “transposed tones”, tones that produce a temporal response in the auditory nerve similar to the auditory-nerve response to a low-frequency tone, dissociating temporal from place cues. Pitch discrimination was poor for the transposed tones, which suggests that temporal information may not be sufficient to produce good pitch perception and that place information may be necessary.

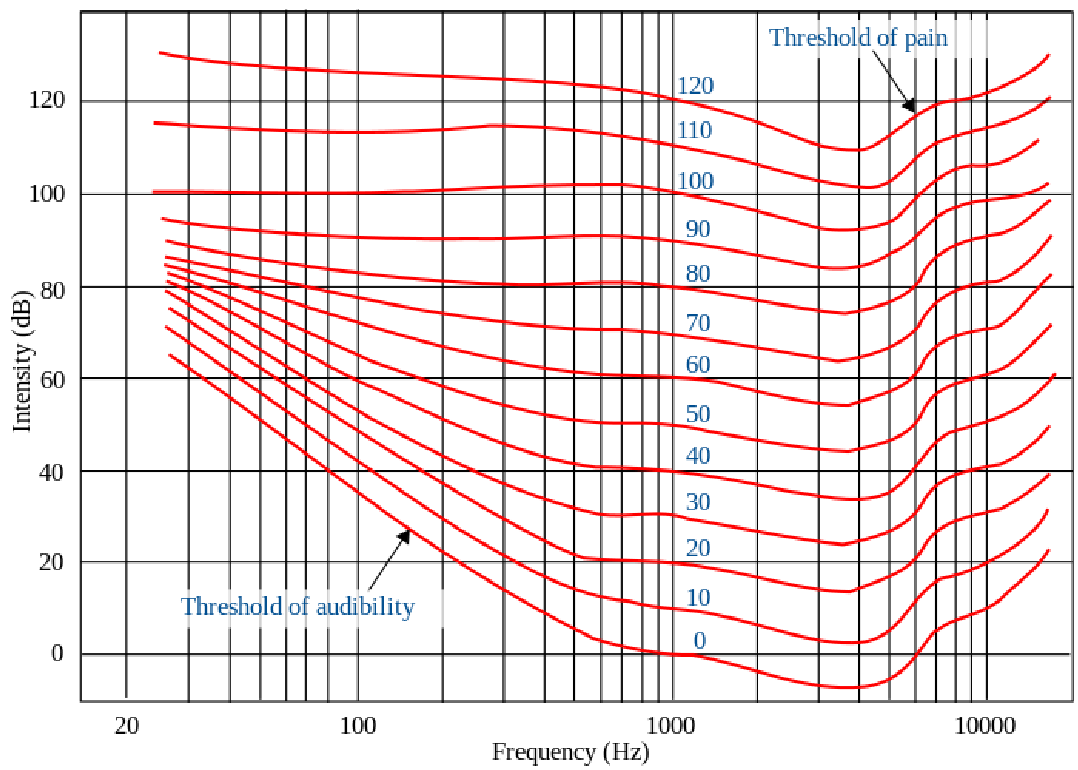


Figure 2. Fletcher-Munson equal loudness contours, showing the perceived loudness of a sound as a function of the frequency and intensity of the sound

Amplitude/Loudness

Loudness is the perceptual measure of the experience of sound intensity (amplitude of the signal). The loudness of sound is dependent on several other acoustical variables, such as frequency and spectral bandwidth. The perceived loudness differs according to the sensitivity to frequency of the auditory system. Tones of the same intensity, but of different frequency, are perceived as being of different loudness (see Figure 2 for representation). Loudness is often represented through equal-loudness contours, a measure of sound pressure (dB SPL), over the frequency spectrum, for which a listener perceives a constant loudness when presented with pure steady tones. Human hearing is most sensitive to frequencies between 2 and 4 kHz. In the logarithmic scale, the range of audible sound to the human ear is from 0 dB SPL (hearing threshold) to 120-140 dB SPL (pain threshold; SCENIHR, 2008). The neural encoding of loudness depends on firing rates and the number of neurons firing. A louder stimulus will cause more neurons to fire and for them to fire at a higher rate.

Source localization

Treatment of pitch is processed separately for each ear, but sounds reach the auditory system through both ears at once. In binaural sound processing, the central auditory system compares the signals arriving to the two ears, detecting and utilizing various imbalances (between the sound levels or frequencies of signal), time of arrival (onset of signal) and phase (peak of signal) to perform source localization in space and signal-to-noise enhancement. Following the dual theories of pitch encoding, spatial localization can also rely on spatial or temporal differences (Shamma, 2001). In binaural models like the Jeffress model (1948), the time delays between the two ears are used to estimate the azimuthal (position in a 360º circle around a central point) location based on the phase-shifts between the responses for equivalent frequencies between the two ears. The correlation between the responses is done by an organised array of neural delays.

Auditory Object Creation

Once the different auditory attributes have been treated separately, they must be bound to create what is known as an “auditory object”. An auditory object is a perceptual construct corresponding to the sound that can be assigned to a particular source. It has spectrotemporal features (pitch, loudness, timbre) that make it separable from other auditory objects (Bizley, 2013).

Auditory organisation of the sensory percept binds sensory features together into coherent structures, and segregates them from dissimilar ones. One of the first to describe the process of auditory organisation was Heise and Miller (1951), who found that subjects can split a signal made of two sine waves into two auditory streams when there is as little as a 15% difference in frequency between the streams, for frequencies from 150 Hz to 7000 Hz. According to Bregman (1990), auditory organisation follows two steps: first, the auditory stimulus is decomposed into segments, and those segments are then grouped into streams following what Bregman calls the ‘‘analysis-synthesis’’ process. Segments can be grouped either within a time-frame or across time (sequentially). Segments can also be grouped according to the bottom-up emergent structure of the stimulus, through a process called primitive grouping, or through a top-down process based on prior learned structures called schema grouping. As an example, tones that belong to a familiar pattern will be grouped together in a stream more easily, through schema grouping (Wang & Chang, 2008).

Several characteristics of the stimulus will cause primitive grouping in accordance with Gestalt theory: 1) proximity in frequency; 2) periodicity, where a segment is a harmonic of another segment frequency; 3) a continuous or discontinuous but smooth transition in frequency, where the temporal modulation in frequency between the segments is similar; 4) temporal proximity, where two segments are presented close in time or have similar onset and offset times; and 5) common source, where two segments are considered to come from the same source (Bregman 1990). Note that the same Gestalt principles apply to other features of the auditory object, such as amplitude, where primitive grouping will be done based on proximity, continuity or temporal proximity. A sequence of auditory objects that unfold over time and that is identified as originating from a unique source forms an auditory “stream” (Bizley & Cohen, 2013).

One of the rules of Gestalt theory is that human perception will tend to interpolate or extrapolate perceptual "objects" if the boundaries of the objects are obscured, and perceive them as continuous even when no such information is available to the auditory system. In audition, this creates a percept called the auditory continuity illusion (Miller & Licklider, 1950). As an example, when a silent period (a gap) is introduced into a target sound, the observer perceives two separate shorter duration sounds. However, if noise is presented during the gap, the target is perceived as being continuous, creating a continuity illusion. Expressed more formally, if a target sound is partially overlapped by noise (called a mask), continuity will occur if the target sound and the mask are sharing a common border (the onset of the noise is the offset of the target sound) and have abrupt transitions from the target to the mask. Moreover, the target sound and the mask must be presented to the same ear (i.e., there is no interaural time difference between the two sounds). The key to the grouping principles underlying both streaming theories and the continuity illusion is the idea of predictability: the auditory system must generate a prediction from current and previously presented sounds to build a model of what is likely to occur next (Winkler, Denham & Nelken, 2009; Bregman, Colantonio & Ahad, 1999).

Auditory streaming

Bregman (1990; Bregman, Ahad, Crum & O'Reilly, 2000) conducted studies on the relationships between frequency and presentation rate in auditory segregation. He presented alternating tones A and B of different frequencies at different rates of presentations. Bregman found that at a high rate of presentation (100 ms onset-to-onset), subjects perceive two different streams (a stream of As and a stream of Bs), while at lower rate (400 ms onset-to-onset), they perceive a single stream (ABABAB), a percept that is consistent with Gestalt laws of grouping.

Dannenbring and Bregman (1976) and Bregman (1990) used different semitone separations to evaluate the effect of frequency separation between the A and B tones. A semitone is equal to a twelfth of an octave or half a tone. When the semitone separation between frequency A and B is small (0.5 semitones), listeners are likely to report hearing one auditory stream. When this separation is large (>10 semitones), listeners report hearing two auditory streams. At intermediate frequency separations (such as 3–7 semitones), the perception of ‘one stream’ and ‘two streams’ alternates over time, creating a ‘bistable’ percept. However, with increased listening time, a stable two-stream percept may develop (Bizley & Cohen, 2013).

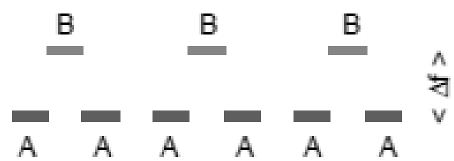


Figure 3. Van Noorden stimuli. The stimuli are composed of A and B pure tones with frequency separations that create different experimental conditions. The stimuli can either be perceived as a single galloping stream or two alternating streams

Streaming is often studied through an ABA paradigm that uses sequences of alternating A and B tones, where A and B are tones of different frequencies, presented following an ABA-ABA-ABA-ABA… pattern (see Figure 3). Van Noorden (1975) carried out a now classic study on the influence of pitch and rate of presentation on auditory streaming. Van Noorden found that when the A and B frequencies are similar or when the rate of presentation is slow, the subjects hear one stream of alternating sounds, termed a “galloping” stream because it mimics the sound of a horse’s gallop. When the rate of presentation is higher or the frequencies farther apart, the subjects hear two streams.

Figure 4 is a summary of the streaming percepts present in the Van Noorden study, as a function of the frequency separation between the A and B tone and the rate of presentation (shown as Stimulus Onset Asynchrony on the graph).

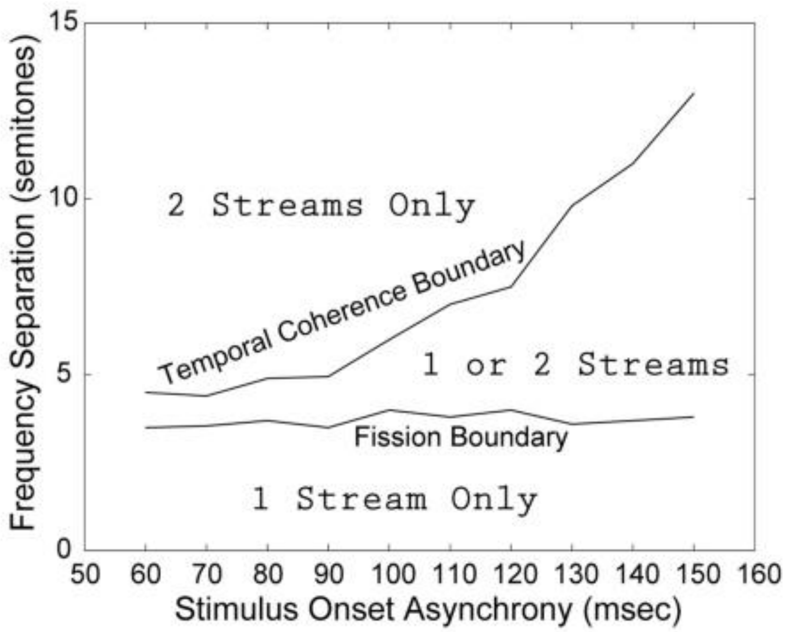


Figure 4. Streaming diagram from Snyder, Gregg, Weintraub and Alain (2012), presenting Van Noorden (1975) results. The streaming percept is presented as a function of the frequency separation between the A and B tone and the rate of presentation (stimulus onset asynchrony). At large frequency separations and slow presentation rate, 2 streams are perceived. The limit where the percept goes from being bistable to two streams is called the temporal coherence boundary, and the limit where the percept goes from bistable to one stream is the fission boundary

Stream segregation progresses with time; thus one might report hearing only a single galloping stream early in a sequence of triplets, and then report hearing two streams later in the sequence (Anstis & Saida, 1985). Van Noorden (1975) explained this result by a bandpass filter. If the tones pass through this filter in succession, a single stream will be heard. The bandpass filter can change its central frequency with time, but only at a certain rate. For fast presented stimuli, the filter cannot change its frequency fast enough and tones are segregated. Anstis & Saida (1985) explained the results through frequency change detectors. When the detectors register a transition, a single stream will be heard. These detectors tire with time, thus explaining why stream segregation takes time. Bregman et al. (2000) posit that activation in a neural circuit that has just registered the frequency of a tone will be extinguished with time unless it is reactivated by another stimulus. The strength of the reactivation is inversely proportional to the onset difference and the frequency separation between the tones.

Streaming in the auditory cortex

Studies of single- and multi-neuron recordings in the primary auditory cortex (A1) have found that different percepts of streaming have a direct influence on cortical activity. By presenting alternating A and B tones, Fishman, Reser, Arezzo and Steinschneider (2001) found that at slower presentation rates (5 or 10 Hz), cortical sites that respond preferentially to the A frequency showed marked responses to both the A and the B tones, so that the temporal pattern of neural activity varied according to the presentation rate. In contrast, at faster presentation rates (20 or 40 Hz), the responses to the B tones were markedly reduced, and the neural activity patterns consisted predominantly of the A tone responses. So, at fast presentation rates, where, according to the psychoacoustical literature, the tone sequences are likely to evoke a percept of two separate streams, the A and B tones excited distinct neural populations in A1. In contrast, at slow presentation rates, where the stimulus sequence was presumably perceived as a single stream, the A and the B tones excited largely overlapping populations.

Elhilali, Ma, Micheyl, Oxenham and Shamma (2009) found that, independent of semitone separation, when tone A and tone B were presented simultaneously, listeners reliably reported one stream. This neural mechanism, temporal synchrony, is an important factor in the formation of perceptual streams, but is not essential, as temporally coherent sounds can be streamed (Micheyl, Kreft, Shamma & Oxenham, 2013). The topographic explanation stipulates that when the frequency separation between tone A and B is small, only one peak of activity is produced in the primary auditory cortex when presented with an ABA stimulus. For a large frequency separation, there are two activity peaks, decoded as two streams. In the case of a bistable percept, the number of peaks is unclear (Micheyl, Carlyon, Gutschalk, Melcher, Oxenham, Rauschecker, Tian & Wilson, 2007).

Pressnitzer, Sayles, Micheyl, and Winter (2008) found that the cochlear nucleus is important for constructing the stream and the perceptual organisation of the auditory scene. Activity in regions in the frontal and parietal lobes are also correlated with the percept of streaming reported by subjects (Micheyl et al., 2007). Gutschalk, Micheyl, Melcher, Rupp, Scherg and Oxenham (2005) recorded magnetic evoked responses to sequences of ABA triplets. The magnitude of the P1m/N1m peaks (occurring from 50 to 100 ms after stimulus onset) evoked by the B tones at different A–B frequency separations were statistically correlated with psychophysical measures of streaming in the same listeners. The magnitude of the P1m/N1m in the response to the B tones was consistently larger when the percept was “two streams” than when it was “one stream” (Micheyl et al., 2007).

Attention

The effects of attention on the perception of streaming have also been studied. Attention is usually studied along a dichotomy, but there has been poor agreement in the usage of the endogenous/exogenous dichotomy and top-down/bottom-up dichotomy, especially when used as a synonym to the voluntary/involuntary attention dichotomy (Schröger, Marzecova & San Miguel, 2015). In this thesis, exogenous attention is taken to be the stimulus driven attention created by salient auditory stimuli, stimuli that are either novel (either in frequency, amplitude, spatial localization or duration) or loud. Salience is an interaction between stimuli and a sensory system that leads to the act of choosing relevant from irrelevant stimuli (Robinson & Petersen, 1992). Endogenous attention is defined here as the conscious focalisation of conscience. The studies discussed here looked at focused attention, the ability to attend to specific stimuli; selective attention, the ability to sustain a response to specific stimuli in the presence of distracting stimuli and alternating attention, the ability to shift the focus of attention between tasks.

Effect of time on task on streaming: behavioural studies

Human behavioural studies have shown dissociation between a pre-attentive, low-level sound segregation process, and a process modulated by the time spent attending to stimuli that can be called into play in forming perceptual objects and streams (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Thompson, Carlyon & Cusack, 2011).

Bregman and Rudnicky (1975) showed that low-level auditory streaming can be carried out without attention in an automatic innate process. They added bracketing tones (X) around the A and B tones, thus presenting a XABX sequence. The frequency of the X tone was chosen so that it could either be grouped with the AB tones if no other tones were presented or be grouped with a second stream of tones of similar frequency when they were presented. The authors found that when the X tones were considered part of a different stream than the AB tones, the subjects were better on the pattern recognition task to be done on the AB tones. The grouping mechanism pulled attention away from the X tones, thereby showing that the grouping mechanism was not part of attention.

Carlyon et al. (2001) employed a paradigm in which they presented ABA triplet stimuli in one ear for 21 seconds. The frequency separation between the A and B tones was manipulated. The A and B tones differed by 4, 6, 8 or 10 semitones in different conditions. Subjects were asked to press a button when they heard 1 stream and a second button when they heard two streams. Streaming builds over time (Anstis & Saida, 1985), and thus it was expected that perception of two streams would be more probable later rather than earlier in the 21 second sequence of triplets. This was found to be the case. In addition, frequency separation between the A and B tones affected stream segregation, with larger frequency separations leading to faster streaming, as Van Noorden had found (1975). The authors also examined the effects of attention on the perception of streaming. The authors presented 400 ms noise bursts over a 10-s period in the ear opposite to the one in which the ABA triplets were presented. The noise bursts either increased or decreased in intensity. In one condition, the noise bursts were irrelevant and were to be ignored. The subjects were asked to attend to the triplet channel to determine whether they perceived one or two streams. In another condition, the subjects attended the noise bursts and were asked to determine whether they were increasing or decreasing in intensity. They thus ignored the ABA triplets in the opposite ear during this 10 seconds (s) interval. Then, at the end of the noise burst sequence, subjects then switched their attention to the opposite channel containing the ABA triplets. Again, they were asked to report whether they now heard one or two streams. The authors were investigating whether the streaming would build up when the triplet channel was not attended. When subjects attended the noise bursts of 10 s, they nevertheless also passively processed the unattended triplet sequence in the opposite ear. If stream segregation occurs in the absence of attention, when attention was subsequently switched to the triplet channel, the subjects should have reported a similar proportion of two stream perception, compared to when they only attended the triplet channel. However, this was not the case. When subjects only attended the ABA triplet sequence, stream segregation was more apparent later in the sequence compared to when attention was later switched from the noise to the triplet channel. The experience of streaming does therefore appear to be dependent on attention being directed to the sequence containing the triplets.

Cusack et al. (2004) followed up on this study using the same paradigm, but presenting the noise bursts in the same ear as the to-be-attended triplets in addition to the condition of presenting the noise bursts in the opposite ear as the triplets. They found that when attention is focused on the noise bursts, there is substantially less build-up in auditory streaming in the ABA triplets, independently of the ear the noise bursts were presented in. They found a small but significant amount of segregation of the ABA sequence, even immediately after the task switch. The authors concluded that either the listeners were imperfectly focusing their attention on the noise bursts and allocating some to the ABA tones, or that streaming can occur in the absence of attention. They also found that the build-up of stream segregation was slower when attention was directed to noise bursts in the same frequency region as the triplets, even if the sounds were in the same ear. The authors also tested how streaming decreases, comparing the leaky integrator model of streaming hypothesis (Beauvois & Meddis, 1996) that posits that streaming decays with a similar time constant to its build-up, and the Bregman (1978) hypothesis that streaming can be reset by abrupt changes. Their results support that the streaming mechanism modulated by attention can be reset by an attentional shift away from the sequence.

A problem with the research presented by Carlyon et al. (2001) and Cusack et al. (2004) was that their measure of streaming was subjective. It relied on subjects subjectively reporting what they were experiencing. Thompson et al. (2011) developed a more objective method to measure streaming. In the ABA triplet sequence, they created a “deviant” by delaying its onset from the time it was expected to occur. This deviant is easier to detect when the sequence is perceived as a single stream rather than two streams (i.e., earlier rather than later in the sequence; Bregman & Dannenbring, 1973; Roberts, Glasberg & Moore, 2008; Vliegen, Moore & Oxenham, 1999; Vliegen & Oxenham, 1999; Warren, Obusek, Farmer & Warren, 1969). Subjects were asked to press a button when they heard a deviant sound, instead of making a streaming judgement. It was expected that detection of the deviant would be easier to make early in the sequence when subjects presumably perceived a single stream, and by evaluating the ease of deviant detection, the authors could measure streaming.

Two experiments were run. In the initial experiment, subjects were presented with 25 ABA triplets. The time between A and B tones was 75 ms and the time between ABA triplets was 125 ms. A deviant stimulus was created by delaying the onset of the B deviant either 30 or 50 ms after its usual onset time (time delay condition) infrequently in the sequence. The frequency difference between the A and B tone was either 4 or 8 semitones. The time at which the deviant B tone was presented also varied in four different conditions. In the No deviant condition, a deviant was not presented at all. In the Early condition, the B deviant occurred early, in the 5th triplet. In the Late condition, it occurred late, in the 20th triplet. In the Both condition, it occurred in both the 5th and 20th triplet. The authors reported a complex interaction between delays in presentation time and frequency. The results for the first experiment showed that subjects were indeed poorer at detecting the deviant in the Late condition when the frequency difference between A and B was large. Conversely, the time of occurrence of the deviant (early and late) had no effect when the frequency difference between A and B was small.

Thompson et al. (2011) carried out a second experiment, in which attention to the triplets was manipulated. The deviant was created by a 50 ms delay to the B tone in an ABA triplet. The authors kept the two frequency separation conditions (4 and 8 semitones). In this experiment, the deviant could only occur early or late. Like in the Carlyon et al. (2004) study, they presented ABA triplets in one ear and noise bursts in the opposite ear, lasting 10 s. In one condition, subjects were asked to attend to the ABA triplets and to signal detection of the deviant. In the other condition, subjects were asked to initially attend the noise bursts for 10 s, and then switch to the ABA triplet channel to detect the possible deviant. As in their first experiment, in the small frequency difference condition, the detection accuracy of the deviant occurring early in the attended condition, late in the attended condition or late in the switched condition did not significantly differ. For the larger frequency difference (8 semitones), when subjects attended to the tones throughout, the detection of a deviant was poorer later rather than earlier in the sequence, which is consistent with streaming having built up over time. The detection of a late deviant was better when subjects switched attention midway through the sequence, compared to when they had attended the triplet sequence for its entire duration. This is again consistent with streaming having built up less during the time the subjects were instructed to ignore the triplet sequence.

Billig and Carlyon (2016) conducted a follow-up study using ABA triplets to further investigate the influence of attention on streaming, combining both the subjective impression of streaming reported in previous Carlyon studies and the deviant detection found in the Thompson et al. (2011) study. In this experiment, there was again either a 4 or 8 semitones frequency difference between the A and B tones. Deviant triplets, where the middle B tone was delayed by 50 ms, could occur in three positions: early (sixth triplet), middle (18th from last triplet), or late (seventh from last triplet) with eight combinations of deviant positions possible (none; early only; middle only; late only; early and middle; early and late; middle and late; early, middle, and late). The sequence length could be 35, 37, or 39 triplets. A series of 400 ms noise bursts was presented to the contralateral ear. There were two experimental conditions. Subjects were either asked to attend to the triplets and ignore the noise bursts in one condition, or to initially attend the noise bursts and then switch their attention to the triplets midway. There were two task conditions. The subjects either had to report their percept of streaming or perform a button press upon detection of the deviants. See Figure 5 for an illustration of the experimental conditions.

In the objective, deviant detection condition, when the subjects had to attend to the sequence, deviants were best detected at the smaller frequency separation and when they occurred at the start of the sequence. Late deviants were better detected in the switched attention condition than the attend condition, in accordance with performance being poorest when stream segregation had occurred. These results confirmed that the removal of attention from the tones during the first part of the sequence improved performance compared to the late attend condition in which the tones were attended throughout.

In the subjective, streaming perception condition, a build-up pattern can be seen during the attend condition, with the probability of segregation increasing over time, and at a faster rate for the greater frequency separation. The proportion of segregated percept for the second, attended section of the switch condition is lower than that the second section of the attend condition. The proportion of segregated percepts was greater following a switch than at the start of a fully attended sequence. The time until the first "two streams" report was also shorter (2.7 s) following a switch than at the start of a sequence. The authors found that switching attention tended to reset perception to a single stream, but that streaming built up again more rapidly after a switch than at the start of a sequence. This would indicate that switching the focus of attention does not remove all trace of earlier perceptual experience. Finally, they confirmed the finding that the detection of occasional temporal shifts in the B tone of an ABA sequence was easier when the pattern is heard as integrated in a single stream, regardless of the frequency difference between the tones and the time in the sequence at which the deviant occurs.

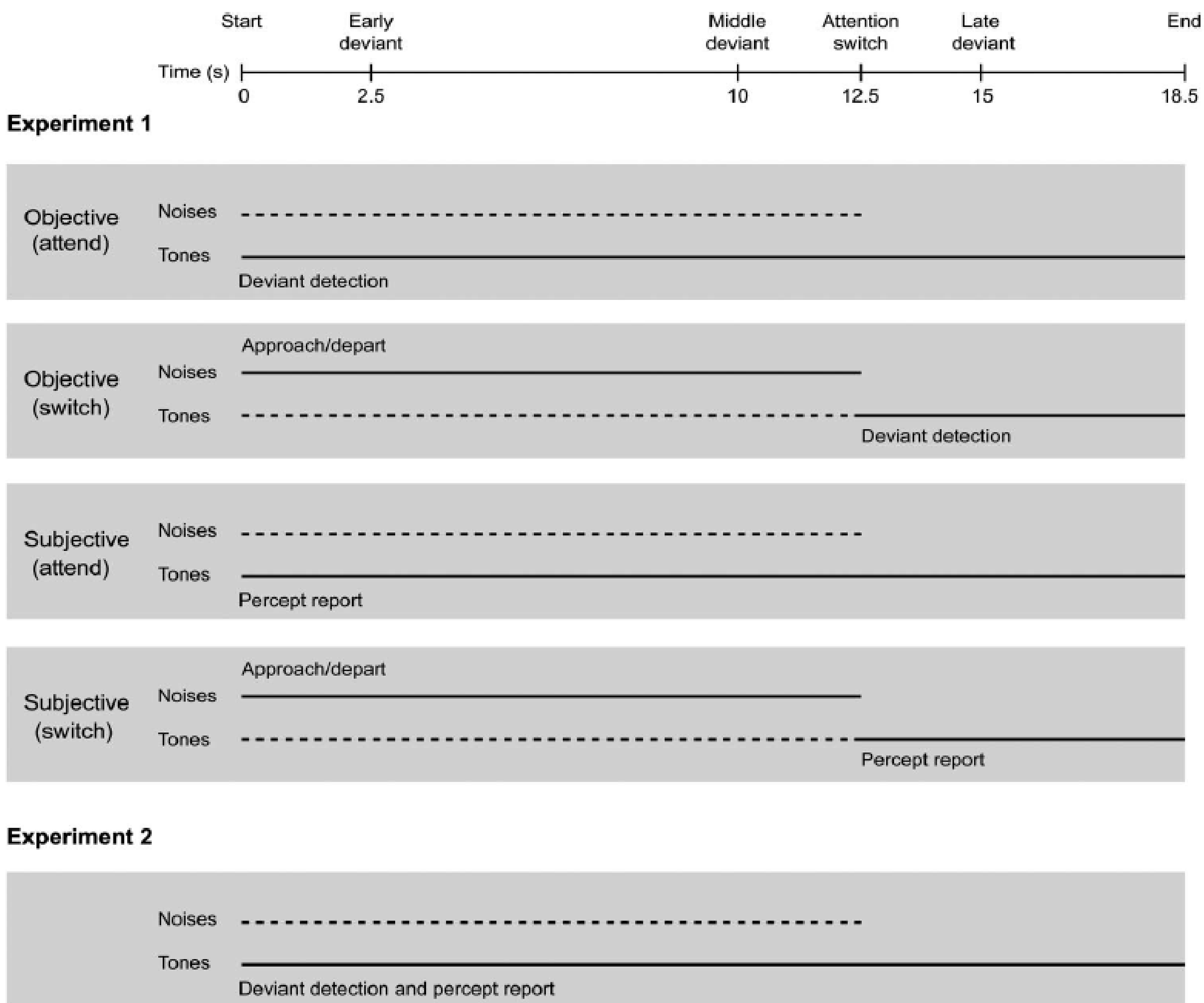


Figure 5. Conditions of the Billig and Carlyon (2016) paper: The attended sounds are shown with solid lines and the ignored sounds with dashed lines. The task for the attended sounds is indicated above/below the corresponding line. The timeline at the top shows the occurrence of potential events, averaged across sequences.

Event-Related Potentials Measures

The previous studies of streaming have relied on either so-called subjective measures (subjects reporting what they are experiencing) or more objective measures of performance (detection of deviant stimulus with the triplet sequence). Many cognitive neuroscience laboratories record brain activity during these types of tasks. Electroencephalography (or EEG) records the electrical activity of the brain over various regions of the scalp. An external stimulus or an internal cognitive event (such as decision-making or a memory comparison) causes a minute change in the ongoing EEG. These minute changes in the electrical activity of the brain are called event-related potentials (or ERPs). ERPs consist of a series of negative- and positive-going "components" where the components are thought to reflect different aspects of cognitive processing, A component is defined based on its polarity, latency (when it occurs), and general scalp distribution (where it is maximum and minimum).

A major problem with the use of ERPs is that their amplitude is much smaller than that of the ongoing EEG in which they are embedded. In the waking adult, the amplitude of the ERPs may be smaller than 1 μV and even larger components are much smaller than that of the ongoing EEG.

Fortunately, the ERP component of interest can be made visible through signal averaging. Signal averaging is a procedure in which a "signal" (the ERP component of interest) occurring in repeated stimulus presentations (or trials) emerges from the random, ongoing background EEG, or "noise" within these trials. With repeated stimulus presentation, the average of random activity will tend to zero. It is assumed that on each trial, the amplitude of the signal will not vary (i.e., is constant). The average of a constant is, of course, the constant. Signal averaging will thus reduce the amplitude of the background EEG, allowing the signal to emerge. The average amplitude of the background noise will however never be zero. In general, in order to “see” the signal with respect to the noise, most researchers indicate that a minimal signal:noise ratio (S:N) of 2:1 or even 10:1 is necessary.

The average noise decreases as a function of the square root of the number of trials. The EEG of an alert human adult varies from 20-100 µV. ERPs vary in amplitude from about 1 to 10 µV. Let us assume that the background EEG on most trials measures 50 µV and the ERP component of interest measures 4 µV. To achieve a S:N of 2:1 requires the background EEG to be reduced to 2 µV. This would require 625 stimulus presentations measures (the square root of 625 is 25). Fortunately, other methods, such as the judicious use of filtering, can be used to reduce the amplitude of the background EEG. It is thus reasonable to create a methodology with 30-60 trials per condition when looking at a large component such as the P3 wave, 150-200 trials per condition when looking at a medium-sized component like the N2 wave, and 400-800 trials per condition when looking at a small component like the MMN wave (Luck, 2004).

ERP measures of deviant detection

Novelty detection is said to be a pre-attentive process, where neurons encode stimulus change from the pattern of preceding stimuli. The deviancy detection system continuously monitors the auditory environment, tracks changes, dynamically updates its representation of the acoustic scene. Change detection can be quantified by looking at certain ERP components that are elicited by deviant stimuli signalling a change from the past. A so-called oddball paradigm is often used in change detection studies. A subject is presented with a stream of frequently-occurring and repeating homogenous “standard” sounds. At rare (or “odd”) times, a feature of the standard is changed to form a “deviant”. All auditory stimuli, whether attended or ignored, will elicit obligatory ERPs, most notably an “N1” occurring at approximately 100 ms followed by a “P2” occurring at approximately 180-200 ms. These are mainly affected by the physical features of the stimulus, particularly its intensity. Because the intensities of the standard and deviant stimuli are identical, the N1-P2 should not vary, apart from timing. The amplitude of the N1 and P2 are also much affected by the rate of stimulus occurrence. In the streaming studies, stimuli are presented very rapidly and as a result, N1-P2 will be very small. In addition, the deviant will also elicit a series of ERPs components depending on whether the subjects are asked to actively detect the deviant (N2b and P3b) or do so passively (MMN and P3a).

The MMN is a negative-going auditory component occurring between 100 ms and 200 ms. Its maximum is over centro-frontal areas of the scalp and it inverts in amplitude at the mastoids. Its amplitude increases as the extent of deviance increases. Importantly, the MMN appears to reflect a pre-attentive process. It can be elicited in the absence of attention and can even be elicited during an unconscious state, natural sleep (Sculthorpe, Ouellet & Campbell, 2009). Although the MMN is present when stimuli are presented outside the focus of attention (Näätänen, 1990; Sussman, Winkler, Huotilainen, Ritter & Näätänen, 2002), MMN generation may be sensitive to top-down control processes (Sussman, 2007; Sussman, Winkler, Huotilainen, Ritter & Näätänen, 2002; Sussman, Winkler & Wang 2003; Woldorff, Hillyard, Gallen, Hampson & Bloom, 1998).

A different negative component is elicited if the subject is asked to actively attend to the oddball sequence and detect the deviant stimulus (called a “target” in these conditions). This is the N2b. The N2b overlaps and summates with the MMN and distinguishing between the two has been a topic of much debate (Müller-Gass, Stelmack & Campbell, 2006; Näätänen, 1991; Woldorff & Hillyard, 1990). Some authors indicate that the MMN and the N2b can be distinguished based on their scalp distribution. The MMN inverts in amplitude at the mastoids, whereas the N2b does not (Alho, Paavilainen, Reinikainen, Sams & Näätänen, 1986). However, these differences are very subtle.

The P3b is a much-studied ERP component (Sutton, 1967). In easy detection tasks, it occurs about 300 ms following the detection of the rare target. This large amplitude positivity is at its maximum over parieto-central regions of the scalp. Several factors affect the P3b (Johnson, 1986). Attention is especially critical. If the subject does not attend to the oddball sequence, the rare target does not elicit a P3b. If the subject does attend the sequence and fails to detect the target, again a P3b is not elicited. Difficulty of target detection also has important effects. Johnson notes that the P3b is affected by equivocation. If the subject is not certain that a rare target has been presented, its amplitude will be smaller. For this reason, the amplitude of P3b is inversely related to task difficulty; it decreases in amplitude as the task becomes more difficult and subjects become more equivocal about their decision (Johnson, 1986). On the other hand, N2b may be larger when the target is difficult to detect.

While a P3b will not be elicited if the subject is not actively attending to the oddball sequence, an earlier positivity, the P3a, may be elicited automatically in the absence of attention. The P3a occurs earlier than the P3b and its scalp distribution is more centro-frontal than the parietal maximum P3b. Its amplitude varies directly with the extent of change and is especially large for highly novel stimuli. The P3a component thus is thought to reflect an involuntary orienting to an unexpected event (Schröger, 1996; Sussman, Chen, Sussman-Fort & Dinces, 2013).

Effect of attention and time on task on streaming: ERP studies

Event-related potential studies (Sussman, 2007; Sussman, Winkler, Huotilainen, Ritter & Näätänen, 2002) have shown a dissociation between a pre-attentive, low-level sound segregation process, and an attention-dependent process that can be called into play in forming perceptual objects and streams.

While attention is not always required for streams to form (Sussman, Horvath, Winkler & Orr, 2007), attention can heavily influence a listener's perception. Attention may be required to select representations in an ambiguous auditory scene. One support for such a theory comes from the fact that two distinct ERP components are evoked with different latencies (Snyder, Alain & Picton, 2006; Winkler, Takegata & Sussman, 2005). The first component, that Winkler et al. (2005) called the earlier frontocentral negative difference (EFND), is thought to occur while there is an initial representation of two alternative interpretations of the sound (one stream versus two streams). In their study it was indexed as an MMN. The later component, that Winkler et al. (2005) called the later centro-frontal negative difference (EFND), reflects the listener's decision (Winkler et al., 2005). It was an N2b, only occurring when the subjects perceived a galloping pattern and noticed the omissions in the pattern. Once an auditory scene has been parsed into its component objects, selective attention can operate on these components to facilitate further processing and resolve competition between multiple sources.

Snyder, Alain and Picton (2006) conducted an ERP study investigating the effect of attention on streaming. Subjects had to listen to 10.8 s sequence of alternating ABA tones report at the end of the sequence whether they heard one or two streams. In a second condition, the subjects had to ignore the tones. The authors found that the amplitude of P1, N1, P2 and N1c components of the auditory-evoked potential increased during a 10.8 second sequence of ABA triplets and correlated with behavioural reports of stream segregation. The increase was greater when subjects attended to the tone sequence than when they ignored it. A positive component peaking about 200ms after the beginning of each ABA triplet increased over several seconds after sequence onset, paralleling the build-up of stream segregation observed psychophysically.

Auditory Continuity Illusion

Streaming is caused by the interplay of auditory stimuli characteristics and top-down process. Another perceptual effect created by the same Gestalt principles that are used to create auditory objects is the continuity illusion (Bregman, Colantonio & Ahad, 1999). When a silent period (a gap) is introduced into a target sound, creating an interrupted tone, the observer perceives two separate shorter duration sounds. However, if noise is presented during the gap, the target is perceived as being continuous, creating an auditory continuity illusion (Miller & Licklider, 1950). This is caused by the fact that human perception will tend to extrapolate perceptual objects if the boundaries of the objects are obscured and perceive them as continuous. As with auditory streams, other factors can modulate the creation of an auditory object through this illusion.

Riecke, Van Opstal, Goebel and Formisano (2008) studied the relation between perceived continuity, and the duration, loudness difference and noise frequency spreads of the interrupting noise for interrupted (discontinuous) and non-interrupted (continuous) tones at different frequencies. Subjects had to report their percept of continuity going from the tone being most likely being continuous to most likely being discontinuous. The tone duration was 2800 ms and the tones could have a frequency of 500, 930, 1732, 3223 or 6000 Hz. Six noise durations were tested: 200, 400, 600, 900, 1,400, and 2,000 milliseconds. The authors found that the threshold for going from “probably continuous” to “probably discontinuous” was at a 1400 ms length, with shorter noises resulting in a higher likeliness of a continuity percept. Six loudness difference were tested, with the noise being presented at -8, -6, -3.5, 0, +6, and +12 dB differences from the tone. The authors found that the threshold for going from ‘’probably continuous to ‘’probably discontinuous’’ was between 0 and 6 dB, with less intense noises resulting in a higher likeliness of a continuity percept. Six noise frequency spreads were tested: 0, 0.25, 0.5, 0.75, 1, and 1.25 octaves. The authors found that the threshold for going from ‘’probably continuous to ‘’probably discontinuous’’ was between 0.5 and 0.75 octaves, with noises closest in frequency to the tone resulting in a higher likeliness of a continuity percept.

Micheyl, Carlyon, Shtyrov, Hauk, Dodson and Pullvermuller (2003) studied ERPs during a continuity illusion study to discover if attention influences the continuity illusion. The authors posited that if the continuity illusion is reflected in the MMN, the illusion is detected relatively early in processing and is preattentive. They employed an oddball paradigm in which the frequently occurring standards and rarely occurring deviants were to-be-ignored while subject watched a silent video film. The two standards were either a continuous tone or a tone interrupted by a silent 40 ms gap. They could have a duration of 120 ms or 240 ms. Both possible standards were followed by a noise burst occurring at a pseudorandomly selected delay after the onset of the standard (but never temporally overlapping with it). The presentation of the noise bursts was done to ensure that the presence of these noise bursts during the deviants would not elicit an MMN simply because they were rare. The two deviants were identical to the standards, but a noise burst was presented either overlapping the middle of the tone or during the silent middle gap. The noise bursts could either be in a frequency region that encompassed the tone, creating a continuity illusion, or be in a frequency region that did not encompass the tone and thus did not create a frequency illusion. There were therefore four conditions: continuous standards and the gap in the deviants filled with noise creating the perception of a continuous tone, continuous standards and the gap in the deviants filled with noise creating the perception of a discontinuous tone, discontinuous standards and the gap in the deviants filled with noise creating the perception of a continuous tone, and discontinuous standards and the gap in the deviants filled with noise creating the perception of a discontinuous tone. The authors posited that the MMN should be larger in when the deviant was perceived as continuous (when noise that created a continuity illusion was presented in the gap) and the standard as interrupted than when both standard and deviant stimuli were perceived as interrupted. They also postulated that the MMN should be larger when the standards were perceived as continuous and the deviants as interrupted (when noise that did not create a continuity illusion was presented in the gap) than when both standard and deviant stimuli were perceived as continuous. Their results support their hypotheses. The authors concluded that the cognitive processes responsible for the continuity illusion occur within the first 200 ms after the onset of the gap filled by the frequency appropriate noise and thus that the continuity illusion occurs in the absence of directed attention.

Kobayashi and Kashino (2012) investigated the effect of the presentation of a sequence of brief flanking sounds (30 ms 50 dB tone pips) on the percept of continuity of a target tone presented in alternation with bandpass noise in a target sound sequence. They varied the frequency separation between the target tones and the flanking tones, as well as the temporal relationships between the target tones, bandpass noise and flanking sounds. The target sounds frequency was 1000 Hz while the flanking sounds frequency was 500, 2073, 4040, or 7737 Hz. There were five stimulus onset asynchronies (SOAs) possible between the flanking sound and inducer: –200, –100, 0, +100, and +200 ms, where negative values indicate that the inducer preceded the flanker. They found that when the flanking sound was synchronised with the noise disruptor (0 ms asynchrony), the apparent continuity was significantly enhanced, independently of the frequency difference between the target tone and flanking sound. They posited that the flanking sound, together with the sudden onsets of the noise disrupters, captures attention that would otherwise have been directed to the target sounds. Compensation for the attentional disruption results in enhanced apparent continuity for the target sounds and the gaps in the target sounds may go undetected. The authors do present the fact that these findings may be explained by a simpler theory: as presented earlier, the default assumption of the auditory system is continuity when one sound is interrupted by another.

Streaming and the continuity illusion

The interaction between streaming and the continuity illusion has been the subject of very little research. Tougas and Bregman (1990) conducted such a study. Their stimuli were overlapping pairs of tones with one element of the pair increasing gradually in frequency and the other decreasing in frequency. The tones could either be perceived as auditory streams that ‘’cross’’ (with the first tone descending for the whole duration and the second tone ascending for the whole duration) or ‘’bounce’’ (with the first tone descending then ascending and the second tone ascending then descending). The frequency range for the tones was between 400 and 1600 Hz. There were 4 conditions. The all-pure condition consisted of two pure-tone glides. The all-rich condition consisted of two glides, one ascending and the other descending, each containing a set of three harmonics. The rich-crossing condition consisted of an ascending rich glide and a descending pure-tone glide. The rich-bouncing condition consisted of pure-tone glides forming the upper half of the X and rich glides forming the lower half. A 200 ms interval, centered at the point at which the glides crossed, was either left intact (continuous interval), replaced by a noise burst (noise interval) or replaced by silence (silence interval). The noise was bandpass filtered between 200-2400 Hz. The subjects had to evaluate on a 7-point scale the clarity of isolation of the standard sound when heard as part of the four conditions. The authors found that the continuous and noise intervals (where a continuity illusion was present) yielded essentially the same stream formation and responded in the same way to the added harmonics. In the presence of a continuity illusion, streaming occurred as if the tones had been continuous. The authors thus posit that auditory streams were formed through heuristics for grouping and then the presence of acoustic information which suggested that part of the two streams had been masked activated a perceptual continuity mechanism which acted to fill in the gaps.

Bregman, Colantonio and Ahad (1999) did a study to evaluate if the continuity illusion and streaming are based on the same Gestalt principles. High-frequency narrow-band noise (NBN) bursts were presented in alternation with a low-frequency NBN bursts. Noise bursts of wide-band noise (WBN) or silent gaps were presented between them. The high and low NBNs could be segregated through streaming and the NBN could be heard as a continuous tone caused by the presence of noise between two NBNs, creating a continuity illusion. The authors manipulated three variables. First, the rate of presentation of the sequence: 110, 150, or 200-ms onset-to-onset, with matching NBN durations: 73 ms, 99 ms and 133 ms, and WBN half the duration of the NBN. This was done to keep sequences of the same lengths. The second variable manipulated was the separation of the center frequencies of the high and low NBNs, which were varied between 3 and 23 semitones. The last manipulation was done on the bandwidths of the high and low NBN, which were varied from 100 to 1000 Hz. There were two tasks. In the first, subjects had to rate their perception of the continuity illusion and in the second they had to give their perception of the number of streams heard. In this streaming perception task, there could either be silent gaps between NBN or noise presented between them. The authors found NBN stimuli respond to the same variables as tonal stimuli, with greater frequency separation between NBN noises and higher rate of presentation leading to greater streaming. The manipulations of temporal and frequency parameters had similar effects on all three tasks: the continuity illusion, stream segregation with WBN separating the NBN bursts and stream segregation with silence separating the NBN bursts. The authors posit that both stream segregation and the continuity illusion possibly share the process of auditory object creation that binds segments of the auditory scene, but that one or two experiments (this one and Tougas and Bregman, 1990) are not sufficient to show that stream segregation and illusory continuity share an underlying mechanism.

Computational Models of Auditory Attention

Models can bring together information coming from different sources and scales, and demonstrate if a proposed mechanism is sufficient to produce an observed phenomenon. Computational models have been increasingly used in the study of cognitive processes, as computer power has allowed for the complex representation of those processes. These models are built to represent a cognitive theory and their performance is evaluated by how well they replicate human performance in cognitive tasks. Given the complexity of the auditory object creation processes, using a model to present a simplified and synthetized representation of the stages involved is often useful. Computational models of the auditory process are multilayer models, where each layer aims to represent a different step of the auditory source localization, stream segregation and auditory object identification process, as introduced previously. Most models are based around the concept of an auditory saliency map that gives a weighted representation of the acoustic environment where elements that are more likely to be detected have the largest weight. This is called a winner-take-all process. One of the contributions of this thesis, presented in chapter 2, is the creation and evaluation of a complete auditory model, and therefore, the next sections will introduce some of the most important features of those models: filters, oscillator banks, saliency maps, and the modelling of attention.

Filters

The first element of computational models of auditory attention is a spectrogram representation of the stimulus, where the intensity of the response of a certain number of filters or other extracted features (spectral and temporal contrast, pitch, spectro-temporal orientation, etc.) is represented as a function of time. This step mimics the information processing stages in the central auditory system. The most studied element of the central auditory system is the frequency separation carried out by the cochlea. Prevalent models were developed by Slaney (1993) and Patterson and Holdsworth (1996). These models represent the cochlea as a filter bank, where each filter corresponds to a point on the basilar membrane that responds preferentially to a certain frequency. The filters are often gammatone filters (Johannesma, 1980), bandpass filters that model the displacement of hair cells in response to the input. They are called gammatone because the response function is the product of a gamma function and the central frequency of the incoming tone.

There are two key output representations that can be used: firing rate representation and spiking neuron representation. The firing rate is the frequency with which a neuron fires. It is therefore an average of the neuron activation spikes for a time period. Spiking neuron representations represent each neuron activation spike. The spiking neuron model is therefore more biologically accurate but is also more demanding in terms of calculation power, and most models use the firing rate representation of the output instead of spiking neurons one. When using a spiking neuron model, the output is often squared or half-wave rectified, or a model of hair cell transduction is added (Hewitt & Meddis, 1994). The hair cell transduction model is based on a representation of the hair cell in the inner ear containing reservoirs of neurotransmitters. In it, the amount of such neurotransmitter emitted by the cell is proportional to the hair cell displacement and to the probability of the postsynaptic cell in the auditory nerve firing.

Oscillator models

The next step in an auditory object creation model is to model a lateral cortical inhibition process to create feature maps. Using this idea, Von der Malsburg and Schneider (1986) proposed a neural network model of auditory segregation based on synchronised oscillators with temporal correlation as the binding factor. Each oscillator represents a neuron and responds to a certain feature. Von der Malsburg and Schneider state that if the cells that respond to certain features fire simultaneously, they will be incorporated into a global description of the object created by temporally correlated oscillators. The oscillatory correlation theory is supported by experimental data obtained through a monitoring of neuronal firing rates in the auditory cortex (40 Hz oscillations, Galambos, Makeig & Talmachoff, 1981; Wang, 2005).

The two main rules in oscillatory correlation are that cells that fire in a synchronised way are said to have onset synchrony and form an object, and that cells with a different firing phase are considered to be responding to another object. This is supported by the Gestalt theory of auditory object segmentation.

Most models use a full lateral connectivity (all units to all units) to create phase synchrony (see LEGION, presented in the next section). Full connectedness in the networks leads to synchronisation of all oscillators firing simultaneously, indiscriminately from proximity. As such, full lateral connectivity models cannot account for frequency proximity but can be used to synchronise different modalities or objects comprised of multiple non-contiguous frequencies. In these models, a global inhibitor or a time-delay between oscillators (Schillen & Konig, 1994) is used to create the asynchrony. Some models use a more local connectivity instead to create a better discriminated synchrony. In the local connectivity option, the oscillators need to be in close-range for the time-delay to work.

LEGION (Locally Excitatory, Globally Inhibitory Oscillatory Network).

In the LEGION model (Wang & Terman, 1995; Wang, 1996; 1999; 2005; 2008), each oscillator represents a spiking neuron and is a feedback loop between an excitatory unit x, representing the neuron's synaptic activation, and an inhibitory unit y, representing the refractory period of the neuron. Figure 6 presents the trajectories of an active and inactive oscillator neuron. The entire ensemble models the spiking process of neurons. For a positive input, the paths for x and y intersect along the middle branch of the cubic function and the oscillator is considered active. The oscillator then goes from active to a silent state that mimics the refractory period of neurons. For a negative input, the oscillator is inactive and the solution is a fixed point along the left branch of the cubic function.

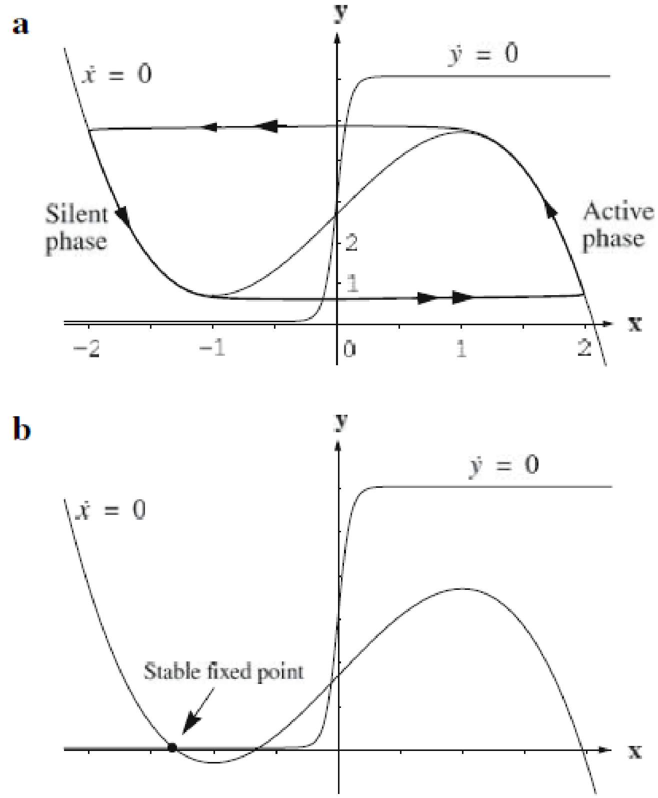


Figure 6. Active (top) and inactive (bottom) oscillators (from Wang and Chang 2008)

The input to each oscillator is a function of the external input, which is the sum of the connections between the oscillators minus the effect of a global inhibitor. Lateral inhibition between receptors is modelled through a central inhibitory unit, which receives input from all oscillators in the oscillator bank and is linked to all oscillators through a weight matrix. All oscillators are connected to all other oscillators to increase connections between oscillators firing at the same time. Local connections represent cooperation between auditory elements while the global inhibitor represents competition. Segmentation of the auditory stimulus into auditory objects is caused by the local connections of oscillators being covered by different objects, while the global oscillator helps desynchronise objects. Through this model segmented objects will have different phase synchrony.

LEGION is a bidimensional network, with frequency and time delay as the dimensions. Once an oscillator enters the active space, the global inhibitor is triggered and starts inhibiting the whole network. Oscillators connected to the active oscillator see their connections reinforced while all other connections are inhibited, and thus the oscillators within the auditory object are synchronised.

The advantage of oscillatory networks is that the presence of oscillators makes it impossible for one object to gain predominance over the others (thereby suppressing other objects) for an extensive length of time, thus modelling the fact that sustained attention has a limit (Sarter, Givens & Bruno, 2001). The global inhibitor corresponds to attentional control in humans. A downside of this model is that in the presence of noise, this model creates many small segments (Wang & Chang, 2008). In this scenario, a more long-range connectivity can aid segregation, namely by using a Gaussian neighbourhood function.

The tones that are grouped together do not need to belong to the same frequency channel, thereby allowing the construction of complex auditory objects. Auditory segregation or grouping depends on the strength of lateral connections, which itself depends on distance. Tones can also show frequency modulation, and the network will connect streams with continuous frequency transitions. These features make the LEGION model very suitable for creating auditory objects bound by temporal proximity.

Saliency maps

The next element aims to combine the auditory objects formed by the oscillators into saliency maps. Itti and Koch (1998, 2000, 2001a, 2001b) presented a framework for how features can be combined in saliency-based attention systems for vision, which can also be used for auditory features. Large changes in acoustic energy may result in a passive and automatic switch of attention from current demands to the "intruding" auditory channel (sudden and/or loud noises), as these noises have been associated with danger throughout human evolution. Attention will be captured by irrelevant sounds in an exogenous manner if they are sufficiently salient. In audition, exogenous (bottom-up) attention keeps track of novelty, either in frequency, amplitude, spatial localization, or duration, as novelty makes elements salient (Bregman, 1990). Itti and Koch equate the feature extraction and subsequent signal moderating process to a fast temporal sequence of localised pattern recognition problems. In the model, each feature is computed by center-surround operations that are sensitive to abrupt changes in intensity of activation and where the center has a fine scale, while the surrounding has a coarse one. After extracting features at multiple scales, "center-surround" differences are calculated, resulting in feature maps. Within these feature maps, saliency of the stimuli is decided by a combination of the saliency of its attributes. Models usually use a simple summation of all feature maps or a linearly weighted combination of all feature maps.

In Itti and Koch's model (2001a), they use the concept of gist as their representation of saliency. Perceptual gist refers to the representation of a scene built during perception and the image properties (spatial frequency orientations and scales, colour, and texture density) that enable an efficient categorization of a real-world scene (Oliva, 2005, Oliva & Schyns, 1997; Oliva & Torralba, 2001). Gist processing is a pre-attentive process and guides attention to focus on a subset of stimuli locations in order to analyse the details of the target locations. In Itti and Koch's model (2001a), saliency/gist is modelled as a feature vector. The cumulative gist vector is obtained by combining the gist feature vectors of each specific dimension. Top-down biases are added to the cumulative gist vectors to enhance the activation of neurons tuned to features of the target stimuli and inhibit the neurons that are not tuned to it (Itti & Koch, 2001a). In Itti and Koch's model, the biases are weights learned in a supervised manner through a multilayer perceptron to enhance specific dimensions of the gist features related to the task, while attenuating the effect of unrelated dimensions.

Kayser, Petkov, Lippert and Logothetis (2005) created an auditory saliency map. The resulting saliency map gives the exact location that is most salient at any given time through a bank of leaky integrate and fire neurons. The activation at the more salient locations rises and fires faster and this activation feeds into a winner-takes-all network that selects only one location to attend. The focus of attention is shifted to the winning location, and the inhibition in the winner-takes-all network prevents any other location in the network from firing and resets their value. A local inhibition is triggered in the saliency map at the location of the focus of attention to implement the inhibition of return. It allows the attention to shift to the next most salient location to be attended and prevents attention from returning to the previously attended location.

The saliency maps are usually normalised to be comparable across modalities when multiple modalities are combined. They are amplified through a non-linear function so the most salient elements retain their saliency when combined with other maps, and also to simulate competition between neighbouring salient elements in each map. These normalised feature maps are combined into the auditory saliency map.

Endogenous attention

There is a second type of attention that is said to arise not from the stimulus but from conscious cognitive processes. James (1890) defined attention as a focalisation and a concentration of consciousness, allowing the mind to take possession of one out of several simultaneously possible events, thoughts, or objects. There are several different types of endogenous attention relevant to the model presented in this thesis: 1) focused attention, which is the ability to attend to specific stimuli; 2) sustained attention, which is the ability to keep attending to said stimuli; 3) selective attention, which the ability to sustain a response to specific stimuli in the presence of distracting stimuli and 4) alternating attention, which is the ability to shift the focus of attention between tasks.

Lavie (1995, 2000) posits that the allocation of attention is influenced by both structural and processing constraints, with an initial filtering mechanism selecting items, and the selected items going on to compete for resources involved in deeper processing. Cognitive resource weights are temporary intents such as top-down directed attention regarding a task, where a conscious shifting of attention to a percept is performed.

Top-down attentional processes, using contextual information and prior knowledge, focus the attentional resources to better perform the tasks needed. The selection of information to pass into the working memory is a highly competitive, hierarchically structured process (Kayser, et al., 2005). Because of limited cognitive resources, there is a competition between auditory objects. At a low level, competition occurs between basic sound features like frequency or temporal structure, while at a higher level, competition occurs between auditory streams. Finally, competition can occur between different sensory modalities if different perceptual objects are present in each. Both bottom-up and top-down cues can direct auditory attention to a particular object, and an ‘object-based’ neural representation modulated by behavioural demands (Shinn-Cunningham & Best, 2008; Shamma, 2008). One of the main models in attention is the spotlight of attention (Posner, Snyder & Davidson, 1980), a model that sequentially illuminates different parts of the visual or auditory scene for analysis. The searchlight model is applicable to the saliency map.

The winner of the competition between features and auditory objects is the element with the greatest strength, through a winner-take-all process that combines the influence of both the top-down and bottom-up processes (Itti & Koch, 2001a). Attention operates at the level of objects (Desimone & Duncan, 1995; Zatorre, Mondor & Evans, 1999), and even when attention is focused on a low-level stimulus feature (i.e. pitch), there is enhanced sensitivity to other features of that source (i.e. location).

It was shown by Posner and Cohen (1984) that subjects detected targets presented in the same location as a previous target slower than targets presented elsewhere. A way to model this process, called inhibition-of-return (IOR), prevents attention from staying permanently at the same (most salient) location. In modelling, this can be done by using the saliency map as the input of a layer of leaky integrate-and-fire neurons. The neurons connected to a salient location will fire faster and then they will return to base level, thus shifting attention. This spiking neurons map then becomes a winner-takes-all neural network that models the concerted effect of both bottom-up and top-down attentional processes.

Research contribution of thesis

As can be seen from the information presented in this chapter, computational auditory scene analysis is a vast field, but one that is dominated by models aimed at efficient processing and not biological plausibility. It is therefore of significant research interest to create a complete model that follows the different known steps of auditory object creation according to neuropsychological and psychological research. While there are many models of saliency, as summarised above, to the best of the author's knowledge, there is no single model that simultaneously incorporates the influence of both endogenous and exogenous attention on saliency while modelling the treatment of multiple auditory features in an integrated manner. One contribution of this thesis is the creation of a model that encompasses the current knowledge of auditory processing in the literature. One important improvement of the model presented in this thesis over previous models is the incorporation of multiple auditory attributes in the model. This model is introduced and discussed in depth in Chapter 2.

While the initial aim was to further refine and modify the model in response to experimental results, the nature of the results obtained meant that it was considered more logical to further modify the model to take account of results found in the literature. The complete model and results, both the initial biologically plausible model and the subsequent refinements, are presented in Chapter 2, and represent a key contribution of this research.

In addition to this, the review conducted in this chapter identified that attention is not always regarded as being strictly necessary for auditory streaming. Various studies by a wider range of researchers have found evidence of an effect of the time spent attending to a task on auditory object creation and streaming, but the extent of the modulatory influence of attention is still a matter of debate. Some studies found an early influence, while others found that attending to the auditory input influences stream creation much later in the process. Building on a proven methodology using ABA triplet tones, used by many other behavioural studies (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Thompson, Carlyon & Cusack, 2011) to study streaming, this experiment follows-up on the initial work of Thompson et al. (2011) by including ERP measures to gain insight on cognitive processing beyond what is generally available using behavioural measures. A problem with the strict use of behavioural measures is that it is methodologically extremely difficult to determine the extent of processing of stimuli that occur in a channel that it is not attended. The subject does not, of course, make a behavioural response to these to-be-ignored stimuli. In the absence of a behavioural response, ERPs can be used to provide a means of measuring the extent of processing of unattended stimuli. Moreover, even when a channel is attended and a behavioural response is made, the cognitive processes leading up to the response must be inferred. ERPs again provide a means of measuring processing prior to, at the time of and following the behavioural response. In this research, the conscious decision was made not to replicate the methodology used by Billig and Carlyon, as using both a subjective and objective condition while also recording brain activity would have required too many conditions. In addition, subjective recording of streaming perception has also been performed multiple times. Objective measures were the most suitable for comparison with the ERPs.

Following up on this initial contribution was a second study that aimed to ascertain how the presence of a continuity illusion would influence deviant detection in an experimental framework shown to induce streaming. As the continuity illusion creates a unified percept of a stimulus, the extent it will modify deviant detection within an ABA triplet and how that modification will differ according to the stream percept is a research question of great interest that has not been comprehensively investigated in the literature. This novel experiment was designed to gather new insight on the interconnections between streaming, the continuity illusion and the time spent attending to the task, and the experiments are presented in Chapter 3.

Chapter 2: Neural Network Model of Auditory Object Creation

Building on the literature review presented previously, this chapter focuses on the development of a model of auditory object creation as influenced by attention. Creating such a model is a good way to illustrate the current knowledge about the formation of auditory objects and the modulatory effect of attention in a simplified and easy to understand manner. The level of abstraction chosen is just above cellular level, in that each neuron in the later stages of the model represents assemblies of cells rather than single ones.

Models of Attention

Models of attention like Miller and Cohen (2001) are made of several interconnected modules that reflect the sequential processing of information from the initial perception of the stimuli to the enhancement of relevant information by attention. Starting with stimuli perceived, the first module consists of salience filters, responsible for enhancing responses to stimuli that are infrequent or are of instinctive or learned biological importance (Koch & Ullman, 1985). Competitive selection determines which information gains access to working memory (Desimone & Duncan, 1995), which then stores the selected information for detailed analysis (Baddeley, 2003). Top-down sensitivity control, bottom-up salience and the quality of the information encoded determine the signal strength. The information with the highest signal strength then re-enters the loop to compete with the information already in working memory.

Working memory includes competitive processes within itself, which allows for stimuli that are attended to reach working memory and stay active in it (Miller & Cohen, 2001). The working memory is also responsible for generating signals that improve the quality of the information being held in memory. One such signal modulates the sensitivity of neural circuits that represent the information, to increase the signal to noise ratio. This modulation by attention is present in many areas of the brain (neocortex, limbic cortex, basal ganglia, superior colliculus, cerebellum) depending on the modality of the stimulus. Activation in that region is taken to represent the relative salience of stimuli (Bisley & Goldberg, 2003). Neurons that have receptive fields that respond to preferentially to the attended stimulus and are currently activated see an increase in their sensitivity, while neurons attuned to other features see an actual decrease in their sensitivity (Chelazzi, Miller, Duncan & Desimone, 1993; Reynolds & Desimone, 2003; Treue & Mansell, 1999; Treue & Martinez-Trujillo, 1999). Top-down sensitivity control thus controls both excitatory and inhibitory connections. The modulations of neural responsiveness can also be precise concerning the timing of stimuli, as there is an increase in firing rate that corresponds with the period of decision (Khayat, Spekreijse & Roelfsema; 2006; Motter, 1994)

Sensory information is processed in a highly parallel and multilevel fashion, which creates a competitive process for the selection of the information that passes into working memory. Competition occurs upon the summative signal strengths of the encoded information, the modulations by top-down bias signals and the influences of bottom-up salience filters. Competition also occurs at each level of modulation and filter, from basic stimulus parameters to higher integrated features, and finally, between modalities. The comparison is between response strengths of multiple, simultaneous and distinct stimuli, a winner-takes-all process that selects the strongest response and nonlinearly inhibits all the other responses within the representation. The efficiency of such a network is increased by the addition of positive recurrent connections (Major & Tank, 2004; Shu, Hasenstaub & McCormick, 2003).

Top-down attention can selectively focus on a limited range of an acoustic feature dimensions, or can focus on the expected (or recalled) features of an auditory target. Endogenous auditory attention is the key to switching attentional focus to different features, objects, or streams of interest within the acoustic scene (Fritz, Elhilali & Shamma, 2007). Top-down signals can adjust attentional filters precisely and rapidly to dynamically reshape receptive fields in the primary auditory cortex in accord with salient target features and task demands (Carlin & Elhilali, 2015).

Proposed Auditory Model

The auditory model proposed in this thesis is supported by a comprehensive range of research. This includes the work of Wang (1995, 1996, 1999, 2008) for the oscillatory neural network, Itti and Koch (1998, 2000, 2001a, 2001b) for the saliency model and Wrigley and Brown (2004) for the architecture of single feature processing streams, inhibition of return of activation and attentional leaky integrate and fire neuron. Figure 7 illustrates the general structure of the proposed model. It is composed of 4 consecutive levels. The first level extracts the perceptual features (amplitude, frequency and spatial location) of the stimuli through filter banks, the second binds these features according to location and timing through oscillator banks, the third is a saliency map that models exogenous attention by showing the elements that are salient, based on the bottom-up characteristics of the stimuli. The saliency is modulated by a top-down attentional marker that represents top down attention: the ear being actively attended. The last level represents the moment in the stimuli sequence where stimuli are salient enough to pass an attentional threshold.

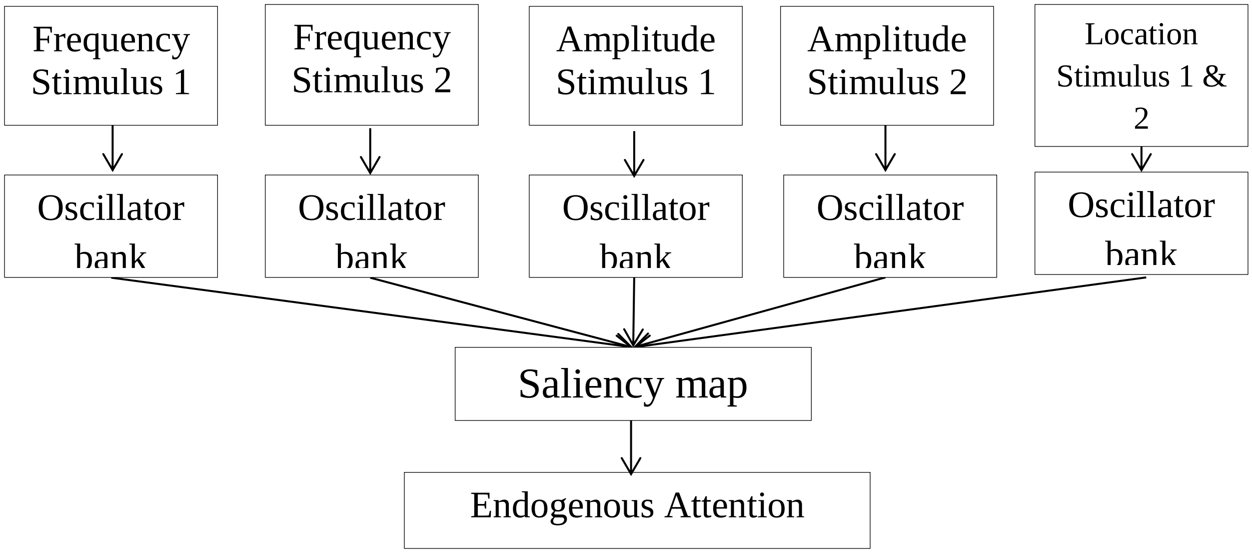


Figure 7. Architecture of the model

Filter banks for frequency, amplitude and location

Following Bregman's analysis-synthesis process (Bregman, 1990), the first component of the model extracts basic perceptual features of the stimuli (the analysis). The stimulus perceptual attributes that are processed are frequency, amplitude and spatial localization, as those are the key attributes found in the literature to have been studied regarding auditory object creation. Therefore, the model presented in this thesis has one feature extraction module for each attribute.

Frequency

The first module (frequency feature extraction) is a representation of the frequency separation performed by the cochlea. Shamma (2001) proposed a model that bridges the temporal and spectral hypothesis of frequency/pitch separation. The first stage is early auditory filtering coupled with temporal enhancement (i.e. by spiking only at the peaks of the input waveform) to produce more highly synchronised response waveforms. The frequency of a tone is encoded in the auditory nerve, both spatially, by its central frequency location, and temporally, by the periodicity of the responses in the fibers that innervate this central frequency. The average firing rate of a fiber reflects the amplitude of the sound signal at its central frequency. The second stage is a matrix of coincidence detectors that compute the average pair-wise instantaneous correlation (or product) between responses from all central frequencies across the input array. As part of the model proposed in this chapter, the frequency segmentation component implements Shamma's proposed approach.

In this model, Slaney's (1993) widely used implementation of the Patterson gammatone filters is used, as it is the most efficient implementation of frequency segregation as it reduces the computational effort of implementation by approximately 50%. Equation 1 shows the response function as the product of a gamma function and the central frequency of the incoming tone (f_c).

$$g_{f_c}(t) = t^{N-1}e^{-2\pi t b(f_c)}\cos(2\pi f_c t + \varphi)u(t) \quad [1]$$

where N is the filter order, φ is the phase of the signal, $u(t)$ is the unit step function ($u(t)$ = 1 for time > 0, 0 else) and $b(f_c)$ is a function that determines the bandwidth for the center frequency.

Patterson and Holdsworth (1996) found that when N = 4, the function fits the purported human auditory filter shape. It is symmetric around f_c and the bandwidth is evenly separated on both sides. Following human experimental data, Glasberg and Moore (1990) found that the equivalent rectangular bandwidth (ERB) of a frequency f_c = 24.7 + 0.108 f_c. The filter bank is spaced according to the bandwidth rather than the center frequencies, as those follow a logarithmic scale: $b(f_c)$ = 21.4 log (0.00437 f_c + 1), and give low frequencies filters narrow bandwidths and high frequencies filters broad bandwidths.

Gains are added to the filters to better reflect the sensitivity of the cochlea, which has optimal response in the 2-4 kHz frequency range. Because low frequency filters have a narrow bandwidth, the time point where they register a peak response will be later than for higher frequency filters. It is useful to phase compensate the filter banks to align the time point of the peak of the envelope by adding a time lead $t_c = \frac{N-1}{2\pi b(fc)}$ to the output of the filters (Patterson, Robinson, Holdsworth, McKeown, Zhang & Allerhand, 1992). Also, a change to the phase $\varphi = -2\pi t_c f_c$ is made to align the peak within the envelope. Equation 2 gives the final expression of the gammatone filter with the time lead and phase modifications giving it greater biological plausibility:

$$g_{f_c}(t) = (t + t_c)^{N-1} e^{-2\pi t b(f_c)(t+t_c)} \cos(2\pi f_c t) u_c(t) \quad [2]$$

As our interest is at a biologically plausible but psychology oriented level, a spiking model was chosen for the auditory object creation model. As the cochlea activation is not the focus of the model, it was elected to half rectify our outputs instead of using the hair cell model.

A number of 32 filters was chosen as it allowed for good frequency separation, while keeping the frequency splitting between filters to a minimum. As different stimuli were presented in each ear, the filter bank was applied separately to each stimulus.

Amplitude

The second feature extraction module uses the envelope of the sound presented to the model. This envelope represents the frequency of the vibration of the basilar membrane that encodes the amplitude of those sounds. The amplitude was split into 32 bins to give an equal representation to amplitude and frequency in the model. The bins were 5 dB SPL wide, meaning that the whole encoder covered the amplitude range from 5 dB SPL to 115 dB SPL. This was a suitable range and was chosen because human hearing covers from 0 dB SPL to approximately 120-130 dB SPL (threshold of pain). The amplitude was calculated for a time window of 100 ms.

As different stimuli were presented in each ear, the filter bank was also applied separately to each stimulus.

Spatial location

Finally, the third feature extraction module is an auditory field composed of 32 contiguous locations, which represents the spatial localization of the sounds. While creating the stimuli, their location in space was determined. Stimuli are presented at their full amplitude at the determined location of their source and a Gaussian neighbourhood is applied to the stimuli so that sounds in contiguous locations to the sound source are present in those locations but with reduced amplitude. Figure 8 presents a simplified 12-channel representation (out of the 32 used in the model) of the stimuli strength at different locations. It showcases the amplitude of the stimulus present at each location and the mingling of them in the center of the diagram (at channel 7). In that figure, the triplet stimulus was presented at channel 5 and the noise burst stimulus at channel 10, starting from the left.

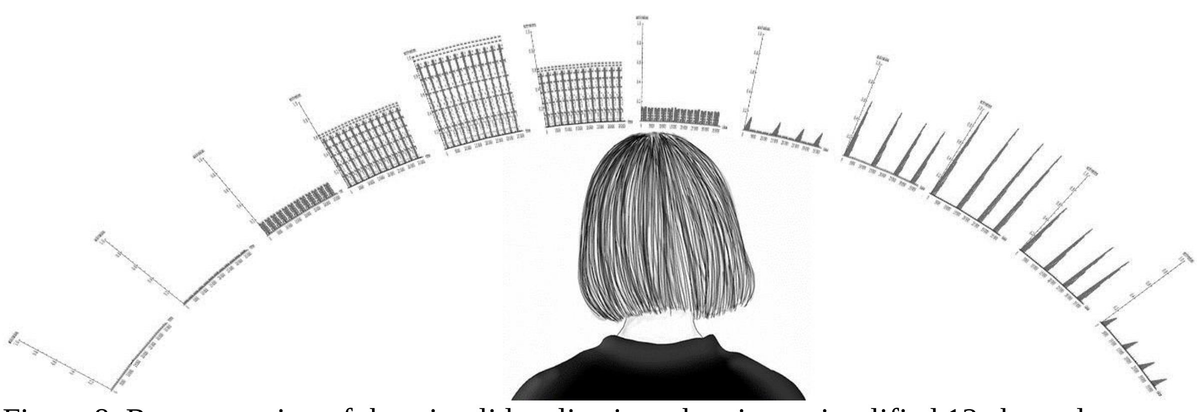


Figure 8. Representation of the stimuli localization, showing a simplified 12 channel representation, with a stimulus present at channel 5, and channel 10 (from left), and attenuated in neighbouring channels.

Oscillator model

The second element of the model groups sensory features together into coherent structures, and segregates them from dissimilar ones. In this model, auditory streams are organised according to proximity in frequency, location, common amplitude, and common onset and offset time. Autocorrelation within the vector is the base of segment formation within a frequency, location or amplitude bin. The oscillatory model used in this model is LEGION (Wang, 1995, 1998, 2005, 2008). Spiking neurons were used to maintain a degree of biological plausibility. The criterion used to create a segment is an autocorrelation over 0.8 for the duration of a 200 ms time window (Wrigley & Brown, 2004). After this is completed, correlations between adjoining frequency bands are carried out to determine if the stimulus encompasses more than one frequency band. If their correlation is above the 0.8 threshold, they are grouped.

An oscillator model is used to bind the neural correlates according to temporal correlation. Each oscillator represents a neuron and responds to a certain feature. The equations of an oscillator are presented in Equation 3 and 4. The nullcline of x is a cubic function, while the nullcline of y is a sigmoid, as seen in Figure 6.

$$\dot{x} = 3x - x^3 + 2 - y + I_0 + \rho \qquad [3]$$

$$\dot{y} = \varepsilon\left[\gamma\left(1 + tanh\frac{x}{\beta}\right) - y\right] \qquad [4]$$

In Equations 3 and 4, ε is the rate of change, γ is the decay parameter, β is the parameter describing the steepness of the sigmoid curve, ρ is a noise parameter and I_0 is the input to the oscillator.

The input to each oscillator (I_0) is a function of the external input (I_r), of the sum of the connections between the oscillators (W_{ik}) minus the effect of the global inhibitor (W_z).

$$I_0 = I_r - W_z S(z, \theta_z) + \sum_{k \neq i} W_{ik} S(x_k, \theta_k) \qquad [5]$$

The *S()* function is a sigmoid function that bounds the oscillator activity to be within a suitable range, as determined by the threshold θ.

$$S(m, \theta) = \frac{1}{1+e^{-K(m-\theta)}} \quad [6]$$

Where *K* is the steepness parameter and *m* the input to the function

The lateral inhibition between receptors is modelled through a central inhibitory unit, *z*, which receives input from all oscillators in the oscillator bank. The threshold θ_{in} for the inhibitor to start inhibiting is 0.1 (Wrigley & Brown, 2004)

$$\dot{z} = H(\sum_k S(x_k, \theta_x) - \theta_{in}) - z \quad [7]$$

Finally, W_z is the weight matrix linking the oscillators to the global inhibitor *z*, W_{ik} is the lateral connections matrix between oscillators and *H* is the Heaviside function, a function that keeps positive values and changes negative values to zeros.

$$W_{ik} = H(x_i)H(x_k)e^{-\left(\frac{(x_{i,t}-x_{j,t})^2}{\sigma_t^2}+\frac{(x_{i,f}-x_{k,f})^2}{\sigma_f^2}\right)} \quad [8]$$

Features whose information is processed in this model are frequency, amplitude, and spatial localization. The model has five oscillator banks each composed of 32 oscillators to follow the number of channels from the previous step (noise burst frequencies, triplet frequencies, noise burst amplitudes, triplet amplitudes and spatial localization). The oscillator bank used has a full lateral connectivity (all units to all units) to create phase synchrony, but modulates the strength of the connections according to distance. Full connectedness in the networks leads to synchronization of all oscillators firing simultaneously, which allows us to bind contiguous features together in a complex auditory object according to Gestalt laws. The auditory segments are formed according to the consistency of the feature displayed and the onset of each feature. Auditory segregation partly depends on the strength of lateral connections, itself depending on distance, as shown in the weight function in Equation [8].

Age trackers (Wrigley & Brown, 2004) that are leaky integrators are used to allow synchronization only between segments with the same onset.

$$Age_k = d_{age}(g_{age}[M_k - Age_k]^+ - [1 - H(M_k - Age_k)]c_{age}Age_k \quad [9]$$

where Age_k is the age of the channel, d_{age} is the parameter controlling the rise of the function, c_{age} is the parameter controlling the decay of the function, g_{age} is a gain factor and M_k is the index determining if this section of the stimulus is part of the segment or not. In the present model, the parameters used are those chosen in Wrigley and Brown (d_{age}=.001, c_{age}=5, g_{age}=3, 2004).

Saliency map

The next element of the model combines the segments into a saliency map (Itti & Koch, 2001a) through a simple summative process. The resulting saliency map gives the segment created by the oscillator map that is most salient at any given time. The model also includes an implementation of inhibition of return. The saliency or overall strength of the sum of the exogenous and the endogenous attention is passed as an input to an attentional leaky integrator (A_{LI}) that fires if the saliency is high enough.

$$\dot{ali} = A_{LI} - ali \quad [10]$$

$$A_{LI} = H(\sum_k H(x_k)[(\frac{\alpha_k}{\theta_\alpha}) - T_k]^+ - \theta_{ALI}) \quad [11]$$

where H is the Heaviside function, x_k is the oscillator activation, θ_{ALI} is the threshold over which an oscillator contributes to the attentional process (0.1 as per Wrigley and Brown's parameters), α_k is the envelope of the filtered stimuli, θ_α is a threshold determining how intense a stimulus needs to be to overcome top-down attentional processes and T_k is the term containing both endogenous and exogenous attentions thresholds.

$$T_k = (1 - A_{ear}A_k)Age_k \qquad [12]$$

The variable A_{ear} is the ear preference in Wrigley and Brown (2004) determining the top-down endogenous attention. In this model, A_{ear} is 1 when endogenous attention is brought to that ear and 0 when it is not. A_k is the exogenous attention and is proportional to the proximity to the most salient channel as determined by a Gaussian neighborhood. Note that for a sufficiently salient stimulus, the stimulus-driven attention would win over the top-down one, as the threshold over which an oscillator contributes to the attentional process is arbitrary.

$$A_k = Max_{A_k}e^{\frac{-(k-p)^2}{2\sigma_{ALI}^2}} \qquad [12]$$

At the end of the process, the attentional leaky integrator either fires or not, depending on if there has been a large enough accumulation of activation to generate combined exogenous and endogenous attention on a particular location. The final output is a time series of activations showing when attention has been captured by the stimuli throughout the sequence.

Testing

The model was tested using a variation of the stimuli by Carlyon et al. (2001) as these were previously used by Wrigley and Brown (2004) to test their auditory attention model, in addition to recent use by Thompson et al. (2011) and Billig and Carlyon (2016) for their behavioural studies. In one ear, which corresponds to one side of the location field (position 28 of the location array), noise bursts are presented. The noise bursts are either increasing in amplitude or decreasing in amplitude over 40 ms and the maximum amplitude is 100 dB. On the other side of the field, at location 5, Van Noorden (1975) triplet stimuli of alternating frequencies are presented. Those stimuli are of the ABA form, with A tones having a frequency of 400 Hz and amplitude of 55 dB and B Tones having a frequency of 635 Hz and amplitude of 55 dB. Figure 9 represents one set of stimuli presented.

There were four conditions for the simulations. First, one where only stimulus-driven attention was present, and where the activation of the attentional leaky integrate and fire neuron depends solely on the saliency of the stimuli. No top-down attention was added. Second was a condition where voluntary top-down attention was brought to the right ear in addition to the stimulus-driven exogenous attention. The stimuli in the left ear were ignored. Third is the reverse of the second condition, where voluntary top-down attention was brought to the left ear and the stimuli in the right ear were ignored. In the last condition, the model had to switch voluntary attention midway through the sequence from the right ear (noise bursts) to the left (triplets), mirroring the experimental condition of Carlyon's study.

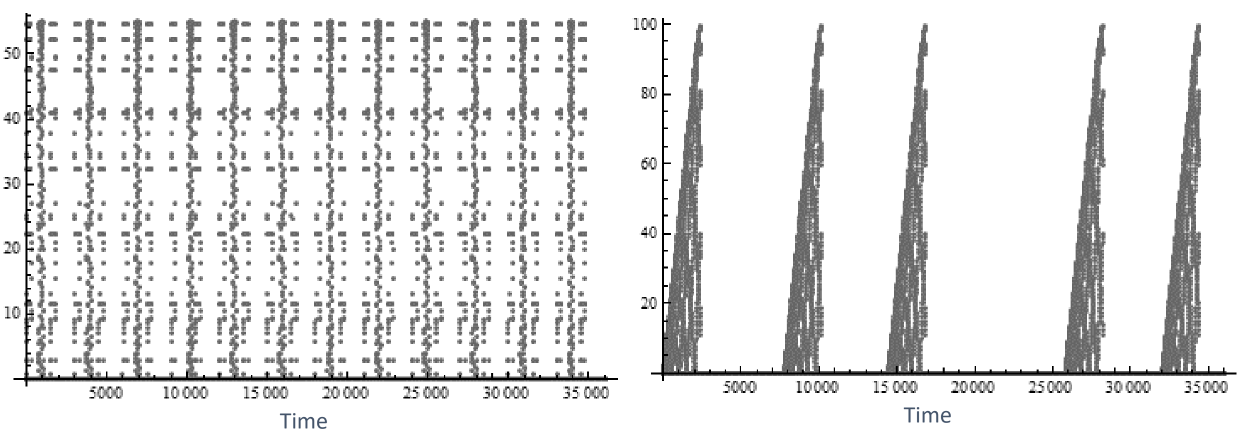


Figure 9. Stimuli. Left. ABA triplet sequence. Right. Noise bursts sequence

Results

Figure 10 presents the segments for all five filter banks. Note that in the graphs, the lower frequencies are at the bottom of the y axis and the highest at the top (and reversed for amplitude).

For the localization module, the rightmost channel is at the bottom and the leftmost channel is at the top. It can be seen on the figure that segments were formed for the two frequencies present in the triplet stimulus (line 16 and 20 in Figure 10a). The segments for frequency for the noise bursts are less well defined, as the burst was filtered white noise, and thus covered a wider band of frequency filters (from 4 to 7 in Figure 10b). The amplitude is the same for all three tones of the ABA triplets as all of them were presented at 55 dB (in Figure 10c). On the contrary, the segments for the noise bursts, that were here increasing in amplitude, sequentially cover filters of higher amplitude as time goes (in Figure 10d). Finally, as stimuli were presented at specific locations, segments were created in the filters corresponding to those locations (in Figure 10e).

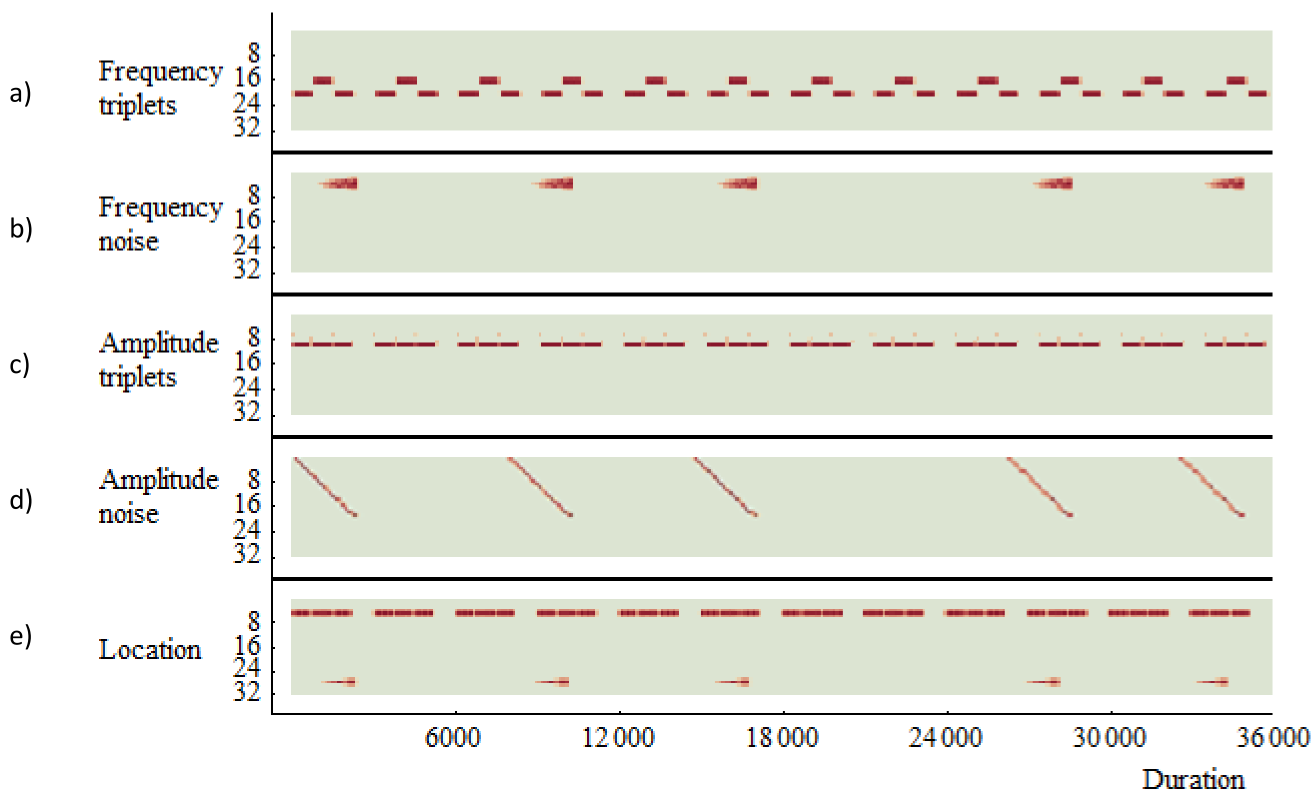


Figure 10. Segments

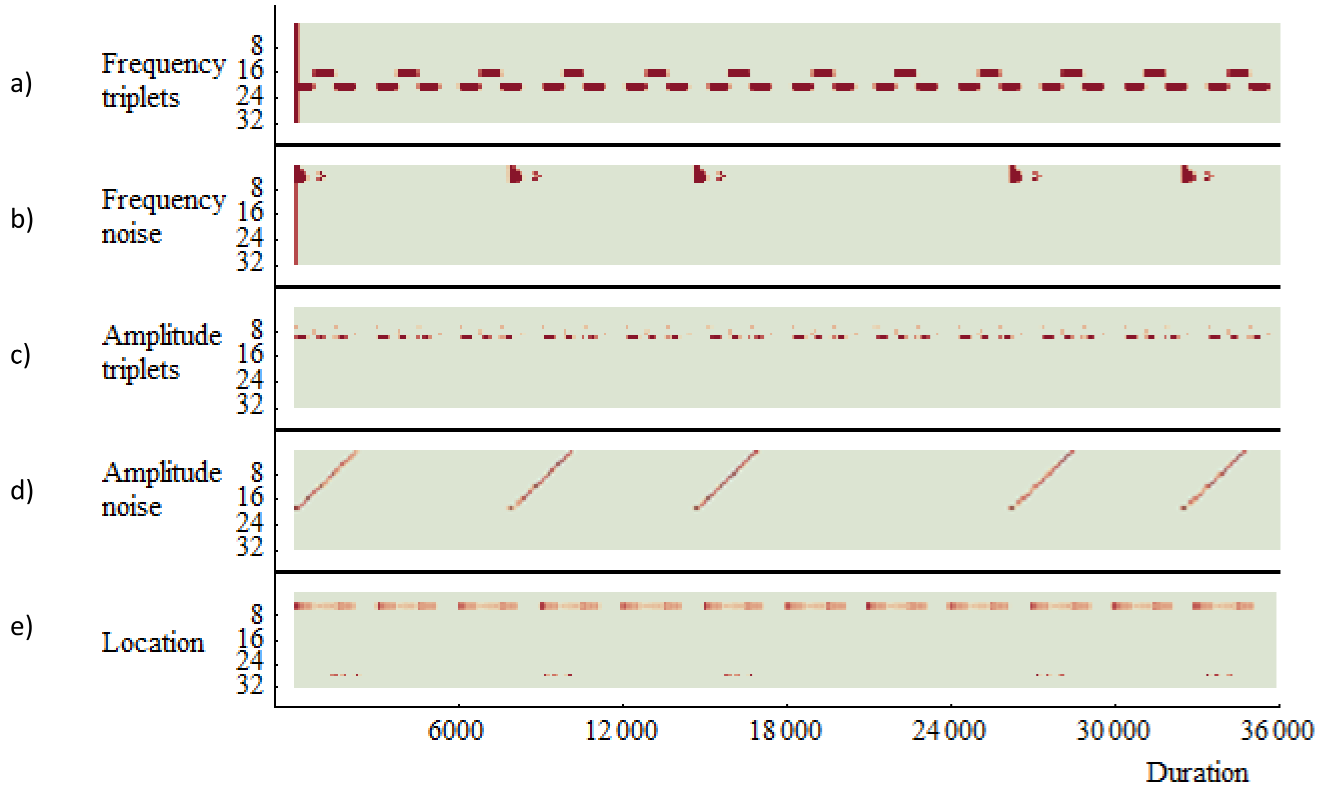


Figure 11. Active oscillators

Figure 11 shows the activation of the oscillators for the same five features through time. The darker the colour in the graph, the higher the activation is. It can be seen from these results that the model does attend to the location where a stimulus is being presented effectively through the range of different stimuli. In Figure 11, the first and third band, Figures 11(a) and 11(c) respectively, describe the activation of oscillators related to the frequency and amplitude bands present in the left ear caused by the ABA sound presentation. The results show that activation for both follow that presentation temporally and as such there was a good temporal binding of the features. The same can be said for the second and fourth bands, Figures 11(b) and 11(d), representing the activation of oscillators related to the frequency and amplitudes bands present in the right ear caused by the noise bursts presentation. Finally, the last section, Figure 11(e) presents the activation of the location specific oscillators and it can be seen that the fifth oscillator responds to ABA sequences while oscillator 28 follows the noise bursts, which corresponds to the correct locations. The activation present at the very beginning of the sequence in a), b) and c) is because the global inhibitor has not yet suppressed activation for channels that are not relevant.

Figure 12 present the activation of the leaky-integrate-and-fire attentional neuron (in blue), which shows the moment the stimulus was salient enough to capture attention. In all graphs of Figure 12, the yellow lines represent the presentations of the triplets and the pink lines represent the presentations of the noise bursts.

The simulations were first performed without adding any top-down attentional processes, and thus the attention presented is strictly stimulus-driven, letting only the characteristics of the stimuli decide where the attention should be drawn. The top left graph in Figure 12 shows the activation of the attentional leaky integrator in that situation. It illustrates that in the absence of endogenous top-down attention, the model gets activated by all stimuli, both the noise bursts and the ABA sequences.

In the second set of simulations, the top-down attention that was present in the human task was added. Attention was drawn only to the stimulus presented in the left ear, the triplets (top right), and in the left ear, the noise bursts (bottom right). Here again, the networks succeed in replicating a condition where top-down attention is directed to one ear and the other one is ignored, suppressing the information coming from that ear.

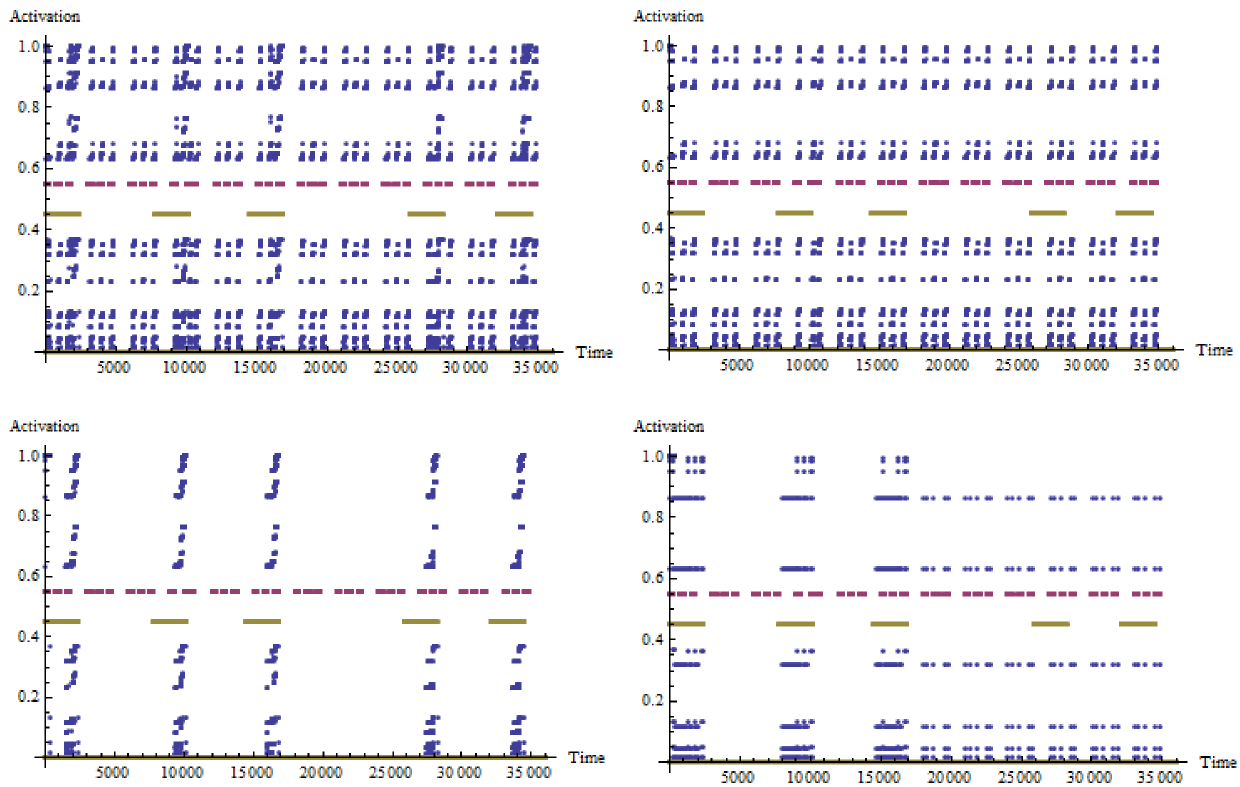


Figure 12. Network outputs. Top left. Exogenous attention. Top right. Attention triplets. Bottom left. Attention on noise bursts. Bottom right. Carlyon task with attention switching. Triplet stimuli is shown in pink and the burst stimuli is shown in yellow

The last set of simulations were similar to Carlyon's attention switching task. In the experiment, the subjects were asked to pay attention to one ear for half the stimuli presentation, then switch to the other ear for the other half. Here again, the network succeeds in replicating a condition where top-down attention is directed to one ear and the other one is ignored. The bottom right graph of Figure 12 presents the activity of the attentional activation of the model when attention was switched midway through the presentation (18000 on the x axis of the graphic). It can be seen that the model fires whenever there is enough activation, as required by theory but that exogenous activation is suppressed when attention is endogenously aimed elsewhere. The attentional leaky integrator thus fires whenever a salient noise burst is present in the first half of the sequence and then only when an ABA sequence is present in the second half, after the attentional switch. It can also be seen that as the noise bursts have a greater intensity, the activation of the saliency detector is greater for the noise bursts than the triplets.

One of the elements present in research by both Carlyon et al. (2001) and Thompson et al. (2011) was the build-up of streaming as time passes. The model presented here was not designed to show a separation of percept between A and B tones in the triplet sequences. What can be seen are the times when attention is captured by an auditory object. As such, the attentional leaky-integrate and fire unit does not display if the A and B tones are considered part of one object or two separate ones. An evaluation of the oscillator banks (Figure 11) indicates that the strength of connections between the A tones and B tones and A tones with themselves remains equivalent thorough the sequence. This model thus cannot model streaming. As the model presented in this chapter was a model of auditory object creation, this is not something that has been implemented, but it would be a valuable future development to refine the model.

This model was able to replicate a number of overt properties of auditory scene perception: the higher saliency of long tones over short tones, the higher saliency of louder tones over quieter tones and the greater response to frequencies between 2-4 kHz.

Model modifications

Once the initial model was created, several refinements were developed and evaluated, based on work in the literature that aimed to make the model more psychologically accurate.

Localization array. The first modification aimed to bring the treatment of the auditory input closer to the distributed process present in information treatment where stimuli can come from multiple azimuthal locations instead of just the two (the left ear or the right ear) initially implemented in the model for the frequency and amplitude attributes (Middlebrooks, 2015). To do this, location attribute was transformed into a more distributed carrier of other attributes. In the initial model, the stimulus localization and the other components were treated separately. The interaction between position and amplitude or frequency was not taken into account. In this modification, the localisation vector is considered a vector of transmission for the frequency and amplitude information. Figure 13 presents the architecture of the modified network, where instead of having one filterbank (and oscillator bank) for amplitude and one filterbank (and oscillator bank) for frequency, there are as many filterbanks and oscillator banks for each of those two attributes as there are locations.

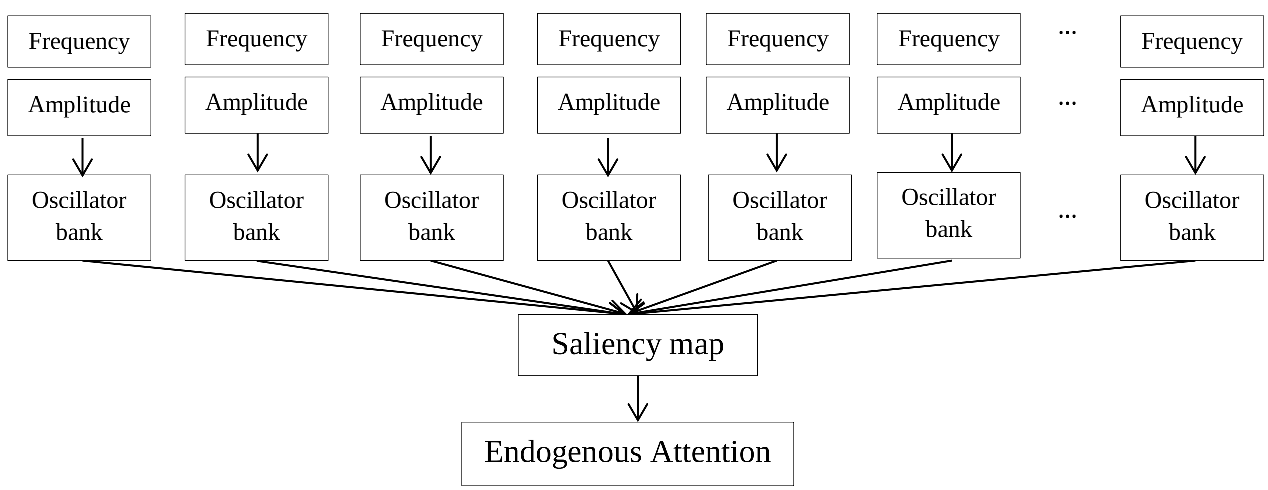


Figure 13. Localization array

A 22-position array was used, similar to the one used for the position attribute in the initial model. The locations are left-right and the array represents a hemisphere array around the listener. The stimuli presented at location 1 are most distant from stimuli presented at location 22. The model was reduced from 32 locations to just 22 since this was considered sufficient to offer a good separation and to also allow for demonstration of the source overlap without needlessly adding excessive computation time. For the same reason (to reduce excessive computation time), the stimuli are also halved in duration. Having a completely spherical array would only mean adding another set of locations bridging the location between position 22 and 1 behind the subject, which was deemed to not be useful for the presented stimuli, but theoretically could be added to the model in future research, to allow for a front-back positioning of stimuli.

Each position represents the frequency and amplitude intensities according to their distance from the "source" location, using a difference of Gaussians. The modified model therefore has a 22 x 22 array of frequency, representing the position and the result from the filter bank, and a similar 22 x 22 array for amplitude. Each of these arrays was then passed to the oscillator banks to create weighted connections between elements occurring at the same time and respecting a threshold for minimal activation. These oscillators were then passed to the saliency network to get the location of the most salient point at each time, as was done for the initial model. In this array, the triplets were presented at location 5 and the noise bursts at location 20.

The activations created by the position array are presented in Figure 14. In the figure, the yellow line is the model's response, the blue line the left ear stimulus and the pink line the right ear stimulus. The tested conditions were the same as for the initial model. Leftmost (a) present the activation when the noise burst sequence was attended, b) present the activation when the triplet sequence was attended, c) presents the Carlyon task where attention being transferred half-way, here at the 9000 timepoint, which was the most important result to replicate and d) is the exogenous attention only. The figure displays that the model correctly diverts attention according to the saliency and top down attention as the first model did.

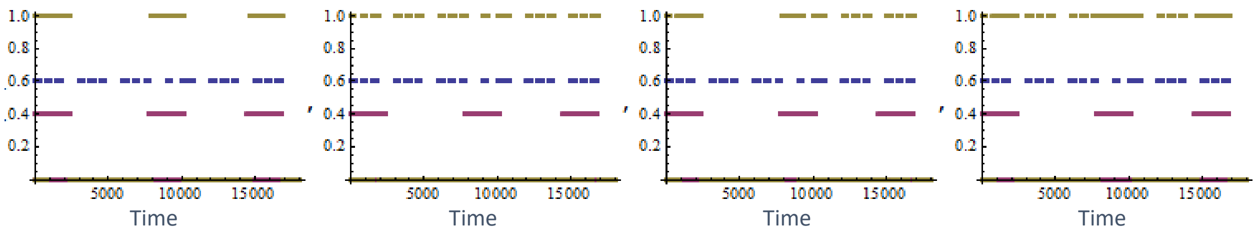


Figure 14. Activation output from the localization array network. a) Attention on right ear (noise bursts) b) Attention on left ear (triplets) c) Carlyon task d) Exogenous attention. Triplet stimuli is shown in blue and the burst stimuli is shown in pink

Those changes did not improve on the initial model in terms of the leaky integrator output. As the reasoning behind these modifications of the initial model was to create a more biologically plausible model, the change did improve the model. The added computational cost is non-negligible but this modified model would be best suited to model the auditory object creation of sounds presented at locations other that far left and far right.

While location specific improvements were investigated, showing that integrating more plausible modifications did not change the saliency detection of the model, there is scope for more refinement. Further research could test other modifications to the model, such as adding other localization dimensions: front-back, up-down. Such changes would allow for a better representation of stimulus localization, but it should be stressed that each addition does add a prohibitively high cost to processing.

Attention array. In addition to location specific modifications, the second modification that was investigated in this thesis was to change the top down attention to an array, building on the change previously carried out on the localization attribute. As attention can be drawn to one ear or the other, it is also possible to draw it to a spatial location if the source location is not confined to the two ears. The initial top down attention was therefore converter to a vector, where the center of attention can be focused one specific location. Following the modified spotlight model of attention (Eriksen & St-James, 1986), the focus of this attention fades quickly outside of the spotlight but does not reach 0 in contiguous locations. A Gaussian was used to represent this gradual spotlight. In this modification, each location had its own saliency map and attentional leaky integrator. However, it was found that this modification did not improve on the initial model. By splitting the saliency activation into several locations, the activation left at each position was too small to pass the threshold for the triplets. Only very salient stimuli, in the present case, the noise bursts, appeared in the final attentional leaky integrator. Figure 15 shows the 22 attentional leaky integrators. Only the integrators covering the locations of the noise burst stimulus were activated (locations 19, 20 and 21) while the less salient triplets (situated around location 5) do not cross the threshold required for attention. As the saliency map works by addition, splitting the activation prevented the module from working as expected. Time constraints prevented further evaluation of this, but for future work, such a modification could be attempted if a non-additive binding of attributes was implemented, and would be expected to have positive results.

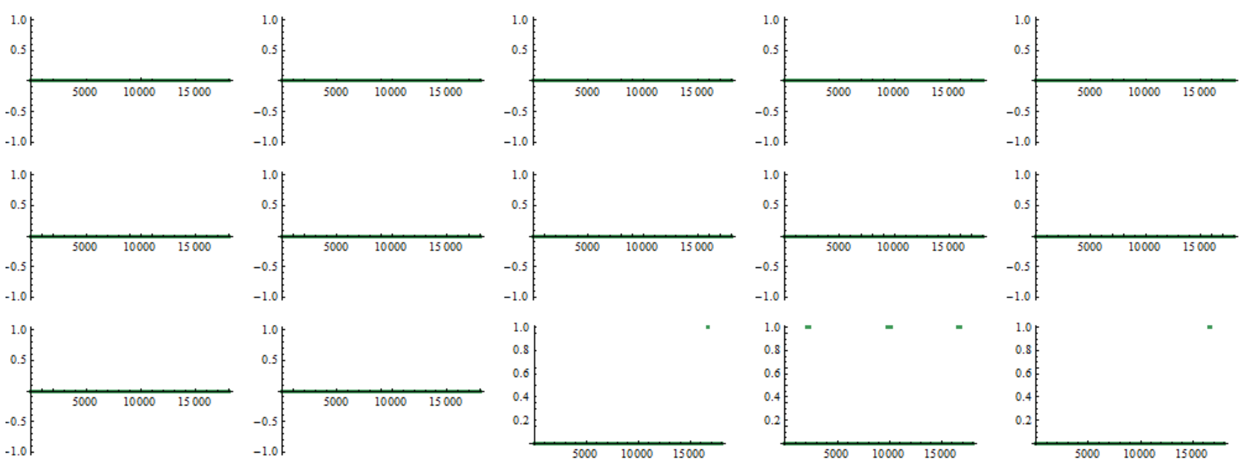


Figure 15. Activation output for the saliency map array network. Only the noise bursts were picked up

Oscillatory binding of features. Building on the conclusions reached by the saliency model, the implementation of non-additive binding was attempted. The idea was to bind the segments formed by the separate attributes into auditory objects by using another oscillatory network (Wang & Chang, 2008). In the initial model discussed previously, the binding of attributes was done by a simple summation. The aim was to create a new input to the saliency map that would give priority to attributes that were temporally bound instead of just adding the different activations.

No conclusive results were achieved with this modification of the model. As the attributes were already in oscillatory pulse form, the time they were activated was not long enough to create an input building enough activation to cause the second binding oscillator to fire, being already shorter than the initial attributes. Also, being spiking neurons, they had a negative phase that mostly canceled out the positive phase. It was therefore required to only make use of the positive values of the signal. The modification was tested on the initial attributes, as the previous discussion concerning modifying the locations showed that the results were not significantly better, and did not warrant the prohibitively long processing time to use the multiple location array for very little improvement. However, tests showed that the binding did not improve the initial model, instead reducing the power of the top down attention module by strengthening connections between elements that were faint but simultaneous, thus removing discrimination of saliency. This addition worsened the quality of the representation obtained with the initial model. Figure 16 shows the activations of the attentional leaky integrator in the same four conditions as the initial model.

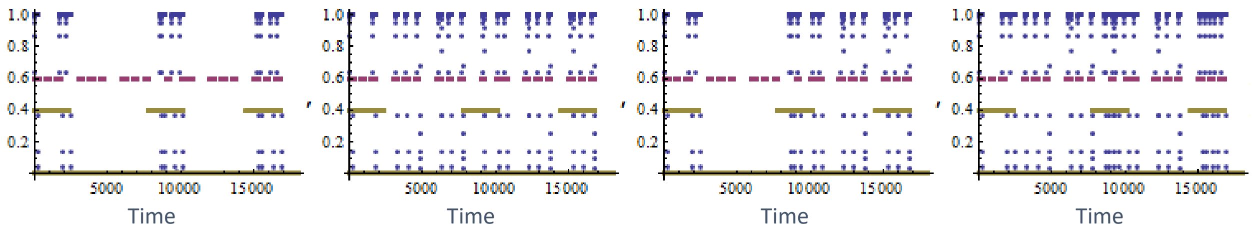


Figure 16. Activation output for the oscillatory binding of features network. a) Attention on triplets b) Attention on noise bursts c) Carlyon task d) Exogenous attention. Triplet stimuli in pink and noise burst stimuli in yellow

Overall Limitations

While the model presented produced positive results overall, there are several areas that can be further developed. Firstly, the model is very sensitive to parameters and contains a large number of them to account for the different possible inputs. Segments are created based on thresholds, as is saliency and the final attentional leaky integrator. Information transmission from one level to another is gated by thresholds and by changing those thresholds, what information is transmitted changes massively. It is possible to optimise it for a set of inputs, but there is the risk of it becoming overspecialised and not able to represent the changeable capabilities of the auditory system.

In addition, the final attention detector output is simple but gives little information on the different inputs and how they contributed to the final activation. Using the sum of activations coming from the oscillators with a threshold would give us sensibly the same information while making the contribution of each attribute more obvious, but would remove some of the biological plausibility that the attentional leaky-integrate-and-fire neuron brought.

Conclusion

This chapter presented a new model of auditory object creation (as published in Morissette & Chartier, 2015). It contains sensory feature extraction based on theories and research by Bregman (1990), sensory feature binding based on work by Wang (1995, 1996, 1999, 2008) for the oscillatory neural network, work by Wrigley and Brown (2004) for the architecture of single feature processing streams and inhibition of return, work by Itti and Koch (1998, 2000, 2001a, 2001b) for the saliency map and Wrigley and Brown (2004) for the attentional leaky integrate and fire neuron. The presented model was the first that simultaneously incorporates the influence of both exogenous and endogenous attention on saliency while modelling the treatment of multiple auditory features in an integrated manner. This initial model was implemented and was found to yield good results, using established stimuli also used by Carlyon et al. (2001) and Thompson et al. (2011), of the interaction of bottom-up and top-down attention in the creation of auditory objects. These positive results represent a contribution to the research in this domain.

To build on this, further modifications of this initial model were proposed. Three modifications were investigated, a modification of the attentional representation to allow for a more distributed focus of attention, a modification of the binding of different sensory features from an additive model to an oscillatory binding and a modification of the localization representation of the features. The location array modification to the model improved the correspondence with the distributed process of source localization compared to the initial model, thus improving the plausibility of the model even if it didn’t change the model’s results. However, further research is needed, especially on the final saliency detector, to be able to make use of the proposed changes, as the lone attentional neuron gives a very narrow representation of when stimuli pass an attentional threshold.

Based on the research presented in this chapter, the best model to use is the distributed array model, keeping the single location saliency map and additive binding process. This model represents a conclusive representation of the auditory object creation process. As discussed previously in the chapter, these can be improved, but this configuration was found to produce the best results. With this model a simplified representation of the process of object creation and its interaction with attention was presented. This process is still very much under investigation and therefore the remainder of this thesis was centered around experimental work aiming to further our knowledge of the effect of the time spent attending to a task on auditory streaming.

Chapter 3

Experiments on the Effect of the Time Spent Attending to a Task on Auditory Object Creation

As discussed in the general literature review, the influence of the time spent attending to a task on auditory streaming is an issue that remains contentious. Human behavioural studies (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Shamma, Elhilali & Micheyl, 2011; Thompson, Carlyon & Cusack, 2011) and event-related potential studies (Snyder et al. 2006; Sussman, 2007; Sussman, Winkler, Huotilainen, Ritter & Näätänen, 2002) have shown dissociation between a pre-attentive, low-level sound segregation process, and an attention-dependent process that can be called into play in forming perceptual objects and streams. While there is a consensus about streaming being at least in part pre-attentive, the influence of the time spent attending to a task is still a matter of debate, especially concerning whether attention speeds up streaming and if so, under what conditions.

As this remains an issue of some debate, it was felt that this merited further investigation in this thesis. Two experiments were therefore designed and performed to investigate this in more depth. The two experiments presented in this chapter use a methodology that enables the study of the influence of the time spent on task on the streaming process through behavioural measures (Billig & Carlyon, 2016; Carlyoon et al., 2001; Thompson et al., 2011). As in previous studies on streaming, triplet stimuli are used (Bendixen, Denham, Gyimesi & Winkler, 2010; Shamma et al., 2011). The triplets are presented in a single auditory channel (a designated ear). In one condition, the subjects are asked to attend to these stimuli. In another, they are asked to ignore the triplets and instead attend to stimuli presented in the opposite ear for a limited period of time. From a methodological point of view, it is very difficult to rely on behavioural (performance) measures to determine the extent of processing of the stimuli presented in a to-be-ignored channel. Auditory event-related potentials (ERPs) provide a means to monitor the extent of information processing of unattended stimulus input. ERPs are the minute changes in the ongoing electrical activity of the brain (the “EEG”) that are elicited by external stimuli or internal psychological events. These ERPs consist of a series of negative- and positive-going “components”, thought to reflect different aspects of information processing. Even when a channel is attended and a behavioural response is made, the cognitive processes leading up to the actual response must be inferred. ERPs thus also provide a means to measure processing prior to, at the time of, and following the behavioural response.

The first experiment reported in this chapter is a follow-up to the auditory deviant detection behavioural task carried out by Thompson et al. (2011) and Billig and Carlyon (2016). Using ABA triplets as stimuli to cause streaming, deviant detection in conditions where time on task varies were compared. ERPs are also recorded to gather more insight into the decision process leading to the deviant detection and as a supplemental mean to compare the effect of the time spent attending to the task on the deviant detection in a context of auditory streaming.

The second experiment to be reported in this chapter investigates the effect of the presence of a continuity illusion and the time spent on task on streaming, using a modification of the Thompson et al. (2011) methodology, which was previously shown to induce streaming. The continuity illusion is an infrequently investigated perceptual effect and while its interaction with streaming was evaluated for a very specific set of stimuli, the interaction between attention and the continuity illusion with regard to streaming in ABA triplets has not previously been investigated. The evaluation of the effect of the continuity illusion on deviant detection in a condition of streaming has also never been investigated with ERP measures.

Experiment 1. Influence of the time spent on task on stream segregation as tested through deviant detection in time-shifted frequency differing triplets

This study examined the influence of the length of time that a sequence of auditory stimuli is attended to on the experience of streaming. Streaming here refers to the separation of the auditory stimuli from an initial unified percept into multiple percepts, split according to auditory sources. The information from each source becomes an independent perceptual stream, and these streams may be divided further into separate objects by the onset and offset of sounds.

Early studies indicated that the experience of streaming occurs pre-attentively and does not benefit from attention. These studies argue that the build-up of streaming occurs very rapidly and accurately by the low-level auditory system before subjects are conscious of this streaming taking place (Bregman, 1978; Bregman, 1990; Bregman & Dannenbring, 1973, Bregman & Rudnicky, 1975). Bregman (1990) presented alternating tones A and B of different frequencies at different rates of presentations. He found that when auditory stimuli are presented very rapidly, subjects perceive two different streams, one of A tones and one of B tones. In contrast, when the same sequence is presented more slowly, subjects report a single stream, composed of alternating A and B tones. In their study, Bregman & Rudnicky (1975) showed that low-level auditory streaming can be carried out without attention in an automatic innate process.

However, more recent studies have demonstrated that directing attention to the auditory sequence can influence the perception of streaming. A series of behavioural studies (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Shamma et al., 2011; Thompson, Carlyon & Cusack, 2011) have shown a dissociation between a pre-attentive, low-level sound segregation process (separation into segments according to source or stimuli characteristics), and an attention-dependent process that may influence the formation of streams. While there is a consensus about streaming being at least in part pre-attentive, the influence of the time spent attending to a task is still a matter of debate, especially concerning whether it speeds up streaming and if so, under what conditions.

One way of examining the influence of attention is to manipulate “time-on-task”, the amount of time that a subject has been attending to the auditory sequence. Deviant detection is often tested through oddball paradigms, where a stream of frequently-occurring and repeating homogenous “standard” sounds are presented and a rare (or “odd”) feature of the standard is changed to form a “deviant”. Studies have found that deviants are detected more accurately when a single auditory stream is perceived rather than two (Bregman & Dannenbring, 1973; Roberts, Glasberg & Moore, 2008; Vliegen, Moore & Oxenham, 1999; Vliegen & Oxenham, 1999).

Based on this result, Carlyon et al. (2001) developed a paradigm to test the influence of attention on streaming. They presented sequences of ABA triplets (in which A and B are different tones) in one ear and a series of noise bursts in the other ear. In one condition, subjects were asked to focus their attention on the channel containing the triplets and to ignore the channel containing the noise. They pressed one button when they heard 1 stream and a different button when they heard 2 streams. In a second condition, the subjects were asked to focus their attention noise burst stimuli for 10 s and determine whether they were increasing or decreasing in intensity. The noise bursts ended after 10 s and subjects were then asked to switch attention to the opposite ear where the ABA triplets were presented. Because stream segregation builds over time (Van Noorden, 1975), the experiment aimed to see whether the 10 s of attending to noise burst had aided with stream segregation in the triplet channel or not. The authors found that attention is necessary for streaming to build up. When subjects ignored the ABA triplets sequence while carrying out the task on noise bursts, subsequent streaming judgments in the ABA triplets sequence revealed less stream segregation of the previously ignored tones than they did when subjects had been attending to them. The authors also found that the speed of the separation in two streams is proportional to the frequency differences, as found by Van Noorden.

One issue with the study by Carlyon et al. (2001) is that the determination of streaming is based on subjects' judgements of what they were experiencing. However, such judgments are highly subjective. Thompson et al. (2011) developed a more objective method to monitor streaming. They presented a "deviant" within the ABA triplet sequence, either early in the sequence or late. This deviant was created by delaying the onset of the B tone in an ABA triplet by 50 ms. Subjects were asked to press a button upon detection of the deviant. This deviant should have been easier to detect when the subject is experiencing the sequence as a single stream. As discussed previously, a single stream is perceived at the beginning of the sequence and two streams are perceived later in the sequence. Two different frequency separations between the A and B tones, 4 and 8 semitones were also used.

Thompson at al.'s (2011) results confirmed this deviant detection hypothesis. Detection of the deviant was better when the deviant occurred early rather than late in the sequence for the large frequency separation. In a second experiment, the time spent attending to the triplet sequence was modified in a different way. The paradigm was similar to that used by Carlyon et al. (2001). The ABA triplet sequence used previously was presented in one ear. In the opposite ear, noise bursts were presented. When the subjects attended to the triplet sequence, the deviant could occur either early or late. When they attended to the noise bursts to determine whether their intensity was increasing or decreasing, the triplets were to be ignored. However, the presentation of the noise bursts ended at the mid-point of the presentation of the triplets, when attention was then to be switched to the triplet channel. In a small frequency separation condition, when subjects attended the triplet channel, accuracy of detection of the early and late deviants did not vary. Time-on-task did influence the results when the frequency separation between A and B tones was wider. When the triplet channel was attended, the early deviant was detected more accurately than the late deviant. Importantly, when attention was only switched to the triplet channel after first attending to the noise channel, then detection of the late deviant was better than the detection of the late deviant in the condition attended throughout.

Billig and Carlyon (2016) conducted a similar experiment. ABA triplets were again presented to one ear and noise bursts to the opposite ear. In this study, the deviant could occur in three positions: early, middle or late in the sequence. As in Thompson's experiment, there was a condition during which the subjects had to continually attend to the triplets and another condition during which they initially attended the noise bursts and then subsequently switched their attention to the triplets. Billig and Carlyon were also concerned that the subjects might not have experienced streaming in the Thompson et al. (2011) study. In a separate condition, they thus had subjects report their percept of streaming. By comparing the deviant detection accuracy after the task switch and after having attending to the whole sequence, they found that auditory stream segregation is facilitated by attention. They also found that the detection of the deviant is easier when the sequence is heard as integrated in a single stream, regardless of the frequency difference between the tones and the time in the sequence at which the deviant occurs.

A major problem in the study of the effects of attention is that when a stimulus sequence is to-be-ignored, in the absence of behavioural measures of performance, it is very difficult to determine the extent of processing of these stimuli. It is possible, for example, that in the streaming studies, while subjects may be asked to attend to the ear containing the noise bursts and ignore the triplets occurring in the opposite ear, they could nevertheless momentarily attend to and sample the triplets (i.e., “eavesdrop). Auditory ERPs provide an exquisite means to monitor the extent of information processing of unattended stimulus input. As discussed in previous chapters, ERPs are the minute changes in the ongoing electrical activity of the brain (the “EEG”) that are elicited by external stimuli or internal psychological events. These ERPs consist of a series of negative- and positive-going “components”, thought to reflect different aspects of information processing. The auditory ERPs consist of a series of components. The auditory oddball sequence is often used, where, as discussed previously, subjects are presented with a series of frequently occurring, homogenous “standard” stimuli. At rare (or “odd”) times, a feature of the standard is changed forming a “deviant”. In the streaming studies when an ABA triplet is frequently presented, it acts as a “standard”. At times, the onset of the B tone is changed and occurs later. This ABA triplet thus serves as a “deviant”. When attention is directed to the sequence, the rare deviant is often called a “target” because it is to-be-detected. In the present study, the ABA sequence may be either attended or ignored. To avoid describing the rare stimulus as a “target” in one condition and a “deviant” in another, it will be consistently referred to in this thesis as a “deviant.” All auditory stimuli, whether attended or ignored, will elicit obligatory ERPs, most notably an “N1” occurring at approximately 100 ms followed by a “P2” occurring at approximately 180-200 ms. These are mainly affected by the intensity of the stimulus. Because the intensities of the standard and deviant stimuli are identical, the N1-P2 should not vary, apart from with regard to timing. In the streaming studies, stimuli are presented very rapidly and as a result, the amplitude of the N1-P2 will be very small as is is affected by the rate of stimulus occurrence. In addition to the N1-P2, the deviant also elicits another negativity occurring from 100-250 ms (depending on the extent of change). This is the mismatch negativity (MMN). Several studies indicate that the MMN is elicited relatively independently of attention (Muller-Gass et al., 2006; Näätänen, 1990; Sussman, Winkler et al., 2002). Thus, when subjects are attending to the ear containing the noise burst, the occurrence of the deviant in the triplet sequence should still elicit an MMN. Similar to the N1, the amplitude of the MMN is very small, often being less than 1 µV. The ongoing EEG background noise in waking, alert adults often exceeds 25 µV, and the small ERPs are embedded in this background noise. To visualise small amplitude ERPs such as the MMN, the stimulus is required to be repeatedly presented. With an averaging of the single trial responses, a reduction of the background noise will occur, allowing the signal (the ERP of interest) to emerge. This may require several repetitions of the stimulus, particularly if the amplitude of the ERP component is small. Luck (2014) indicates that as many as 400 deviant presentations will be required to allow the small amplitude MMN to emerge from the background noise. With the current experimental paradigm, it is therefore not expected to see a clear MMN.

Attention does affect later ERP components. When the deviant is to be actively detected, a large N2b is elicited (The MMN was initially labelled as an N2a). The N2b appears to be affected by the difficulty of a task. It is larger for deviants that are difficult to detect, perhaps reflecting the extra attentional effort requited to make the detection. N2b is largest at central scalp locations present around 200-300 ms after the stimulus onset and is only present after correct detection of a rare stimulus during conscious attention.

A much-studied ERP component associated with detection of deviants is the so-called P3b. When the subject actively attends to a sequence and detects the deviant, a large amplitude positivity, maximised over parietal scalp sites, is elicited. This is the P3b. It can be distinguished from an earlier positivity, the P3a, occurring from 200-300 ms. Both the P3a and P3b are elicited by rare, deviant stimuli. However, the P3a occurs when the subject ignores the sequence containing the deviant, providing the extent of change is large. The P3b is not elicited if the sequence is ignored. The scalp distributions are also different. The P3a is maximised over centro-frontal regions of the scalp, while the P3b is largest over parieto-central sites. In easy detection tasks, the P3b often occurs at around 300 ms, thus the alternative label, the “P300”. Several factors influence the P3b (Johnson, 1988). It only occurs if the rare target is detected. Its amplitude varies indirectly with the probability of target occurrence. This amplitude is also affected by uncertainty, or what has been called “equivocation”. In difficult tasks, when the subject is uncertain that the target has been presented, the amplitude of the P3b is reduced. If the subject is uncertain that a deviant has been presented, its amplitude will be smaller. For this reason, the amplitude of P3b is inversely related to task difficulty; it decreases in amplitude as the task becomes more difficult and subjects become more equivocal about their decision (Johnson, 1986).

The experiment presented in this chapter employs a paradigm similar to that used by Thompson et al. (2011). It investigates the difference that the time spent attending to a task has on the speed of auditory stream formation by presenting two experimental conditions manipulating the time spent on task. This experiment examines whether deviant detection will be easier to perform when a single stream or two streams are perceived. Crucially, it also investigates the effect of a task switch midway through the sequence. The stimuli used in this experiment is perceived as a single stream of alternating tones early in the sequence and perceived as two streams of distinct tones later in the sequence. In this experiment, following established work in the literature, subjects were presented with frequently occurring “standard” triplets. At rare times, the triplet pattern changed, in that the central tone was shifted forward temporally, creating a deviant triplet.

As the methodology of this study followed Thompson et al. (2011), but with the addition of ERP results to the standard behavioural measures, a replication of their behavioural results was expected. So, for a sequence of sounds attended throughout, the early deviant detection performance was expected to be better than for the late deviants, as less streaming would have had time to build up early in the sequence. For the same reason, the late deviant detection performance in a sequence where the task to be completed was switched midway (i.e. initially attending to noise bursts, and then switching to deviant detection) was expected to be better than for a sequence where the same task was to be completed throughout, as the time spent on the task increases the speed of the streaming process, and by switching tasks midway, less streaming would have had time to build up. It was also of great interest to compare the early deviant in the sustained task condition to the late deviant in the switched task condition to see if the time spent on task from the beginning of the sequence differed from the time spent on task after a task switch. This condition was not present in Thompson et al. (2011). The expectations were for the early deviant in the sustained condition to be easier to detect, as some streaming of the ABA sequence should occur, even when the subjects attend to a different task.

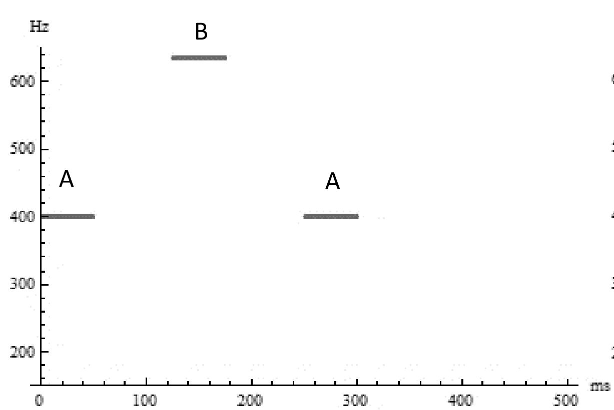
As Thompson's methodology contained both actively attended and to-be-ignored presentation of the deviants, the N2b and P3b components were analysed. No event-related potential recording has ever been done with the Thompson et al. (2011) methodology, making the experiments presented in this thesis the first example of this. The expected results were for the N2b to be smaller and P3b to be bigger for the easier to detect early deviant, than for the late deviant in the sustained attention task. It was also expected that the late deviant in the attention switching task would show a smaller N2b and bigger P3b than the late deviant in the sustained attention task. Finally, the expected results were for the N2b to be smaller and P3b to be bigger for the easier to detect early deviant in the sustained task condition than for the late deviant in the switched task condition, but by a very moderate amount as those occur at similar moments after the sequence has started to be attended.

Methods

Subjects. Eleven subjects (three female, eight male) between the ages of 18 and 34 years took part in the study. None reported a history of neurological or psychiatric disorder and all had self-reported normal hearing. Subjects signed a consent form and received an honorarium for participation in this study. The study was conducted according to the guidelines of the Canadian Tri-Council (Medical, Natural and Social Sciences) on ethical conduct involving human subjects. Of the eleven subjects, the data of one had to be removed as he had failed to detect any deviants in the second experimental condition. His results were kept for the first experimental condition as they did not differ significantly from the other subjects.

Procedure. Auditory stimuli were presented binaurally through Sony MDR-V6 headphones while the subject sat in a sound-attenuated room. The instructions were given orally and then displayed on a computer screen at the beginning of the experiment. During the experiment, the instructions were replaced by a static image of clouds and the subjects were asked to attend to the auditory stimuli presented in one ear at a time.

The auditory sequence in the left ear was pure tones grouped in an ABA triplet pattern, the triplets being separated by silent periods. The A and B tones forming the triplets were 400 and 635 Hz pure tones respectively (Cusack et al., 2004), having a total duration of 50 ms and a 5 ms rise-and-fall time. The tonal separation represented an 8-semitone difference, a separation that was shown to produce a clear streaming effect (Carlyon et al., 2001; Thompson et al., 2011). The tones were presented at an intensity of 55 dB SPL. In the standard triplet, the individual tones forming the ABA triplets were separated by a 75 ms silent period. Deviance from the standard triplet was created by delaying the onset of the B tone by 50 ms (i.e., 125 ms after the offset of the A tone, Thompson et al., 2011; Vliegen & Oxenham, 1999). Only the larger delay of the Thompson et al. study was used, because this deviant is relatively easier to detect. The time between triplets was 125 ms (Thompson et al., 2011). Thus, the time from the onset of one triplet to the onset of the next triplet was 500 ms. See Figure 17 for representation of the standard and deviant triplets. Each sequence of triplets contained 24 ABA triplets and therefore had a total duration of 12 seconds. The next sequence began 1 second after the end of the last one.



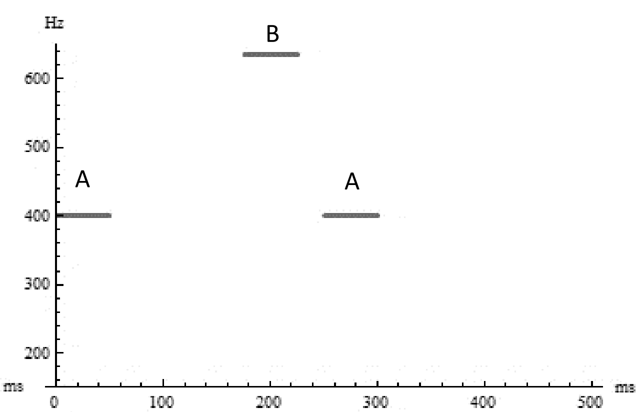


Figure 17. Triplet stimuli for first experiment. Left. Standard triplet. Right. Deviant triplet.

Auditory stimuli were also presented in the opposite, right ear. This consisted of a sequence of noise bursts centered around 2500 Hz (white noise filtered between 2000 and 3000 Hz). The duration of one noise burst was 400 ms. It either increased linearly in intensity from 0 to 90 dB SPL or decreased from 90 to 0 dB SPL over the 400 ms duration, giving the perception of “approaching” and “retreating” sounds respectively. The noise bursts were presented at random intervals varying from 250 to 1000 ms after the offset of the previous noise burst. They could thus overlap with the presentation of the triplets in the opposite ear. The number of noise bursts that were presented varied depending on the experimental condition. A total of 12 noise bursts were presented in a 12 second sequence for the sustained task experimental condition and 6 noise bursts presented during a 6 second sequence for the switched task experimental condition. The stimuli presented in each ear was not counterbalanced because the ear of presentation has no influence on perception (Cusack et al., 2004).

Two experimental conditions were presented. In the “sustained” attention condition, subjects attended to the triplet sequences occurring in the left ear for the entire 12 s duration of the triplets. They were asked to signal their detection of a deviant triplet by pressing the spacebar of a keyboard as quickly and accurately as possible. The deviants could occur early or late in the sequence, in both positions or not at all. The reason a No deviant condition was presented was so that subjects would not develop the expectation that if the deviant did not occur early in the sequence, then it surely must occur later. Similarly, the Both condition was presented so that subjects would not develop the expectancy that once the early deviant had been presented, it could not be presented later, which may have caused the remaining triplet sequence to be ignored upon detection of an early deviant. Processing of the first and second half of the standard triplets should therefore have been similar. In the sustained attention condition, attention was directed continuously to the triplet sequence occurring in the left ear and the noise bursts presented in the right ear were to-be-ignored. The noise bursts were presented for the full 12 s duration of the attended triplet sequence.

In the “switched’’ attention condition, noise bursts were presented for only a 6 second period. Subjects were asked to attend to the noise bursts in the right ear and to signal the detection of the “approaching” or “retreating” noise by pressing one of two keys on the keyboard. The triplets presented in the left ear were to-be-ignored during this period. Note that an early occurring deviant triplet pattern should not have been actively detected. When the noise bursts ceased (after 6 seconds), subjects were told to switch attention to the left ear (to the channel containing the triplet sequence) to now actively detect the deviant. The experimental conditions were counterbalanced, with half the subjects performing the sustained task condition first and half the subjects performing the switched task condition first.

In the Early deviant sequence, a deviant pattern occurred after the 4th standard (i.e., during the 5th triplet). The onset of the deviant B tone was thus 2.175 s after the start of the sequence. In the Late deviant sequence, a deviant pattern occurred after the 19th standard triplet (i.e. during the 20th triplet). The onset of the B tone thus occurred 9.675 s after the beginning of the sequence. In the Both deviant sequence, a deviant pattern occurred following the 4th standard triplet (i.e., during the 5th triplet) and again following the 19th standard triplet (i.e., during the 20th triplet). Therefore, two sequences presented deviants during the 5th triplet and two sequences presented deviants during the 20th triplet. In the switched task condition, the late deviant pattern occurred following the 7th standard triplet after the task switch. See Figure 18 for representation of the four deviant sequences.

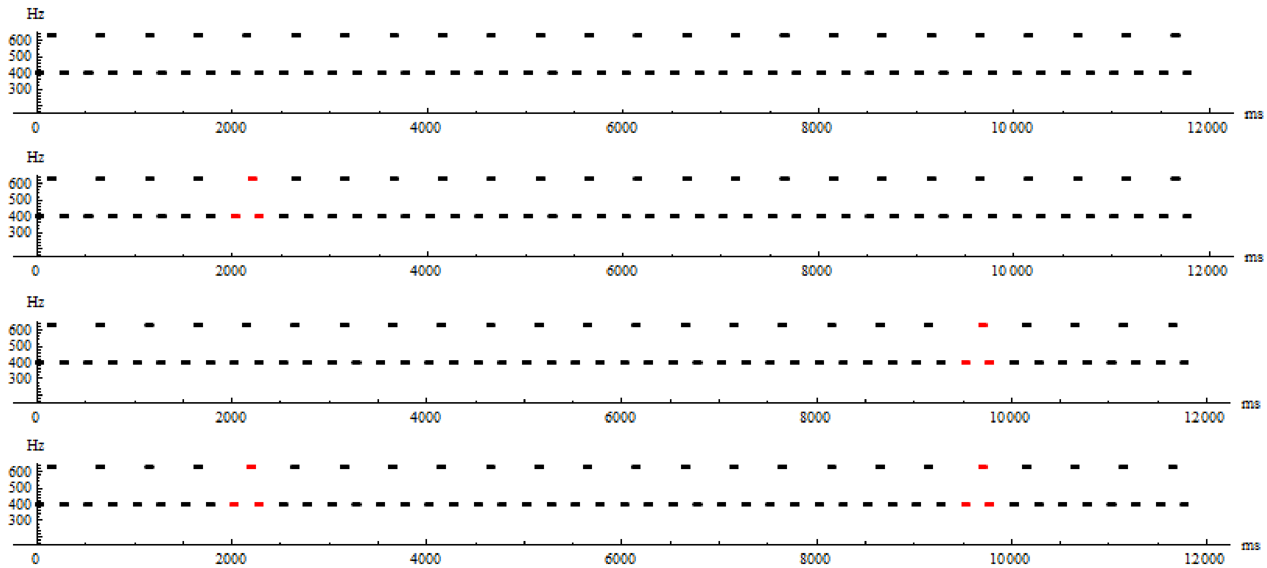


Figure 18. Deviant presentation conditions. Top: No deviant sequence. Second: Early deviant sequence. Third: Late deviant sequence. Bottom: Early and late deviant sequence

In each experimental condition, 50 of each of the four possible sequences were presented. A total of 4800 triplets were therefore presented in each experimental condition, consisting of 4600 standard triplets and 200 deviant triplets. 100 deviant triplets occurred early in a sequence (in the 5th triplet, the Early condition). 100 deviant triplets occurred late in a sequence (in the 20th triplet, the Late condition). Of those, 50 occurred in a sequence containing only a late deviant and 50 occurred in a sequence containing both an early and a late deviant. The order of the sequences was pseudo-randomised so all sequences would be presented 50 times.

Quantification of Behavioural Performance. Performance was measured by accuracy of detection (‘‘hits’’) of deviants and noise bursts, correct rejections of non-deviant stimuli and reaction times (RTs). A response had to be made within 1500 ms after stimulus onset of the first A tone of the triplet (1325 ms after the deviant B onset) to be considered a correct detection. Performance data (i.e., accuracy, RT) was analysed using t-tests to determine if the performance differed depending on if the deviant was presented in isolation (only an early deviant or only a late deviant) or in both positions. Finally, performance data were analysed using an 3x2 ANOVA with repeated measures for the three different conditions in which deviant triplets had to be actively detected (Early and Late deviants in the sustained task condition and Late deviants in the switched task condition) and the sequence type (One deviant presented (Early or Late) or 2 deviants (Both)). Accuracy and response times were not evaluated for the early deviants in the switched task condition as they were to-be-ignored and no response should have occurred to the deviants.

Physiological Recording. The EEG was recorded from 32 scalp sites representing frontal (Fp1, Fp2, AFz, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6), central (C3, Cz, C4, C8, CP5, CP1, CP2, CP8), parietal (P7, P3, Pz, P4, P8), temporal (T7 and T8) and occipital (Oz, O1, O2) areas of the scalp, using active silver/silver chloride electrodes attached to an electrode cap (Brain Products, GmbH, Munich, Germany). Two additional channels were recorded from individual electrodes placed on the left and right mastoids (TP9 and TP10). The nose served as a reference but the recordings were later re-referenced to the mean of the activity at the left and right electrodes. Eye movements and blinks were recorded from an electrode placed at the infra orbital ridge of the left eye. The ground electrode was located between the Fz and FPz sites. Inter-electrode impedances were kept below 50 kΩ. The physiological signals were digitised continuously at a 500 Hz sampling rate and stored on hard disk.

The continuous EEG data was later digitally filtered using a bandpass of 0.05–20 Hz with a notch filter at 60 Hz to filter the fluorescent lights. Vertical EOG activity was later computed by subtracting data recorded at FP1 from that at the infra-orbital ridge. Horizontal EOG activity was later computed by subtracting data recorded at FT9 from that recorded at FT10.

Independent Component Analysis (Makeig, Bell, Jung & Sejnowski, 1996) was then used to identify eye movements and blinks that were statistically independent of the EEG activity. This algorithm was trained to recognise each subject's blink signature, and this factor was then partialled out of the EEG traces. The EEG was visually inspected for channels containing high levels of noise and segments containing too much noise were rejected from further analysis.

The subject's continuous EEG data were then segmented into 1000 ms epochs, time-locked to the onset of the B tone of the triplets, which included a 50 ms pre-stimulus baseline period and a 950 ms post-stimulus period. The single trial segments were subsequently baseline corrected. Segments containing EEG activity greater than $\pm$ 100 μV were rejected from further analysis. Only segments for correct detections of deviant triplets were analysed further. The segments were then sorted and averaged for electrode site of interest (Fz, Cz or Pz), experimental condition (sustained task, switched task), Stimulus (Standard, Deviant) and position in the sequence (Early, Late).

Quantification of the MMN, N2b and P3b Components. In this experiment, the physical features of the standard and deviant B tones were identical, with the deviant being created by its delayed onset. The N1 component elicited was thus expected to be similar. The average of all data points in the 50 ms period before the presentation of the B tone of the triplet served as a zero-amplitude baseline from which all amplitudes were measured. When the triplet channel was to-be-ignored, a MMN should still have been elicited by the deviant. However, as discussed previously, one limitation with the MMN is that the amplitude is usually very small, therefore requiring a large number of deviants to be presented for the component to be visible. For practical reasons, in this experiment it was decided to only present 100 early and 100 late deviants in a single condition. This meant that as expected, the MMN was thus very difficult to observe in the background noise of individual subjects, and was therefore not quantified.

When the triplet channel was attended, an N2b and P3b were elicited. In most studies, they are normally measured in the "raw" deviant waveforms. An alternative is the so-called difference wave computed by subtracting the standard from the deviant waveforms. This difference wave removes processing that is common to both the standard and deviant leaving only processing that is unique to the deviant. The use of difference waveforms when the condition manipulation results in latency shifts of components can lead to an amplitude variation in the difference waveform, at times appearing as an artificial modulation of a component, or even a spurious (non-existent) difference-related component (Muller-Gass & Schröger, 2007). The problem in the use of the deviant-standard wave in the present study is that the physical onset of the deviant occurred 50 ms later than the standard. The ERPs elicited by the physical onset of the stimulus would thus occur somewhat later for the deviant than the standard. In such cases, the use of a difference wave can be questioned. These difference waves were however computed and appear in Appendix 4. The standard used for the difference waves are the 100 standards that were presented in the exact same position in the sequence as the deviant they were compared to.

The grand averaged "raw" deviant waveforms (average of all subjects' averages) were used to identify the N2b and P3b in each condition. The N2b was largest over centro-frontal sites and was thus quantified at Cz and Fz. The P3b was maximum over parietal regions and was thus quantified at Pz. The N2b appeared more as a distinctive peak, while the P3b appeared more as a long-lasting slow wave. The N2b was measured as the mean amplitude within ± 50 ms of the peak within the grand average. The P3b was measured as within ± 75 ms of the peak in the grand average.

The ERPs were compared with a 4x2 ANOVA with repeated measures on the four conditions (sustained attention early, sustained attention late, switched attention early (to-be-ignored), switched attention late) and sequence type (One deviant presented (Early or Late) or 2 deviants (Both)) at for the N2b at Fz and Cz and the P3b at Pz.

A deviant could occur both early and late in the sequence, only early in the sequence, only late in the sequence, or not at all in each experimental condition. An initial comparison of the ERPs of the early deviants, when late deviants were not presented (i.e., early only condition) was made with the early deviants when late deviants were also presented (i.e., both condition). The ERPs of the early deviant were not affected by whether a later deviant was presented or not. A similar comparison was made for the late deviant. Again, the ANOVA revealed that the ERP to the late deviant was not affected by whether an early deviant was presented or not. For this reason, single trial ERPs to all early deviants and all late deviants were collapsed and sorted according to condition and scalp site. See Figure 31 and Table 4 in Appendix 2 for details of these comparisons.

Results

The data for one of the experimental condition had to be rejected for one subject as the subject affirmed that he failed to detect any deviant during the sustained condition. His EEG data was thus also rejected for the same condition. The subject’s performance on the second (switched attention) condition was however similar to others. These data were therefore used in subsequent analyses.

Performance Data. The reaction times are calculated from the onset of the first A tone of the triplet. It is important to note that the deviant B tone occurred 175 ms after this onset for the deviants. In the case of the early deviants occurring in the sustained task condition, the deviant detection occurred after the subjects had been doing the task for 2 seconds (4 triplets). For the late deviants occurring in the sustained task condition, the deviant detection occurred after the subjects had been doing the task for 9.5 seconds (19 triplets). In the case of the late deviants occurring in the switched task condition, their detection would have been made 3.5 seconds (7 triplets) after the task change.

Effect of Deviant Presentation Timing (Early, Late or Both). There was no significant difference in detection accuracy or RTs for early deviants presented in sequences where only an early deviant was present and for early deviants presented in sequences where both an early and a late deviant were presented, for both the sustained task condition and the switched task condition. The accuracies and RTs were therefore grouped. See Table 1 in Appendix 1 for the stimuli detection accuracies, Table 2 for the RTs and Table 3 for the correct rejections.

Attended Deviant Stimuli Performance. Subjects correctly detected 74.6% (SD = 20.0%, 95% CI: 58.3%-90.8%) of the early deviants presented in the sustained condition, 66.5% (SD = 29.6%, 95% CI: 43.7%-91.6%) of the late deviants presented in the sustained task condition and 80.4% (SD = 19.7%, 95% CI: 63,4%-94,65%) of the late deviants presented in the switched task condition. The differences in accuracy across conditions did follow the trends in the literature ($F(2,36)=3.82$, $p=.03$, $r^2=0.19$). The accuracy for the late deviant in the switched condition is significantly higher than the accuracy for the late deviant in the sustained condition (p=.03). The accuracy for the early deviant in the sustained condition is not significantly different from the accuracy for the late deviant in the sustained condition (p=.06) or from the accuracy for the late deviant in the switched condition (p=.30). The average RTs for the detection of early deviants in the sustained task condition was 729 ms (554 ms after B tone onset, SD = 99 ms, 95% CI: 658 ms-800 ms), 755 ms (580 ms after B tone onset, SD = 131 ms, 95% CI: 661 ms-849 ms) for the late deviants in the sustained task condition and 695 ms (520 ms after B tone onset, SD = 80 ms, 95% CI: 638 ms-752 ms) for the late deviants in the switched task condition. The three deviant detection RTs were not significantly different ($F(2,36)=2.97$, $p=.07$, $r^2=0.16$) but the trend is the same as with the accuracy. The response times to the late deviant in the switched condition (p=.06) and to the early deviant in the sustained condition (p=.07) are faster than for the late deviant in the sustained condition.

Non-deviant Stimuli Performance. Subjects correctly withheld a response on 99.6% (SD = 0.3%) of triplet presentations where a deviant was not present in the sustained task condition and on 99.7% (SD = 0.3%) of triplet presentations in the switched task condition. Subjects correctly withheld a response on 99.5% (SD = 0.9%) of the to-be-ignored early deviants during the first half of the switched task condition.

The detection accuracy of the noise burst detection in the switched task condition was very low (9.9%, SD = 9.2%). This may have been caused by the difficulty of discriminating between the approaching and departing noise bursts. If this were the case, the false alarm rates would have been high. Subjects responded on average to 48.8% (SD = 35.5%) of the noise bursts. The mean of the d’ for the task was -0.85 (SD=1.30). Thus, subjects made more false alarms than correct detections. The rate of responses does however provide evidence that the subjects were attending to the task and not thus attending to the opposite triplet channel.

ERP Results. In all subsequent figures, the average of the activity in the 0 to 50 ms period was used to compute the zero-voltage baseline from which components are measured. As the deviant is a time-shifted B tone, occurring in the middle of the triplet, this baseline represents the 50 ms period prior to the physical onset of the middle B tone of the triplet. In all Figures, the onset of the B tone is indicated by the vertical dotted line (i.e., at 50 m). As the deviant is time-shifted, the onset of the standard B tone occurs 75 ms after the offset of the A but 125 ms after the offset of the A tone for the deviant. In the Figures, 3 EEG channels are illustrated, Fz, Cz and Pz from top to bottom respectively. The ERPs are a "grand" average (average of all subjects' averages).

There was a significant difference between the four conditions (early in the sustained attention condition, early in the switched attention condition, late in the sustained attention condition and late in the switched attention condition) for the N2b (calculated between 250 and 350 ms) at Fz ($F(3,27)=7.35$, $p<.005$, $r^2=0.45$) and Cz ($F(3,27)=5.59$, $p<.005$, $r^2=0.38$). There was a significant difference between the four conditions for the P3b (calculated between 350 and 500 ms) at Pz ($F(3,27)=3.34$, $p=.03$, $r^2=0.27$). The inter-condition differences are presented later in this chapter. A comparison of the later positive deflection (calculated between 500 and 650 ms) at Pz yielded no significant difference ($F(3,27)=1.35$, $p=.28$, $r^2=0.14$). A comparison of the N1 (calculated between 100 and 200 ms) between the three attended deviants also found no significant difference at Fz ($F(2,18)=1.11$, $p=.35$, $r^2=0.11$) and Cz ($F(2,18)=0.53$, $p=.60$, $r^2=0.06$). There were no significant differences for the N2b at Cz and the P3b at Pz for early deviants presented in sequences where only an early deviant was present and for early deviants presented in sequences where both an early and a late deviant were presented for the sustained task condition. Similarly, there were also significant differences for the N2b at Cz and the P3b at Pz for the late deviants. The ERPs were therefore grouped according to their position in the sequence (early or late) and the task (sustained attention or switched attention). See Table 4 in Appendix 2 for the statistics.

Effects of Attention on the Processing of the Early Deviants. In the sustained attention condition, subjects attended to the triplet channel and ignored the channel containing the noise bursts. Evidence of attention to the triplet sequences was provided by the actual detection of the deviants. In the switched attention condition, subjects performed a detection task on the noise bursts for the first half of the sequence and were instructed to ignore the triplet channel containing the early deviant. Evidence of attention to the noise bursts was provided by the actual detection of the bursts.

The grand average ERPs following presentation of the correctly detected early deviant triplets when they were attended (in the sustained task condition) and when they were to be ignored (switched task condition) are illustrated in Figure 19. The deviant-standard difference waves can be found in Appendix 4.

In Figure 19, a centro-frontal positive deflection is apparent at about 200 ms (150 ms after the onset of the early deviant) at Fz and Cz following the to-be-ignored deviant in the switched task condition. This positivity is not apparent when the triplet channel was attended. This positivity is therefore possibly a P3a. The signal-to-noise ratio was large enough to permit its analysis in individual subjects. The absent P3a in the sustained attended condition was quantified at the same latency range (125 to 275 ms) as the P3a in the ignore condition. There was a significant difference at Fz ($t(9)= 2.95$, $p=.02$, $d=1.12$), Cz ($t(9)= 3.04$, $p=.01$, $d= 1.13$) and Pz ($t(9)= 3.69$, $p<.01$, 1.54) between the two early deviants.

A small negativity, the N2b, can be seen at Cz at about 300 ms (250 ms after the onset of the early deviant) for the sustained task condition. The amplitude of N2b for the attended early deviant at Cz was significantly larger than for the unattended early deviant ($p=.03$, $d=0.32$) but not at Fz ($p=.12$, $d=0.27$).

A late positivity, the P3b, was elicited by the deviants in the sustained task condition at Pz at about 350 to 600 ms (300 to 550 ms after the onset of the early deviant). The difference was not significant between the two early deviants for the P3b ($p=0.05$, $d=0.79$), but the trend shows a larger P3b for the attended deviant. Surprisingly, a large positivity from 200 to 600 ms is also apparent when the triplet sequence was to-be-ignored and attention directed to the detection of the intensity of the noise burst. The difference was not significant at Pz between the two early deviants for this late positivity ($p=0.49$, $d=0.36$).

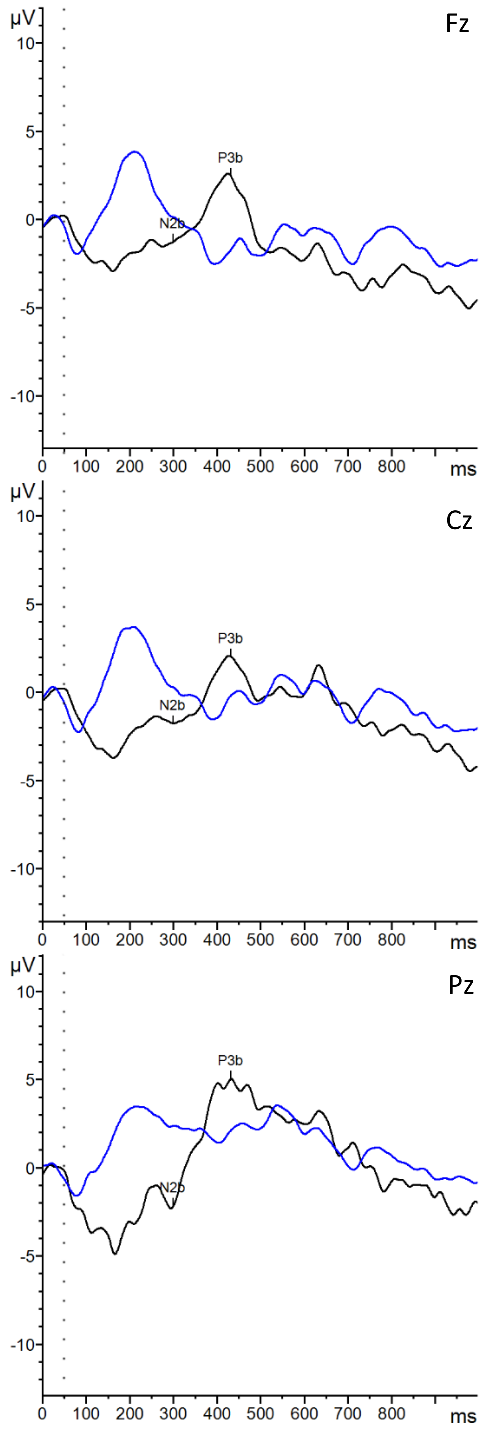


Figure 19. Grand averaged ERPs following presentation of the early deviants in the sustained and switched task conditions (Switched task condition - blue line and Sustained task condition - black line). The ERPs are time-locked to the physical onset of the deviant (i.e., onset of B tone 50 ms later than the expected onset, dotted vertical line). A very small negativity, the N2b, is apparent at about 250 ms for the attended deviant. A positivity, the P3a, is apparent at about 200 ms at Fz and Cz for the to-be-ignored deviant and a P3b is apparent at Pz from 350 to 550 ms for the attended deviant.

Building up of Streaming. The building up of streaming over time was examined by comparing the early and late deviants in the sustained task condition. Previous research has indicated that detection of the deviant is better accomplished early rather than later in an attended sequence of triplets. This was replicated in the current study.

Figure 20 illustrates the ERPs following the deviants presented early (in black) and late (in red) during the streaming sequence. The difference waves can be found in Figure 33 in Appendix 4.

A large N2b was apparent around 250 ms (200 ms after the onset of the deviant) for the late deviant and was much reduced in amplitude for the early deviant. The differences at Fz ($p=.01$, $d=0.77$) and Cz were significant ($p<.01$, $d=0.73$).

Following the detection of the early deviant, a large amplitude P3b is apparent at Pz, lasting from 350 to 600 ms (300 to 550 ms after the onset of the deviant). This P3b displayed a wide scalp distribution being clear also at frontal sites. This positivity was therefore measured at Fz, Cz and Pz as the mean of all data points between 350 and 450 ms. It was significantly larger for early than late deviants at all electrode sites, Fz ($p<.005$, $d=0.68$), Cz ($p<.005$, $d=0.75$) and Pz ($p=.02$, $d=0.58$).

A later positivity (a P3b) from 500 to 650 ms (450 to 600 ms after the onset of the deviant) at Pz was also apparent for both the early and late deviants. The difference between the early and late deviants was not significant ($p=.92$).

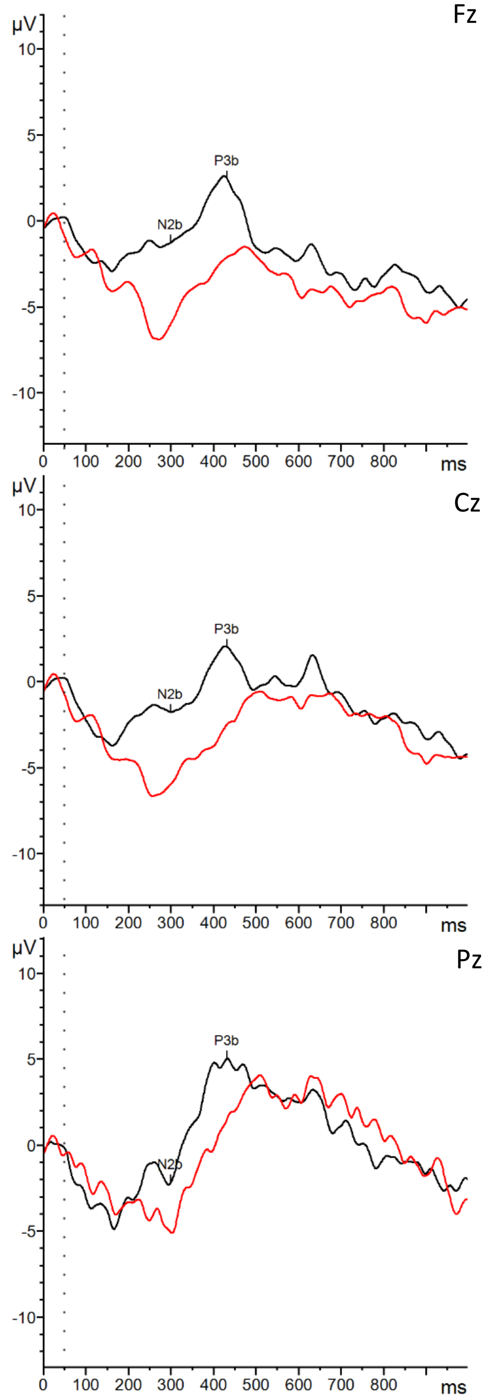


Figure 20. Grand averaged ERPs following the presentation of the deviants of the sustained task condition. (Early deviant - black line, Late deviant - red line). A negativity, the N2b, is apparent at about 250 ms for the late deviant. A positivity, the P3b, is apparent from 400 to 600 ms at Pz for both deviants

Effect of Task Switching on Streaming. The influence of task switching on deviant detection was examined using the late deviants in the switched task and compared to the sustained task. In the sustained task condition, the subjects had to attend to the triplet channel for the whole duration of the sequence while in the switched task condition, the subjects were to initially complete a task on the noise bursts and after 6 s subsequently attend to the triplet channel. It was expected that the detection of the late deviant would be easier after the subjects had just recently switched attention to the triplet channel (Thompson et al., 2011).

The grand average deflections following presentation of the late deviants in the sustained and switched condition are illustrated in Figure 21. The difference waves can be found in Figure 34 in Appendix 4.

A very distinct large (> 5μV) negativity was apparent at about 150 ms (about 100 ms after the onset of the deviant) at frontal and central sites for the late deviant in the switched task condition. This was an unexpected finding, which will be discussed in a later section. The differences between 100 and 200 ms at Fz (p=.49., d=.22) and Cz (p=.58, d=.21) were not significant between the two late deviants.

An N2b was apparent between 225 to 350 ms (175 to 300 ms after the onset of the deviant) for both the late deviant in the sustained task condition and the late deviant in the switched task condition. The amplitude of this N2b was significantly larger for the late deviant in the sustained task condition at Cz (p=.046, d=0.83). The difference at Fz was not significant (p=.06, d=0.85).

A large P3b was observed between 350 and 700 ms (300 and 650 ms after the onset of the deviant) for both late deviants. The difference at Pz between 350 and 500 ms was significant (p=.03, d=0.99). The difference at Pz between 500 and 650 ms was not significant p=.27).

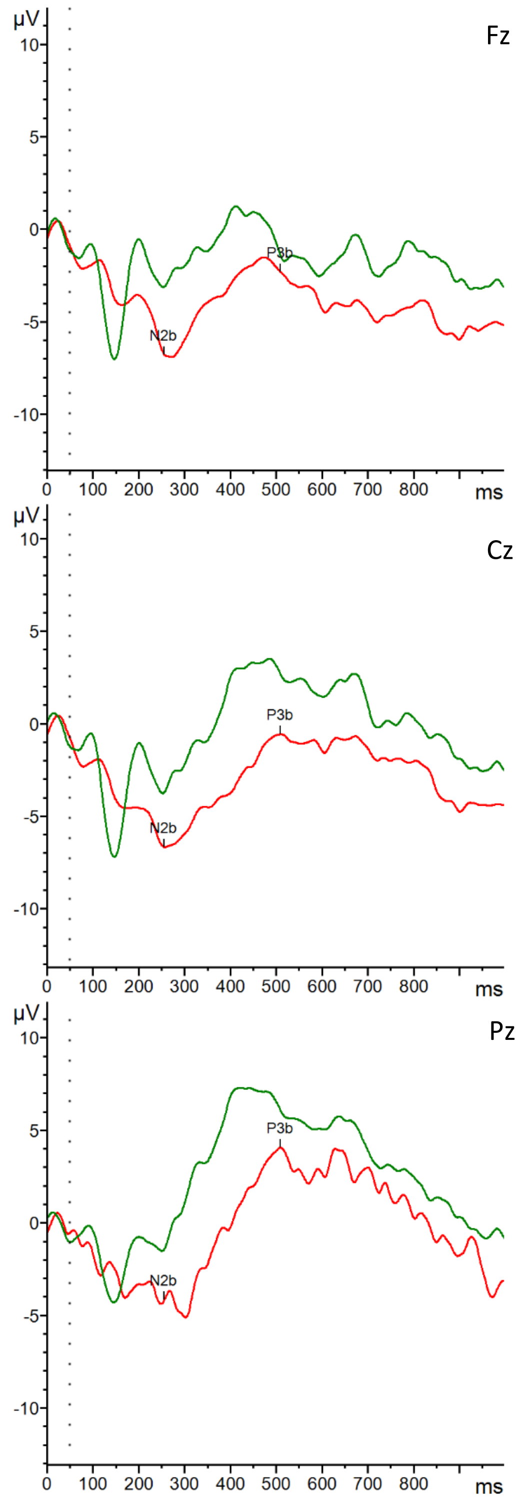


Figure 21. Grand averaged ERPs following presentation of the late deviants (sustained task – red line, switched task – green line). A negativity is apparent at about 150 ms at Fz and Cz for the late deviant in the switched condition. A later negativity, the N2b, is apparent at about 250 ms at Cz for the late deviant in both conditions. A positivity, the P3b, is apparent from 350 to 700 ms at Pz for both deviants.

Effect of Passive Build-Up on Streaming The effect of passive build-up on streaming was also examined by comparing the N2b and P3b elicited by the early deviants in the sustained task condition to those elicited by the late deviants in the switched task condition. In the switched condition, while the deviant physically occurred late in the sequence, it was experienced to occur early in the sequence. However, when the deviant occurred early in the sustained attention condition, it was presented after the 4^{th} triplet while in the switched attention condition, it physically occurred after the 7^{th} triplet.

An N2b was apparent between 225 to 300 ms (175 to 250 ms after the onset of the deviant) for the late deviant in the switched task condition while the N2b was very small for the early deviant in the sustained task. They were not significantly different in amplitude at Fz $p=.73$, $d=0.11$) and Cz ($p=.75$, $d=0.13$). A large P3b was observed between 350 and 700 ms (300 and 650 ms after the onset of the deviant) for both deviants (see Figure 22). The late deviant in the switched task condition was not significantly different from the early deviant in the sustained task condition at Pz between 350 and 500 ms ($p=.27$, $d=0.43$) and between 500 and 650 ms ($p=.32$). The difference waves can be found in Figure 35 in Appendix 4.

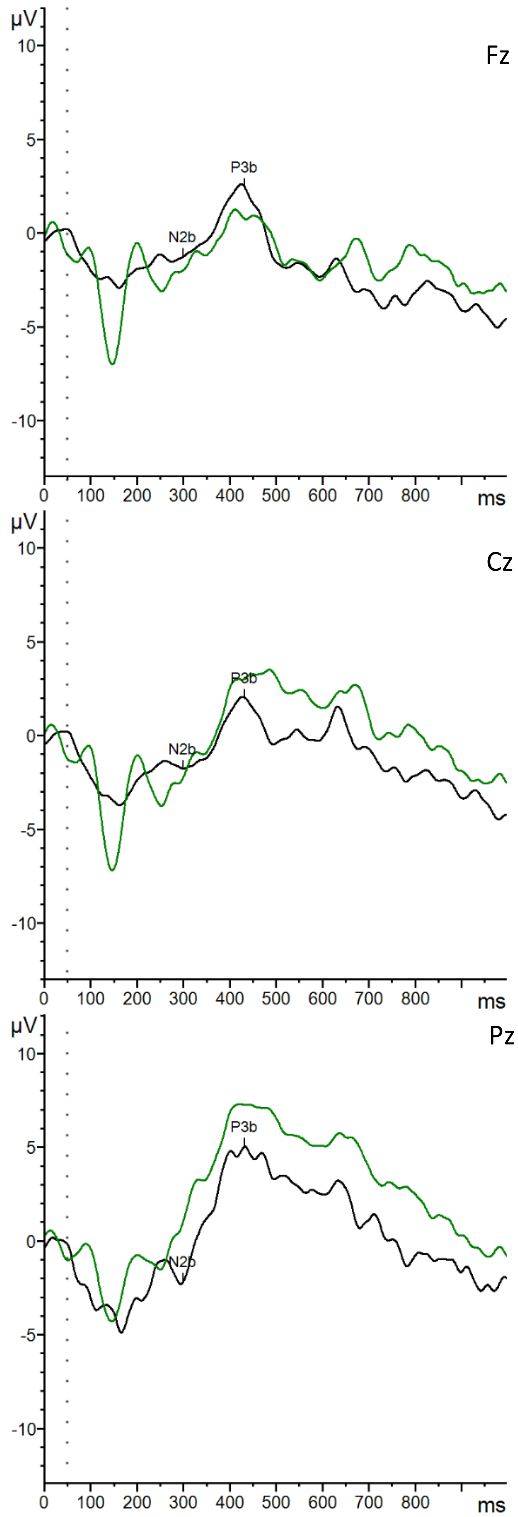


Figure 22. Grand averaged ERPs following presentation of early deviant in the sustained condition (black line) and late deviant in the switched condition (green line). An N2b is apparent at about 250 ms at Fz and Cz for the late deviant in the switched condition. A positivity, the P3b, is apparent from 350 to 700 ms at Pz for both deviants.

Ignored triplet channel. When subjects attended to the noise channel, the triplet sequence was to-be-ignored. Thus, processing of these triplets would thus have occurred passively. The early deviant should have elicited an MMN. The MMN is usually computed in a difference wave. In Figure 23, the red line represents the ERPs to the standard B tone while the blue line represents the ERPs to the deviant. The deviant-standard difference wave is illustrated as a black line. As may be observed, a negative deflection maximum in amplitude over front-central regions is apparent at about 100 ms. This could potentially be an MMN. However, it would then occur 50 ms after the physical onset of the deviant, much too early for a “true” MMN. Yet, it also occurs 100 ms after the expected time of onset of the B tone, and MMNs have been observed to occur at this latency. However, due to practical design considerations, only a limited number of deviants (100) were presented to individual subjects, and therefore the MMN was challenging to identify in the noisy waveforms. See Figure 31 in Appendix 3 for the individual difference waves between early to-be-ignored deviants and standards. Therefore, it was concluded that in this case, no MMN can be reliably detected. In contrast, the grand average is based on the average of almost 1000 deviant presentations, allowing the low amplitude possible MMN to emerge.

A later positivity is also visible in Figure 23 at about 200 ms. This has a centro-frontal scalp distribution and is thus consistent with a P3a, occurring to a deviant present in an unattended channel. This P3a was quantified in comparison to the attended early deviant and presented in an earlier section.

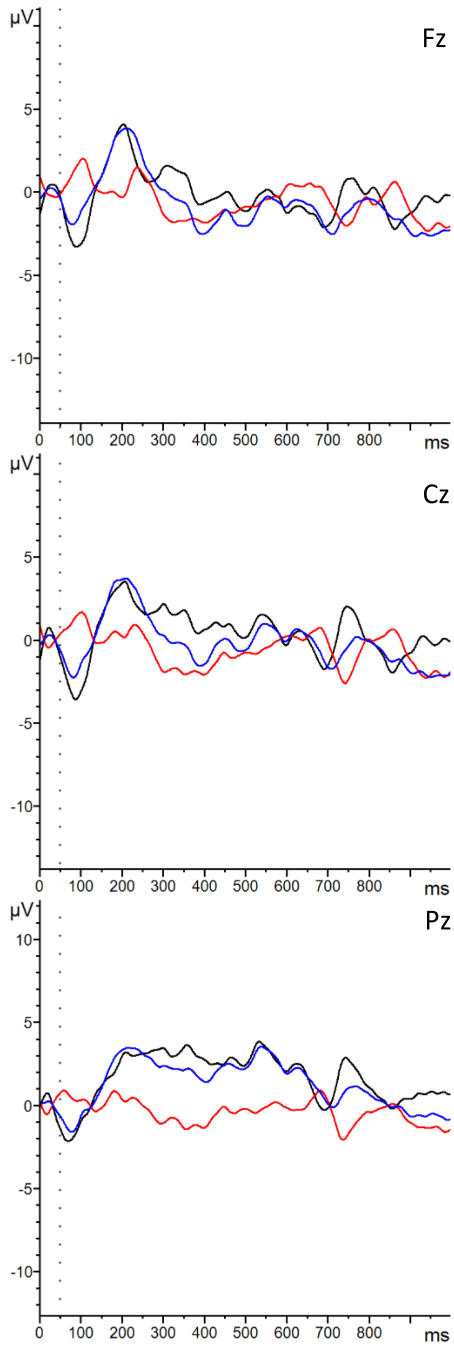


Figure 23. Grand average difference wave of the early ignored deviant in the switched task condition. Standard in red, early deviant in blue and deviant-standard difference wave in black. A negativity occurring at about 100 ms in the difference wave may potentially be a MMN. This is followed by a positivity at about 200 ms, which represents P3a activity.

Discussion

The results obtained for the behavioural performance did successfully replicate the results of Thompson et al. (2011) and previous studies (Carlyon et al., 2001; Cusack et al., 2004). There was a significant difference between the deviant detection accuracy results between the three attended deviant detection conditions while the difference between the deviant detection RTs was not statistically significant. This is probably because of the small sample size and high variability between subjects. In the sustained attention condition, the subjects detected the early deviants better than the late deviants. Similarly, when attention was switched to the triplet channel after having attended to the noise bursts, the late deviants were detected better than when the same late deviants occurred when attention had been sustained exclusively to the triplet channel. Thus, even though the late deviant physically occurred at the identical time, its ease of detection was dependent on how long attention had been focussed on the triplet channel. The RTs show a similar pattern.

The key outcome of this experiment is that the ERP results support the behavioural results. The late deviant in the sustained task condition was expected to be more difficult to detect after a build-up of streaming, and to show a larger N2b and smaller P3b than the other two conditions (early deviant in the sustained attention task, late deviant in the switched attention task). The results obtained supported those expectations. In the sustained attention condition, the N2b was indeed smaller and the P3b larger for the early than the late deviant. Similarly, the N2b was smaller and P3b larger to the late deviant was larger when attention was switched to the triplet channel compared to when it had been sustained to it.

The overall response rate (correct and incorrect decisions) to the noise task was 48% and the hit rate is very low. This suggests that the task was very difficult with subjects unable to reach a decision on more than half the trials. It is difficult to know why previous studies using a similar paradigm did find much better performance. It is possible that subjects did not sustain attention to the noise channel. Rather they may have sampled ("eavesdropped" on) the triplet channel. The ERP data do not support this hypothesis, however. Had subjects overtly attended the triplet channel, a large N2b and P3b would have been expected to be elicited by the deviant stimuli in the triplet channel. A short latency positivity was nevertheless observed to the deviant in the triplet channel. Its short latency and more anterior scalp distribution are more consistent with a P3a than a P3b. A P3a can be passively elicited by an unattended deviant if the extent of deviance is large.

The small N2b deflection found for the early deviant in the sustained task condition may have been a result of a temporal and spatial summation with the later P3b positivity. The N2b and P3b occurred at a somewhat similar time and had a similar central scalp distribution.

The large negative deflection at Fz and Cz for the late deviant in the switched attention condition might reflect an object-related negativity (ORN). The ORN component has been shown to follow the listener's perception of two concurrent sounds (Alain, Arnott & Picton, 2001). If this is the case, it should also have been elicited by the late deviant in the sustained attention condition, when the subject perceived two separate auditory streams. This negativity was not apparent to the late deviants in this condition.

The P3b elicited by the deviants presented itself as a long deflection, often showing multiple peaks. In this study, the deviant was created by delaying the presentation of the B tone of the triplet. Subjects were asked to determine when the onset of this B tone was delayed. It is possible that subjects made a double decision, an initial decision being made when the expected B tone was not presented at its expected onset and a second decision when it was presented. By observing the figures (see Figure 20 to 22), it can be seen that the first P3b deflection is larger than the second one. This might indicate that the decision was made on the absence of the deviant, only to be confirmed by the subsequent presence of said deviant.

Time on task was examined in two ways. When attention was sustained, processing of the early and late deviants varied. Time on task was also compared in a second way. When attention was initially directed to the noise burst channel and subsequently switched to the triplet channel, the detection of the "late" deviant occurred only after attention had been directed to the triplet channel for a brief period of time. Thus, while the deviant physically occurred late in the sequence, psychologically it occurred "early". Thus, processing of the deviant that physically occurred early when attention was directed to the triplet channel was compared to that of the "late" deviant when attention was switched to the triplet channel. Accuracy of detection of the early in the sustained condition and of the late deviant in the switched condition did not significantly differ, which is also the results obtained regarding the N2b and P3b of those two conditions.

Overall, building on the work of Thompson et al (2011), this study was the first example of this type of experiment to be carried out with ERP measures, making it an interesting contribution to the literature. The behavioural results replicated other studies in the literature, and significantly, the behavioural results obtained were supported with ERP results.

Experiment 2. Influence of the time spent on task and the presence of a continuity illusion on stream segregation as tested through deviant detection in noise-overlapped triplets

The second experiment reported in this chapter investigates the effect of the time spent on task and the presence of a continuity illusion on streaming, using a modification of the methodology used by Thompson et al. (2011), which was previously shown to induce streaming.

Several factors influence streaming. Streaming is often studied through an ABA paradigm that uses sequences of alternating A and B tones, where A and B are tones of different frequencies, presented following an ABA-ABA-ABA-ABA… pattern. Van Noorden (1975) and Bregman et al. (2000; Bregman, 1990) showed that speed of presentation influence streaming. They found that when alternating A and B tones are presented very rapidly, subjects perceive two different streams, one of A tones and one of B tones. In contrast, when the same sequence is presented more slowly, subjects report a single stream, composed of alternating A and B tones. Bregman (1990) also tested the influence of the frequency separation between the A and B tones. He found that when separation between frequency A and B is small (0.5 semitones), listeners report hearing one auditory stream. When this separation is large (>10 semitones), listeners report hearing two auditory streams. At intermediate frequency separations (such as 3–7 semitones), the perception of ‘one stream’ and ‘two streams’ alternates over time, creating a ‘bistable’ percept. This effect of frequency was found by other authors (Carlyon et al., 2001; Dannenbring & Bregman,1976; Thompson et al., 2011; Van Noorden, 1975).

Stream segregation also progresses with time. A single stream is perceived early in a sequence of triplets and two streams are perceived later in the sequence (Anstis & Saida, 1985).

This effect the “bistable percept”, as with increased listening time, a stable two-stream percept may develop (Bizley & Cohen, 2013).

Carlyon et al. (2001) developed a paradigm to test the influence of attention on streaming. They presented sequences of ABA triplets in one ear and a series of long duration noise bursts in the other ear. In one condition, subjects were asked to focus their attention on the channel containing the triplets and to ignore the channel containing the noise. They pressed one button when they heard 1 stream and another when they heard 2 streams. In a second condition, the subjects were asked to focus their attention noise burst stimuli for 10 s and determine whether they were increasing or decreasing in intensity. The noise bursts ended after 10 s and subjects were then asked to switch attention to the opposite ear where the ABA triplets were presented. Because stream segregation builds over time, the experiment aimed to see whether the 10 s of attending to noise burst had aided with stream segregation in the triplet channel or not. The authors found that attention is necessary for streaming to build up. When subjects ignored the ABA triplets sequence while carrying out the task on noise bursts, subsequent streaming judgments in the ABA triplets sequence revealed less stream segregation of the previously ignored tones than they did when subjects had been attending to them.

Streaming has also been shown to influence deviant detection. Deviant detection is often tested through oddball paradigms, where a stream of frequently-occurring and repeating homogenous “standard” sounds are presented. A rare (or “odd”) a feature of the standard is changed to form a “deviant”. Studies have found that deviants are detected more accurately when a single auditory stream is perceived rather than two (Bregman & Dannenbring, 1973; Roberts, Glasberg & Moore, 2008; Vliegen, Moore & Oxenham, 1999; Vliegen & Oxenham, 1999).

One issue with the study by Carlyon et al. (2001) is that the determination of streaming is based on subjects' judgements of what they were experiencing. However, such judgments are highly subjective. Thompson et al. (2011) developed a more objective method to monitor streaming, based on the effect of steaming on deviant detection. They presented a "deviant" within the ABA triplet sequence, either early in the sequence or late. This deviant was created by delaying the onset of the B tone in an ABA triplet by 50 ms. Subjects were asked to press a button upon detection of the deviant. This deviant should have been easier to detect when the subject is experiencing the sequence as a single stream, at the beginning of the sequence. Two different frequency separations between the A and B tones, 4 and 8 semitones were also used. Thompson at al.'s (2011) results confirmed this deviant detection hypothesis. Detection of the deviant was better when the deviant occurred early rather than late in the sequence for the large frequency separation. In a second experiment, the time spent attending to the triplet sequence was modified in a different way. The paradigm was similar to that used by Carlyon et al. (2001). The ABA triplet sequence used previously was presented in one ear. In the opposite ear, noise bursts were presented. When the subjects attended to the triplet sequence, the deviant could occur either early or late. When they attended to the noise bursts to determine whether their intensity was increasing or decreasing, the triplets were to be ignored. However, the presentation of the noise bursts ended at the mid-point of the presentation of the triplets, when attention was then to be switched to the triplet channel. In a small frequency separation condition, when subjects attended the triplet channel, accuracy of detection of the early and late deviants did not vary. Time-on-task did influence the results when the frequency separation between A and B tones was wider. When the triplet channel was attended, the early deviant was detected more accurately than the late deviant. Importantly, when attention was only switched to the triplet channel after first attending to the noise channel, then detection of the late deviant was better than the detection of the late deviant in the condition attended throughout.

Streaming is the main form of auditory scene organisation presented in this thesis. As presented in the General Introduction and investigated in Bregman et al. (1999), the same Gestalt principles that guide initial steaming also guide the perceptual effect called the continuity illusion. When a silent period (a gap) is introduced into a sound, the observer perceives two shorter duration sounds. If noise is presented during the gap, the sound is now perceived as being continuous, creating an illusion of continuity (Miller & Licklider, 1950).

Riecke, Van Opstal, Goebel and Formisano (2008) studied the relation between perceived continuity, and the duration, loudness difference and noise frequency spreads of the interrupting noise for interrupted and non-interrupted tones at different frequencies. Subjects had to report their percept of continuity going from the tone being most likely being continuous to most likely being discontinuous. The authors found that shorter noises, less intense noises and noises closest in frequency to the tone resulted in a higher likeliness of a continuity percept.

Micheyl, Carlyon, Shtyrov, Hauk, Dodson and Pullvermuller (2003) studied ERPs during a continuity illusion study to discover if attention influences the continuity illusion. They employed an oddball paradigm in which stimuli had to be ignored. The two standards were either a continuous tone or a tone interrupted by a silent gap. Both were followed by a noise burst occurring at a pseudorandomly selected delay after the onset of the standard. The two deviants were identical to the standards, but a noise burst was presented either overlapping the middle of the tone or during the silent middle gap. The noise bursts could either be in a frequency region that encompassed the tone, creating a continuity illusion, or be in a frequency region that did not encompass the tone and thus did not create a frequency illusion. There were therefore four conditions: continuous standards and the gap in the deviants filled with noise creating the perception of a continuous tone, continuous standards and the gap in the deviants filled with noise creating the perception of a discontinuous tone, discontinuous standards and the gap in the deviants filled with noise creating the perception of a continuous tone, and discontinuous standards and the gap in the deviants filled with noise creating the perception of a discontinuous tone. The MMN was larger in when the deviant was perceived as continuous (when noise that created a continuity illusion was presented in the gap) and the standard as interrupted than when both standard and deviant stimuli were perceived as interrupted. The MMN was also larger when the standards were perceived as continuous and the deviants as interrupted (when noise that did not create a continuity illusion was presented in the gap) than when both standard and deviant stimuli were perceived as continuous.

Kobayashi and Kashino (2012) investigated the effect of the presentation of a sequence of brief flanking sounds (tone pips) on the percept of continuity of a target tone presented in alternation with bandpass noise in a target sound sequence. They varied the frequency separation between the target tone and the flanking tone, as well as stimulus onset asynchronies between the target tones, bandpass noise and flanking sounds. They found that when the flanking sound was synchronised with the noise disruptor (0 ms asynchrony), the apparent continuity was significantly enhanced, independently of the frequency difference between the target tone and flanking sound. They posited that the flanking sound, together with the sudden onsets of the noise disrupters, captures attention that would otherwise have been directed to the target sounds and enhanced apparent continuity for the target sounds.

Several studies showed that the continuity illusion can improve performance in tasks. Carlyon, Deeks, Norris and Butterfield (2002) conducted a study where subjects had to identify vowels composed of two formants. When the two formants were presented as pulses and not at the same time, performance was very poor. When bursts of low-frequency (50–624 Hz) noise that made the formant sound continuous were presented in the silent gaps between each instance of the first formant, and bursts of high-frequency (624–4000 Hz) noise in the gaps between instances of the second formant, performance was improved. The presence of the continuity illusion also helped in a frequency discrimination task (Plack & White, 2000). In this study, the authors tested fundamental frequency discrimination for single-burst tones and for stimuli containing two tone bursts separated by gaps. The performance in the gapped condition was worse than that for the single-burst condition with the same cumulative duration (e.g. two 20-ms bursts separated by 8 ms compared to one 40-ms burst). Adding a bandpass noise in the same frequency band as the tone in the gap between the two-tone bursts improved performance to the level of the single-burst condition.

The interaction between streaming and the continuity illusion has been the subject of very little research. Tougas and Bregman (1990) used stimuli that were gliding tones (either increasing gradually in frequency or the reverse) meeting at the mid time point. The tones could either be perceived as auditory streams that “cross” (with the first tone descending for the whole duration and the second tone ascending for the whole duration) or “bounce” (with the first tone descending then ascending and the second tone ascending then descending). The frequency range was between 400 and 1600 Hz. There were 4 conditions. The all-pure condition consisted of two pure-tone glides. The all-rich condition consisted of two glides, one ascending and the other descending, each containing a set of three harmonics. The rich-crossing condition consisted of an ascending rich glide and a descending pure-tone glide. The rich-bouncing condition consisted of pure-tone glides forming the upper half of the X and rich glides forming the lower half. A 200 ms interval, centered at the point at which the glides crossed, was either left intact (continuous interval), replaced by a noise burst (noise interval) or replaced by silence (silence interval). The noise was bandpass filtered between 200-2400 Hz. The subjects had to evaluate on a 7-point scale the clarity of isolation of the standard sound when heard as part of the four conditions. The authors found that the continuous and noise intervals yielded essentially the same stream formation and responded in the same way to the added harmonics. In the presence of a continuity illusion, streaming occurred as if the tones had been continuous. The authors thus posit that auditory streams were formed through heuristics for grouping and then the presence of acoustic information which suggested that part of the two streams had been masked activated a perceptual continuity illusion mechanism which acted to fill in the gaps.

Bregman, Colantonio and Ahad (1999) did a study to evaluate if the continuity illusion and streaming are based on the same Gestalt principles. High-frequency narrow-band noise (NBN) bursts were presented in alternation with a low-frequency NBN bursts. Noise bursts of wide-band noise (WBN) or silent gaps were presented between them. The high and low noise bursts could be segregated through streaming and the NBN could be heard as a continuous tone caused by the presence of noise between two NBNs, creating a continuity illusion. The authors manipulated three variables: the rate of presentation of the sequence, the separation of the center frequencies of the high and low NBNs, and the bandwidths of the high and low NBNs. There were two tasks. In the first, subjects had to rate their perception of the continuity illusion and in the second they had to give their perception of the number of streams heard. In this streaming perception task, there could either be silent gaps between NBN or noise presented between them. The authors found NBN stimuli respond to the same variables as tonal stimuli, with greater frequency separation between NBN noises and higher rate of presentation leading to greater streaming. The manipulations of temporal and frequency parameters had similar effects on all three tasks: the continuity illusion, stream segregation with WBN separating the NBN bursts and stream segregation with silence separating the NBN bursts. The authors posit that both stream segregation and the continuity illusion possibly share the process of auditory object creation that binds segments of the auditory scene.

Current study

Experiment 1, presented previously in this chapter, examined how the time spent on task influenced deviant detection in a sequence of ABA triplets perceived as a single stream or as two streams. Confirming the work of previous studies with additional objective measures, the results of the study show that a sequence is perceived as only one stream at the beginning and as two streams later in the sequence. The results of this study also confirm that deviant detection is easier when the ABA triplets are perceived as one stream, in the beginning of the sequence.

Tougas and Bregman (1990) showed that there was no difference in streaming between conditions where a continuity illusion was present and when the tones were continuous. The gliding tones they used as stimuli were presented concurrently, and thus the streaming occurred between tones presented simultaneously. Bregman et al. (1999) obtained streaming with sequentially presented frequency differing noise burst stimuli. The ABA tones used in Experiment 1 are presented sequentially. Given the very limited research conducted in this area, it is thus of great interest to investigate if the continuity illusion will have the same effect on streaming with such stimuli. As previous research also showed that the continuity illusion often leads to similar performance on task as continuous tones, it is of interest to investigate if the presence of such an illusion will hinder or help the detection in the context of a deviant detection task in a paradigm that has been shown to induce streaming (the ABA tones sequence).

Tougas and Bregman (1990) used a broadband noise (200-2400 Hz frequency range) and obtained a continuity illusion regardless of the frequency of the tone. In the current study, presenting noise in a gap presented in the center of the B tone of a triplet should induce a strong continuity illusion if the noise is bandpass filtered around the frequency of the B tone, and not if the noise is bandpass filtered around the frequency of the A tone, as a noise of similar frequency to a tone presented to fill a gap creates a stronger continuity illusion (Micheyl et al. 2003). A larger frequency difference between the A and B tones in the triplet was shown to induce faster streaming. It is hypothesised that a B frequency bandpass filtered noise would support streaming and an A frequency bandpass filtered noise would either create no streaming or hinder it.

Bregman et al. (1999) showed that the continuity illusion is impacted by the same variables as streaming. Tougas and Bregman (1990) postulated that auditory streams were formed then information suggesting that part of the two streams had been masked activated a continuity illusion. Several studies (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Thompson, Carlyon & Cusack, 2011) have shown a dissociation between a pre-attentive, low-level sound segregation process, and a process modulated by the time spent attending to stimuli that can be called into play in forming perceptual objects and streams. This study thus endeavors to identify the interaction between the initial pre-attentive streaming, the continuity illusion and the streaming created by the time spent attending to a task.

In this new experiment, the interaction of the continuity illusion and streaming on deviant detection is examined. Both the streaming process and the continuity illusion follow the same initial Gestalt rules for grouping attributes into an auditory object (Bregman, Colantonio & Ahad, 1999). This study aims to investigate if a continuity illusion within a tone will induce streaming as if the physical characteristics of the tone had been continuous, and how it will influence deviant detection. This experiment was designed as a modification of the Thompson et al. (2011) experimental paradigm that was used in Experiment 1. ERPs were again recorded to provide a measure of the extent of processing of the various stimuli. The experiment again employed the ABA triplet sequence, with noise being presented during the central portion of the B tones. The noise frequency was filtered in different conditions. In one condition, it overlapped with the 635 Hz B tone, creating the ‘’B’’ frequency bandpass filtered noise, and in the other, it overlapped with the 400 Hz A tone, creating the ‘’A’’ frequency bandpass filtered noise. The B tone was occasionally changed to form a deviant. In this study, the deviant was created by inserting a silent period (a “gap”) into the center of the B tone. The standards are therefore continuous B tones with noise (either “A” or “B” frequency bandpass filtered) overlapping the central part of the tone and the deviants are interrupted B tones with noise (either “A” or “B” frequency bandpass filtered) filling the gap. Again, a deviant could occur early or late in the sequence.

In the absence of a continuity illusion, an early deviant is easier to detect in a stream of ABA tones than a late deviant (Billig & Carlyon, 2016; Carlyon, Cusack, Foxton & Robertson, 2001; Cusack, Deeks, Aikman & Carlyon, 2004; Thompson, Carlyon & Cusack, 2011, previous study). A noise that has a frequency bandpass filtered around the A tone will not create a continuity illusion when presented in the gap in the deviant B tone, as the A noise does not overlap the frequency of the B tone. The resultant deviant B tone will therefore be perceived as interrupted. This should delay streaming as the absence of a continuity illusion will prevent the building of streaming into a steam of A tones and a stream of B tones, remaining a single stream of ABA triplets. For such a percept, deviant detection should be easier for a late deviant as the ABA pattern will be strengthened with time. When presented in the gap in the deviant B tone, the B frequency bandpass filtered noise should induce a continuity illusion, making the deviant interrupted B tone being perceived as continuous. This will allow streaming to happen, as in Tougas and Bregman (1990). In this condition, similar results as in Experiment 1 are expected, with the deviants presented early in the sequence being easier to detect than the late deviants. Deviant detection is expected to be easier for an A frequency bandpass filtered noise than a B frequency bandpass filtered noise presented at the same position, as the continuity illusion creates the percept of a continuous B tone (Micheyl et al., 2003), making deviant detection harder.

Methods

Subjects

Eleven subjects (five female, six male) between the ages of 18 and 24 years volunteered to participate in the study. None of the subjects participated in Experiment 1. None reported a history of neurological or psychiatric disorder and all reported normal hearing. Subjects signed a consent form and received an honorarium for participation in this study. The study was conducted according to the guidelines of the Canadian Tri-Council (Medical, Natural and Social Sciences) on ethical conduct involving human subjects.

Procedure

Different auditory sequences were presented to the left and right ears through Sony MDR-V6 headphones while the subject sat in a sound-attenuated room. The instructions were given orally and then displayed on a computer screen at the beginning of the experiment. During the experiment the instructions were replaced by a static image of a landscape so not to provide any visual distraction and the subjects were asked to attend to the auditory stimuli presented in one ear at a time.

The auditory sequence in the left ear consisted of pure tones grouped in an ABA triplet pattern, with bandpass filtered noise being presented in the middle of the B tone. The A stimulus was a 400 Hz pure tone having a 150 ms duration. The B stimulus was a 635 Hz pure tone with a duration of 400 ms. In this study, there was a constraint on the duration of the triplets as they needed to be presented rapidly to induce a streaming effect (Van Noorden, 1975; Bregman, 1990), but the noise duration had to be long enough to create a continuity illusion. Bregman (1990) and Warren, Obusek, Farmer and Warren (1969) used 200 ms duration tones and obtained streaming. The tonal separation represented an 8-semitone difference, a separation that was shown to produce a clear streaming effect (Carlyon et al., 2001; Thompson et al., 2011). The A and B tones were presented at an intensity of 55 dB SPL.

In order to create the continuity illusion, a silent period (a gap) must be introduced in a stimulus and this gap must be filled with noise. In some trials, a change (a deviant) from the standard triplet was created by inserting a gap in the centre of the B tone, lasting 200 ms, which coincides with the noise duration and onset. The noise burst then “filled” this gap. In the standard B tone, a gap was not introduced into the B tone. The noise burst thus overlapped (or was “mixed”) with the continuous B tone.

Filtered noise with a mean intensity of 50 dB SPL was presented over the central part of the B tone for a duration of 200 ms. A lower intensity was chosen for the noise than for the tone as Riecke and al. (2007) have shown that a lower intensity noise level leads to a stronger continuity illusion. The duration of the noise was chosen as the shortest shown to induce a continuity illusion: 200 ms (Riecke et al., 2007). This noise duration was also used in Tougas and Bregman (1990) study of streaming and the continuity illusion. The noise was filtered in different conditions so that in one condition, its frequency bandpass included the 635 Hz B tone while in the second condition, its frequency bandpass included the 400 Hz A tone. Thus, the white noise bandpass filter was set between 300 and 500 Hz, creating the ‘’A’’ bandpass filtered noise, or between 500 and 700 Hz, creating the ‘’B’’ bandpass filtered noise respectively. See Figure 24 for illustration of the standard and deviant triplets, overlapped by A or B noise.

The experience of streaming is also affected by the time between triplets. To minimise this time (300 ms offset to onset), the delay between the ABA tones was 0 ms. Each sequence contained 12 triplets and therefore had a total duration of 12 seconds. The next sequence began 1 second later. Sequences contained deviant triplets either in the 5th triplet (Early deviant condition, deviant onset 4.25 s after the start of the sequence) or in the 10th triplet (Late deviant condition, deviant onset 9.25 s after the start of the sequence). Only those two positions in the sequence were used as no effect of having both early and late deviants or only early deviants (and the same for late deviants) in the sequence had been found in Experiment 1, findings similar to those of similar to Thompson at al. (2011).

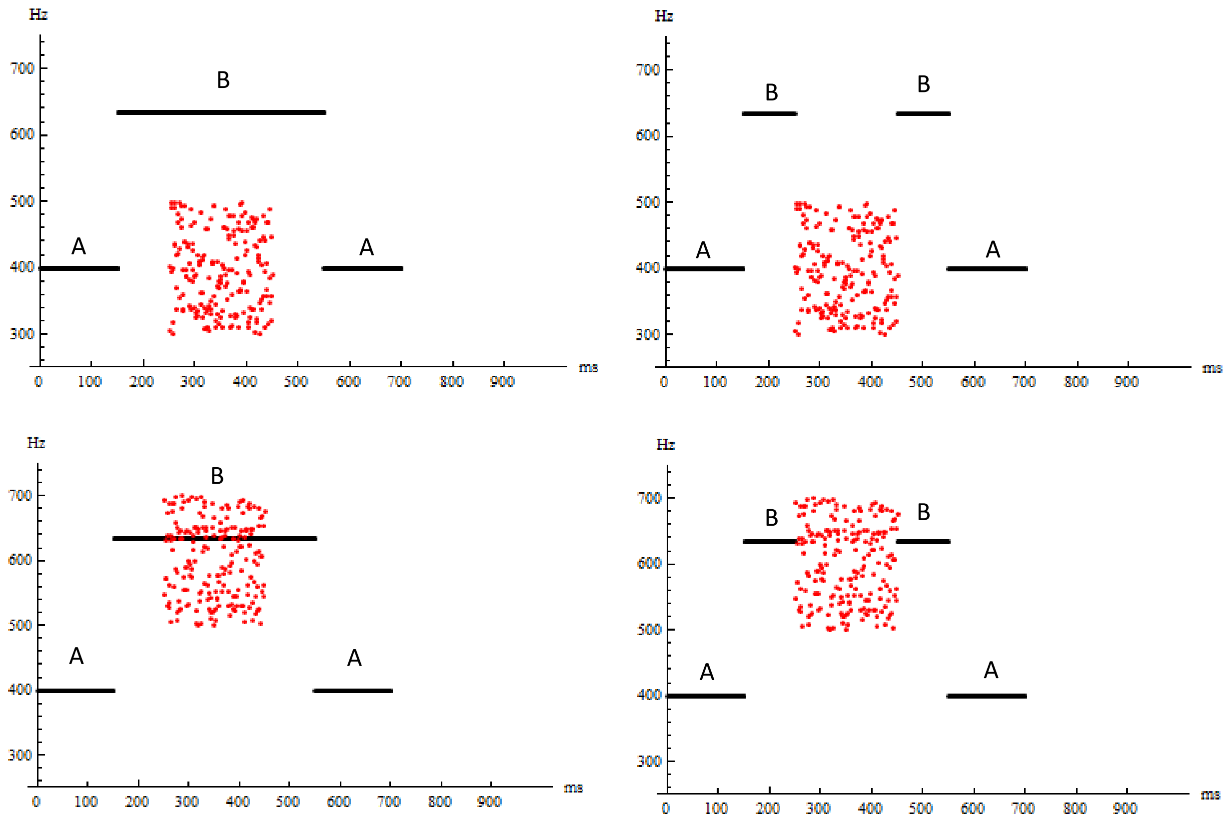


Figure 24. Triplets stimuli used in Experiment 2. Upper Left. Standard triplet, A bandpass filtered noise presented in during continuous B tone. Upper Right. Deviant triplet, A bandpass filtered noise presented in the gap of interrupted B tone. Lower Left. Standard triplet, B bandpass filtered noise presented in during continuous B tone. Lower Right. Deviant triplet, B bandpass filtered noise presented in the gap of interrupted B tone. The pure tones are in black and the filtered noise in red

Auditory stimuli were also presented in the opposite right ear. This consisted of a sequence of noise bursts centered around 2500 Hz (white noise filtered between 2000 and 3000 Hz). The duration of the noise burst was 400 ms. They either increased linearly in intensity from 0 to 90 dB SPL or decreased from 90 to 0 dB SPL over the 400 ms duration, giving the perception of “approaching” and “retreating” sounds respectively. The noise bursts were presented at random intervals varying from 250 to 1000 ms after the offset of the previous noise burst. They could thus overlap with the presentation of the triplets in the opposite ear. The number of noise bursts that were presented varied depending on the experimental condition. A total of 12 noise bursts were presented in a 12 seconds sequence for the sustained attention experimental condition and 6 noise bursts were presented during a 6 seconds sequence for the switched attention experimental condition.

Two experimental ("task") conditions were presented. In the sustained attention condition, subjects attended to the triplet sequences occurring in the left ear for the entire 12 s duration of the triplets. They were asked to signal their detection of a deviant triplet by pressing the spacebar as quickly and accurately as possible. The deviants could occur either early or late. The noise bursts presented in the right ear were to-be-ignored.

In the switched attention condition, noise bursts were presented for a 6 seconds period. Subjects were asked to attend to the noise bursts and to signal the detection of the "approaching" or "retreating" noise by pressing one of two keys on the keyboard. The triplets in the left ear were to-be-ignored during this period. Note that an early occurring deviant triplet pattern should not have been detected actively. When the noise bursts ceased (after 6 seconds), subjects were told to switch attention to the left ear to the channel containing the triplet sequence to now actively detect the deviant pattern. The experimental conditions were counterbalanced, with half the subjects performing the sustained attention condition first and half the subjects performing the switched attention condition first. This followed the experimental design presented in the previous experiment, as this had been found to be a good methodology, as well as being used by other researchers in the literature.

In each task condition, a total of 2400 triplets were therefore presented, consisting of 2200 standard triplets and 200 deviant triplets. 100 deviant triplets occurred early in a sequence (in the 5th triplet, the Early condition). Of those, 50 presented an "A" noise and 50 presented a "B" noise. 100 deviant triplets occurred late in a sequence (in the 10th triplet, the Late condition). Of those, 50 presented an “A” noise and 50 presented a “B” noise. The order of the sequences was pseudo-randomised so all sequences would be presented the same amount of times

The experiment had a 2x2x2 design, with deviants being presented either early or late in the sequence, the noise being filtered to have a bandpass encompassing the frequency of the A tone or the B tone and attention either being held on the triplet sequence for its entire duration or initially held on the noise bursts and then switched to the triplet sequence.

Quantification of Behavioural Performance. Performance was measured by deviant detection and noise bursts (‘‘hits’’) accuracy, and correct rejections. Reaction times (RT) were also measured. A response had to be made within 1500 ms after stimulus onset of the first A tone of the triplet (1250 ms after the deviant B onset) to be considered a correct detection. Performance data (i.e., accuracy, RT) was analysed using a 3x2 mixed ANOVA with repeated measures for the three different conditions in which deviant triplets had to be actively detected (Early and Late deviants in the sustained attention condition and Late deviants in the switched attention condition) and 2 groups according to noise frequency. A Greenhouse-Geisser correction was used when sphericity was violated. Subsequent comparisons were performed between the three deviant conditions. Accuracy and RTs were not evaluated for the early deviants in the switched attention condition as it was to-be-ignored and no response was required.

Physiological Recording

The EEG was recorded from 32 scalp sites representing frontal (Fp1, Fp2, AFz, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6), central (C3, Cz, C4, C8, CP5, CP1, CP2, CP8), parietal

(P7, P3, Pz, P4, P8), temporal (T7 and T8) and occipital (Oz, O1, O2) areas of the scalp using active silver/silver chloride electrodes attached to an electrode cap (Brain Products, GmbH, Munich, Germany). Two additional channels were recorded from individual electrodes placed on the left and right mastoids (TP9 and TP10). The nose served as a reference. Eye movements and blinks were recorded from an electrode placed at the infra orbital ridge of the left eye. The ground electrode was located between the Fz and FPz sites. Inter-electrode impedances were kept below 50 kΩ. The physiological signals were digitised continuously at a 500 Hz sampling rate and stored on hard disk.

The continuous EEG data was later digitally filtered using a bandpass of 0.05–20 Hz with a notch filter at 60 Hz to filter the fluorescent lights. Vertical EOG activity was later computed by subtracting data recorded at FP1 from that at the infra-orbital ridge. Horizontal EOG activity was later computed by subtracting data recorded at FT9 from that recorded at FT10. Independent Component Analysis (Makeig, Bell, Jung & Sejnowski, 1996) was then used to identify eye movements and blinks that were statistically independent of the EEG activity. This algorithm was trained to recognise each subject’s blink signature, and this factor was then partialled out of the EEG traces. The EEG was visually inspected for channels containing high levels of noise and segments containing too much noise were rejected from further analysis.

The subject’s continuous EEG data were then segmented into 1000 ms epochs, time-locked to the onset of the first A tone of the triplets, which included a 50 ms pre-stimulus baseline period and a 950 ms post-stimulus period. The single trial segments were subsequently baseline corrected. Segments containing EEG activity greater than ±100 μV were rejected from further analysis. Only segments for correct detections of deviant triplets were analysed further. The segments were then sorted and averaged for electrode site (Fz, Cz and Pz) and experimental condition (sustained attention, switched attention), Stimulus (Standard, Deviant), Frequency of Noise (A frequency bandpass filtered, B frequency bandpass filtered) and position in the sequence (Early, Late).

Quantification of the N2b and P3b Components

The average of all data points in the 50 ms period before the presentation of the first A tone of the triplet served as a zero-amplitude baseline from which all amplitudes were measured. When the triplet channel was to-be-ignored, an MMN should still have been elicited by the deviant. As discussed in the results section of the previous experiment, the amplitude of the MMN is usually very small requiring a large number of deviants to be presented. Only 100 early to-be-ignored deviants were presented in a single condition, meaning that MNN was thus very difficult to observe in the background noise of individual subjects. However, as the focus of this experiment was on the attended deviants, the MMN was therefore not quantified.

When the triplet channel was attended, an N2b and P3b were elicited. They can be measured through a difference wave computed by subtracting the standard from the deviant waveforms. This difference wave removes processing that is common to both the standard and deviant leaving only processing that is unique to the deviant.

Grand averaged difference waves between the deviants and standards were used to identify the N2b and P3b in each condition. The standards used in the comparison were the 100 standards presented in the same position in the sequence as the deviant. The N2b was largest over centro-frontal sites and was thus quantified at Fz and Cz. The P3b was maximum over parietal regions and was thus quantified at Pz. The N2b appeared more as a distinctive peak, while the P3b appeared more as a long-lasting slow wave. The N2b was measured as the mean amplitude within ± 50 ms of the peak within the grand average, between 100-200 ms after the deviant onset (350-450 ms after the triplet onset). The P3b was measured within ± 75 ms of the peak in the grand average, between 300-450 ms after the deviant onset (550-700 ms after the triplet onset). The conditions were compared with a 4x2 Repeated-Measures ANOVA for both the N2b and P3b, with the four experimental conditions (early sustained attention, early switched attention, late sustained attention, late switched attention) and Frequency of Noise (A frequency bandpass filtered, B frequency bandpass filtered) forming the groups.

Results

Performance Data.

Reaction times (RTs) were calculated from the onset of the first A tone of the triplet. It is important to note that the change in the B tone occurred 250 ms after this onset. In the case of the early deviants occurring in the sustained condition, the deviant presentation would have been detected after attention had been sustained to the preceding 4 triplets. For the late deviants occurring in the sustained condition, their presentation would have been detected after attention had been sustained to the preceding 9 triplets. In the case of the late deviants occurring in the switch condition, their detection would have been made 3 triplets after attention had been switched from the noise channel.

Effect of Noise Bandpass. Very unexpectedly, as the frequency of the noise presented in the gap within a tone was previously shown to affect the continuity illusion, there was no significant difference in the accuracy of detection ($F(1,20)=1.46$, $p=.24$) and RTs ($F(1,20)=0.15$, $p=.71$) of the deviants whether A bandpass filtered noise or B bandpass filtered noise was presented. See Table 5, 6 and 7 in Appendix 5 for statistical analyses of the accuracies, RTs and correct rejections.

Attended Deviant Stimuli Performance. Subjects correctly detected 83.1% (SD = 11.7%, 95% CI: 75.2-90.9%) of the early deviants presented in the sustained attention condition, 90.4% (SD = 8.9%, 95% CI: 84.4-96.3%) of the late deviants presented in the sustained attention condition and 87.5% (SD = 11.4%, 95% CI: 79.8-95.1%) of the late deviants presented in the switched attention condition. The differences just failed to attain significance ($F(1.34,26.75)=3.78$, $p=.052$, $r^2=.16$). The detection accuracy of the late deviants in the sustained attention condition was higher than the detection accuracy of the early deviants in the sustained attention condition ($p<.001$). This is the reverse of the results obtained in Experiment 1. There is also a lot of variability, as seen by the large standard deviations, between different participants on their response patterns.

The mean reaction time (RT) for the detection of early deviants in the sustained attention condition were 805 ms (SD=116 ms, 95% CI: 727 ms-883 ms) for the early deviants in the sustained attention condition, 674 ms (SD=106 ms, 95% CI: 602 ms-745 ms) for the late deviants in the sustained attention condition and 798 ms (SD=148 ms, 95% CI: 699 ms-883 ms) for the late deviants in the switched attention condition. The three deviant detection RTs were significantly different ($F(1.33,26.52)=11.36$, $p<.005$, $r^2=0.36$). RTs for the late deviants in the sustained attention condition were significantly faster than the RTs for the early deviants presented in the sustained attention condition ($p<.001$) and the late deviants presented in the switched attention condition ($p<.01$). The accuracy and RT results show a pattern in which the late deviant in the sustained attention condition was the easiest to detect.

Non-Deviant Stimuli Performance. Subjects correctly withheld a response on 99.9% (SD = 0.2%) of triplet presentations when a deviant was not presented in the sustained attention condition and on 95.5% (SD = 3.9%) of triplet presentations in the switched attention condition. Subjects correctly withheld a response on 81.7% (SD = 25.4%) of the to-be-ignored early deviants presented during the first half of the switched attention condition.

The detection accuracy of the noise bursts in the switched attention condition was very low (13.2%, SD = 12.5%). Subjects responded on average to 19.6% (SD = 10.4%) of the noise bursts. The mean of the d' for the task is 0.405 (SD=0.51).

ERP Results. In all the following figures, the mean of the activity in the -50 ms to 0 ms period was used to draw the zero-voltage baseline from which the different components were measured. This baseline represents the 50 ms period before the onset of the first A tone of the triplet, (i.e., during the silent period between the onset of the triplets). In the Figures, the onset of the A tone is indicated by a black vertical line (i.e., at 0 ms). The onset of the deviant, the noise-filled gap in the B tone, is indicated by a grey dotted line (i.e., at 250 ms after the onset of the A tone). Grand averaged waveforms (average of all subjects' individual averages) are illustrated (except in Figure 30). Significantly, these grand averages represent the deviant-standard difference waveforms. In the text, the latencies after the onset of the deviant will be presented. The amplitude N2b was not significantly different between the four experimental conditions (early sustained attention, early switched attention, late sustained attention, late switched attention) at Fz ($F(3,30)=1.22$, $p=.32$, $r^2=.11$) and Cz ($F(3,30)=1.00$, $p=.41$, $r^2=.09$). There was a significant difference between the four experimental conditions for the P3b at Pz ($F(3,30)=14.33$, $p<.001$, $r^2=.59$). There was no effect of the frequency of filtered noise (presented in next section) or any significant interactions between the experimental conditions and noise frequency (Statistical findings are presented in Table 8 in Appendix 5).

Effect of Filtered Noise on the Continuity Illusion and Deviant Detection. Figures 25a and 25b illustrate the ERPs elicited by the deviant stimuli for the A (black lines) and B bandpass filtered noise (red lines). Figure 25a illustrates the results for the sustained attention task, and the switched attention task is shown in Figure 25b. The early deviants are shown left and the late deviants right. A negative deflection, the N2b is apparent at about 150 ms after the deviant onset (400 ms after the triplet onset) at Fz and Cz. This is followed by a large parietal maximum positivity, the P3b occurring at about 300 ms after the deviant onset (550 ms after the triplet onset).

ANOVAs were run to compare the N2b and P3b for the A frequency bandpass filtered deviants to the B frequency bandpass filtered deviants. No significant differences between A frequency bandpass filtered noise and B frequency bandpass filtered noise were found for the N2b at Cz ($F(1,10)=2.64$, $p=.13$, $r^2=.21$) and P3b at Pz ($F(1,10)=0.03$, $p=.87$, $r^2=.003$). A non-significant difference for the N2b at Fz ($F(1,10)=4.93$, $p=.05$, $r^2=.33$) shows that there is a tendency for the B frequency bandpass filtered noise to have a larger deflection than the A frequency bandpass filtered noise. This can be most easily seen in Figure 25b. Although the difference is non-significant, it does indicate that the deviant was more difficult to discriminate when the noise was closer to the B tone, and thus creates a greater continuity with the pure tone. It was nevertheless surprising not to obtain more markedly different results for the two noise frequencies. As there were no statistically significant differences between A and B frequency bandpass noise, the ERPs are collapsed in the following sections presenting the effect of the experimental conditions.

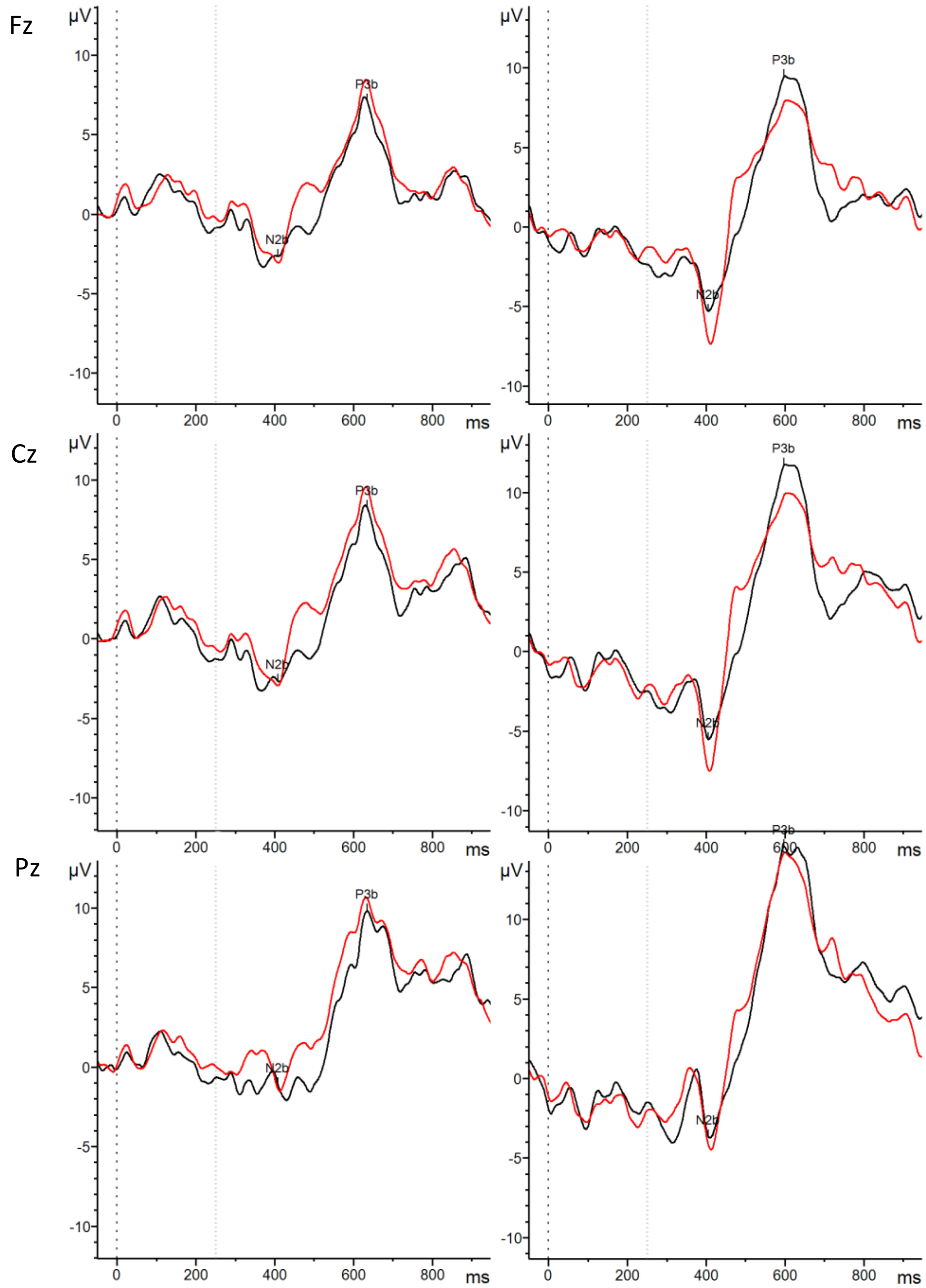


Figure 25a. Difference waves as influenced by noise frequency in the sustained attention task. Left: Early deviant. Right: Late deviant. A frequency bandpass filtered noise (black lines) and B frequency bandpass filtered noise (red lines). A negativity, the N2b, is apparent at about 400 ms at Fz and Cz and a positivity, the P3b, is apparent from 550 to 700 ms at Pz for in all difference waves.

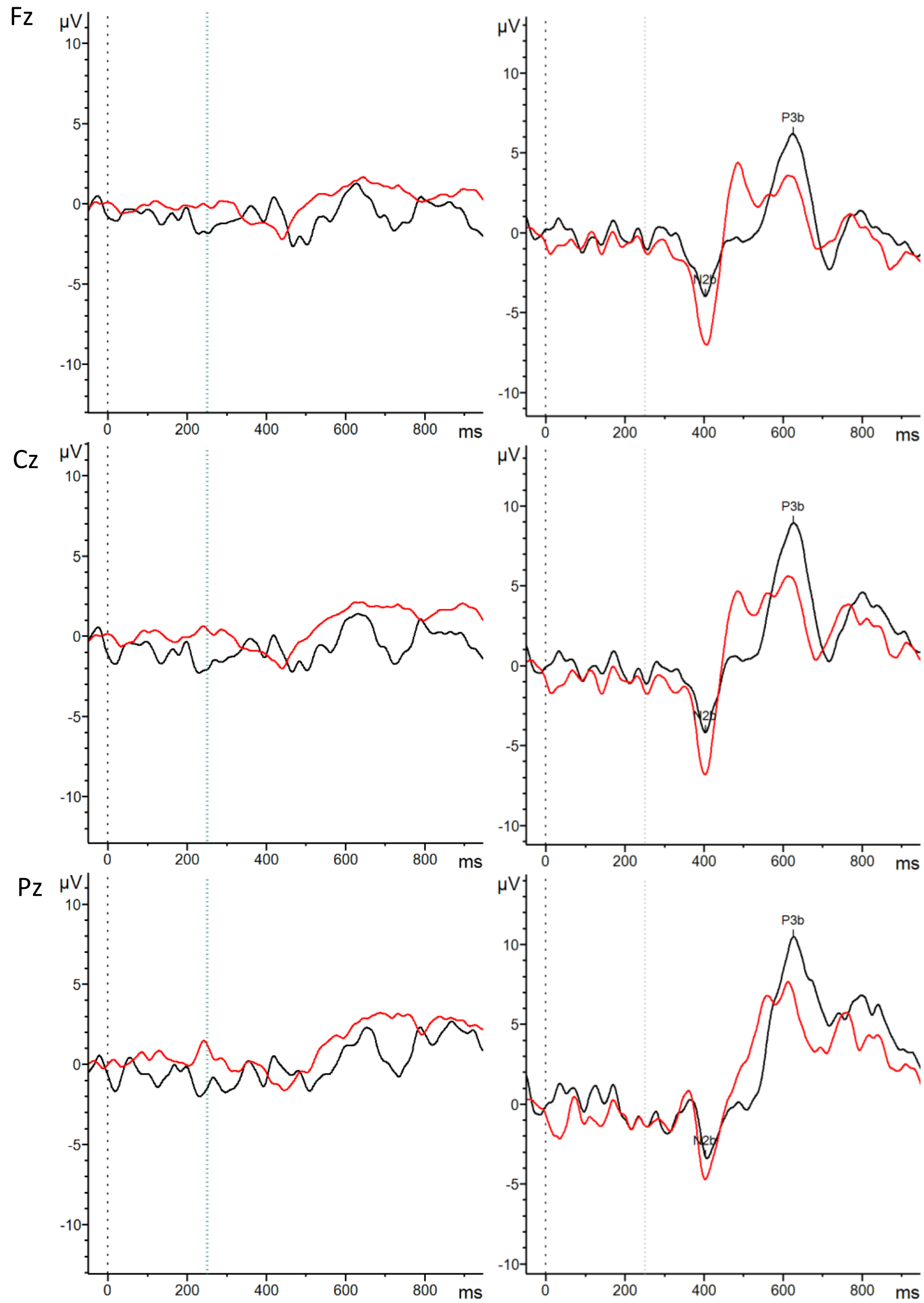


Figure 25b. Difference waves as influenced by noise frequency in the switched attention task. A bandpass filtered noise (black lines) and B bandpass filtered noise (red lines). Left. Early deviant. Right. Late deviant. A negativity, the N2b, is apparent at about 400 ms at Fz and Cz and a positivity, the P3b, is apparent from 550 to 700 ms at Pz for the late deviant. The positive deflection of the A and B tones differ at Fz and Cz between 450 and 550 ms, with the B tone showing an early positive deflection.

Effects of Attention on the Processing of the Early Deviants. In the sustained attention condition, subjects attended to the triplet channel and ignored the channel containing the noise bursts. Evidence of attention to the triplet sequences was provided by the actual detection of the deviants. In the switched attention condition, subjects attended the noise channel and performed a detection task on the noise bursts for the first half of the sequence and were instructed to ignore the triplet channel containing the early deviant. Evidence of attention to the noise bursts was provided by the actual detection of the noise bursts.

The grand average early deviant-standard differences wave for the sustained and switched conditions are illustrated in Figure 26. In the case of the switched condition, subjects were not attending to the triplet channel containing the early deviant. When subjects did attend to the triplet channel, the correctly detected early deviant elicited an N2b around 150 ms followed by a P3b at about 350 ms, maximum at Pz.

The amplitude of the attended early deviant N2b was not significantly different between the attended and the ignored condition at Fz (p=.15) and Cz (p=.16). At Pz, the P3b following presentation of the early deviant was significantly when the triplets were attended compared to when they were ignored (p=.02). When the noise burst channel was to-be-attended, it was possible that the occurrence of the early deviant in the to-be-ignored channel resulted in a passive switch of attention away from the noise burst channel, as in Experiment 1. This switch of attention to the triplet channel should then have resulted in the elicitation of a fronto-central maximum P3a occurring from 200-300 ms, as was found in Figure 19 of Experiment 1. As is apparent in Figure 26, a P3a was not observed.

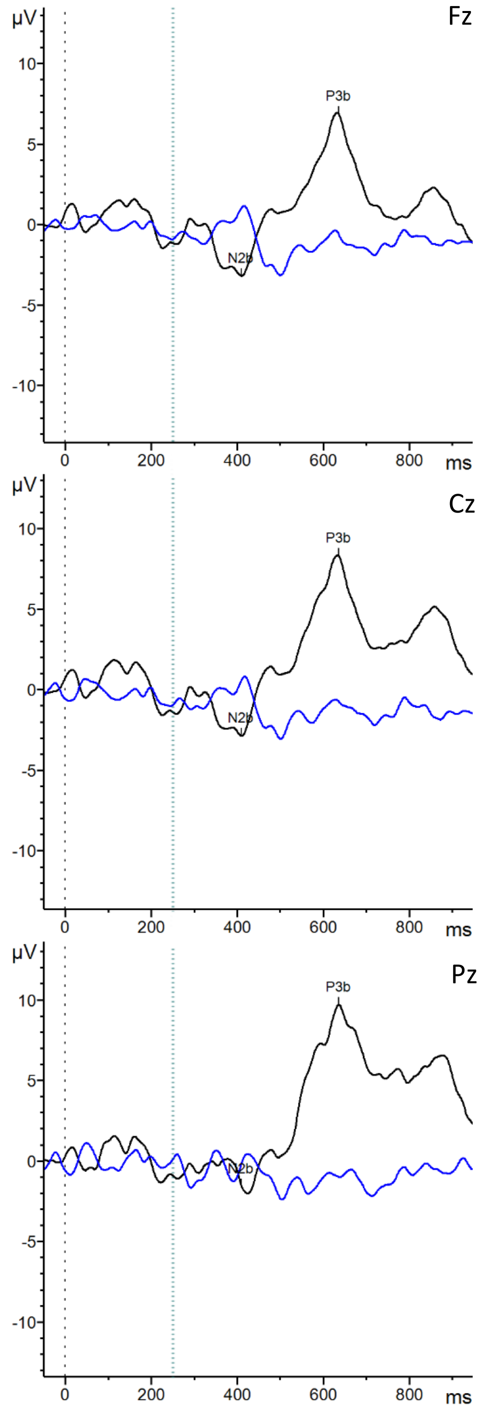


Figure 26. Deviant-standard difference wave following the presentation of the early deviants (Sustained condition - black line, Switched condition - blue line). A negativity, the N2b, is apparent at about 400 ms at Fz and Cz for the attended deviant. A positivity, the P3b, is apparent from 550 to 700 ms at Pz for the attended deviant

Building up of Streaming. The building up of streaming over the time spent attending to the task was examined by comparing the N2b and P3b elicited by the early and late deviants in the sustained attention condition. Previous research has indicated that detection of the deviant is better accomplished early rather than later in an attended sequence of triplets. An N2b was apparent around 150 ms post B tone onset ms for both the early and the late deviants (see Figure 27) at Fz and Cz. The N2b was not significantly different at Fz ($p=.15$) and Cz ($p=.24$) between the early and late deviants. Unexpectedly, the amplitude of the P3b at about 350 ms was significantly larger for the deviant occurring late compared to that occurring early ($p<.01$). This appear to indicate that the late deviant was easier to detect the early deviant.

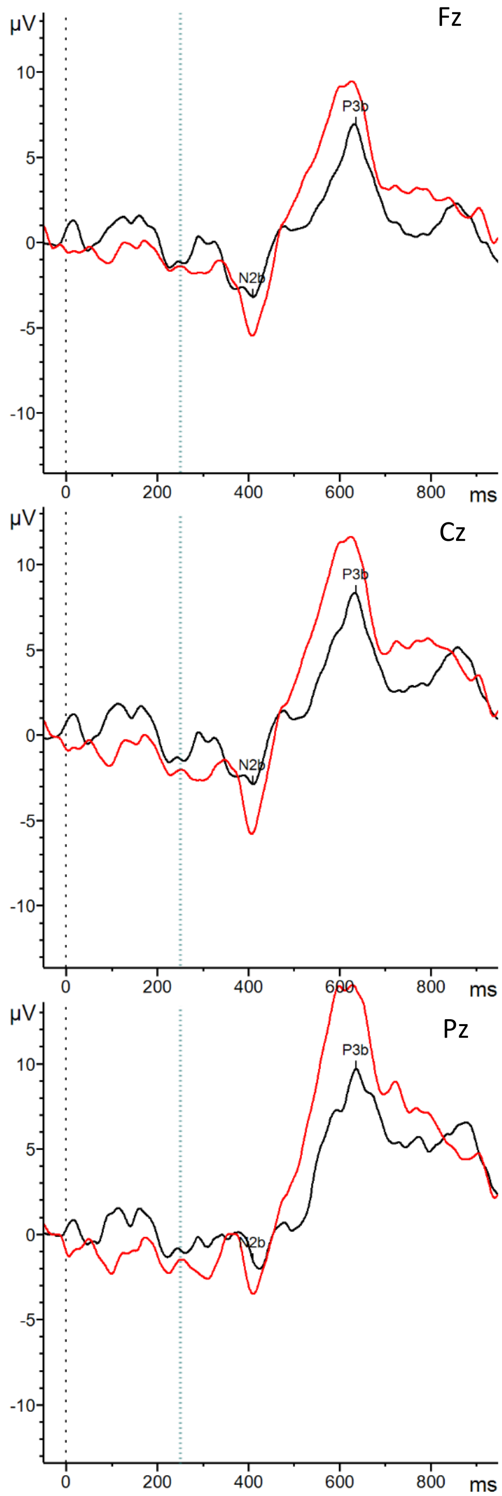


Figure 27. Difference waves between deviants and standards after presentation of the deviants in the sustained attention condition. (Early deviant - black line, late deviant - red line). A negativity, the N2b, is apparent at about 400 ms at Fz and Cz for both deviants. A positivity, the P3b, is apparent from 550 to 700 ms at Pz for both deviants, with the deflection to deviant in the sustained attention condition being the largest

Effect of Task Switching on Streaming. The influence of task switching on deviant detection was examined by comparing the late deviants in the switched attention condition to the late deviant in the sustained attention condition. In the sustained attention condition, the subjects had to attend the triplet channel for the whole duration of the sequence while in the switched attention condition, the subjects were to initially complete a task on the noise bursts and after 6 s subsequently attend to the triplet channel. Again, this deviant should have been easier to detect in the switched than in the sustained attention condition because while the deviant physically occurs "late" in the triplet sequence, it is experienced as occurring "early" after the switch from the noise burst channel. The grand average difference waves following presentation of the late deviants in the sustained attention and switched attention conditions are illustrated in Figure 28. The amplitude of the N2b was not significantly different at Fz (p=.76) and Cz (p=.88) between the sustained attention and switched attention conditions. Unexpectedly, the amplitude of the P3b was significantly larger for the late deviant in the sustained attention condition than in the switched attention condition (p<.01), which would indicate that the deviant detection was easier for the deviant in the sustained attention condition.

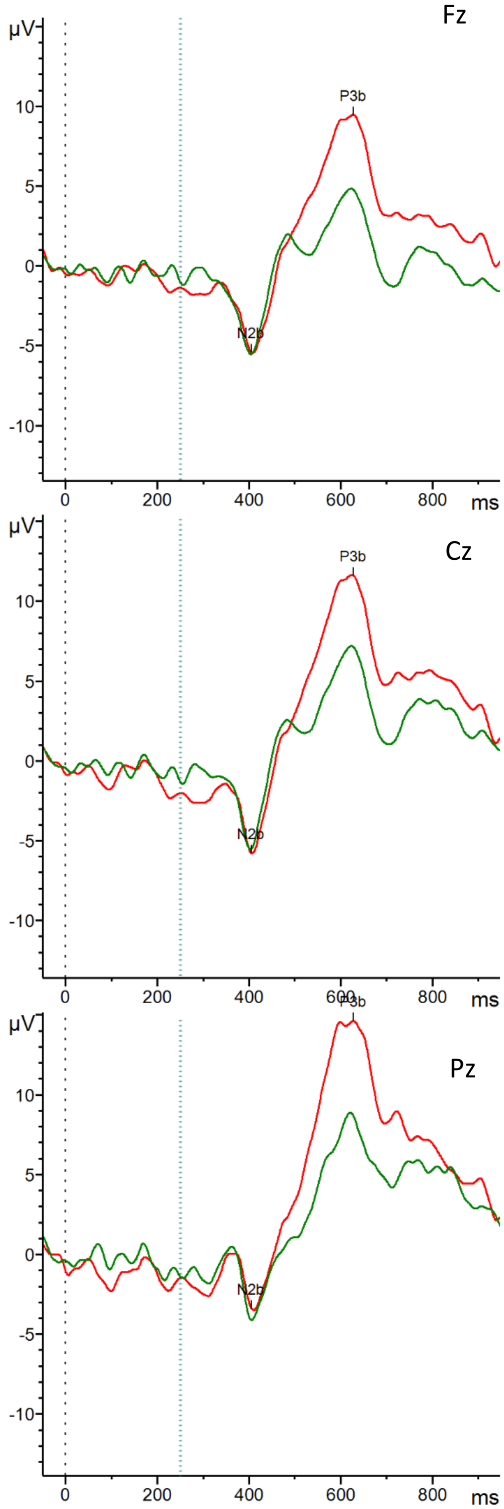


Figure 28. Difference waves between late deviants and standards after presentation of the late deviants (Sustained condition - red line, Switched condition - green line). A negativity, the N2b, is apparent at about 400 ms at Fz and Cz for both deviants. A positivity, the P3b, is apparent from 550 to 700 ms at Pz for both deviant, with the deflection to deviant in the sustained attention condition being the largest

Effect of Passive Build-Up on Streaming. The effect of task switching on streaming was also examined by comparing the N2b and P3b elicited by the Early deviants in the sustained attention condition to those elicited by the Late deviants in the switched attention condition. In the switched condition, while the deviant physically occurred late in the sequence, it was experienced to occur early in the sequence. The difference waves following presentation of the early and late deviants are illustrated in Figure 29. The amplitude of this N2b was not significantly different at Fz ($p=.58$) and Cz ($p=.72$) between the two deviants. The amplitude of the P3b at Pz was also not significantly different between the two deviants ($p=.76$).

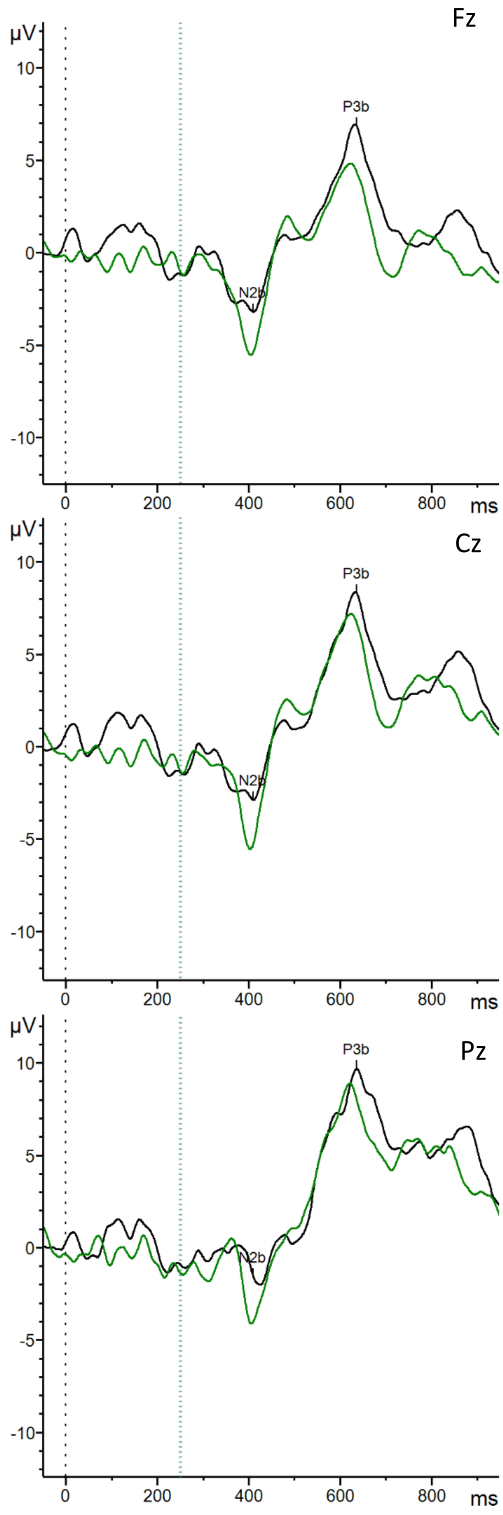


Figure 29. Difference waves between deviants and standards after presentation of the early deviants in the sustained condition (black line) and the late deviants in the switched condition (green line). A negativity, the N2b, is apparent at about 400 ms at Fz and Cz for both deviants. A positivity, the P3b, is apparent from 550 to 700 ms at Pz for both deviants. There are no significant differences between the two deviant conditions.

Ignored triplet channel. Figure 30 presents the difference between the standard and deviant in the to-be-ignored deviant condition. Similar to the previous experiment, a deviant stimulus should elicit an early negativity, a potential MMN, peaking at about 100-150 ms, regardless of whether the subject attends to the channel or not. However, this is not visible in Figure 30. As discussed, the amplitude of the MMN is very small (often less than 1 μV) relative to the background noise of the EEG, and more deviants needed to be presented in order for the MMN to emerge from the background EEG.

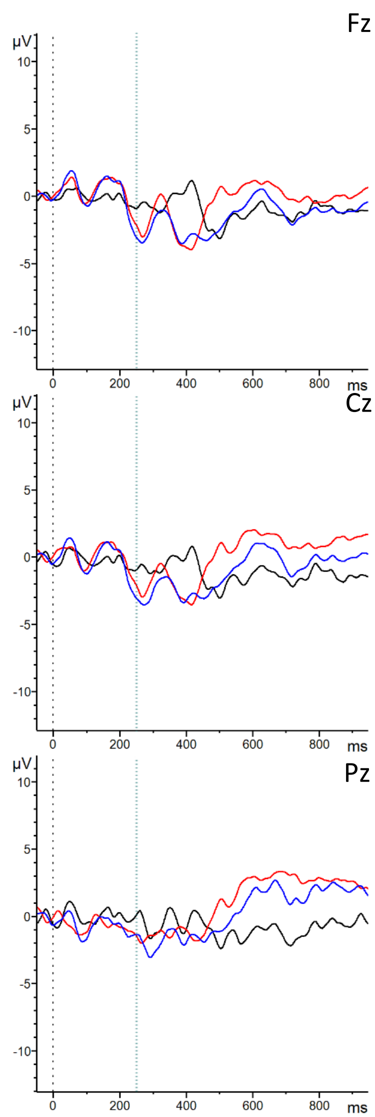


Figure 30. Early ignored deviant in the switched attention condition. Standard in red, Early deviant in blue and deviant-standard difference wave in black.

Discussion

Firstly, despite using two different types of noise, A frequency bandpass filtered and B frequency bandpass filtered, there was unexpectedly no significant difference between them in terms of results, so the combined results for the comparison between the position condition (early or late) and task condition (sustained or switched) were presented henceforth.

That the manipulation of the noise frequency did not yield any statistically significant results is very surprising as such a difference had been obtained by Micheyl et al. (2003). However, it is particularly interesting that the N2b at Fz to B frequency bandpass filtered noise was larger than for A frequency bandpass filtered noise (even if that difference is not statistically significant, $p=.05$). This would suggest that the B frequency bandpass filtered noise creates a more uniform percept, through the continuity illusion, making it harder to perceive the presence of a deviant. The P3b were not significantly different, which means that the decision on the presence of the deviant was equivalent for both noise conditions.

In conditions where the continuity illusion was to be elicited, similar results as Experiment 1 were expected. In Experiment 1, the performance and ERP results indicated that the late deviant in the sustained attention condition was harder to detect than the early deviant in the sustained attention condition and than the late deviant in the switched attention condition. In this experiment (Experiment 2), performance data indicated that the late deviant in the sustained attention condition was the easiest to detect. The late deviants in the sustained attention condition had a higher accuracy than the late deviants in the switched attention condition. The RTs followed the same pattern, with the late deviant in the sustained attention condition having the fastest RTs. The late deviant in the sustained attention condition also elicited the larger P3b. Such a pattern shows that the longer the subjects attended to the triplet sequence, the better they were at the task and the faster they responded. This was the expected result for the “A” bandpass filtered noise tones that break the continuity illusion, thus preventing streaming and with the percept of an “ABA” pattern building up with time.

The EEG results obtained in Experiment 2 corroborated the findings of the behavioural results. The N2b deflections were not significantly different between the deviants, indicating that no deviant was significantly different in term of perceptibility. A larger deflection was observed for both late deviants (sustained and switched) in comparison to the early deviants in the sustained attention condition (Figure 27 and 29). This would normally be taken to indicate that the late conditions were harder to discriminate, but as the N2b is modulated by the difference of the deviant to the standard, if the entire triplet was considered the pattern and streaming did not occur, the deviant might have been more perceptually different from the standard the longer the sequence went on.

The P3b showed that the late deviant in the sustained attention condition was the easiest to detect, as it was statistically larger than the P3b for both the early deviant in the sustained condition and the late deviant in the switched attention condition. This result agrees with the behavioural results. As the amplitude of P3b is inversely related to task difficulty (Johnson, 1986), this result shows that subjects were more certain their decision in the case of the late deviant in the sustained attention condition.

In this experiment, the absence of a P3a is a positive sign as it shows that the triplet was not considered task relevant while it was to be ignored and is good evidence that the subjects did concentrate on the noise burst task. Even if the participants did not perform well on the noise burst detection time, this was not because they were “listening in” on the triplet sequence.

As this section shows, the experimental findings did not match the initial expectations. Given the extremely limited number of existing studies, there are very few equivalent experiments for us to compare the obtained results to, meaning that these results are both a new contribution, and unexpected. There are a number of possible reasons for these unanticipated results. First, it could be hypothesised that the addition of a noise segment in the middle of the B tone of all triplets may have prevented streaming from occurring and forced the percept to remain one stream throughout. Bregman, Colantonio and Ahad (1999) had used noise burst stimuli and still obtained steaming, so there is previous work that has managed to create streaming with noise bursts in the stimuli. For a consistent one stream sequence, deviant detection would improve the longer the subjects have been attending to the sequence, as it makes the deviant more salient compared to the standards. The rate of stimulus presentation might also have contributed to streaming being harder to achieve. Warren, Obusek, Farmer and Warren (1969) and Bregman (1990) had successfully used a 200 msec onset to onset stimuli to induce streaming but they had no noise presentation in their stimuli.

A second potential theory is that even if streaming built up successfully, for ABA triplets, the continuity illusion can only be induced within the B tone when the A tones are perceived as being part of the same auditory object. The A tones may capture attention away from the gap in the B tone as discussed by Kobayashi and Kashino (2012). Thus, when the percept was of one stream at the beginning of the sequence, the continuity illusion was present and the deviant harder to detect. Later in the sequence, when the streaming had occurred, the A tones were perceived as being part of another stream and could not contribute to strengthening the continuity illusion, making the deviant easier to detect.

However, it should be noted, that although there were a number of unexpected findings, interestingly, the results obtained support the idea that attention does not influence the continuity illusion. If attention had influenced the continuity illusion, the illusion would have been stronger later in the sequence, making the deviant appear more similar to a standard and thus making it harder to discriminate the deviant later in the sequence. The findings of this study do not indicate this, and therefore, it can be concluded on the basis of those findings that attention does not influence the continuity illusion.

Based on the obtained results, it is most likely that the passing of time in the presence of attention strengthened the percept of the triplet as a unified stream, making it easier to detect the deviant the longer the sequence was attended. The triplet presentation rate (1 each second), A and B tone durations and frequency separation were expected to create streaming but it can be hypothesised that the addition of a noise section in the middle triplet interfered with the streaming process. For a subsequent study to build on the results of this study, using standard stimuli that do not contain noise would be prescribed, as it can be posited that this noise prevented streaming from occurring normally.

Conclusion

This thesis investigated the influence of the time spent attending to a task on auditory streaming and the formation of auditory objects by carrying out comprehensive experiments from two different angles. First, based on the literature that described the processes involved in the exogenous and endogenous influence of attention on the separation of an auditory scene into separate objects, a functioning multilayer model, published in Morissette and Chartier (2015), was designed and constructed. This model encompasses multiple sensory attributes such as frequency, amplitude and spatial location, building on less sophisticated models that only make use of frequency. The model presented is a modest improvement on the model created by Wrigley and Brown (2004). This comprehensive model was designed to fully cover the process of auditory object creation, beginning from the initial perceptual levels, through to the unification of this information into more informative objects through saliency and Gestalt rules of grouping, and then finally the influence of consciously driven attention on what objects will retain attention. This resulted in the creation of a robust and easily modifiable model, with solid results that matched those reported in the literature. Further refinements were then applied to this functioning model to experiment with adding further levels of psychological plausibility to it. These were fully discussed in the latter part of Chapter 2, and included experimentation with better representation of spatial localization, more biologically plausible bindings between attributes, and more flexible representation of the top-down attention. These experimental changes had a solid justification based on work in the literature, and were found to be partially successful. Some results proved their viability, such as the distributed location array, while others did not yield an improvement on the initial model, such as the oscillatory binding of features and distributed attentional integrate-and-fire neurons. Overall though, the model was found to achieve solid results, supporting and validating other reports in the literature.

This model represents a synthetized version of known findings from previous studies on auditory object creation. However, as was seen in the literature review presented in Chapter 1, the extent of the influence of the time spent attending to a task on streaming is still a subject being investigated. Given the first part of this thesis consisted of creating a synthetic model, it was felt appropriate to then approach this issue from a different angle, and gather new information through empirical studies on the influence of the time spent attending to a task on streaming. Based on a detailed and thorough survey of the literature, the reported results of Thompson et al. (2011) were taken as the starting point. This meant that the initial expectations of whether attention is needed for streaming to occur and to what extent, as measured through deviant detection, were defined by their results. This was justified by evidence from multiple researchers (Billig & Carlyon, 2016; Carlyon & Cusack, 2005; Carlyon et al., 2001; Cusack et al., 2004; Shamma et al., 2011; Snyder et al., 2006; Sussman, 2007; Sussman et al. 2002; Thompson et al., 2011) having shown a dissociation between a pre-attentive, low-level sound segregation process, and an attention-dependent process. Based on this evidence, the methodology of Thompson et al. (2011) was carefully followed, but significantly, with the addition of ERP measures to allow us more detailed insight into the cognitive processes leading up to the responses.

The objective results replicated the subjective behavioural findings of Thompson et al. (2011), with late deviants in the sustained condition being harder to detect than the early deviants in the sustained attention condition, and the late deviants in the sustained condition being harder to detect than the late deviants in the switched attention condition. This is in accordance with previous results showing that streaming builds with time, that it builds faster with attention and that deviants are easier to detect when heard in a single perceived stream of tones. The ERPs provided interesting results to support the behavioural results. Easier to detect deviants when less streaming had occurred elicited a smaller N2b and bigger P3b. The ERP results also gave us a confirmation that the deviant presented during the part of the experimental condition the subjects were to do a task on the noise bursts was salient enough to be detected, as exemplified by the presence of a P3a. One interesting tentative result obtained is that in the scenario of a deviant created by a delayed onset of the central B tone, the deviant detection showed a double decision. It was posited that the decision leading to the response was made on the detection of absence of the deviant. Given the small number of subjects, this would need to be investigated further before anything definitive can be affirmed based on that result, but the replication of subjective measures with objective ERP measurements represents another key contribution of this research.

This initial experiment was followed-up with a much more challenging experiment investigating the influence of the continuity illusion on deviant detection in a task that had been shown to create streaming. This had not been comprehensively studied in the literature using ABA triplets, and so there were extremely limited points of comparison. As in the previous experiment, a methodology based on existing similar (but not identical studies) was used. However, it was interesting that the results did not match the initial hypotheses. There was no effect of the frequency of the noise presented in the gap in the deviant tone, contrary to previous results (Micheyl et al., 2003; Riecke, Van Opstal, Goebel & Formisano, 2008). It was expected, based on previous research, that bandpass filtered noise encompassing the frequency of the B tone would create a continuity illusion when presented in the gap, while bandpass filtered noise encompassing the frequency of the A tone would not. The deviant detection results show that the longer the subjects attended to the triplet sequence, the easier it was to detect the deviant, in clear contradiction with the previous experiment. Both the behavioural and ERP results support that conclusion. The results obtained are coherent with the strengthening of a pattern in time. Various research had shown that in the presence of the continuity illusion, performance was similar to performance when continuous tones were presented (Carlyon et al., 2002; Micheyl et al., 2003; Plack & White, 2000; Tougas & Bregman, 1990). The second did not show such a result.

While a number of clear research contributions were made, there are some important limitations to the studies presented in this thesis that should be emphasised. It should be noted that very large variability was identified for all conditions in the two experiments, both in terms of accuracy and ERPs, showing that subjects sometimes behaved in very different ways to each other. This happened despite being given clear and consistent instructions and following existing methodologies closely. A larger subject pool would have helped to reduce this variability. A sample size of 11 was chosen, as this was a common group size for EEG studies. However, such a sample size is not large enough to have enough power for significant differences for some behavioural results. The power analyses are in Appendix 6. Despite following a clear and established methodology, it could be argued that given the limited time and resources available, there may have been some benefits to completing only one empirical study with 20 subjects.

The subject pool consisted of volunteers from the University of Ottawa and the repeated deviant detection task is arguably rather long and repetitive. As such, there were some small issues with guaranteeing that subjects completed the task correctly. By questioning the subjects after completion of the experiments, it was identified that one of the initially chosen subjects for the pilot (data not included) fell asleep during the experiment. It was also found that another subject failed to respond to one experimental condition altogether. To prevent these extreme cases (as they were obvious cases of invalid data) from affecting the results, measures were taken to control this data, in this example (not responding to the task) only taking correct detections of the deviants for further analyses. While this showed that it was possible to apply suitable controls to the data, and the instructions given to the subjects carefully followed the methodology discussed in numerous studies in the literature, these issues do indicate that when the experiments are tasks where adherence to protocol regarding attention is reliant entirely on the goodwill and best intentions of the subjects, there are potential issues. However, this is a common issue with many studies of this nature.

As part of the experiments in this study, a distractor task was used, in the form of noise bursts. Issues were identified with this task. The experiments used the exact same noise burst detection task as Thompson et al. (2011), as the intention was to replicate their methodology as closely as possible while adding objective measures (ERP data) to verify their results. The results found that the subjects in the results reported in this thesis performed poorly in comparison to other results in the literature, with accuracies being below 15% for both experiments. It was confirmed that the subjects were attempting the task (48% of responses), albeit failing, in the first experiment, however, no such solid conclusion can be drawn in the second experiment (20% response rate). This does then bring into question whether the subjects were truly ignoring the triplets during the first part of the switched attention task, as it was not possible to guarantee that they were completing the distractor task effectively. The presence of a P3a in the first experiment would show that the triplet sequence was not properly ignored. No P3a was seen for the second experiment, which indicates that the to-be-ignored triplet sequence was not considered task relevant. Despite this being a recommended method that has been used in the literature, these findings make a strong case that a different distractor task could arguably add more certainty to the results. This is especially true for the second experiment, where noise bursts were presented in in each ear. Even if they were of very different frequency ranges, they might have interfered with the completion of each of the tasks.

Finally, the choice of presenting noise in standard triplets in Experiment 2 may have prevented streaming to occur as expected. The noise was expected not to hinder steaming as Bregman, Colantonio and Ahad (1999) had used noise burst stimuli and still obtained steaming. Micheyl et al. (2003) had used noise bursts in their deviant detection experiment in the presence of a continuity illusion. As an experimental paradigm that was robust and successfully used in both Experiment 1, and in other studies in the literature, given the time limitations present with our study, it was chosen to rely on robust findings in the literature and not to carry out a pilot study, thus missing this problematic stimulus choice.

With this particular experiment, a balance had to be sought between presenting short tones rapidly to induce streaming, while keeping the tones long enough to be able to introduce a noise in their middle and for the tone to be presented long enough to induce a continuity illusion. The initial hypothesis was that the presentation of the triplet tones (150 ms onset-to-onset A tones bracketing a 400ms onset-to-onset B tones) would elicit streaming, even if it was less rapid than in Experiment 1. This was supported by research by Bregman, who had found that at a high rate of presentation (100 ms onset-to-onset), subjects perceive two different streams, while at lower rate (400 ms onset-to-onset), they perceive a single stream. His research had used 200 ms onset to onset and managed to successfully induce streaming, as did Warren, Obusek, Farmer and Warren (1969). The frequency separation chosen (8 semitone) had also shown clear streaming in previous studies (Carlyon et al., 2001; Thompson et al. 2011; Experiment 1). The duration of the noise, 200 ms, was chosen as it was the most likely to induce a continuity illusion in Riecke et al. (2007). This noise duration had also been used in Tougas and Bregman (1990) study of streaming and the continuity illusion. In most previous studies of the continuity illusions, tones were long in comparison to the noise creating the continuity illusion (2800 ms tone/200 ms noise (shortest), Micheyl et al. 2003; 133ms tone/66 ms noise, Riecke et al., 2008; 120 ms and 240 ms tones/40 ms noise, Bregman et al., 1999; 1640 ms tone/200 ms noise, Tougas & Bregman, 1990). However, while the obtained results show streaming did not occur, this should be verified with different tone lengths and speeds in the same ABA stimuli in future research, to see if the result was obtained because of the noise chosen or because streaming does not occur in ABA triplets in the presence of a continuity illusion.

In conclusion, this thesis makes a valuable research contribution and consolidates current knowledge about the interaction of attention with auditory object creation into a new auditory model created and presented in this thesis. This is further reinforced by two additional experiments, reported in the later chapter, make an interesting addition to a research area which is still very contentious, using objective ERP measures in addition to the more conventional behavioural measures. The research done in the second experiment is also one of very few that have looked at the influence of the presence of a continuity illusion on streaming, and the first to look at deviant detection in that context.

References

Alain, C., Arnott, S.R. & Picton, T.W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from brain potentials. *Journal of Experimental Psychology, 27*, 1072–1089.

Alho, K, Paavilainen, P, Reinikainen, K, Sams, M. & Näätänen, R. (1986) Separability of different negative components of the event-related potential associated with Auditory stimulus processing. *Psychophysiology*, 23, 613-62.

Anstis, S & Saida, S. (1985) Adaptation to Auditory Streaming of Frequency-Modulated Tones. *Journal of Experimental Psychology: Human Perception and Performance. 11(3)*, 257-271.

Baddeley, A. (2003) Working Memory: Looking back and looking forward, *Nature Neuroscience, 4*, 829-839

Bendixen, A., Denham, S.L., Gyimesi, K. & Winkler, I. (2010) Regular patterns stabilize auditory streams. *J Acoustic Society of America, 128*(6), 3658-3666.

Billig, A.J. & Carlyon, R.P (2016) Automaticity and Primacy of Auditory Streaming: Concurrent Subjective and Objective Measures. *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 339–353

Bisley, J.W. & Goldberg, M.E. (2003) Neuronal activity in the lateral intraparietal area and spatial attention, *Science, 299*(5603), 81-6.

Bregman, A.S. (1978) Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4*(3), 380-387.

Bregman, A., Ahad, P., Crum, P. & O'Reilly, J. (2000) Effects of time intervals and tone durations on auditory stream segregation. *Perception & Psychophysics, 62(3)*, 626-636.

Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Massachusetts: The MIT Press, 1990

Bregman, A., Colantonio, C. & Ahad, P. (1999) Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Perception & Psychophysics, 61(2)*, 195-205

Bregman, A.S. & Dannenbring, G.L. (1973) The Effect of Continuity on Auditory Stream Segregation. *Perception & Psychophysics*, *13*, 308-312.

Bregman A.S. & Rudnicky A.I. (1975). Auditory segregation: Stream or streams. *J. Exp. Psychol. 1*, 263–267

Carlin, M. A. & Elhilali, M. (2015) Modeling attention-driven plasticity in auditory cortical receptive fields. *Frontiers in Computational Neuroscience, 9*, 106.

Carlyon, R.P. & Cusack, R. (2005) Effects of attention on auditory perceptual organization, in *The Neurobiology of Attention*, Eds. L. Itti, G. Rees & J.K. Tsotsos, Academic Press.

Carlyon, R. P., Cusack, R., Foxton, J. M. & Robertson, I. H. (2001) Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol. Human, 27*(1), 115–127

Carlyon, R.P., Deeks, J., Norris, D. & Butterfield, S. (2002) The continuity illusion and vowel identification, *Acust. Acta Acust*., 408–415.

Carlyon, R.P., Micheyl, C., Deeks, J.M. & Moore, B.C.J. (2004) Auditory processing of real and illusory changes in FM phase. *J. Acoust. Soc. Am., 116*, 3629–3639.

Chelazzi, L., Miller, E.K., Duncan, J. & Desimone, R. (1993) A neural basis for visual search in inferior temporal cortex. *Nature, 363*, 345–347.

Cusack, R., Deeks, J., Aikman, G. & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance, 30*(4), 643-656.

Dannenbring, G.L. & Bregman, A.S. (1976) Stream segregation and the illusion of overlap. Journal of Experimental Psychology: Human Perception and Performance, 2(4), 544-555.

Desimone, R. & Duncan, J. (1995) Neural Mechanisms of Selective Visual Attention. *Annual Review in Neuroscience*, *18*, 193-222.

Eriksen, C.W & St. James, J.D. (1986) Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics. 40*(4), 225–240

Fishman, Y.I., Reser, D.H., Arezzo, J.C. & Steinschneider, M. (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research, 151(1-2)*, 167-187.

Fritz, J.M., Elhilali, S. & Shamma, S. (2007) Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiology. 98*, 2337–2346.

Galambos, R., Makeig, S. & Talmachoff, P.J. (1981) A 40-Hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci USA, 78,* 2643–2647.

Glasberg, B.R. & Moore, B.C.J. (1990) Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*, 103-108.

Gutschalk, A., Micheyl, C., Melcher, J., Rupp, A., Scherg, M. & Oxenham, A. (2005) Neuromagnetic Correlates of Streaming in Human Auditory Cortex. *Journal of Neuroscience, 25(22)*, 5382-5388.

Heise, G.A. & Miller, G.A. (1951) An experimental study of auditory patterns. *American Journal of Psychology, 64*, 68-77.

Helmholtz, (1863) On the Sensations of Tone as a Physiological Basis for the Theory of Music, trans. by Alexander J. Ellis from the fourth (1877) edition, New York: Dover Publications.

Hewitt, M.J. and Meddis, R., (1994) A Computer Model of Amplitude-Modulation Sensitivity of Single Units in the Inferior Colliculus. *J.Acoust. Soc. Am., 95(4)*, 2145-2159.

Itti, L. & Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506.

Itti, L. & Koch, C. (2001) Feature combination strategies for saliency-based visual attention systems, *Journal of Electronic Imaging, 10*(1), 161–169.

Itti, L. & Koch, C. (2001) Computational Modelling of Visual Attention, *Nature, 2,* 194-203.

James, W. (1890) *The Principles of Psychology*

Itti, L., Koch, C. & Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on pattern analysis and machine intelligence, 20*(11), 1-5.

Jeffress, L.A. (1948) A place theory of sound localization. *J Comp Physiol Psychol, 41*, 35-39.

Johannesma, P.I.M. (1980) Narrow band filters and active resonators in *Psychophysical, physiological, and behavioural studies in hearing*, Delft: Delft University Press, 62-63.

Johnson R. (1986) A triarchic model of P300 amplitude. *Psychophysiology. 23*(4), 367-84.

Kayser, C., Petkov, C., Lippert, M. & Logothetis, N. (2005) Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology, 15*(21), 1943-1947

Khayat, P.S, Spekreijse, H, Roelfsema, P.R. (2006) Attention lights up new object representations before the old ones fade away. *J Neurosci, 26*, 138–142.

Kobayashi, M. & Kashino, M. (2012) Effect of Flanking Sounds on the Auditory Continuity Illusion, *PLoS ONE, 7*(12).

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Langner, G. (1992). Periodicity coding in the auditory system. *Hear. Res., 60*, 115-142.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 451–468.

Lavie, N. (2000). Selective attention and cognitive control: Dissociating attentional functions through different types of load. In S. Monsell & J. Driver (Eds.*), Attention and performance XVIII* (pp. 175–194). Cambridge, MA: MIT Press

Licklider, J.C. (1951) A duplex theory of pitch perception. *Experientia. 7*(4), 128-34.

Luck, R.L. (2005). Ten simple rules for designing ERP experiments. In T. C. Handy (Ed.), event-related potentials: a methods handbook. Cambridge, MA: MIT Press.

Major, G & Tank, D. (2004) Persistent neural activity: prevalence and mechanisms. *Curr Opin Neurobiol. 14,* 675–684.

Makeig, S, Bell, AJ, Jung T.P & Sejnowski, TJ (1996) Independent Component Analysis of Electroencephalographic Data. In: *Advances in Neural Information Processing Systems* (Touretzky D, Mozer M, and Hasselmo M, eds), 8: 145-51.

Meddis, R. & Hewitt, M.J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Acoust. Soc. Am., 89*(6), 2866-2882.

Micheyl, C., Carlyon, R.P., Cusack, R & Moore, B.C.J. (2005) Performance measures of auditory organization. In: Pressnitzer D, de Cheveigne A, McAdams S, Collet L, editors. *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*. Srpinger; New York, NY: 203–211.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B. & Wilson, E.C. (2007) The role of auditory cortex in the formation of auditory streams. Hear. Res. 229, 116–131.

Micheyl, C., Carlyon,R.P., Shtyrov, Y., Hauk, O., Dodson, T. & Pullvermuller, F. (2003) The neurophysiological basis of the auditory continuity illusion: a mismatch negativity study. *Journal of Cognitive Neuroscience, 15*, 747-758.

Micheyl, C., Kreft, H., Shamma, S. & Oxenham, A.J. (2013) Temporal coherence versus harmonicity in auditory stream formation*. Journal of the Acoustical Society of America*, *133(3)*, 188-94.

Middlebrooks, J.C. (2015) Sound localization. *Handbook of clinical neurology*, *129*, 99-116

Miller, E.K. & Cohen, J.D. (2001) An integrative theory of prefrontal cortex function, *Annu Rev Neuroscience, 24,*167-202.

Miller, G.A & Licklider, J.C.R. (1950) The intelligibility of interrupted speech. *Journal of the Acoustic Society of America*, *22*(2), 167–173.

Moore, B.C.J & Ernst, S.M. (2012) Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *J. Acoust. Soc. Am., 132*, 1542–1547.

Morissette, L. & Chartier, S. (2015) Saliency model of auditory attention based on frequency, amplitude and spatial location, *Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN)*, p.1-5.

Motter, B.C. (1994) Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J Neuroscience, 14*(4), 2178-89.

Müller-Gass, A., Stelmack, R.M. & Campbell, K. (2006) The effect of visual task difficulty and attentional direction on the detection of acoustic change as indexed by the Mismatch Negativity. *Brain Research, 1078*(1), 112-130.

Müller-Gass, A. & Schröger, E. (2007) Perceptual and cognitive task difficulty has differential effects on auditory distraction, *Brain Research*, 169-177.

Näätänen, R. (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci.*, *13*, 201–288.

Näätänen, R. (1991). Mismatch negativity outside strong attentional focus: a commentary on Woldorff et al., *Psychophysiology, 28*, 478–484.

Näätänen, R., Simpson, M. & Loveless, N.E., (1982) Stimulus deviance and evoked potentials. *Biol. Psychol. 14*, 53-98.

Oliva, A. (2005) Gist of the scene. In: Itti, L., Rees, G. and Tsotsos, J.K. (Eds.), *Neurobiology of Attention*. Elsevier, San Diego, CA, pp. 251–256.

Oliva, A. & Schyns, P.G. (1997) Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 72–107.

Oliva, A. & Torralba, A. (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comp. Vis., 42: 145–175.

Oxenham A.J. (2012) Pitch perception. *J Neurosci., 32(39)*,13335–13338.

Oxenham A.J. (2014) Revisiting place and temporal theories of pitch. *Acoust Sci Technol., 34(6)*, 388–396.

Oxenham AJ, Bernstein JGW & Penagos H. (2004) Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. USA.*, *101*, 1421–1425

Oxenham A.J., Micheyl C., Keebler M. V., Loper A & Santurette S. (2011). Pitch perception beyond the traditional existence region of pitch, *Proc. Natl. Acad. Sci. U.S.A.*, *108*, 7629–7634.

Patterson, R.D. and Holdsworth, J. (1996) "A functional model of neural activity patterns and auditory images," *Advances in Speech, Hearing and Language Processing*, (W. A. Ainsworth, ed.), Vol 3. Part B. JAI Press, London, 547-563.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. & Allerhand M. (1992) Complex sounds and auditory images. In: Cazals, Y., Demany, L., Horner, K. (Eds), *Auditory physiology and perception, Proc. 9th International Symposium on Hearing*. Pergamon, Oxford, 123-177.

Plack, C.J. & White, L.J. (2000). Perceived continuity and pitch perception. *Journal of the Acoustical Society of America, 108,* 1162-1169.

Plaut, D.C. (2000). Methodologies for the computer modeling of human cognitive processes. In F. Boller, J. Grafman and G. Rizzotti (Eds.), *Handbook of neuropsychology*, 2nd Edition, Volume 1. New York: Elsevier.

Posner, M.I. & Cohen, Y. (1984) Components of Visual Orienting. In: Bouma, H. and Bouwhuis, D.G., Eds., *Attention and Performance X: Control of Language Processes*, Erlbaum, Hillsdale, 531-556.

Posner, MI, Snyder, CR & Davidson, BJ. (1980) Attention and the detection of signals. *J Exp Psychol. 109(2),* 160-74.

Pressnitzer, D., Sayles, M., Micheyl, C. & Winter, I.M. (2008) Perceptual Organization of Sound Begins in the Auditory Periphery. *Current Biology*, *18(15)*, 1124-1128.

Reynolds, J.H. & Desimone, R. (2003) Interacting Roles of Attention and Visual Salience in V4. *Neuron, 37*(5), 853-63.

Riecke, L., van Opstal, A.J., Goebel, R. & Formisano, E. (2007) Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J Neurosci., 27*(46), 12684-12689.

Roberts, B., Glasberg, B.R. & Moore, B.C. (2008) Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *J Exp Psychol Hum Percept Perform., 34*(4), 992-1006

Sarter, M., Givens, B. & Bruno, J.P. (2001) The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Res Brain Res Rev., 35*(2), 146-60.

Schröger, E. (1996) A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of cognitive neuroscience, 8(6),* 527-539.

Schröger, E., Marzecová, A. & SanMiguel, I. (2015) Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur J Neurosci. 41(5),* 641-64.

Scientific Committee on Emerging and Newly Identified Health Risks (2008) Potential health risks of exposure to noise from personal music players and mobile phones including a music playing function. 81 p.

Schillen, T.B. & Konig, P. (1994) Binding by temporal structure in multiple feature domains of an oscillatory neuronal network. *Biological Cybernetics*, *70,* 397–405.

Sculthorpe, L.D., Ouellet, D.R. & Campbell, K. (2009) MMN elicitation during natural sleep to violations of an auditory pattern. *Brain Res.*, *1290*, 52-62.

Seebeck, A. (1841) Beobachtungen über einige Bedingungen der Entstehung von Tönen [Observations on some conditions for the formation of tones] Ann. Phys. Chem. 53, 417–436.

Shamma, S. (2001) On the role of space and time in auditory processing. *Trends in Cognitive Sciences, 5*(8), 340-348.

Shamma, S. (2008) On the Emergence and Awareness of Auditory Objects. *PLoS Biology, 6(6),* 1141-1143.

Shamma, S., Elhilali, M. & Micheyl, C. (2011) Temporal doherence and attention in auditory scene analysis. *Trends in Neuroscience, 34*(3), 114-123.

Shinn-Cunningham, B.G & Best, V. (2008) Selective Attention in Normal and Impaired Hearing. Trends in Amplification, 12(4), 283-299.

Shu, Y., Hasenstaub, A. & McCormick, D.A. (2003) Turning on and off recurrent balanced cortical activity. *Nature. 423*, 288–293

Slaney, M. (1993) Auditory Toolbox Version 2, *Technical Report #1998-010*, Interval Research Corporation

Slaney, M. & Lyon, R.F. (1993). On the importance of time-a temporal representation of sound. *Visual representations of speech signals*. 95116

Snyder, J. S., Alain, C. & Picton, T.W. (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18,* 1–13.

Snyder, J.S., Gregg, M.K, Weintraub, D.M. & Alain, C. (2012) Attention, Awareness, and the Perception of Auditory Scenes, *Frontiers in Psychology, 3*, 1-17.

Sussman, E. (2007). A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology, 21*(3-4), 164-175.

Sussman, E., Chen, S., Sussman-Fort, J. & Dinces, E. (2014) The five myths of MMN: redefining how to use MMN in basic and clinical research. *Brain Topogr. 27*(4), 553-64.

Sussman, E., Horváth, J., Winkler, I. & Orr, M. (2007) The role of attention in the formation of auditory streams. *Perception & Psychophysics, 69(1),* 136-152.

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

Sussman, E., Winkler, I., Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review, 10*(3), 630-637.

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

Thompson, S.K., Carlyon, R.P. & Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1253-62.

Tougas, Y. & Bregman, A. (1990) Auditory streaming and the continuity illusion. Perception & *Psychophysics. 47(2),* 121–126.

Treue, S. & Martinez-Trujillo, J.C. (1999) Feature-based attention influences motion processing gain in macaque visual cortex, *Nature*, *399*(6736), 575-579.

Treue, S. & Maunsell, J.H.R. (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience. 19*, 7591-7602.

van de Par, S & Kohlrausch, A. (1997) A new approach to comparing binaural masking level differences at low and high frequencies. *J. Acoust. Soc. Am., 101*, 1671–1680.

Van Noorden, L.P.A.S. (1975) *Temporal coherence in the perception of tone sequences*, Ph.D. Thesis, Eindhoven University of Technology, The Netherlands.

Vliegen, J., Moore, B.C. & Oxenham, A.J., (1999) The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J Acoust Soc Am., 106*(2), 938-45.

Vliegen, J. & Oxenham, A.J. (1999) Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am., 105*(1), 339-46.

Von der Malsburg, C. & Schneider, W. (1986) A neural cocktail-party processor. *Biol Cybern. 54*(1), 29-40.

Wang, D.L. (1996) Primitive auditory segregation based on oscillatory correlation. *Cognitive Science, 20*, 409-456.

Wang D.L. (1999): Object selection based on oscillatory correlation. *Neural Networks*, *12*, 579-592.

Wang, D.L. (2005) The time dimension for scene analysis. *IEEE Transactions on Neural Networks, 16*, 1401-1426.

Wang, D.L. & Chang, P.S. (2008) An oscillatory correlation model of auditory streaming. *Cognitive Neurodynamics, 2*, 7-19.

Wang, D.L. & Terman, D. (1995) Locally excitatory globally inhibitory oscillator networks. *IEEE Trans. Neural Net. 6*(1), 283-286.

Warren, R.M, Obusek, C.J., Farmer, R.M. & Warren, R.P. (1969) Auditory sequence: confusion of patterns other than speech or music. *Science. 164*(3879), 586-7.

Winkler, I., Takegata, R. & Sussman, E. (2005) Event-related brain potentials reveal multiple stages in the perceptual organization of sound, 25(1), 291-299.

Woldorff, M.G. & Hillyard, S.A. (1990) Attentional influence on the mismatch negativity. *Behav. and Brain Sciences,* 13: 258-260

Woldorff, M.G., Hillyard, S.A., Gallen C.C., Hampson S.R. & Bloom F. E. (1998) Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology 35*, 283–292.

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

Wrigley, S.N. & Brown, G.J. (2004) A Computational Model of Auditory Selective Attention, *IEEE Transactions on Neural Networks*, *15*(5), 1151-1163.

Zatorre, R.J., Mondor, T.A. & Evans, A.C. (1999) Auditory Attention to Space and Frequency Activates Similar Cerebral Systems. *NeuroImage*, *10(5)*, 544-554.

Appendices

Appendix 1. Supplemental statistics for Experiment 1

	Early		Late	
	Only	Both	Only	Both
Sustained attention	75.2 (20.2)	74.0 (20.4)	69.0 (29.1)	64.0 (30.5)
Switched attention	---	---	81.2 (19.9)	79.6 (20.0)
Interaction between experimental condition and sequence type: $F(2,36)=0.13$, $p=.88$				
Effect of experimental condition: $F(2,36)=3.83$, $p=.03$, $r^2=.19$				
Effect of sequence type: $F(1,18)=0.03$, $p=.87$				

Table 1. Accuracy (in %) on deviant detection by sequence types

	Early		Late	
	Only	Both	Only	Both
Sustained attention	735 (106)	722 (101)	743 (146)	767 (120)
Switched attention	---	---	698 (73)	692 (88)
Interaction between experimental condition and sequence type: $F(2,36)=0.34$, $p=.71$				
Effect of experimental condition: $F(2,36)=2.97$, $p=.07$, $r^2=.16$				
Effect of sequence type: $F(1,18)= .01$, $p=.94$				

Table 2. Reaction times (in ms) on deviant detection by sequence types

	Early		Standard			
	Only	Both	Only early	Only late	Both	None
Sustained attention	---	---	99.7 (0.2)	99.6 (0.3)	99.5 (0.4)	99.5 (0.4)
Switched attention	99.4 (1.0)	99.6 (0.8)	99.8 (0.2)	99.8 (0.2)	99.5 (0.5)	99.5 (0.8)
	t(18)=		Interaction between experimental condition and sequence type: $F(3,36)=0.20$, $p=.90$			
			Effect of sequence type: $F(3,36)=2.07$, $p=.12$			
			Effect of experimental condition: $F(1,36)= 0.34$, $p=.56$			

Table 3. Correct rejections (in %) by sequence types

Appendix 2. Effect of Deviant Presentation Timing (Early, Late or Both)

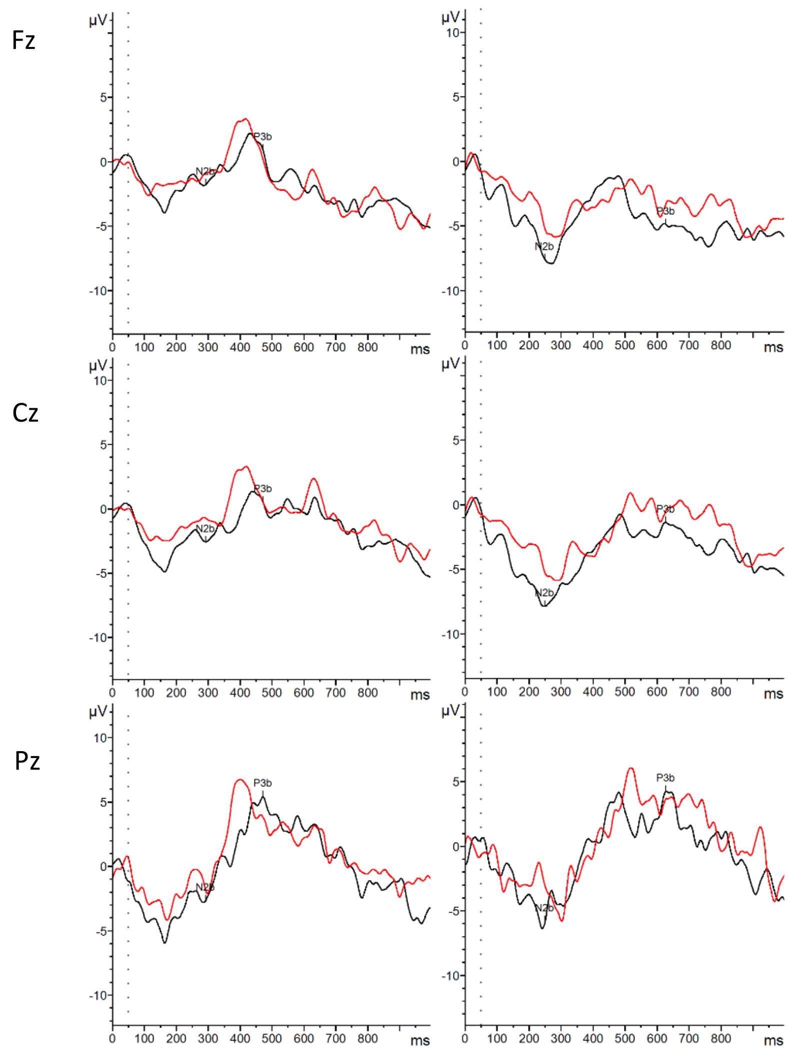


Figure 31. Effect of deviant presentation timing on ERPs. Left. Early deviants (Only – black, Both – red). Right. Late deviants (Only– black, Both – red). A negativity, the N2b, is apparent at about 250 ms for both deviants. A later positivity, the P3b, is apparent from around 350 to 650 ms for the Early deviants and from 400 to 750 ms for the Late deviants

Early-Both		Late-Both	
N2b (Cz)	P3b (Pz)	N2b (Cz)	P3b (Pz)
t(10)=0.95, p=.37, d=0.18	t(10)=0.06, p=.95, d=0.02	t(10)=1.83, p=.10, d=0.78	t(10)=0.96, p=.36, d=0.21

Table 4. ERP differences by sequence types

Appendix 3. Individual difference waves between early to-be-ignored deviants and standards

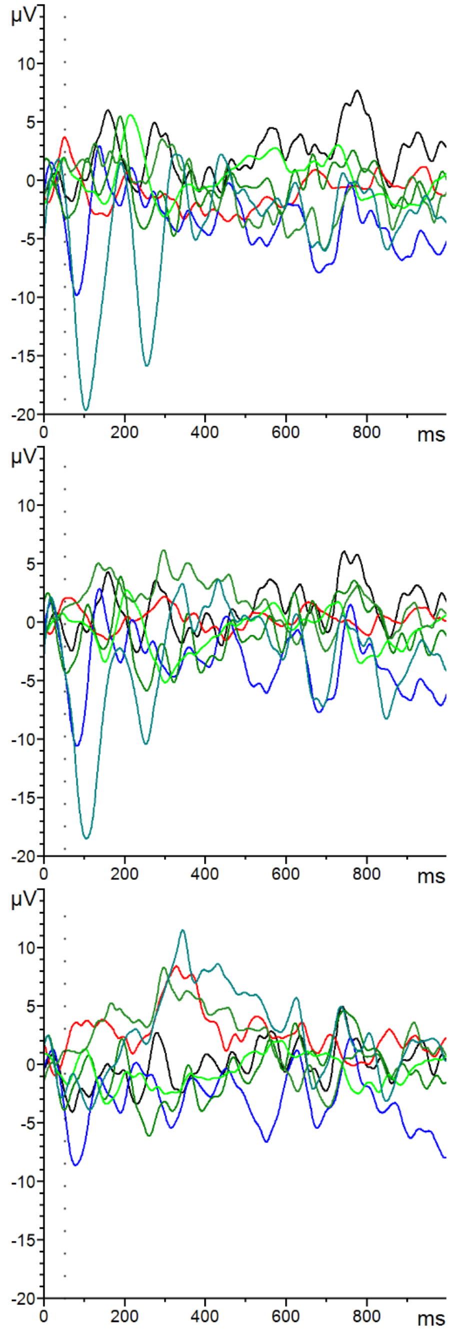


Figure 32. Individual difference waves between early to-be-ignored deviants and standards

Appendix 4. Difference waves for Experiment 1

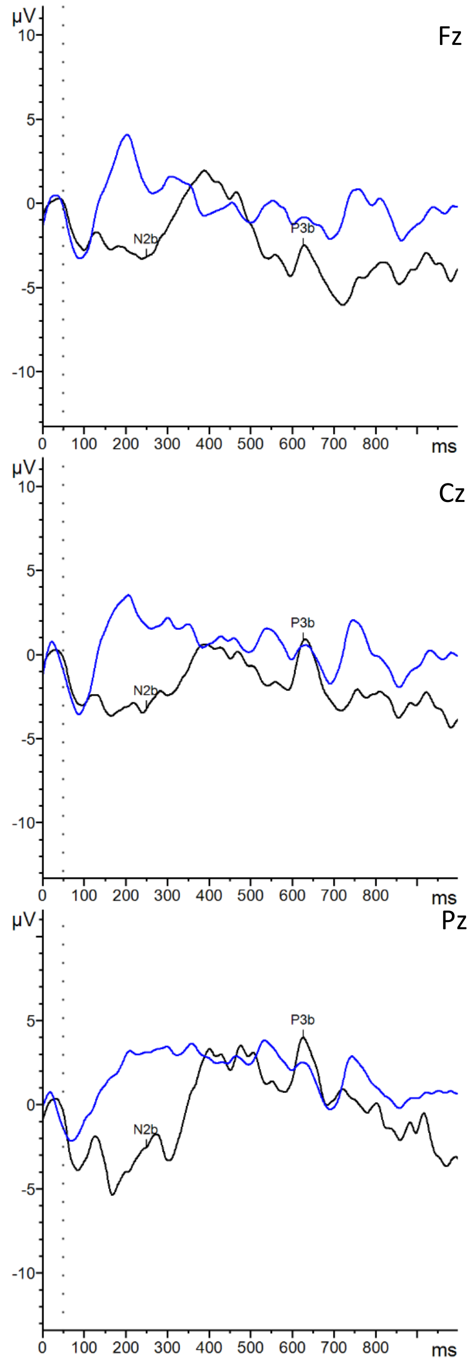


Figure 33. Difference waves between standards and deviants following presentation of the early deviants in the sustained and switched task conditions (Switched task condition - blue line and Sustained task condition - black line). The ERPs are time-locked to the physical onset of the stimulus (i.e., onset of B tone). The N2b, is apparent at about 250 ms for the attended deviant. A positivity, the P3a, is apparent at about 200 ms for the to-be-ignored deviant and the P3b apparent from 350 to 675 ms for the attended deviant. A large sustained positivity is present from 200 to 600 for the to-be-ignored deviant.

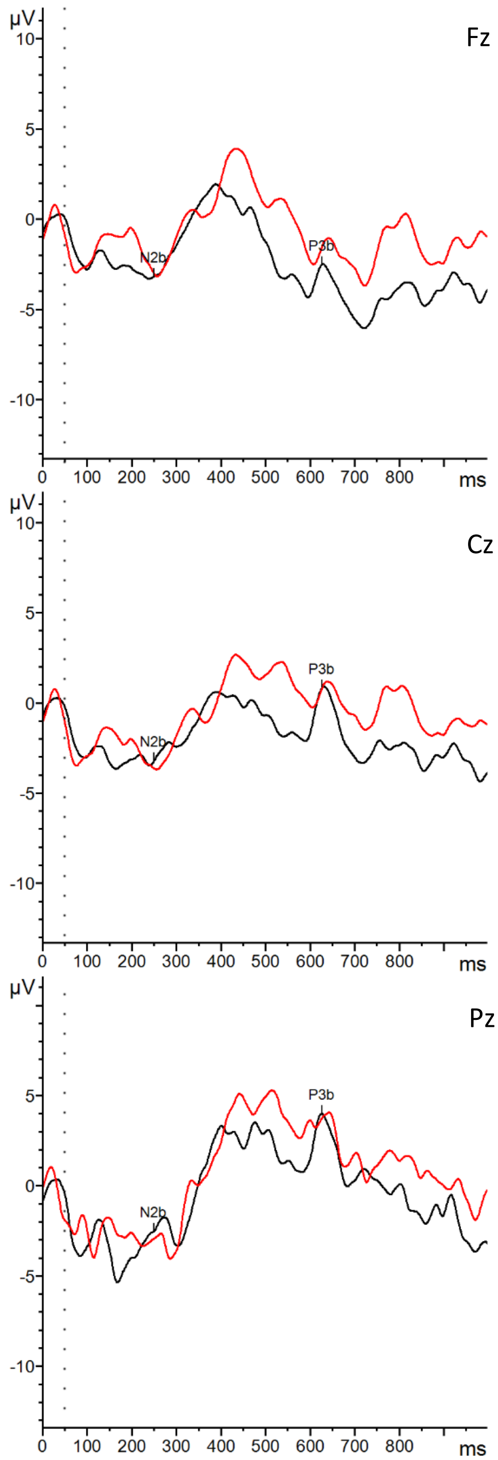


Figure 34. Difference waves between standards and deviants following the presentation of the deviants of the sustained task condition. (Early deviant - black line, Late deviant - red line). A negativity, the N2b, is apparent at about 250 ms at Fz and Cz for both deviants. A positivity, the P3b, is apparent at Pz from 350 to 675 ms at Pz for both deviants

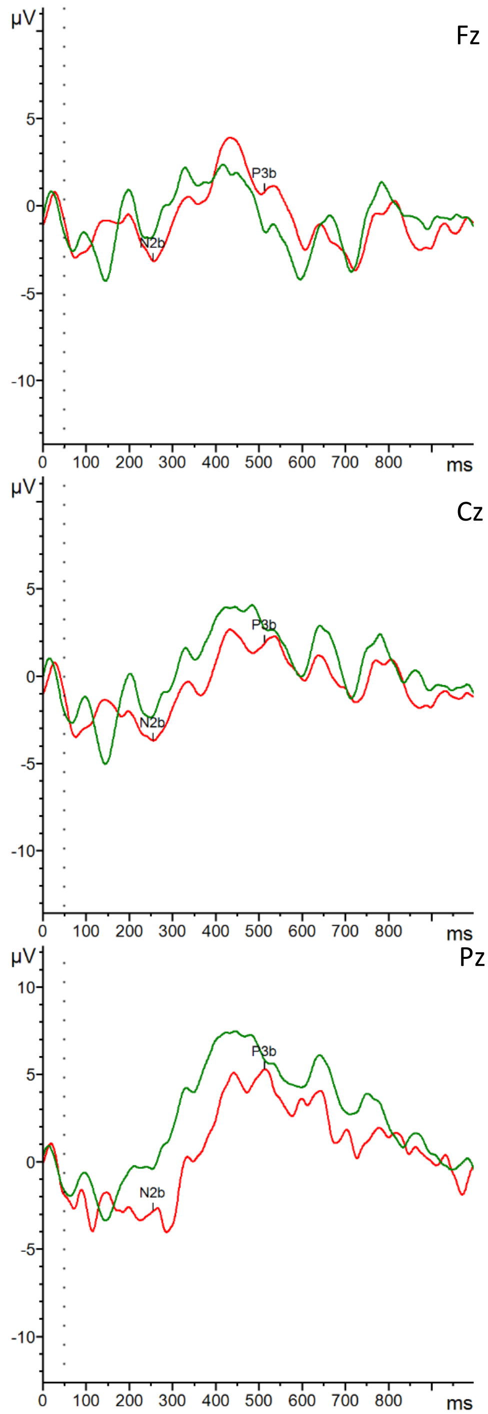


Figure 35. Difference waves between standards and deviants following presentation of the late deviants (sustained task – red line, switched task – green line). A negativity is apparent at about 150 ms at Fz and Cz for the late deviant in the switched condition. A later negativity, the N2b, is apparent at about 250 ms at Cz for both deviants. A positivity, the P3b, is apparent from 350 to 700 ms at Pz for both deviants.

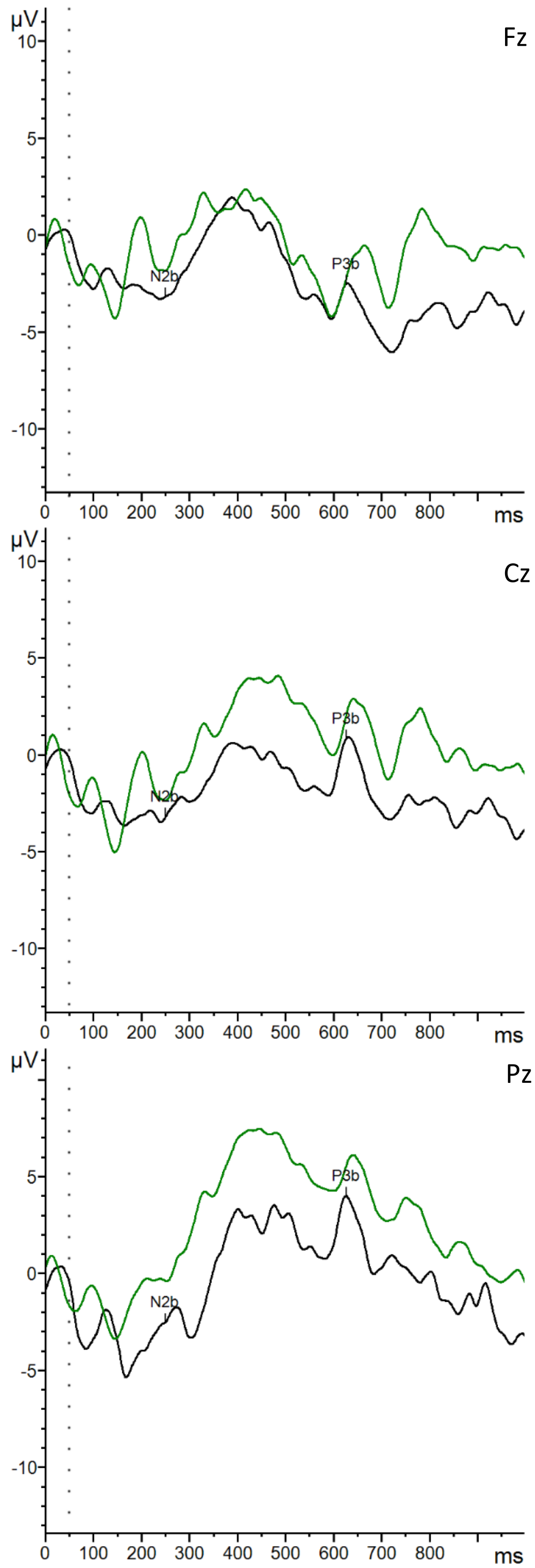


Figure 36. Difference waves between standards and deviants following presentation early deviant in the sustained condition (black line) and late deviant in the switched condition (green line). A N2b is apparent at about 250 ms at Fz and Cz for both deviants. A positivity, the P3b, is apparent from 350 to 700 ms at Pz for both deviants.

Appendix 5. Supplemental statistics for Experiment 2

	Early		Late	
	A frequency bandpass filtered	B frequency bandpass filtered	A frequency bandpass filtered	B frequency bandpass filtered
Sustained attention	80.2 (14.7)	86.0 (10.2)	87.6 (10.0)	93.1 (8.1)
Switched attention	---	---	86.4 (12.4)	88.6 (11.2)
Interaction between task and noise frequency: $F(1.34,26.75)=0.28$, $p=.67$				
Effect of deviant condition: $F(1.34,26.75)=3.78$, $p=.052$, $r^2=.16$				
Effect of noise: $F(1,20)=1.46$, $p=.24$				

Table 5. Accuracy (in %) on deviant detection by sequence types

	Early		Late	
	A frequency bandpass filtered	B frequency bandpass filtered	A frequency bandpass filtered	B frequency bandpass filtered
Sustained attention	820 (118)	791 (120)	679 (113)	669 (101)
Switched attention	---		802 (135)	794 (172)
Interaction between task and noise frequency: $F(1.33,26.52)=0.07$, $p=.86$				
Effect of deviant condition: $F(1.33,26.52)=11.36$, $p=.001$, $r^2=.36$				
Effect of noise: $F(1,20)=0.15$, $p=.71$				

Table 6. Reaction times (in ms) on deviant detection by sequence types

	Early to-be-ignored deviant		Standard			
	A frequency bandpass filtered	B frequency bandpass filtered	Early A frequency bandpass filtered	Early B frequency bandpass filtered	Late A frequency bandpass filtered	Late B frequency bandpass filtered
Sustained attention	---	---	99.9 (0.3)	99.9 (0.3)	100 (0.0)	99.8 (0.4)
Switched attention	84.7 (24.6)	78.7 (26.8)	96.5 (0.3)	94.5 (0.5)	95.2 (0.5)	96.0 (0.3)
	$t(20)=0.36$, $p=.73$		Interaction between task and noise: $F(3,40)=0.56$, $p=.65$			
			Effect of noise: $F(3,40)=0.49$, $p=.69$			
			Effect of task: $F(1,40)=53.63$, $p<.001$			

Table 7. Correct rejections (in %) by sequence types

Groups compared	Statistics		
	N2b at Fz	N2b at Cz	P3b at Pz
Interaction between noise frequency and experimental condition	F(3,30)=0.30, p=.82	F(3,30)=0.30, p=.82	F(3,30)=1.54, p=.23
Effect of noise frequency	F(1,10)=4.93, p=.051 r^2=.33	F(1,10)=2.64, p=.13 r^2=.21	F(1,10)=0.03, p=.87 r^2=.003
Effect of experimental condition	F(3,30)=1.22, p=.32 r^2=.11	F(3,30)=1.00, p=.41 r^2=.09	F(3,30)=14.33, p<.001 r^2=.09

Table 8. Statistics for ERPs of Experiment 2

Appendix 6. Power analyses

RTs Experiment 1: A sample size of 27 would have been needed to find a significant difference between the RTs, for an effect size of 0.16 (found effect size), an alpha of 0.05, a power of 0.80 (the observed power was 0.54) and a correlation of 0.60 (mean correlation between the 3 repeated variables).

Accuracies Experiment 2: A sample size of 36 would have been needed to find a significant difference between the accuracies, for an effect size of 0.16 (found effect size), an alpha of 0.05, a power of 0.80 (the observed power was 0.53) and a correlation of 0.43 (mean correlation between the 3 repeated variables).

Effect of noise on accuracy Experiment 2: A sample size of 268 would have been needed to find a significant difference between the two bandpass filtered noises, for an effect size of 0.07 (found effect size), an alpha of 0.05, a power of 0.80 (the observed power was 0.21) and a correlation of -0.12.