Effects of Attention on Change Deafness

by

Kristina Carol Backer

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Graduate Department of Psychology
University of Toronto

© Copyright by Kristina C. Backer 2010
Effects of Attention on Change Deafness

Kristina C. Backer
Master of Arts
Graduate Department of Psychology
University of Toronto
2010

Abstract

Detecting acoustic changes in our environment, such as a rattlesnake’s sudden approach, can be essential for survival. Although auditory change detection has been intensively investigated using sequentially-presented sounds, very little is known about how we detect changes in a natural, complex scene comprised of multiple concurrent sounds. The present study used a location-switch change deafness paradigm; on each trial, participants listened to two consecutive auditory scenes, consisting of three natural sounds played in distinct locations, and reported if the two scenes were the same or identified the two sounds that switched locations. Directing a listener’s attention to a changing sound improved accuracy and decreased reaction time, relative to uncued trials. However, when participants’ attention was invalidly directed to a non-changing sound object, performance suffered. Further analyses showed that these effects could not be explained by energetic masking. Thus, attention is necessary for change identification in complex auditory scenes.
Acknowledgments

I would like to thank my supervisor, Dr. Claude Alain, for his support and expert guidance in the experimental design and data analysis, as well as Patricia van Roon for technical assistance.
# Table of Contents

Acknowledgments.......................................................................................................................... iii  
Table of Contents.......................................................................................................................... iv  
List of Tables ................................................................................................................................. vi  
List of Figures ............................................................................................................................... vii  
Chapter 1 Background Literature.................................................................................................. 1  
  1 Auditory Scene Analysis ............................................................................................................ 1  
  2 Visual Change Detection ............................................................................................................ 2  
  3 Auditory Change Detection ....................................................................................................... 9  
    3.1 Oddball and Mismatch Negativity ...................................................................................... 9  
    3.2 Change Deafness .............................................................................................................. 11  
Chapter 2 Experiment 1 ............................................................................................................... 15  
  4 Experiment 1 Methods ............................................................................................................. 15  
    4.1 Participants ..................................................................................................................... 15  
    4.2 Stimuli ............................................................................................................................ 15  
    4.3 Procedure ....................................................................................................................... 16  
    4.4 Data Analysis ................................................................................................................... 17  
  5 Experiment 1 Results ............................................................................................................... 17  
Chapter 3 Experiment 2 ............................................................................................................... 19  
  6 Experiment 2 Methods ............................................................................................................. 19  
    6.1 Participants ..................................................................................................................... 19  
    6.2 Stimuli ............................................................................................................................ 20  
    6.3 Procedure ....................................................................................................................... 21  
    6.4 Data Analysis ................................................................................................................... 22  
  7 Experiment 2 Results ............................................................................................................... 22
List of Tables

Table 1. Experiment 1 Performance.
List of Figures

Figure 1. Experiment 1 Trial Structure. The first auditory stimulus was an auditory scene comprised of three distinct sounds each played in a different location, indicated by the different colors on the schematic. The second auditory stimulus was a noise burst, shown in grey. Tick marks along the time axis show the onset and offset of the auditory scene (1005 ms), noise burst (500 ms), and visual semantic cue (500 ms).

Figure 2. Experiment 2 Trial Structure. This schematic shows an example of a valid Location cue Change trial. Like Figure 1, tick marks along the time axis indicate the onset and offset of the visual cue (500 ms), auditory scenes (1005 ms), and noise burst (500 ms).

Figure 3. Change Deafness Effect. Panel a) shows accuracy results, and b) RT results from Experiment 2. For this figure and subsequent figures, asterisks reflect Bonferroni-corrected post-hoc comparisons while error bars reflect standard error of the mean.

Figure 4. Benefit of Valid Cue. Directing attention to a changing sound results in a) improved accuracy and b) faster RTs relative to uncued trials.

Figure 5. Cost of Invalid Cue. Directing attention away from a changing sound results in a) decreased accuracy and b) slower RTs relative to uncued trials.
Chapter 1
Background Literature

In this chapter, I will first provide a brief introduction to auditory scene analysis. Next, I will give an overview of visual, followed by auditory, change detection. Throughout the chapter, I will discuss the roles working memory (WM) and attention play in change detection.

1 Auditory Scene Analysis

Our auditory environment often comprises a myriad of sounds from various sources such as multiple talkers, reverberations, and other background noise that usually occur while riding the subway to work or dining at a restaurant. Over the last 30 years, researchers have identified numerous cues that help listeners to sort the incoming acoustical data into distinct sound sources, hereby referred to as auditory objects or streams. For instance, sounds with common spatial locations, onsets, and spectral profiles usually originate from the same physical source and therefore are usually assigned (i.e., perceived) as belonging to the same perceptual object (Bregman, 1990). The sounds that surround us often change in a predictive manner such that a certain sound may lead us to expect the next one, such as in music or during a conversation. Listeners capitalize on stimulus invariance and predictability to detect changes in the auditory environment.

The psychological and neural mechanisms underlying the detection and perception of changes in our auditory environment have been extensively investigated using the oddball paradigm in which infrequent oddball stimuli disrupt regularities within an auditory sequence. Although this paradigm has enhanced our understanding of the processes involved in detecting and orienting our attention to rare or novel acoustic events, it does not tell us how listeners notice the appearance or disappearance of sounds or changes in the configuration of our auditory scene. In everyday listening situations, our acoustic environment changes constantly, and observers must be able to keep track of multiple sound objects. Once an auditory scene has been parsed into its component objects, selectively attending to one stream (e.g., your friend’s story), ignoring all the other talkers or background noise, is crucial for effective communication, especially in acoustically adverse or “cocktail party” environments (Cherry, 1953). The current research uses a change deafness task to examine the role attention plays in noticing changes in our acoustic surroundings. Since change blindness, the visual analogue of change deafness, has been
extensively studied, it is helpful to review visual change detection before discussing auditory change detection in more detail.

2 Visual Change Detection

In vision, change detection is often studied using a phenomenon known as change blindness. Change blindness is an observer’s surprising failure to notice salient or large changes to a naturalistic picture (e.g., landscape photograph or indoor scene), familiar objects (e.g., faces), or to one object within an array of abstract objects (e.g., rectangles), when the change occurs during some sort of brief interruption such as a saccade, eye blink, mudsplash, or movie cut (see Simons & Rensink, 2005 for a review). Usually motion transients, which occur in the absence of an interruption, draw our attention to the change, enabling us to easily notice it (Klein, Kingstone, & Pontefract, 1992).

Two types of paradigms are commonly used to study change blindness: the one-shot paradigm and the flicker paradigm (for reviews, see Rensink, 2000a; Simons & Rensink, 2005). In the one-shot, or delayed match-to-sample, paradigm, two visual scenes are briefly shown in sequence with a brief interruption (e.g., blank screen, mudsplash, etc.) presented between the two scenes to prevent any motion transients that would facilitate change detection. Usually, in about 25 to 50% of the trials, the two scenes are identical, and the remaining proportion of trials includes some sort of change in one item of the picture or stimulus array (e.g., color change, location change, one item replaced by another, etc.). Participants report if they thought the two scenes were the same or different, and accuracy and, occasionally response time (RT), are used as the outcome measures. Since the one-shot method may not give observers enough time to encode the scenes into WM, Rensink and colleagues developed the flicker paradigm. In the flicker paradigm, the one-shot sequence is shown repeatedly with a blank screen separating each presentation of the visual scene until the observer sees the difference between the two scenes (Rensink, O'Regan, & Clark, 1995, 1997). Thus, response time, instead of accuracy, is usually used as the outcome measure. Even using the flicker paradigm, which should provide the observer with adequate time to form robust representations of the two scenes, participants often remain blind to the change, commonly for up to 50 seconds (Rensink, 2000a).

In order for an observer to successfully detect a change between two scenes, it is presumable that the following processes must occur: 1) encoding of at least the changing item within a visual
scene into WM (Jolicoeur & Dell'Acqua, 1998; Vogel, Woodman, & Luck, 2006), 2) retention of
the WM contents throughout the interruption phase and the presentation of the second scene
(Gold & Green, 2005; Spencer & Hund, 2002), 3) perceptual representation of the second scene
(e.g., sensory or WM), and 4) comparison between the pre- and post-change representations
(Mitroff, Simons, & Levin, 2004). Currently, it is unclear at which stage(s) of WM processing
(i.e., encoding, retention, and comparison) change detection fails. Furthermore, it is well-known
that attention plays a significant role in change detection, but how attention and WM interact to
yield successful change detection remains unresolved.

Several hypotheses, related to WM and attention, have been presented to explain why change
blindness occurs (see Simons & Rensink, 2005, for a review). The first is the encoding-failure
hypothesis, which postulates that if the neural representation of the pre- and/or post-change
scenes is of poor quality, then change detection will fail (Brady, Konkle, Oliva, & Alvarez, 2009;
Levin, Simons, Angelone, & Chabris, 2002; O'Regan & Noe, 2001; Rensink, 2002a). This
encoding failure can result from directing attention away from the changing object, leading to a
weak representation of the changing object in at least one of the scenes (Koivisto & Revonsuo,
2005; Rensink, 2002a). This logic draws upon evidence that observers can search for change in
up to four or five objects (Luck & Vogel, 1997; Pashler, 1988; Rensink, 2000b), which is in
accordance with attentional (Pylyshyn & Storm, 1988; Rensink, 2000b; Scholl, Pylyshyn, &
Feldman, 2001) and WM (Cowan, 2001) capacity estimates. However, even though we can
distribute our attention across approximately four objects, only one change can be noticed at any
moment, known as change simultagnosia (Rensink, 2002b).

Other change blindness studies have demonstrated that focused attention is necessary to notice a
change. For example, change detection improves when the change relates to a semantically
central or “interesting” item in the scene (Rensink, et al., 1997), even when the changes are
matched for salience (Kelley, Chun, & Chua, 2003). These findings suggest that objects that
exogenously attract the observer’s attention have a more detailed representation in WM (Simons
& Rensink, 2005) and subsequently decrease change blindness (Scholl, 2000). Likewise, valid
endogenous cues directing an observer’s attention to the changing item have been shown to
increase accuracy and decrease reaction time on a change blindness paradigm in both humans
(Becker, Pashler, & Anstis, 2000; Koivisto & Revonsuo, 2005; Landman, Spekreijse, & Lamme,
2003) and macaques (Cavanaugh & Wurtz, 2004). Also, expertise may influence change
detection, as football experts see changes in football scenes more quickly than non-experts, presumably because experts can scan their attention over the image more quickly and efficiently than non-experts (Werner & Thies, 2000). The role of attention in change detection has also been revealed through several studies employing scalp recording of event-related brain potentials (ERPs). These studies have revealed a positive wave at parietal sites, referred to as the P3 wave, which is enhanced in amplitude during change detection relative to change blindness trials (Fernandez-Duque, Grossi, Thornton, & Neville, 2003; Koivisto & Revonsuo, 2003; Niedeggen, Wichmann, & Stoerig, 2001; Pourtois, De Pretto, Hauert, & Vuilleumier, 2006).

To further understand how attention, encoding, and change detection relate, it is necessary to discuss how we create a mental representation of our visual world. The mainstream theories on this issue involve the construction of a finite number of fairly complex structures (see Rensink, 2002, for a review) such as object files (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992) which are then used by the change detection process. An object file, a temporary, abstract representation of an object in the visual environment, is created when attention is focused on that particular object. The physical feature information in each object file is updated when any aspect of that object changes (e.g., position change, occlusion by another object, etc.). Furthermore, each object file contains information about both attended and unattended properties of an object, and this physical information can be compared to object representations in long-term memory for identification purposes (Kahneman, et al., 1992). Object files, once formed by focused attention, can exist for a limited time even in the absence of attention (Kahneman, et al., 1992). Specifically, two or three previously-attended objects, in addition to the currently-attended item, can be retained in visual WM (Irwin, 1992; Irwin & Andrews, 1996; Irwin & Zelinsky, 2002).

Focused attention surely plays a role in change detection, but it does not completely explain the existence of change blindness. Observers still fail to notice changes even when the objects are to be attended (Caplovitz, Fendrich, & Hughes, 2008; Levin & Simons, 1997; Simons & Rensink, 2005), especially when the change is unexpected (Triesch, Ballard, Hayhoe, & Sullivan, 2003; Williams & Simons, 2000). This failure can be explained by assuming that the representation of a complex visual scene consist of several object files, which contain sparse information, thereby leaving an incomplete representation of the visual environment (Becker & Pashler, 2002; Kahneman, et al., 1992; Levin & Simons, 1997). Furthermore, the observer can make a task-
based selection of features to input into WM, screening out the irrelevant feature dimensions (Rensink, 2000b; Triesch, et al., 2003). Luck and Vogel’s (1997) results suggested that features are held in WM as bound entities, rather than represented by parallel processes. Wheeler and Treisman (2002) were unable to replicate this finding, but instead found evidence supporting the idea that features along different dimensions (e.g., color, location, or shape) are maintained in WM as parallel, disparate pieces that can only be bound together by focused attention, creating our coherent, rich visual experience. They hypothesized that if different feature dimensions are maintained bound in WM, then detecting changes to two conjoined feature dimensions (e.g., two shapes switch colors) should be as accurate as change detection of only one feature dimension (e.g., the colors of two shapes are replaced by two novel colors). In a series of change blindness experiments, observers were less accurate in detecting changes in the conjunction between two features than in detecting changes to only one feature dimension, supporting the notion that feature dimensions are represented in parallel streams in WM (Wheeler & Treisman, 2002).

In summary, change blindness due to an encoding failure may occur if focused attention is drawn away from the changing item, especially at the time of the change, or if our visual representations are sparse or incomplete as some studies have suggested (Becker & Pashler, 2002; Kahneman, et al., 1992; Levin & Simons, 1997; Levin, et al., 2002). Both of these cases can explain the existence of change blindness to a change in one or more feature dimensions, due to an encoding failure. Failed binding can explain change blindness when the change involves a conjunction between two feature dimensions (Wheeler & Treisman, 2002), assuming that both feature dimensions are accurately encoded. However, if one or both features are improperly encoded, then a binding failure is merely a downstream effect of a WM encoding failure. Thus, there is strong evidence showing that change blindness can result from inaccurate encoding, reflecting the interaction between WM and focused attention needed to create an internal representation of our visual environment and the nature of the encoded contents.

The retention- and comparison-failure hypotheses have also been used to explain change blindness. The retention-failure hypothesis postulates that the pre-change representation is lost during the interruption between the two scenes or is overwritten by the post-change representation (e.g., Beck & Levin, 2003; Becker, et al., 2000; Brawn & Snowden, 1999; Landman, et al., 2003; Shapiro, Caldwell, & Sorensen, 1997; Tatler, 2001). Two studies (Becker, et al., 2000; Landman, et al., 2003) used attentional cueing paradigms to test the
retention-failure hypothesis. Both compared change detection performance on a delayed-match-
to-sample change blindness paradigm when observers were given either no cue or a cue at
various time points from the onset of the first scene to the onset of second scene. Landman and
colleagues (2003) showed that cueing up to approximately 600 ms after the offset of the first
scene can improve change detection. Both studies concluded that if attention is not employed
within a short time after the offset of the first scene, the contents in iconic memory (which is a
volatile form of short-term memory with contents decaying within a few hundred milliseconds
after the disappearance of a visual stimulus (Coltheart, Laming, Routh, & Broadbent, 1983;
Gegenfurtner & Sperling, 1993; Sperling, 1967)) will not be transferred to WM, and the second
scene will overwrite the brain’s representation of the first scene.

Furthermore, both studies found that a cue presented simultaneously with the second scene (i.e.,
post-change cue) did not enhance change detection performance relative to a no-cue condition
provides evidence against the comparison-failure hypothesis, which states that the comparison
process fails due to limited processing resources even though both pre- and post-change
representations exist (e.g., Hollingworth, 2003; Hollingworth & Henderson, 2004; Mitroff, et al.,
2004; Simons, Chabris, Schnur, & Levin, 2002). With a post-change cue, the observer would
only need to compare the cued object, rather than each object in both arrays, from the pre- and
post-change representations, assuming that this comparison should require few processing
resources, contradicting the comparison-failure hypothesis (Landman, et al., 2003). However,
another study found the opposite results; a post-change cue facilitated change detection, by
constraining retrieval and comparison processes to a target (Hollingworth, 2003). It is important
to note, as Landman and colleagues (2003) remarked, that Hollingworth’s study used pictures of
real visual scenes (e.g., a room) rather than abstract stimuli (letters in Becker et al., 2000;
rectangles in Landman et al., 2003). Natural stimuli, which tend to include redundant
information and rich details, may result in a much more stable, high-level representation
compared to artificial stimuli (Hollingworth, 2003), and the efficacy of a post-change cue may
increase with the stability of the visual representation, leading to these contradicting results.

Several other studies have shown evidence in favor of the comparison-failure hypothesis
(Angelone, Levin, & Simons, 2003; Mitroff, et al., 2004; Simons, et al., 2002). In a clever
experiment involving a real-life situation, a researcher disguised as an athlete holding a white
and red basketball asked a pedestrian for directions to the gym (Simons, et al., 2002). A crowd of confederates then interrupted the conversation and the ball was conspicuously removed. Upon debriefing, some participants, who did not notice that the ball had disappeared, could still accurately describe the ball, suggesting that they had encoded and retained a visual representation of the ball, which was only retrieved upon receiving an effective retrieval cue (i.e., a specific question about what the athlete was carrying) (Simons, et al., 2002). In a similar study, a camera cut separated pre- and post-change videos (Angelone, et al., 2003). Observers who failed to detect the change could still identify the pre-change item from a line-up of photographed items in the video at the same above-chance level as those who had detected the change. This suggests that participants were able to encode and retain a representation of items in the scene, but failed to compare this representation with the post-change video footage (Angelone, et al., 2003). In another study, Mitroff et al. (2004) provided further evidence for the comparison-failure hypothesis. The authors presented either four, six, or eight line drawings of familiar objects to participants and asked them to answer two types of questions on each trial: a change-detection question and an encoding question. The encoding question was a 2AFC one, in which participants had to choose which of two objects was present in either the pre- or post-change scene, or both scenes. The results demonstrated that even when participants failed to detect a change, they were still able to answer the object encoding questions above chance levels, providing evidence that a comparison failure can explain change blindness, since the pre- and post-change representations were accessible (Mitroff, et al., 2004).

In a series of experiments, Hyun and colleagues (2009) measured manual (i.e., via a button press) and saccadic reaction times, as well as event-related potentials (ERPs) to characterize the neural mechanisms underlying the comparison process during change detection. Their results suggested that an unlimited-capacity, feature-specific comparison process may play a role in change detection, which in turn directs either overt or covert attention to a potentially-changing object. They posit that a second limited-capacity comparison process is needed to confirm that the now-attended object is truly different from its representation in WM before making a manual response (Hyun, Woodman, Vogel, Hollingworth, & Luck, 2009). However, it is currently unclear at which comparison stage change detection fails, and future research should be conducted to clarify this issue.
The studies reviewed above suggest that change blindness could occur either at a specific stage of WM processing (i.e., encoding, retention, or comparison/retrieval), or as several EEG studies suggest, change blindness could also result from failures at multiple processing stages. Because of its high temporal resolution, EEG (or magnetoencephalography, MEG) is an appropriate method for investigating neural activity during change blindness to determine the processing stage at which change detection fails. Most of the published change detection EEG studies have reported differences in activity between change-detected and change-blindness trials after the post-change scene (e.g., Eimer & Mazza, 2005; Schankin & Wascher, 2007; Turatto, Angrilli, Mazza, Umilta, & Driver, 2002), which likely reflect the perceptual consequences of change blindness, rather than the processing failure that leads to change blindness. Instead, key differences in brain activity between change-detected and change-blindness trials may arise prior to the change and provide deeper insight into the processing stage(s) at which change detection fails.

Koivisto and Revonsuo (2005) recorded ERPs during a change blindness task, in which the pre-change display was accompanied by either no attentional cue, a valid cue, or an invalid cue. Participants reported if any of the eight rectangles in the display changed orientation from the pre- to post-change display. ERPs prior to the post-change display (i.e., during encoding and retention of the first scene) showed effects of focused attention, which predicted change blindness on uncued and invalid cue trials and successful change detection on the validly cued trials (Koivisto & Revonsuo, 2005).

In another study, Pourtois and colleagues (2006) analyzed ERPs in response to a change blindness task, in which participants reported if one of two simultaneously-presented faces changed from the first to second display. Like Koivisto & Revonsuo (2005), they also found an early effect of attention prior to the post-change scene. Specifically, the P1 wave (a positive deflection at occipital sites that peaks at ~100 ms after stimulus onset) elicited by the first pair of faces was enhanced when the change was detected relative to change-blindness trials. This P1 effect likely reflects better sensory encoding of the whole display via distributed attention (Pourtois, et al., 2006). They also found a sustained negative CNV-like (contingent negative variation) activity at lateral frontal electrodes prior to the change that was more negative during change blindness trials than either no change or detected change trials. This effect could reflect better encoding of the pre-change display into WM, better preparation for detecting changes in
the post-change scene, or the retention of the pre-change display in WM (Pourtois, et al., 2006). This study also revealed a larger N170 (negative deflection at ~170 ms over the occipital-parietal scalp regions) during change blindness and a larger P3 during change detection relative to change blindness, which was also found in Koivisto and Revonsuo’s study (2005).

Another recent EEG study examined oscillatory data from a change detection task using human face stimuli and found enhanced gamma activity throughout the entire trial, as well as increased theta activity during the encoding period of the first face on trials in which the change was successfully detected (Park, Min, & Lee, in press). The authors attributed the enhanced gamma and theta oscillations to attention and memory encoding processing, respectively. Also, decreased beta activity occurred during the retention and comparison stages of the change-detection trials (Park, Min, & Lee, in press). Together, these three studies show that pre-change differences in neural activity are predictive of successful change detection, but it is not clear if change blindness results as a combination of the pre-change effects or if the earliest effect (e.g., P1 in Pourtois et al., 2006) initializes a cascade of deficits, evidenced by the subsequent effects. Additional EEG or MEG studies are needed to better understand differences in neural processing that underlie change blindness.

3 Auditory Change Detection

3.1 Oddball and Mismatch Negativity

Contrary to visual change detection, most of the work on auditory change detection has focused on pre-attentive, rather than explicit, change detection. Auditory change detection is most commonly studied using an oddball paradigm, in which rare deviant or oddball stimuli are embedded within a repetitive sequence of standard sounds, usually during EEG or MEG recording. Deviant sounds typically elicit a mismatch negativity (MMN or MMNm, the MEG equivalent), which is a sustained negativity that usually peaks at 150 to 250 ms after deviant sound onset, thereby overlapping with the N1 and P2 components (see (Naatanen, Paavilainen, Rinne, & Alho, 2007) for a review). Furthermore, the MMN latency usually decreases as the magnitude of deviance increases. The MMN is evoked by a change in any repeating feature during auditory stimulation, and deviations in many different acoustic properties, such as intensity, frequency, timbre, location, and direction of frequency glide, have been shown to elicit an MMN. Also, changes in temporal features, including rise time, ISI, stimulus order, stimulus
omission, and sound duration, can evoke an MMN response. Other studies have even reported an MMN to violations of abstract rules, deviant speech stimuli, and pattern deviations in musical stimuli (see (Naatanen, et al., 2007; Picton, Alain, Otten, Ritter, & Achim, 2000) for reviews).

Current theories posit that the MMN is triggered by a mismatch between the current sensory input and the prediction of what it should be, based on the implicit sensory memory representations of the past inputs. The MMN is thought to index a relatively automatic process because it can be recorded during passive listening (no response required) (e.g., Alain, Woods, & Ogawa, 1994; Naatanen, Gaillard, & Mantysalo, 1978; Naatanen & Michie, 1979). However, studies that have manipulated listener’s attention have shown that the MMN was attenuated (e.g., Alain & Woods, 1997; Alain, et al., 1994; Naatanen, Paavilainen, Tiitinen, Jiang, & Alho, 1993) or even abolished (Arnott & Alain, 2002; Woldorff, Hackley, & Hillyard, 1991) when participants focused on a concurrent stream of sounds. Many recent studies have focused on discovering the neural underpinnings of this automatic change-detection system. The MMN has at least two distinct generators, one in the supratemporal plane, which is associated with an implicit change detection process, and another in the frontal lobe, which is thought to underlie a subsequent attention switch to the novel or irregular sound (Naatanen, et al., 1978; Naatanen & Michie, 1979; Rinne, Alho, Ilmoniemi, Virtanen, & Naatanen, 2000). Furthermore, a lesion study showed that a cortical network, including the posterior auditory cortex contralateral to the stimulated ear as well as the bilateral dorsolateral prefrontal cortex, maintains the storage of transient sensory memory traces, thereby underlying MMN elicitation (Alain, Woods, & Knight, 1998). A recent EEG/fMRI study provided evidence for a hierarchical neural network underlying this novelty-detection system (Schonwiesner, et al., 2007). This study proposed that first activity in the primary auditory cortex (or earlier) reflects the presence of a change, perhaps via a stimulus-specific adaptation mechanism (Ulanovsky, Las, & Nelken, 2003), followed by more detailed feature analysis in higher auditory areas (i.e., posterior superior temporal gyrus and planum temporale). Finally, the mid-ventrolateral prefrontal cortex is thought to play an important role in determining whether sufficient novelty has been detected to allocate attentional resources to this deviant sound (Schonwiesner, et al., 2007). Therefore, the MMN reliably reflects the detection which can initiate attention shift (indexed by the P3a) to a novel or rare acoustic event, using patterns of past sensory inputs to predict future events in a probability-based or Bayesian manner. However, it is unknown if a similar neural mechanism underlies the
detection of unpredictable changes (i.e., sounds presented in the absence of any particular pattern) in the acoustic environment, such as changes in the configuration of an auditory scene and the sudden disappearance or appearance of sounds within the acoustic environment—that is, the type of changes used in change deafness paradigms.

3.2 Change Deafness

Change deafness is a listener’s failure to detect rather “obvious” changes in the acoustic environment. Thus far, surprisingly few behavioral (and no electrophysiological or neuroimaging) change deafness studies have been published. The notion of change deafness was first conveyed in Cherry’s seminal work on the cocktail party problem (1953). In one experiment, continuous speech passages were presented in one channel to participants, who were required to repeat them out loud, or shadow them. They ignored a concurrent passage played in the opposite headphone channel, in which various changes (e.g., speaker identity change, switch from English to German, etc.) occurred throughout the shadowing task. These changes in the unattended channel generally went unnoticed, hinting at the role of attention in auditory change detection (Cherry, 1953); also see (Moray, 1959; Treisman, 1960).

Follow-up work, using similar dual task paradigms, has been conducted in recent years. Vitevitch (2003) used a shadowing paradigm, in which participants were required to repeat lexically easy and difficult words, to further investigate the role of attention in change detection. He found that participants who focused on the lexical aspect of the words did not detect a change in the unattended indexical dimension (i.e., a voice change). However, participants who attended the indexical dimension of the words noticed a voice change. These results suggest that paying attention to a particular feature dimension is necessary to detect a change within that feature dimension (Vitevitch, 2003). Sinnett and colleagues (2006) presented participants with two rapidly-presented streams simultaneously, and participants were asked to monitor one of the two streams (i.e., either the sounds or common words). Participants later underwent a recognition test, and those who attended the words performed worse on the sounds and vice versa (Sinnett, Costa, & Soto-Faraco, 2006). Although these studies show the importance of attention in noticing changes in our acoustic environment, they are more analogous to visual studies of inattentional blindness (Mack & Rock, 1998), in which participants are not expecting and searching for a change, than of change blindness.
Other recent change deafness studies closely parallel change blindness experiments by using an auditory version of the delayed-match-to-sample paradigm, in which the first auditory scene, comprised of typically three to eight concurrently-presented sounds, is played, followed by a noise burst to “erase” sensory memory, and finally the second auditory scene, which is either identical to or different from the first scene. Changes used by various change deafness studies include the following: sound deletion or addition, the replacement of one sound by a novel sound, and a location switch between two sounds.

Eramudugolla and colleagues (2005) used a delayed-match-to-sample paradigm to more directly explore the role of attention in change deafness. Participants listened to two 5-second long auditory scenes, consisting of four, six, or eight realistic sounds (e.g., piano, bird chirping, etc.) on each trial. In a “nondirected attention” condition, listeners reported if the two scenes were the same or different (i.e., one sound disappeared or two sounds switched locations from the first to second scene). In a “directed attention” condition, they were visually cued to a particular sound and indicated if the cued sound was present or absent in the second auditory scene or, in the location switch experiment, if the cued sound switched locations with another sound. On both the sound deletion and location switch experiments, accuracy decreased as scene size increased when attention was not directed to any particular sound. However, when given a cue, participants performed at ceiling in both experiments, even when the scene consisted of eight sounds. These results suggest that directed attention is necessary to detect changes in complex auditory scenes, but the stage of auditory scene analysis (i.e., pre- or post-stream segregation) at which directed attention modulates change detection is unclear (Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005).

In this study, energetic and informational masking may have contributed to the observed change deafness effect, especially when attention was not guided to any particular sound. Masking arises when sounds overlap in acoustic properties, such as frequency and temporal envelope, and/or spatial location and can cause concurrent sound segregation to fail. If concurrently-presented sounds cannot be segregated from one another, a listener cannot accurately represent the sound objects within an auditory scene, which would often lead to a failure to detect changes between two scenes. Since change detection accuracy decreased with increasing scene size, change deafness can be easily confounded by masking among stimuli, which tends to increase with the number of concurrently-presented sounds. Furthermore, selective attention to a
particular sound can enhance concurrent sound segregation, such that attention acts as a filter to facilitate foreground-background decomposition of the scene and allows for the detection of changes in the attended foreground stream. In other words, the attentional effects observed in this study may originate from a facilitation of concurrent sound segregation by selectively attending to the cued sound.

Two additional recent studies examined the role of auditory WM in change deafness and reported conflicting results. Pavani and Turatto (2008), using a delayed-match-to-sample paradigm, presented participants with either three or four animal sounds in the first scene. From the first to the second scene, either a sound was added (i.e., from three to four sounds), a sound was deleted (i.e., from four to three sounds), or the two scenes were identical. In one experiment, participants indicated if the two scenes were the same or different. If they perceived a difference, they were then required to indicate if the change involved an addition or a deletion and also to identify which specific animal call had changed. When participants detected a change, they were very accurate (90% correct) in reporting if a sound had been added or deleted. However, participants could more accurately identify the change on addition (94%) than on deletion (68%) trials. The authors attributed this performance difference to the number of items presented in the first scene that participants were required to encode and maintain in WM throughout the trial (i.e., three on addition trials and four on deletion trials) (Pavani & Turatto, 2008). These results could also partly reflect energetic masking, since there is a smaller probability of masking among three concurrently-presented sounds than four. Furthermore, the addition of a novel sound to a scene can cause a change in neural firing (i.e., a release from adaptation) and can exogenously capture attention, thereby providing an alternate explanation for the more accurate change identification performance on addition trials.

However, Gregg & Samuel (2008) reported opposite results, in that change deafness did not result from a failure to encode the auditory scenes in WM. Participants were again presented with a one-shot change deafness paradigm, in which auditory scenes were comprised of four, five, or six familiar sounds and either one sound was replaced by a novel sound from the first to second scenes or the two scenes were identical. Like Mitroff, Simons, and Levin’s (2004) change blindness study, participants were required to answer both a change-detection and an object-encoding question on each trial. Similar to Mitroff and colleagues’ (2004) findings, participants consistently experienced change deafness even though the auditory scenes were
usually encoded accurately, supporting a comparison-failure explanation of change deafness (Gregg & Samuel, 2008). Furthermore, Gregg and Samuel (2008) showed that differences in acoustic cues can facilitate change detection. Specifically, when one sound was replaced by another that differed along both harmonicity and fundamental frequency conditions, participants were better at detecting the change than on trials in which the two sounds differed in only one acoustic feature. However, this acoustic manipulation had no impact on object encoding performance (Gregg & Samuel, 2008), since the introduction of a new sound can capture attention. Therefore, the cause of change deafness is not clear from the few published change deafness studies, but it seems that both attention and WM play a role, similar to change blindness accounts.

The present study consists of two experiments aimed first to demonstrate that change deafness cannot be explained by a failure to segregate an auditory scene due to energetic masking, and second to further investigate the role attention plays in auditory change detection, using a location switch task similar to that used in Eramudugolla et al.’s (2005) study.
Chapter 2
Experiment 1

The first experiment was designed to identify complex sounds that could be combined and yet yield unambiguous identification. This was important to ensure that change deafness does not arise as a result of energetic masking, or a failure to disentangle one sound from another. In other words, Experiment 1 ensures that all stimuli used in the subsequent change deafness experiment (i.e., Experiment 2) could be properly segregated and identified.

4 Experiment 1 Methods

The protocol described below was approved by the Toronto Academic Health Sciences Network (TAHSN) Research Ethics Board. All participants provided informed, written consent before participating in this study and were paid for their participation.

4.1 Participants

Eight (2 males) young (mean age = 24.4 years, range: 18 to 30 years), healthy individuals recruited from the local community participated in this experiment. No participants reported a history of neurological illness and were not currently taking any psychoactive medications. Before beginning the experiment, participants completed an audiogram to ensure normal pure-tone air conduction thresholds (≤ 25 dB HL) from 250 to 8000 Hz in both ears. Next, participants completed four blocks of the Etymotic Research QuickSIN Speech-in-Noise Test (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004) as a secondary hearing assessment. The QuickSIN Speech-in-Noise Test measures the signal-to-noise ratio (SNR) between a target sentence and multitalker babble that is required for a listener to understand half of the key words spoken by the target talker. To perform at or below the average SNR level (2 dB SNR) of healthy-hearing listeners, one must be able to successfully segregate concurrent sounds and be fluent in English. Overall, the group’s mean SNR was 1.22 ± 0.54 (standard error).

4.2 Stimuli

Each trio stimulus consisted of three concurrently-played familiar sounds (one non-speech human, one animal, and one musical sound), chosen from a large sound database. Each sound lasted 1005 ms and had a sampling rate of 12.21 kHz. To control for energetic masking, the
power spectral density (psd) of each individual sound was calculated using Welch’s method in MATLAB 7.1 (pwelch.m). For every possible pair of sounds within each possible stimulus trio, a correlation was done on the psds (i.e., three correlations per stimulus trio) in MATLAB. If the p-value for any of the three correlations per stimulus trio was less than or equal to 0.20, the sound combination was discarded. This process generated a list of over 28,000 unique, usable stimulus trios. From this list, 270 trios were randomly chosen, controlling for the number of times an individual sound was chosen, such that an individual sound could be used no more than six times. Also, no two sounds could occur together more than once. 198 unique individual sounds were used to make these 270 trios. To approximately equate perceived loudness, all sounds were normalized to the mean level and equalized based on equal loudness contours in Adobe Audition 1.5. The individual sounds were then combined into stimulus trios using in-house MATLAB code. Each of the three sounds was played in a distinct location (left, centre, and right) with the central location simulated by adding the central sound to both channels after reducing it by 6 dB to approximately match the loudness of the left and right sounds. Each sound type (i.e., human, animal, musical) was played in each location an equal number of times (i.e., 90). Also, a 500 ms white noise burst trio was created, using three uncorrelated noise signals placed at the left, central, and right locations.

4.3 Procedure

Participants were seated in a comfortable chair approximately 82 cm from a computer monitor. The chair was positioned so that two Grason-Stadler GSI speakers were positioned at ±45° relative to the participant’s head. Stimuli were played at approximately 59 dBA sound pressure level, but some sounds peaked at 83 dBC sound pressure level.

All participants completed six blocks of 45 trials per block. Visual and auditory stimuli were presented using Presentation software (version 13.0, NeuroBehavioral Systems). The trial structure is shown in Figure 1 and comprised: 1000 ms silent pre-stimulus period, then the stimulus trio (1005 ms), followed by 750 ms silence, then a 500 ms noise burst, and finally a visual cue (“human”, “animal”, or “music”) which appeared on the screen for 500 ms. Throughout each trial (except when the cue was shown), a fixation cross was displayed on the screen. Participants were instructed to report the location (left, centre, or right) of the cued sound by pressing one of three buttons. Participants were asked to respond as quickly and
accurately as possible. However, the task was self-paced, in that the computer waited for the participant’s response before presenting the next trial.

### 4.4 Data Analysis

Data analysis was done using in-house MATLAB code to calculate the mean RT and accuracy for each cue type (i.e., human, animal, music) across all participants. RT was calculated relative to the onset of the visual cue. An additional analysis was done to identify the trio stimuli that all participants answered correctly, all but one responded correctly, all but two were correct, etc. This was done to determine which trios were easily segregated and should be used in the change deafness experiment (Experiment 2).

### 5 Experiment 1 Results

In general, participants performed well on this task. The mean accuracy and RT for each cue type is shown in Table 1 below.

<table>
<thead>
<tr>
<th>Cue Type</th>
<th>Accuracy Mean ± SE</th>
<th>RT Mean ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>76.8 ± 2.9%</td>
<td>866.0 ± 65.2 ms</td>
</tr>
<tr>
<td>Animal</td>
<td>81.2 ± 3.0%</td>
<td>894.6 ± 64.4 ms</td>
</tr>
<tr>
<td>Music</td>
<td>84.8 ± 3.1%</td>
<td>823.4 ± 50.7 ms</td>
</tr>
<tr>
<td>Overall</td>
<td>80.9 ± 2.9%</td>
<td>861.1 ± 57.6 ms</td>
</tr>
</tbody>
</table>

All participants responded accurately on the same 84 stimulus trios, and seven of the eight participants were correct on the same 63 stimulus trios. Hence, a list of 147 trios was generated, and 126 of the stimuli from this list were used in Experiment 2. I listened to each of the 147 stimuli and hand-selected the most dissociable 126 trios (i.e., the trios, in which all three sounds, not only the cued sound, were the most clear). These 126 stimuli consisted of 82 trios that all participants could segregate and 44 that all but one person answered correctly. Therefore, the results of Experiment 1 showed that participants were able to successfully segregate all of the
stimuli that were used in Experiment 2. It is thus unlikely that any change deafness effects observed in Experiment 2 arise from energetic masking.
Chapter 3
Experiment 2

Experiment 2 was designed to extend findings by Eramudugolla and colleagues (2005), who showed that directing one’s attention to a changing sound via a visual cue can extinguish change deafness. In the current experiment, the change deafness task employed a delayed-match-to-sample paradigm, in which the two auditory scenes on each trial were identical or two sounds switched locations from the first to second auditory scene. Two types of endogenous visual cues were used: 1) location and 2) semantic, and I examined how these cues can facilitate change detection relative to trials on which no cues were presented. First, I hypothesized that validly-cued change trials (~71%) would result in faster RTs and increased accuracy compared to no cue trials. On the other hand, invalid cues (~29%) would result in slower RTs and poorer performance than no cue trials. Second, if acoustic scenes are encoded semantically, as shown by Gregg and Samuel (2009), then the valid semantic and location cues should result in an equivalent increase in accuracy and decrease in RT, relative to uncued trials. However, if location and semantic information is encoded separately, in accordance with the “what” and “where” auditory pathways (Alain, Arnott, Hevenor, Graham, & Grady, 2001; Rauschecker, 1998), then location cues should facilitate spatial change detection to a greater extent than semantic cues.

6 Experiment 2 Methods

The protocol described here was approved by the Toronto Academic Health Sciences Network (TAHSN) Research Ethics Board. All participants provided informed, written consent before participating in this study and were paid for their participation.

6.1 Participants

Seventeen young (6 males; 18 to 32 years), healthy individuals recruited from the local community participated in this experiment. None of these individuals participated in Experiment 1. All participants reported no history of neurological illness and were not currently taking any psychoactive medications. Participants were required to complete two sessions on two different days within the same week. Before beginning each session, participants completed an audiogram to ensure normal pure-tone air conduction thresholds (≤ 25 dB HL) from 250 to 8000
Hz in both ears. One male and one female were excluded from the study due to mild hearing loss. A second male was excluded after the first session because he did not follow task instructions. Of the participants who completed both sessions, two were excluded: one male misunderstood the instructions on the cue blocks, and one female performed at ceiling (above 95% correct on No Cue blocks), leaving no room for improvement on the Cue blocks. Consequently, all analyses include twelve participants (3 males; mean age = 21.4 years, range: 18 to 28 years).

6.2 Stimuli

Each stimulus trio consisted of three concurrently-played familiar sounds (one non-speech human, one animal, and one musical sound), as described in Experiment 1. 126 easily-segregated stimulus trios (comprised of 166 unique individual sounds) were chosen from Experiment 1. Since some participants in Experiment 1 remarked that the central sound was slightly quieter than the left and right sounds, the central sound was reduced by 4 dB, instead of 6 dB, to better equate loudness across all locations. A 500 ms white noise burst trio was created, using three uncorrelated noise signals placed at the left (-45°), central (0°), and right (+45°) locations. All stimuli were presented at 59 dBA SPL (peaking at 83 dBC SPL).

As described below, each trial consisted of two stimulus trios, or auditory scenes. The two auditory scenes could be either identical or different. If they were different, two sounds would switch locations from the first to the second scene. Thus, there were four types of trials: 1) Same (i.e., catch trials), 2) Left-Center (45°) switch, 3) Right-Center (45°) switch, and 4) Left-Right (90°) switch. Each of the 126 stimulus trios were used for each trial type, resulting in 504 trials (126 same and 378 different) per block type (No Cue, Location Cue, and Semantic Cue). Invalid cues were presented on 36 trials within each change condition during the cue blocks (totaling 108 invalid and 270 valid cues on change trials per block type). The location of each sound within each stimulus trio was randomly chosen, so that the same sounds were not presented in the same locations during all trial types and block types. However, within each block type, each sound category (i.e., human, animal, music) was played an equal number of times in each location (i.e., left, centre, and right).
6.3 Procedure

Participants were seated in a comfortable chair approximately 82 cm from a computer monitor. The chair was positioned so that two Grason-Stadler GSI speakers were positioned at ±45° relative to the participant’s head. The QuickSIN Speech-In-Noise test and a digit-span test (both forwards and backwards) were administered to participants, and each test was given during separate sessions. The order of administration was counter-balanced across participants. All participants completed 18 blocks of 42 trials per block (~5 minutes per block) during each session. Blocks were organized into clusters of three (one of each block type: No Cue, Location Cue, and Semantic Cue), and the block types within each cluster were randomly ordered. Also, the blocks (12 per block type) were randomly ordered, so that participants did not receive the same stimulus order within each block type.

Visual and auditory stimuli were presented using Presentation software (version 13.0, NeuroBehavioral Systems). As shown in Figure 2, trials were structured as followed: 1000 ms silent pre-stimulus period, then the first auditory scene (i.e., stimulus trio, 1005 ms), followed by 750 ms of silence, then a 500 ms noise burst, then another 750 ms of silence, and finally the second auditory scene. Throughout each trial (except when the cue was shown), a fixation cross was displayed on the screen. During cue blocks, the trial sequence was the same except that either a semantic (i.e., human, animal, music) or a location (i.e., left (<<), center (++), or right (>>)) visual cue was presented at 250 ms into the trial for 500 ms. As mentioned above, approximately 71% of the cues were valid and 29% invalid. Cues were assigned to stimuli randomly, so that the same stimulus trios were not always tied to invalid cues. Furthermore, within both the Location and Semantic Cue block types, each location and semantic category was cued an equal number of times.

Participants were told that they would hear two auditory scenes on each trial, and that these two scenes would either be exactly the same or two sounds would switch locations. Before beginning each block, they were also informed of the upcoming block type. During cue blocks, participants were also instructed to pay attention to the cued sounds. They were asked to report if the two scenes were the same or, if they noticed a difference, to identify the location switch that occurred (i.e., left and center, right and center, or left and right) by pressing one of four buttons. Participants were asked to respond as quickly and accurately as possible and to guess if
they were not sure. They were given 2 seconds after the offset of the second auditory scene to respond before the start of the next trial.

6.4 Data Analysis

Mean RT and accuracy for Same and Change trials within each block type (i.e., No Cue, Location Cue, and Semantic Cue) across all participants were computed using in-house MATLAB code. RT was calculated relative to the offset of the second auditory scene and included only correct trials. All statistical analyses (i.e., paired t-tests, repeated measures ANOVAs, post-hoc Bonferroni tests, and a multiple regression analysis) were done using PASW Statistics (version 18.0.0) software. The p-values reported for the repeated measures ANOVAs were corrected for sphericity using Greenhouse-Geisser corrections, and p-values below 0.05 were considered significant.

7 Experiment 2 Results

The main objectives of this change deafness experiment include the following: 1) Demonstrate that listeners fail to identify a location switch in a complex auditory scene, 2) Examine the benefit of directing a listener’s attention to the changing sound, 3) Quantify the cost of directing attention away from the changing sound, and 4) Ensure that change deafness cannot be explained by similarities in the acoustic properties (i.e., spectral power and temporal envelope) of the stimuli.

The first six blocks (i.e., two of each block type) were not included in the analysis since on average both accuracy (t = -2.49, p < 0.05) and RT (t = 3.40, p < 0.01) on the uncued change trials were worse on these six blocks (Means = 69.37% and 749.39 ms, respectively) relative to the following blocks (74.89% and 567.42 ms). Also, subjective comments from most participants indicated that it took about six blocks before they felt completely accustomed to the task. Two-tailed Pearson correlations showed that accuracy on uncued change trials (collapsed across left-center, right-center, and left-right conditions) was not correlated with hearing thresholds (r = -0.26, p = 0.41), digit span forward scores (r = 0.032, p = 0.92), digit span backwards scores (r = 0.39, p = 0.22), or total digit span scores (r = 0.24, p = 0.46). There was no significant difference in accuracy between the two types of 45° change trials within the
uncued blocks (left-center: 73.7% and right-center: 71.7%), so these conditions were collapsed for all subsequent analyses \(t_{(11)} = 0.847, p = 0.415\).

A one-way repeated measures ANOVA was conducted to confirm that listeners’ performance on uncued 45° and 90° change trials was significantly worse than that on uncued Same trials, indicating a change deafness effect. Figure 3 shows the group mean accuracy and RT for the uncued same and change trials for the 45° and 90° angle. There was a main effect of condition on percent correct \(F_{(2, 22)} = 35.21, p < 0.001\). Post-hoc Bonferroni comparisons showed that participants were significantly more accurate on Same trials (93.9%) than both uncued 45° (72.7%, \(p < 0.001\)) and 90° (79.3%, \(p < 0.001\)) change trials, but there was no difference in RT (Same: 506.6 ms; 45° Change: 592.8 ms; 90° Change: 517.0 ms; \(p > 0.05\)). Furthermore, on uncued trials, there was a marginally significant difference between performance on 45° and 90° switch trials (\(p = 0.0502\)). Also, separate one-way repeated measures ANOVAs showed that accuracy \(F_{(2, 22)} = 0.65, p = 0.52\) and RT \(F_{(2, 22)} = 1.50, p = 0.25\) on Same trials did not differ among the uncued (93.9%, 506.6 ms), Location cue (94.0%, 476.5 ms), and Semantic cue (94.8%, 461.1 ms) blocks.

Separate 3 (block type) x 2 (change angle) repeated measures ANOVAs were conducted to determine the benefit of a valid cue on accuracy and RT on the Change trials. There was a main effect of block type on both accuracy \(F_{(2, 22)} = 9.72, p < 0.005\) and RT \(F_{(2, 22)} = 12.44, p < 0.001\), see Figure 4). Bonferroni post-hoc tests showed that accuracy was greater for both Location (82.0%; \(p < 0.05\)) and Semantic (82.1%; \(p < 0.01\)) valid cue change trials than uncued change trials (76.0%), but there was no significant difference between change identification performance with a valid location or semantic cue. Bonferroni post-hoc tests showed the same pattern for RT, in that both valid Location (492.8 ms; \(p < 0.01\)) and Semantic (495.8 ms; \(p < 0.01\)) cues facilitated RTs than uncued change trials (567.4 ms). Like accuracy, there was no difference in RT between valid Semantic and Location cue change trials. For both accuracy \(F_{(1, 11)} = 28.02, p < 0.001\) and RT \(F_{(1, 11)} = 8.38, p < 0.05\), there was also a significant main effect of change angle, such that participants identified 90° switches more accurately and quickly (83.2%, 479.26 ms) than 45° switches (76.84%, 549.76 ms). There was no significant interaction between change angle and block type for either accuracy or RT data. Therefore, directing a listener’s attention to a changing sound enhanced change identification accuracy while speeding up RT.
In contrast to this facilitation effect, directing attention away from the changing sounds resulted in a performance cost, as shown in Figure 5. Separate 3 (cue type) x 2 (change angle) repeated measures ANOVAs were conducted on the accuracy and RT data. (Note that one participant did not correctly identify any switches on invalid Location cue 45° change trials, resulting in no RT values for this condition, so the remaining 11 datasets were included in the RT analysis.) There were main effects of both cue type ($F_{(2,22)} = 11.72, p < 0.005$) and change angle ($F_{(1,11)} = 14.75, p < 0.005$) on accuracy, but there was no interaction between cue type and change angle ($F_{(2,22)} = 0.83, p = 0.45$). Post-hoc Bonferroni comparisons showed that participants performed worse on trials with an invalid Location cue (54.93%) than either uncued (76.01%, $p < 0.005$) or invalid Semantic cue trials (63.32%, $p < 0.05$). There was also a trend in which listeners identified changes more accurately on uncued than invalid Semantic cue trials ($p = 0.084$). Furthermore, participants were more accurate on 90° (68.99%) than 45° (60.51%) change trials.

The RT data showed a similar pattern of results, in that there was a significant main effect of cue type ($F_{(2,20)} = 9.90, p < 0.005$), and there was a trend related to the angle of change ($F_{(1,10)} = 4.30, p = 0.065$), such that participants tended to be faster on 90° than 45° switches. Post-hoc Bonferroni tests confirmed that on correctly-identified change trials, participants were faster on uncued trials (562.56 ms) than either invalid Location (699.62 ms, $p < 0.05$) or Semantic (666.89 ms, $p < 0.05$) cue trials. RTs on invalid Location and Semantic cue trials were not significantly different ($p = 0.84$). Like accuracy, there was no interaction between cue type and change angle ($F_{(2,20)} = 1.83, p = 0.195$). Therefore, the data revealed a cost (i.e., decreased accuracy and increased RT) of directing one’s attention away from a changing sound, and this was slightly affected by the spatial expanse over which the change occurred, in that participants tended to respond faster and more accurately on 90° than 45° change trials. Also, invalid Location cues had a greater impact on accuracy than invalid Semantic cues, even though there was no difference in RT depending on cue type.

It is possible, though unlikely in light of Experiment 1, that a listener’s ability to detect and identify a change in the auditory environment is a product of energetic masking among the sounds within each stimulus trio. Energetic masking occurs when acoustic properties among concurrent sounds overlap, and depending on the extent of the overlap, it can cause concurrent sound segregation to fail. Thus, an additional analysis was done to ensure that the degree of similarity of power spectra and/or temporal envelopes within each stimulus trio were not
predictive of change deafness. For each participant, the number of incorrect responses was
tallied for each stimulus across the Change trials in all block types. The number of incorrect
responses for each of the 126 stimuli was divided by the total number of errors for each
participant to calculate the proportion of total mistakes each participant made on each stimulus
trio. These proportions were averaged across all participants to attain the mean proportion of
errors for each stimulus trio. Then, for each stimulus trio the magnitude squared coherence
between each pair of sounds (i.e., sounds 1 and 2, sounds 1 and 3, sounds 2 and 3) was computed
in MATLAB (mscohere.m) to determine the coherence of two sounds as a function of frequency.
Spectral coherence analysis first computes the psd for each sound within the pair using Welch’s
method, and then computes a coherence value, which ranges from 0 to 1 (perfect coherence), at
each frequency bin. For each stimulus trio, the coherence values from each comparison were
averaged to determine the mean coherence within the trio. Similarities between the temporal
envelopes of the three sounds within each stimulus trio were also quantified. First, a low-pass
fifth-order Butterworth filter was created in MATLAB. Then, zero-phase digital filtering was
applied with the Butterworth filter coefficients to a Hilbert transform of each individual sound in
order to extract its temporal envelope. Three Pearson correlations per stimulus trio were done on
the temporal envelopes, and the mean value of the three rho values was calculated for each
stimulus trio.

A multiple regression analysis was done to determine if similar power spectra, temporal
envelopes, and their interaction (i.e., psd coherence * envelope rho for each trio) predict a given
stimulus’ error rate. In other words, if the sounds within a particular stimulus trio have more
similar psd’s and/or envelopes (and thus higher coherence and rho values) than another stimulus
trio, then that first trio should have a higher rate of errors if change deafness is caused by
energetic masking. The multiple regression model used the group mean proportion of errors on
each stimulus trio as the outcome criterion and the psd mean coherence values, envelope mean
rho values, and the interaction between psd coherence and envelope rho values as the predictors.
The overall model was not significant ($F_{(3, 122)} = 2.47, p = 0.065, R^2 = 0.057$), and none of the
predictor coefficients were significant (psd: $p = 0.39$ envelope: $p = 0.065$, interaction: $p = 0.13$).
This suggests that the change deafness results observed in the present experiment were not
caused by masking due to similar acoustic properties among concurrently-presented stimuli.
Chapter 4
General Discussion

In this study, I examined the role attention plays in identifying a location switch between two sounds presented within a complex auditory scene. Relative to trials on which attention was not guided to any particular sound, performance improved (i.e., increased accuracy and decreased RT) when a listener’s attention was directed to one of the changing sounds via a visual semantic or spatial cue. However, when attention was focused on the third, unchanging sound, listeners were remarkably poor at noticing the location switch between the two unattended sounds. Furthermore, on correct, invalidly-cued trials, participants took much longer to make a response than on correct uncued trials. The results of Experiment 1 and a multiple regression analysis confirmed that these change deafness effects were not a by-product of a failure to segregate the sounds within each scene due to similar acoustic properties, or energetic masking.

These results replicate previous findings demonstrating that focused attention plays an important part in detecting changes in our acoustic and visual surroundings. As previously mentioned, Eramudugolla and colleagues (2005) used a similar change deafness paradigm and also found that relative to uncued trials, change detection performance (i.e., accuracy) improved when listener’s attention was directed to a potentially-changing sound. Unlike the present design, Eramudugolla and colleagues’ (2005) study included only valid cues, such that participants reported if only the cued sound changed, and consequently, cued trials led to performance at ceiling, even in scenes consisting of up to eight sounds. In the current paradigm, approximately 29% of the cues were invalid. As a result, participants generally stated during debriefing that they could not completely trust the cue, which likely led to less focused attention and more mistakes on validly-cued trials than in Eramudugolla and colleagues’ experiments. Thus, if a smaller proportion of invalid cues were used, it is likely that participants would have performed closer to ceiling on the validly-cued trials, and thus larger attentional effects would have been observed.

Furthermore, the current study is the first change deafness study to report attentional effects on RT. Likewise, several change blindness studies showed that a valid endogenous cue decreased RT and increased accuracy (Becker, et al., 2000; Koivisto & Revonsuo, 2005; Landman, et al., 2003). Koivisto and Revonsuo (2005) also demonstrated the converse effect—that is, invalid
cues decreased accuracy and increased RT. Since similar results have been attained in both vision and audition, similar or supramodal attentional mechanisms may underlie both change blindness and deafness. In both Eramudugolla et al.’s (2005) experimental design and the present paradigm, the processing stage (e.g., pre- or post-stream segregation) at which attention modulates change detection was not directly addressed. Focusing one’s attention on a particular sound may enhance concurrent stream segregation, or it may facilitate the transfer of the attended object’s sensory memory trace into WM, resulting in better encoding and retention. Several change blindness studies provided evidence that valid attentional cues affected later processing stages, such as WM retention (Becker, et al., 2000; Landman, et al., 2003) and retrieval (Hollingworth, 2003), and future experiments will be conducted to examine if attention interacts with these later stages in the auditory modality.

The change detection processing stage that intersects with attentional processes may also be modulated by the scene duration as well as number of objects within each scene. For instance, when the number of objects presented within a scene exceeds the WM and attention capacity limits (i.e., four or five objects) or when scene duration is short, attention may play a larger role in sound segregation processes. On the other hand, when scene duration is longer (e.g., 5 versus 1 second long) or scene size is at or below the brain’s processing capacity limits, attentional processes may subserve later processing stages, such as the transfer of attended objects from sensory memory to WM, to a greater extent. Further studies are needed to compare the attentional mechanisms that are involved in auditory and visual change detection to determine if a supramodal neural network underlies change detection in complex scenes, especially when attention is guided to a particular object within a scene.

Guiding listeners’ attention via either a valid visual spatial or semantic cue led to similar behavioral effects in the present study. Although the participants’ task was largely a spatial one (i.e., identify a location switch between two sounds), participants must also encode the acoustic profiles or semantic category of at least two sounds in conjunction with their locations within the scene to correctly identify the switch. A recent study used a change deafness task to examine the nature of representations of complex auditory scenes and concluded that auditory objects are represented at a semantic, rather than an acoustic, level (Gregg & Samuel, 2009). Since both semantic and spatial features must be conjoined to accurately identify a location switch between two sounds, valid semantic cues should enhance behavioral performance to the same level as
valid location cues, as observed in the present experiment. However, performance on a purely spatial or semantic task, in which a single object changes location or is replaced by another sound, may be modulated by the type of attentional cue, in accordance with the ventral “What” and dorsal “Where” processing streams (Alain, et al., 2001; Rauschecker, 1998). In other words, valid location cues may enhance performance on a purely spatial task more so than a semantic cue and vice versa.

However, performance on invalidly-cued trials was better for semantic than for spatial cues. Since the location of each semantic category was randomized across trials, participants could not predict the location of a semantically-cued sound. Thus, in order to find a cued animal sound, for instance, one would have to search the first scene until the cued sound is localized. As a result of this search process, a listener may actually encode semantic or acoustic properties of the uncued sounds in the first scene, leading to increased accuracy without a loss in RT, relative to invalid Location cue trials.

Unlike previous change deafness studies which used a delayed match-to-sample change detection paradigm, the present study instead employed a change identification task. Recently, several studies have shown that change detection and identification may be mediated by independent processes. Rensink (2004) used a flicker change blindness paradigm and instructed participants to press one button when they “sensed” a change and a different button when they could identify the change. Within the 30% of observers who could consciously sense a change before identifying it, on average the difference in RT between detection and identification was equal to the RT for identification without detection, suggesting that detection and identification are mediated by different mechanisms (Rensink, 2004). However, Simons and colleagues (2005) argued that the results seen in Rensink’s (2004) study arose because some participants take extra time to verify their first conscious detection of a change. Instead, they provided evidence that change identification and detection are mediated by the same perceptual process, in which detection precedes identification (Simons, Nevarez, & Boot, 2005). Busch et al. (in press) recorded EEG signals from participants, who were presented with a 4 x 4 matrix of colored pictures of familiar objects. On each trial, participants indicated if they detected a change, and if so, they were asked to identify which of 8 subsequently-presented objects had changed. They found electrophysiological evidence that reflected different neural processes when a change was
both detected and identified relative to when a change was detected without correct identification (Busch, Frund, & Herrmann, in press).

A recent behavioral study also provided evidence for independent processes supporting change detection and identification using auditory stimuli. McAnally and colleagues (in press) used a change deafness paradigm, in which participants reported first whether or not they detected a change and then chose one of eleven sounds that most likely disappeared from the first to second scene. They created high threshold theory and signal detection theory models to predict change detection and identification accuracy and compared these models to receiver-operator characteristics (ROCs). Their results suggested that some changes in the auditory environment can be detected by a mechanism distinct from a change identification process, as a dual-process model most accurately predicted their observed data (McAnally, et al., in press). Thus, since change identification appears to be more difficult than change detection, it is likely that performance on the present paradigm would have been slightly enhanced if a change detection task had been used. Nonetheless, the results presented here nicely parallel behavioral effects of attention in change blindness (Becker, et al., 2000; Koivisto & Revonsuo, 2005; Landman, et al., 2003) and deafness (Eramudugolla, et al., 2005) studies, in which participants were required to detect, but not identify, a change. Therefore, assuming that change detection and identification are in fact mediated by independent processes, it appears that attention can modulate either process. Electrophysiological studies are needed to further investigate the timing of attentional modulation on change detection and identification processing.

In the present study, listeners tended to respond more quickly and accurately when the switch occurred over 90° than 45° whether or not a cue was presented. First, recall that a 45° switch always involved the central sound, whose location was simulated, which could cause it to be harder to localize compared to the left and right sounds. As a result, detecting and identifying a 45° switch may be slightly more difficult than a 90° switch. Although the change deafness results reported here were not affected by energetic masking, informational masking may affect performance because of greater spatial proximity between changing sounds during 45° switches and because the central sound is surrounded by acoustic energy on both sides, whereas the left and right sounds are only adjacent to the central sound. Alternatively, in contrast to a 45° switch, a 90° switch involves a greater change in the pattern of acoustic stimulation on the ears from the first to second scene. A greater degree of acoustic deviance or saliency may draw a
listener’s attention to the change, resulting in a more accurate identification of the 90° change compared to the more subtle 45° switch.

Therefore, the current study emphasizes the role attention plays in change identification in naturalistic, complex auditory scenes, replicating previous change blindness and deafness studies. Directing a listener’s attention to a changing sound facilitates identification of the change, evidenced by increased accuracy and decreased RT. However, guiding attention away from a changing object results in a significant performance cost, even when the auditory scene is comprised of only three sounds. Further analyses ensured that these attentional effects on performance could not be explained by energetic masking among concurrently-presented stimuli. Future experiments will more directly investigate the change detection processing stage that intersects with these attentional effects, as well as the contribution of “What” and “Where” processing streams to the successful detection of changes in the auditory environment to determine how the brain represents complex auditory scenes.
References


