Does Incidental Auditory Learning Facilitate Memory-guided Attention?  
A Behavioural and Electroencephalogram (EEG) Study

by

Manda Fischer

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

© Copyright by Manda Fischer 2019
Does Incidental Auditory Learning Facilitate Memory-guided Attention?

A Behavioural and Electroencephalogram (EEG) Study

Manda Fischer

Master of Arts

Department of Psychology
University of Toronto

2019

Abstract

Can implicit associations facilitate auditory target detection? Participants were exposed to audio-clips (half included a lateralized pure tone). Participants then took a surprise memory test in which they detected a faint lateralized tone embedded in each audio-clip and indicated if the clip was (i) old or new; (ii) recollected or familiar, and; (iii) if the tone was on the left, right, or not present when they heard the clip at study. The results show good explicit memory for the clip, but not for tone location. Target detection was faster for old than for new clips but did not vary according to the target-context associations. Neuro-electric activity revealed an old-new effect at midline-frontal sites as well as a significant difference between clips that had been associated with the location of the target compared to those that were not. The implications of these findings in the context of memory-guided attention are discussed.
Acknowledgments

Thank you, Professor Morris Moscovitch and Professor Claude Alain, for your enthusiasm and genuine excitement for the unknown. Your curiosity, care, and sharing of knowledge has made this year a truly stimulating and enriching one.

Thank you to my lab mates in the Alain and Moscovitch Labs for your feedback and camaraderie. I feel very fortunate to be in such a welcoming and kind environment. A special shout-out to Vanessa Chan for her help and patience with the Presentation code (and her delicious baked goods that kept us all alive at the lab!) and to the volunteers in the Alain Lab, Karishma Ramdeo and Shahier Paracha, for their very efficient help with recruitment and EEG set-up, respectively.

A big thank you, also, to my family and friends have been there with me every day and who provide the sunlight that encourages me to constantly push myself and grow.
Table of Contents

Acknowledgments .................................................................................................................. iii
Table of Contents .................................................................................................................. iv
List of Tables ........................................................................................................................ vi
List of Figures ....................................................................................................................... vii
List of Appendices ................................................................................................................ viii

1 Introduction ....................................................................................................................... 1
   1.1 Auditory Scene Analysis: A Model for Auditory Perception .................................... 1
   1.2 Evidence for schema-based ASA ............................................................................. 3
      1.2.1 The effect of attention on ASA ................................................................. 3
      1.2.2 The effect of prior experience and memory on ASA ............................. 3
   1.3 Attention as a mediating factor for the effect of long-term memory on perceptual organization ................................................................. 4
      1.3.1 In vision .................................................................................................. 4
      1.3.2 In audition ............................................................................................ 5
   1.4 Current study ............................................................................................................ 6

2 Methods ........................................................................................................................... 7
   2.1 Participants .............................................................................................................. 7
   2.2 Stimuli ..................................................................................................................... 7
   2.3 Procedure ................................................................................................................ 9
      2.3.1 Exposure phase (encoding) ...................................................................... 10
      2.3.2 Test phase (retrieval) ............................................................................. 10

3 Statistical Analyses ........................................................................................................ 12
   3.1 Predictions ............................................................................................................. 12
      3.1.1 Behavioural ............................................................................................ 12
   3.2 Electroencephalogram (EEG) ............................................................................. 13
List of Tables

Table 1. Summary of the channel level cluster-based permutation statistics.

Table 2. Probe audio-clip: Summary of the channel level cluster-based permutation statistics.

Table 3. Target tone: Summary of the channel level cluster-based permutation statistics.
List of Figures

Figure 1. Paradigm Overview.

Figure 2. Butterfly plot of group mean event-related potentials (ERPs).

Figure 3. Memory accuracy at test phase.

Figure 4. Reaction times for detecting the faint pure tone at test.

Figure 5. Memory accuracy of tone-detection for neutral compared to memory-cue conditions.

Figure 6. Group mean event-related potentials (ERP) for the cue audio-clip: Old vs. new.

Figure 7. Group mean event-related potentials (ERP) for the cue audio-clip: Memory-cue vs. neutral.

Figure 8. Group mean event-related potentials (ERPs) for the probe audio-clip: Old vs. new.

Figure 9. Group mean event-related potentials (ERPs) for the probe audio-clip: Memory-cue vs. neutral.

Figure 10. Group mean event-related potentials (ERPs) for the target: Old vs. new.

Figure 11. Group mean event-related potentials (ERPs) for the target: Memory-cue vs. neutral.
List of Appendices

Appendix A: List of audio-clips used.
1 Introduction

Experimental Psychology is shifting focus from studying cognitive processes in isolation to considering them as components in a complex, interactive system (Romero & Moscovitch, 2015). Making sense of everyday auditory scenes requires hearing the appropriate signal from the incoming sound mixture. Previous research suggests that both low-level acoustic-driven and high-level schema-driven processes are necessary for successful auditory perception (Bregman, 1990). Schema-driven scene analysis involves the use of learned schemas that are stored in memory in order to guide perceptual organization of acoustic input. There is very little empirical work, however, to elucidate the psychological and brain mechanisms, enabling schema-driven scene analysis.

More recent evidence supports the notion that attention and memory, two important and tightly related phenomena in cognitive neuroscience, play a vital role in disambiguating complex acoustic signals in everyday listening situations (Backer & Alain, 2013). Most research has focused on effects of attention on memory; few studies have investigated the converse, that memory may guide attention. The proposed study seeks to examine this, focusing on the relation between memory and attention when a person listens to a complex auditory scene.

1.1 Auditory Scene Analysis: A Model for Auditory Perception

Sound is everywhere in our lives. The difficulty in auditory perception, however, is that sounds originating from separate sources arrive at each eardrum in the form of a single complex pressure wave. This is what Albert Bregman referred to as the “scene analysis problem” (Bregman, 1990, chap.1). What one perceives depends on the grouping of acoustic material into coherent mental representations and on the selection of information for further processing (Bregman, 1990, chap.1; McAdams & Bregman, 1979). For example, at dinner, making sense of what one hears requires that the incoming sound mixture be organized into separate mental representations (Bregman, 1990). This process is known as auditory scene analysis (ASA).

Traditionally, research in the field of ASA has focused on understanding how external properties in sound affect perceptual grouping. Recently, however, there has been a paradigm shift, allowing for a broadening of ASA research. In particular, there is growing interest in understanding the way in which low-level processes interact with high-level processes during
auditory scene analysis. Before empirical evidence existed to support the notion that high-level factors play an important role in ASA, a number of studies yielded results that were inconsistent with the traditional model of auditory perception, suggesting that low-level factors may not be sufficient to facilitate auditory perception. Cutting (1975) demonstrated that conflicting acoustic cues that would predict perceptual segregation could be combined into a coherent mental representation for the purpose of speech recognition. The experiment used phoneme pairs in a dichotic listening task to examine phonological fusion, in which phonemes from two separate speech stimuli are combined to form a new percept that is lengthier and more linguistically complex than each stimulus component alone. In order to isolate a single attributable acoustic cue, stimuli were matched in terms of pitch, intensity, and duration. Therefore, spatialization of each phoneme pair provided the sole acoustic cue for perceptual organization. A purely acoustic-driven model of scene analysis would predict perceptual segregation between phoneme pairs. The study, however, found that phoneme pairs, such as “pay” and “ray” yielded the percept “pray”. This finding implies that perceptual organization must be promoted by factors other than external properties alone and that bottom-up processes are not sufficient to facilitate successful auditory perception. In support of this argument are findings from more recent work that has shown that scene analysis can occur at both early cortical processing stages, as well as central ones (Moore & Gockel, 2002). Taken together, these studies confirm that auditory perception requires the complex interaction of both low and high-level factors at multiple stages of processing (Bey & McAdams, 2002; Bregman, 1990; Snyder & Alain, 2007).

What would a more comprehensive model of auditory perception look like? What mechanisms may be necessary to facilitate the perception of complex auditory stimuli? Bregman (1990) theorized that there are two independent systems underlying the process of ASA and that they are involved jointly in analyzing a complex auditory scene: primitive or sensory-driven and schema or knowledge-driven. In an effort to make consistent interpretations about the environment, the former process was thought to use a set of innate heuristics that would group auditory features based on evolutionarily-dictated principles about how the world behaves (Bregman, 1990, chap.2; Bregman & Pinker, 1978; Koffka, 1935, chap.4). The latter process was thought to use a more sophisticated knowledge of the signal to group incoming auditory input. In this case, acoustic features are selected on the basis of criteria matching, in that those acoustic
features that are consistent with stored schemas for familiar environmental sounds are selected from the incoming sound mixture.

1.2 Evidence for schema-based ASA

When schema-based ASA was first proposed, there was little empirical research to support its legitimacy. Although it was clear that a purely stimulus-driven model of ASA could not explain perception on a variety of auditory tasks, the internal mechanism of knowledge-based scene analysis had not yet been established. To date, a number of studies provide evidence for high-level factors, such as attention and prior experience, in auditory perception.

1.2.1 The effect of attention on ASA

Recent work suggests that ASA is tightly related to attentional mechanisms. One aspect of this relationship is attention’s ability to enhance auditory processing. An event-related potential (ERP) study by Alain & Woods (1994) examined the effect of selective attention on auditory stream segregation, revealing that selective attention has both facilitatory and suppressive effects on the attended and unattended streams, respectively. This finding is consistent with the gain model that argues that attention has the ability to enhance sensory processing of attended streams, leading to a neural processing “gain,” and to suppress processing of unattended streams. Enhanced processing enables stronger representation of auditory objects that are selectively attended, while suppressed processing facilitates weaker representation of auditory objects that are unattended.

For Bey & McAdams (2002), schema-based scene analysis may be mediated by attentional factors. They have argued that a primary distinction between low-level perceptual mechanisms and high-level ones is that the former parses sensory information, while the latter selects information. Research by Dowling et al., (1987) supports this claim, finding that knowledge of a melody facilitates task performance, by enabling attentional focus.

1.2.2 The effect of prior experience and memory on ASA

It is now well established that prior knowledge affects perceptual organization (Dowling, 1973). A study by Bey & McAdams (2002) compared melody recognition of a series of target tones embedded in a distractor sequence when the target melody was presented before the interleaved
melody and after it. They found that recognition was much better when the target was presented alone before, rather than after, suggesting that prior knowledge enables enhanced sensory processing and melody extraction. This study provides further evidence of the effect of high-level factors on auditory scene analysis.

In another experiment, Bey & McAdams (2002) found that melody recognition was not possible when the mean frequency was the same for melody and distractor tones. This observation indicates that bottom-up parsing of sensory information may be a necessary prerequisite for schema-based analysis. Dowling (1973), however, showed that, in the complete absence of acoustic cues for segregation, it is still possible to perceptually segregate a sequence of tones so long as listeners have prior experience with the target tones that are to be separated. Despite these seemingly contradictory results, it must be noted that the two studies differed in terms of their stimuli. Dowling (1973) tested highly familiar children’s songs that were supposedly stored in long-term memory while Bey & McAdams (2002) used melodies that were unfamiliar before the test and were more likely to be encoded in short-term memory. This discrepancy suggests that benefits of prior knowledge and familiarity may be present only when schemas are stored in long-term memory. Research supports this hypothesis and shows that different types of features are encoded in short-term and long-term memory, in that the former stores gist-like contour information and the latter stores precise interval information (Dowling, 1978; Dowling & Harwood, 1986).

1.3 Attention as a mediating factor for the effect of long-term memory on perceptual organization

1.3.1 In vision

Chun & Jiang (1998) were the first to demonstrate memory’s ability to guide attention to facilitate target detection in vision. In their study, sensory processing associated with detection of a target object in a remembered location was directly mediated by attention. This finding provides evidence for enhanced perceptual processing in vision that is directly attributable to higher-order processes. In their study, they investigated contextual cueing, an effect in which a visual scene provides context that, in turn, influences and constrains expectation and visual search behaviour. It is thought that this context biases attention to facilitate target detection. To examine this effect, the study embedded a distractor in a complex visual scene and repeated its
spatial arrangement over many trials. Although participants were not aware of the distractor and its repeated spatial location, they more quickly detected the target when it was presented in the same location as the distractor than when the target was placed elsewhere. These findings suggest that 1) long-term memory for spatial location can facilitate target detection; 2) attention may be a direct mediating factor for facilitating the effect of memory on target detection; and 3) this visual memory-guided attention effect may be supported by implicit memory systems. Furthermore, this effect is observable in the brain. A number of studies have identified the neural correlates underlying this effect in vision (Chaumon et al., 2009; Summerfield et al., 2006). Additionally, Negash et al. (2007) found that contextual cueing relies on support from the medial temporal lobes to guide attention during visual search.

1.3.2 In audition

Vision studies suggest that attention is the primary mediating factor that enhances sensory processing when familiarity of stimuli is manipulated. Research has shown that memory can guide attention in the visual domain (Chun & Jiang, 1998), but, as previously noted, little is known whether it can do so in the auditory one (Zimmermann et al., 2016). Only recently has enough evidence accumulated to start elucidating the potential internal mechanism that supports schema-based auditory scene analysis.

In a dichotic listening experiment, Cherry (1953) demonstrated that once a perceptual stream is formed, top-down effects can facilitate attentional capture of one’s name. Although these results were later contested, the study provides seminal evidence for an effect of auditory memory on attention. Zimmermann et al. (2017) were the first to show that deliberately forming associations between audio-clips and tones enables memory to bias attention to auditory stimuli (Zimmermann et al., 2017). Participants were instructed to pay attention to, and form an association between, an audio-clip and the spatial location of an embedded lateralized pure tone target. The results from a subsequent test on participants’ memory for learned the associations exhibited enhanced target detection. Importantly, this benefit was observed despite the fact that participants were unable to consciously report the association between audio-clip and spatial location. The authors interpreted these findings as evidence for a separate implicit system at play, independent from any explicit knowledge. This conclusion, however, may be problematic. In the absence of explicit recall, performance above chance may still occur without the need for this
dissociation between recognition and priming. As previous studies have indicated, statistical anomalies and noise introduced into a model for explicit learning can yield these dissociative results (Shanks, 2004; Ostergaard, 1992; Plaut, 1995; Poldrack et al., 1999). Therefore, whether implicit learning can indeed facilitate memory-guided attention in hearing remains an open question.

1.4 Current study

The project intends to fill this gap, by investigating whether the effect of memory-guided attention holds in natural listening situations in which incidental associations are formed through unsupervised learning (exposure). The validity of this effect will be tested and the mechanisms that subserve this memory-retrieval process will be identified. To do so, the experiment will draw behavioural and neural indices of the incidental formation and retrieval of long-term associations between audio-clip and spatial location of embedded pure tone. Specifically, the project will (1) further examine the link between auditory memory and attention, by testing whether incidental associations between tone and scene can guide auditory attention; and (2) reveal the mechanism of retrieval that underlies this potential effect, by examining the neural correlates at both encoding and retrieval to see their role in guiding attention.

It is expected that reaction time for detecting a pure tone embedded in a complex scene will vary, as a function of the formation of learned associations between spatial location and auditory scene. It is hypothesized that one should observe both memory and attention components of the memory-guided effect in performance and neural response.
2 Methods

2.1 Participants

Twenty-six healthy young adults participated in the experiment ($M = 26.1$ years $SD = 4.3$). All participants had normal hearing and normal or corrected vision. Each participant was required to pass an audiometric pure tone test, in which their tone hearing thresholds were assessed at octave frequencies between 250 Hz and 8000 Hz. Criteria for normal hearing were 1) to have thresholds lower than, or equal to, 25 decibels Normal Hearing (dB NH) and 2) to have a 15dB difference or less between the two ears at each octave frequency. All participants had no history of psychiatric, neurological, or other major illnesses. Participants were recruited from the Rotman Research Institute participant database and received monetary compensation for their participation. All participants provided written consent; the study is certified for ethical compliance by the Research Ethics Board at Baycrest.

2.2 Stimuli

A total of 108 audio-clips, selected from "http://www.freesounds.org/", were included in the experiment. Each audio-clip was comprised of a complex auditory scene and was selected to have semantic relevance. An initial pilot test confirmed that each clip in the set was nameable and that there was considerable consistency among participants. That each clip can be given a semantic label (e.g., coughing) increases the likelihood of an association between the clip and the tone to be formed and stored in long-term memory (Cohen, et al., 2011; Snyder & Gregg, 2011). All stimuli were adjusted to have levels between -0.5 and +0.5 amplitude (measured using root mean square (RMS)). Clips were cut from their original length to a duration of 2500 ms with a 100 ms rise and fall time. In addition, all clips were down-sampled to a standard sampling rate of 44100 Hz. All stimuli were presented through insert earphones (EARTONE 3a), at a listening volume of 60 dB SPL on average across stimuli, with some sounds peaking at about 80 dB SPL.

Four of the audio-clips were used as buffers to address the primacy and recency effect. Each exposure block always began with two of the buffer audio-clips and ended with the remaining two buffer audio-clips. Data for these clips were not taken into account for the analysis. Eighty clips were used during the exposure phases, in which unintended associations were made between clip and spatial location of the target. These clips were also presented at the surprise
memory-test phase. In the test phase, these clips will be referred to as “old” clips. In addition to the “old” clips, twenty-four “new” clips were introduced at the memory-test phase. These clips did not have any prior associations between them and the target location.

A pure tone target (500 Hz, 500 ms in duration, 50 ms rise/fall time with RMS amplitude between -0.5 and +0.5) was embedded in 50% of the clips at exposure. At test, all clips had a pure tone target embedded in them. In all cases, the pure tone was embedded 2s after clip onset. The lateralization of the tone was counterbalanced and randomized for each participant. The sound level at which the pure tone target was played within the audio-clip varied across participants, according to their individualized SNR ratio. Acoustic stimuli and visual cues were presented, using Presentation software (version 13, Neurobehavioral Systems, Albany, CA).
2.3 Procedure

Participants performed two different types of tasks.

a) Experiment overview

![Experiment overview diagram]

b) Individual trial overview by condition

![Individual trial overview by condition diagrams]

*Figure 1.* Paradigm overview a) Complete experiment overview, comprising of both exposure and test phase. The exposure phase had participants form incidental associations between the sound-clip and target location. The surprise test phase examined the memory of the scene and elicited lateralization of attention. b) Overview of a single trial for exposure and test phased by condition. The memory-cue condition had a target embedded in the left or right ear, at exposure. At test, the target was then presented at the same location. Reaction time was compared to neutral cue condition trials, in which there was no target embedded at exposure (i.e., no tone-clip association formed during the exposure phase). At test, neutral and new clips had the target embedded in the left or right ear (randomized and counterbalanced). Note: Cue audio-clip and
probe audio-clip at test represent the first and second repetition of the audio-clip, separated by an ISI of 1000 ms. Only the probe audio-clip contained an embedded pure tone target, 2000 ms after probe audio-clip onset.

2.3.1 Exposure phase (encoding)

First, participants underwent four exposure phases in which audio-clips were lateralized. Each trial consisted of either the audio-clip alone or the audio-clip and a lateralized (left or right hemisphere) pure tone stimulus. The SNR between the pure tone and audio-clip was high in this phase, allowing implicit learning, via mere exposure, to proceed.

To ensure that the task truly involved the formation of incidental associations, participants were told only to classify audio-clips as natural or manmade. We made no reference to any memory test (Wolters & Prinsen, 1997). To respond, participants used the “left” and “right” arrow keys. Key-response was pseudo-randomized across participants, in that half of the participants had “left” and “right” keys corresponding to natural and manmade, respectively and the other half had “left” and “right” keys corresponding to manmade and natural, respectively. Clip order was randomized across exposure phases and participants. In addition, left-lateralized, right-lateralized, and neutral trial pairings with clip was randomized across participants. This ensured that any effects uncovered would not be attributable to key-response, order-of-scenes, or specific pairing between clip and spatial location, respectively.

2.3.2 Test phase (retrieval)

After the exposure phases, participants were given a surprise memory test (retrieval) on the learned associations. The test contained old audio-clips from the exposure phase and new ones. For old clips that contained a lateralized target, memory-cue trials contained a pure tone at the spatial location learned during the exposure phases, while catch trials contained a pure tone at the spatial location opposite to the one that was learned. The ratio between memory-cue and catch trials was 80:20. For old trials that did not contain a target (neutral) and new clips, a target was embedded at a pseudo-random spatial location. Therefore, all clips contained a pure tone embedded in them. Participants were required to indicate 1) if, at test, the tone was presented on the left, right, or not at all; 2) if the audio-clip was old or new; 3) when old, if they recollected the audio-clip or if it was merely familiar from acquisition; 4) if, at acquisition, the tone was on
the left or right. The intensity of the pure target was adjusted based on a pilot study to allow for approximately 80-90% correct detection of the target within audio-clips. A low SNR was chosen for the test phase in order to ensure that all participants engaged in effortful listening (Alain et al., 2018; Pichora-Fuller et al., 2017).

Audio-clips in this phase were presented twice. The first presentation (S1) served as a cue to orient attention and did not include the target tone. The second presentation (S2) occurred immediately after the offset of S1 and included the target tone. An initial pilot test was conducted by Zimmerman (2018), demonstrating that a single cue was not sufficient to allow participants to activate auditory memory for the target location. This finding highlights the interesting differences between auditory and visual domains.

To respond, participants used the “left” and “right” arrow keys. Key response was pseudo-randomized across participants for questions 2 and 3, in which half of the participants had “left” and “right” keys corresponding to old and new, as well as recollect and familiar, respectively and the other half of participants had “left” and “right” keys corresponding to new and old, as well as familiar and recollect, respectively. Clip order was randomized across participants. This procedure ensured that any effects uncovered would not be attributable to key-response and order-of-scenes, respectively.
3 Statistical Analyses

Catch trials were not analyzed, as there were too few correctly-detected trials (at most 4 left and 4 right per participant) to calculate a reliable average reaction time for this condition. Therefore, assessing the potential cost/benefit of the memory-guided attention effect, by comparing old memory-cue trials to old catch trials, was not possible. Originally, it was expected that participants would have slower reaction times and lower percent accuracy for old catch trials, reflecting the cost associated with the memory-guided attention effect, and that performance would be better for old memory-cue clips, reflecting the benefit associated with the memory-guided attention effect.

3.1 Predictions

We posited that if auditory associations were formed via mere exposure, participants would detect the tone better if it occurred on the same side at test as at it did at exposure. Participants’ responses to questions 2 to 4 at test (see Procedure) would allow us to determine if the memory-cue advantage was related to recollection or familiarity, and if it depended on explicit memory of where the tone was presented at acquisition.

3.1.1 Behavioural

Reaction time (RT) and percent accuracy was measured at test phase. RT and percent accuracy for natural and manmade responses at exposure was recorded and documented for potential future analyses.

3.1.1.1 Old-new effect

In order to assess the contextual component of memory-guided attention, old clips (memory-cue and neutral) were compared to new clips, using a paired two-sample t-test. The first analysis used RT as the dependent variable and clip type (new or old) as the independent variable. It was expected that responses on trials that contain old clips would be faster than those that contain new ones, as the former should have already had an existing long-term representation of the clip.
3.1.1.2 Memory-guided attention effect

To assess the memory-guided attention component of this effect, old memory-cue clips were compared to old neutral clips, using a paired two-sample t-test. The first analysis used RT as the dependent variable and clip type (old memory-cue or old neutral) as the independent variable. It was predicted that there would be more of a benefit in RT in trials that contain old memory-cue clips than in those that contain old neutral clips, as old memory-cue clips would have already had formed associations between clip and spatial location that may aid in biasing attention. Conversely, old neutral clips were not expected to have an association between clip and spatial location. Therefore, it was predicted that target detection for these types of clips would be at chance.

3.1.1.3 Recollect, familiar, or missed effect

To investigate the potential memory system that underlies this memory-guided effect, the question of whether recollection or familiarity for the target’s location is necessary for associative retrieval was examined. Clips that were deemed “missed” were those that were old but were judged by the participant as new. Clips that were deemed “false alarms” were those that were new but were judged by the participant as old. If the memory-guided attention effect was observed, we planned to compare RT and percent accuracy for recollected, familiar, and missed clips to first see if there were any marked differences between the three groups. If gains in RT and percent accuracy were observable for clips that were missed, it would indicate that memory-guided attention is possible under an independent implicit learning process.

3.2 Electroencephalogram (EEG)

For statistical analyses, the ERP waveforms were exported into BESA Statistics 2.0. BESA Statistics 2.0 software automatically identifies clusters in time, frequency, and space, using a series of t-tests that compared the time-frequency data between experimental conditions at every time point. This preliminary step identified clusters both in time (adjacent time points) and space (adjacent electrodes) where the ERPs differed between the conditions. The channel diameter was set at 4 cm which led to around four neighbours per channel. A cluster alpha of .05 for cluster building was used. A Monte-Carlo resampling technique (Maris & Oostenveld, 2007) was then used to identify those clusters that had higher values than 95% (one-sided t-test) of all clusters derived by random permutation of the data. The number of permutations was set at 1,000.
Time-domain analyses (i.e. ERPs) compared new with old audio-clips and memory-cue with neutral audio-clips at three different time windows: cue audio-clip, probe audio-clip, and target. For old and new audio-clips, it was expected that event-related potentials (ERPs) would show a difference during the sustained potential, indicating that information related to the clip was being maintained (Bae & Luck, 2018; Awh and Jonides, 2001).

For memory-cue and neutral clips, it was expected that ERPs would be lateralized after the onset of the cue-audio-clip, indicating that information about the location of the target, associated with the audio-clip, was encoded and represented in the neural signature.

3.2.1 EEG recording and analysis

The EEG was recorded continuously during each exposure phase and during the memory-guided attention task, using a 76-channel acquisition system (BioSemi Active Two, Amsterdam, The Netherlands). Sixty-six EEG electrodes were positioned on the scalp using a BioSemi headcap, according to the standard 10/20 system, with a Common Mode Sense (CMS) active electrode and Driven Right Leg (DRL) ground electrode. Ten additional electrodes were placed below the hairline (both mastoid, both pre-auricular points, outer canthus of each eye, inferior orbit of each eye, two facial electrodes) to monitor eye movements, as well as to cover the whole scalp evenly. After low-pass filtering at 100 Hz, the EEG was sampled at rate of 512 Hz, digitized, and stored continuously for offline analysis using the Brain Electrical Source Analysis software (BESA, version 6.1; Megis GmbH, Gräfelfing, Germany).

3.2.2 EEG preprocessing

Data pertaining to five participants were not used because of excessive eye movements and/or muscle artifacts. The EEG data were visually inspected to identify segments contaminated by defective electrode(s). No more than ten electrodes were interpolated, using values from the surrounding electrodes. The EEG was then re-referenced to the average of all electrodes. The continuous EEG was digitally filtered with 0.1 Hz high-pass filter (forward, 6dB/octave) and 20 Hz low-pass filter (zero phase, 24 dB/octave).

For each participant, a set of ocular movements was identified from the continuous EEG recording and then used to generate spatial components that best account for eye movements. The spatial topographies were then subtracted from the continuous EEG to correct for lateral and
vertical eye movements as well as for eye-blanks. After correcting for eye movements, all experimental files for each participant were then scanned for artifacts; epochs including deflections exceeding 120 µV were marked and excluded from the analysis. The remaining epochs were averaged according to electrode position and experimental conditions. Each average was baseline-corrected with respect to a 200 ms pre-stimulus baseline interval. Approximately 2-15% of trials were rejected for each participant.

The data were parsed into three sets of epochs: Cue audio-clip, probe audio-clip, and target. The cue audio-clip epochs started 200 ms before clip onset (0ms) and ended 2000 ms after the cue onset. We also used the same epoch length and filter settings to examine neuroelectric activity preceding the target tone (i.e., probe scene). For the target-related analysis, the continuous EEG was filtered with 0.1 Hz high-pass filter (forward, 6dB/octave) and 40 Hz low-pass filter (zero phase, 24 dB/octave). The target epochs started 200 ms before target onset (0ms) and ended 800 ms after target onset.
4 Results

4.1 Behavioural results

4.1.1 Memory accuracy for the scene is good, but memory accuracy for the location of the tone is not

Figure 3. Memory accuracy at test phase. A) Memory accuracy for new and old clips. B) Memory accuracy for scene-tone association for memory-cue clips. Chance level of 33% was determined based on the three choices provided in question 3 (left, right, none).

Overall memory accuracy for audio-clips (question 1: Old versus New) and the scene-tone association (question 3: Left, Right, None) were assessed. The results reveal that memory for new ($M = 77.3\%, SE = 2.45$) and old ($M = 78.2\%, SE = 2.34$) scenes are significantly better than chance (50%), $t(24) = 11.13$, $p < 0.0001$, and $t(24) = 12.04$, $p < 0.0001$ (Figure 3 A). Hits minus false alarms for scene accuracy ($M = 55.5\%, SE = 2.95$) was also significantly above chance (0%), $t(24) = 18.84$, $p < 0.0001$, suggesting that participants were capable of distinguishing between old and new items.

Accuracy for reporting the correct target location of memory-cue clips was compared to chance accuracy (33%). The results revealed that participants performed at chance level when asked to report the tone’s location relative to the particular scene, $t(24) = 1.62$, $p > 0.05$ (Figure 3 B).
4.1.2 Memory for the scene enhances processing of the tone by speeding responses to it, but memory does not facilitate lateralized allocation of attention.

![Graph](image)

**Figure 4.** Reaction times for detecting the faint pure tone at test. A) Reaction time for detecting the faint pure tone at test for correctly remembered new and old scenes. Old scenes were comprised of both memory-cue and neutral scenes. B) Average tone-detection reaction time for correctly remembered memory-cue and neutral clips.

The average reaction time for correctly remembered new scenes ($M = 734.62$ ms, $SE = 42.40$) was compared to the average reaction time for correctly remembered old scenes ($M = 690.82$ ms, $SE = 34.07$). The contrast yielded a significant difference between the two conditions, $t(24) = 2.56$, $p < 0.001$, suggesting that memory for the scene facilitates target detection.

Contrasts between memory- and neutral-cue conditions did not yield a significant difference ($p > 0.05$).
4.1.3 Interference during exposure: Memory accuracy is better for neutral-cue clips compared to memory-cue clips

![Bar chart showing memory accuracy for neutral-cue vs. memory-cue conditions.](image)

**Figure 5.** Memory accuracy of tone-detection for neutral compared to memory-cue conditions.

The suprathreshold tone during exposure may have caused interference, eliciting a misallocation of attention. This effect may have disrupted the memory for both the sound and the sound-tone association (Figure 6). To examine this hypothesis, we contrasted memory accuracy for neutral clips ($M = 77.60\%$, $SE = 2.62$) to memory accuracy for memory-cue clips ($M = 64.17\%$, $SE = 2.10$). The analysis yielded a significant difference between the two conditions, $t(24)= 8.98$, $p < 0.0001$, with memory accuracy being better for neutral clips that did not contain a suprathreshold tone than memory-cue clips.
4.2 Electrophysiological results at test phase (retrieval)

EEG was used to identify the neural correlates that underlie the observed effect.

*Figure 2.* Butterfly plot of group mean event-related potentials (ERPs). Grey boxes depict the sub-events. The grey lines show ERPs from all scalp electrodes. The black line shows ERPs at the midline fronto-central electrode (FCz). For illustration purposes, the data were filtered (high-pass filter = 0.4 Hz and low-pass filter = 10 Hz). The sustained potential was filtered out to emphasize the transient onset and offset responses of each sub-event.

4.2.1 Cue audio-clip (0ms – 2000 ms)

In this analysis we examined whether there were neural indices underlying the behavioural difference in new and old clips. ERP averages for old scenes included both neutral and memory-cue scenes. Generally, the cue audio-clips generated transient onset responses that were largest at midline central electrodes (not shown), which were followed by a sustained potential over the left central, temporal and temporal-parietal sites. The transient responses were followed by a sustained potential that was largest over central-parietal and parietal areas.

4.2.1.1 Old vs. new

The contrast assessing whether ERPs differed between old and new clips was first performed on the cue audio-clip (0 – 2000 ms). Cluster permutations yielded five significant clusters with more
sustained negativity for new scenes compared to old ones. The most significant cluster was located over fronto-central, central, and parietal areas of the scalp at 777-1072 ms (Table 1, Figure 6). The second cluster followed the first one in time and was located over frontal scalp area. The third cluster also overlapped with the first cluster in time and showed modulations over the frontal scalp area. The fourth and fifth cluster overlapped one another and preceded all other clusters in time and showed modulations in the central-parietal and parietal areas of the scalp.

Table 1. Cue audio-clip: *Summary of the channel level cluster-based permutation statistics for Old vs. new audio-clips.*

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Cluster</th>
<th>Latency (ms)</th>
<th>P value</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old vs. new</td>
<td>1</td>
<td>777 - 1072</td>
<td>$p &lt; 0.0001$</td>
<td>FC3, FC1, C1, C3, C5, CP5, CP3, CP1, P1, P3, PO3, POz, Pz, CPz, FCz, Cz, C2, TP8, CP6, CP4, CP2, P2, P4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1307 - 1998</td>
<td>$p = 0.001$</td>
<td>F4, F6, F8, FC6, FC4, FC2, FCz, C4, CP4, CP2, P2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>885 - 1004</td>
<td>$p = 0.02$</td>
<td>FP1, FPz, FP2, AF8, FT10, F10, LO2, IO2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>662 - 742</td>
<td>$p = 0.002$</td>
<td>CP5, CP3, CP1, P1, P3, P7, PO3, Oz, POz, Pz, P2, PO4, O2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>506 - 666</td>
<td>$p = 0.02$</td>
<td>FC1, C1, C3, CP3, CP1, P1, Iz, Oz, POz, P2, PO8, O2</td>
</tr>
</tbody>
</table>
Figure 6. Group mean event-related potentials (ERP) for the cue audio-clip: Old vs. new. ERP traces for old and new clips during the cue epoch at test. For this and subsequent figures, the iso-contour maps show the group mean amplitude over a predefined window around the peak amplitude. The time window for Cluster 1, 2, and 4 were 900-1000 ms, 1300-1400 ms, and 900-1000 ms, respectively.

4.2.1.2 Memory-cue vs. neutral-cue

Contrasts assessing whether ERPs differed between memory-cue and neutral clips were first performed on the cue audio-clip (0 – 2000 ms). Cluster permutation yielded one significant cluster over frontal and fronto-central areas, with neutral clips eliciting greater sustained negativity than memory-cue clips (Table 2, Figure 7).

We also tested whether ERP activity would differ when audio-clips were paired with a left compared to a right-lateralized auditory target. No significant difference was observed ($p > 0.05$).
Table 2. Cue audio-clip: Summary of the channel level cluster-based permutation statistics for Memory-cue vs. neutral audio-clips.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Cluster</th>
<th>Latency (ms)</th>
<th>P value</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory-cue vs. neutral</td>
<td>1</td>
<td>1020 - 1100</td>
<td>$p = 0.03$</td>
<td>AF3, F1, F3, AF4, Fz, F2, FC2, FCz</td>
</tr>
</tbody>
</table>

Figure 7. Group mean event-related potentials (ERP) for the cue audio-clip: Memory-cue vs. neutral. ERP traces for memory-cue and neutral clips during the cue epoch at test. The iso-contour maps show the group mean amplitude for the 1000-1100 ms interval.
4.2.2 Probe audio-clip (0 ms – 2000 ms)

4.2.2.1 Old vs. new

The contrast assessing whether ERPs differed between old and new clips was performed from 0 to 2000 ms relative to probe onset. The ERP analyses yielded two significant clusters. The most significant cluster was located over right central, temporal, and parietal areas of the scalp, with old scenes eliciting greater sustained negativity compared to new scenes due to a polarity reversal from anterior to posterior sites (Table 3 and Figure 8). The second cluster overlapped with the first cluster in time and occurred over left and midline frontal and central areas of the scalp, with greater sustained negativity for new scenes than for old ones (Table 3 and Figure 8).

Table 3. *Probe audio-clip: Summary of the channel level cluster-based permutation statistics for Old vs. new audio-clips.*

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Cluster</th>
<th>Latency (ms)</th>
<th>P value</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old vs. new</td>
<td>1</td>
<td>516 - 1998</td>
<td>$p = 0.0007$</td>
<td>AF8, FT8, FC6, C6, T8, TP8, CP6, CP4, P8, P10, PO8, CB2, TP10, FT10, F10, LO2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1311 - 1645</td>
<td>$p = 0.04$</td>
<td>F1, F3, FC1, C1, CP3, CP1, Fz, FCz, Cz</td>
</tr>
</tbody>
</table>
Figure 8. Group mean event-related potentials (ERPs) for the probe audio-clip: Old vs. new. ERPs elicited by the probe stimulus for old and new clips at test. The grey shaded box denotes significant differences between old and new clips. The iso-contour maps show the group mean amplitude for cluster 1 and 2, with 1900-2000 ms and 1500-1600 ms intervals, respectively.

4.2.2.2 Memory-cue vs. neutral-cue

Next, we compared memory- and neutral-cue clips. The ERP analyses yielded two significant clusters. The most significant cluster was located over frontal areas of the scalp, with memory-cue scenes eliciting greater sustained negativity than for neutral scenes (Table 4 and Figure 9). The second cluster followed the first cluster in time and occurred over right parietal areas of the scalp, with greater sustained negativity for neutral scenes than for memory-cue scenes due to a reversal in polarity from anterior to posterior sites (Table 4 and Figure 9).
Table 4. *Probe audio-clip: Summary of the channel level cluster-based permutation statistics for Memory-cue vs. neutral.*

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Cluster</th>
<th>Latency (ms)</th>
<th>P value</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory-cue vs. neutral</td>
<td>1</td>
<td>21 - 201</td>
<td>$p = 0.01$</td>
<td>FP1, AF7, F7, FT7, FC5, T7, FPz, FP2, AF8, AF4, AFz, Fz, F2, AF8, AF4, AFz, Fz, F2, F4, F6, FT9, F9, LO1, IO1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>104 - 172</td>
<td>$p = 0.04$</td>
<td>P1, PO3, O1, Iz, Oz, POz, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2, CB2</td>
</tr>
</tbody>
</table>

Figure 9. Group mean event-related potentials (ERPs) for the probe audio-clip: Memory-cue vs. neutral. ERPs elicited by the probe stimulus for memory-cue and neutral clips at test. The grey shaded box denotes significant differences between memory-cue and neutral clips. The iso-contour maps show the group mean amplitude for the 100-200 ms interval.
4.2.3 Target (0 ms – 800 ms)

4.2.3.1 Old vs. new

Next, we examined the effects of cuing on processing the auditory target. First, the ERPs were re-baseline corrected according to target onset (i.e., from 1800 to 2000 ms, -200 ms before target onset, Figure 1). The analysis revealed four different clusters. The most significant cluster was located over left frontal areas of the scalp, with greater sustained negativity for old scenes than for new ones (Table 5, Figure 10). The second cluster overlapped with the first cluster in time and was located over central-parietal, and parietal areas of the scalp. The difference in ERP amplitude between old and new conditions occurred between 363 and 593 ms after target onset, with greater sustained negativity for new scenes than for old ones. The third cluster preceded the first two clusters in time and occurred over central-parietal and parietal areas of the scalp, with greater sustained negativity for new scenes than for old ones. The fourth cluster preceded all other cluster in time and occurred over right and midline central areas of the scalp, with greater sustained negativity for old scenes than for new ones.

Table 5 Target tone: Summary of the channel level cluster-based permutation statistics for Old vs. new audio-clips.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Cluster</th>
<th>Latency (ms)</th>
<th>P value</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old vs. new</td>
<td>1</td>
<td>357 - 801</td>
<td>$p = 0.001$</td>
<td>FP1, AF7, AF3, F3, FT7, FC5, FPz, FP2, AF8, FT9, F9, LO1, IO1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>363 - 594</td>
<td>$p = 0.0007$</td>
<td>FC1, C1, C3, CP5, CP3, CP1, P1, P3, P5, PO3, O1, Oz, Pz, CPz, DC2, C2, C4, CP4, CP2, P4, PO8, PO4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>246 - 338</td>
<td>$p = 0.03$</td>
<td>C1, CP5, CP3, CP1, P1, P3, P5, PO3, O1, POz, Pz, CPz, CP4, CP2, P2, PO4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>143 - 230</td>
<td>$p = 0.03$</td>
<td>AF3, F1, F3, F5, F7, FC5, FC3, C1, C3, FPz, AFz, Fz</td>
</tr>
</tbody>
</table>
Figure 10. Group mean event-related potentials (ERPs) for the target: Old vs. new. ERPs elicited by the target stimulus for old and new clips at test. The grey shaded box denotes significant differences between old and new clips. The iso-contour maps show the group mean amplitude for cluster 2 and 3, with 500 - 600 ms and 200 - 300 ms intervals, respectively.

4.2.3.2 Memory-cue vs. neutral-cue

Following this analysis, we examined the effects of cuing on processing the auditory target. The ERPs were re-baseline corrected according to target onset (i.e., from 1800 to 2000 ms, -200 ms before target onset). The analysis revealed one cluster located over frontal and fronto-central areas of the scalp, with greater sustained negativity for memory-cue than for neutral scenes (Table 6, Figure 11).
Table 6 Target tone: Summary of the channel level cluster-based permutation statistics for Memory-cue vs. neutral audio-clips.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Cluster</th>
<th>Latency (ms)</th>
<th>P value</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory-cue vs. neutral</td>
<td>1</td>
<td>420 - 566</td>
<td>$p = 0.02$</td>
<td>AF3, F1, FC3, FC1, C1, AFz, Fz, F2, F6, FC6, FC4, FC2, FCz, Cz, C2, C4</td>
</tr>
</tbody>
</table>

**Figure 11.** Group mean event-related potentials (ERPs) for the target: Memory-cue vs. neutral. ERP traces for memory-cue and neutral clips for the target at test. The iso-contour maps show the group mean amplitude for the 500 - 600 ms interval.
5 Discussion

The current study sought to examine whether mere exposure can facilitate memory-guided attention in real-world sounds. Importantly, this project aimed to clarify the relationship between memory and attention in hearing, an area that remains largely unexplored. This study will fill a gap in the literature and provide a direct measure to investigate whether memory-guided attention holds in naturalistic, everyday listening situations. Memory is a dynamic system in that emphasis is placed on the connections between the system and the representations with which it interacts (Moscovitch, 1992, Romero & Moscovitch, 2015; Schacter et al., 2004). Therefore, the study contributes in several ways to what is known about contextual-cueing in natural listening situations and helps us understand how memory guides attention.

5.1 Old-new effect

The fact that old scenes elicited shorter response times compared to new ones suggests that familiarity with the context may speed responses, by allowing attentional resources to be allocated to processing the target. This conclusion is consistent with previous studies that have shown facilitation of RTs for old relative to new arrangements when engaging in visual (Chun & Jiang 1998) and tactile (Assumpção et al., 2015; Nabeta, et al., 2003) search tasks. In this paradigm, endogenous control may have facilitated the use of the context for subsequent target detection. Habitual search and allocation of attention, a form of endogenous control, is comprised of habits. Habits are developed over time as a result of repeated exposure. When one encounters the same context in the future, habitual search enables attention to be guided towards relevant information, while requiring fewer cognitive resources and allowing for resources to be used for processing the target (Le-Hoa Võ & Wolfe, 2015; Thompson & Sabik 2018, Trick et al., 2004).

The behavioural old-new effect is supported by the neural data. The ERP analyses revealed a number of significant modulations at cue audio-clip, probe audio-clip, and target. Two separate effects were observed.

The first effect revealed that new stimuli elicited greater sustained negativity than did old stimuli over midline and frontal sites. This effect was mirrored in lower inferior areas of the scalp as a polarity reversal. This finding is consistent with previous research that has shown greater
positivity for old compared to new items (Mograss et al., 2006). The early modulation may correspond to familiarity for the audio-clip. The left parietal old-new effect has been associated with memory processing (Curran, 2000; Donaldson & Rugg, 1998; Friedman & Johnson, 2000; Rugg & Curran, 2007; Vilberg, Moosavi, & Rugg, 2006; Vilberg & Rugg, 2007; Wilding, 2000).

In addition, the later sustained activity may index maintenance of the memory in short-term memory, as participants were not instructed to make a response during the cue audio-clip.

The second effect was comprised of two modulations during the target period. Old clips at target elicited greater sustained negativity than did new clips. This finding was surprising, as we did not have any a priori hypotheses pertaining to this effect. The significant difference in ERP amplitude elicited by new and old clips suggests that exposure to the preceding context affects how the target itself is processed.

5.2 Memory-guided attention: Memory-cue vs. neutral-cue scenes in detecting lateralized targets

5.2.1 Behavioural results

The behavioural results regarding the difference in the speed of lateralized target detection between old memory-cue clips and old uncued neutral clips were not consistent with the ERP data. Although we did not find any benefits in RT of target detection for memory-clips compared to neutral clips, we did find a significant difference in ERP amplitudes comparing the two conditions. This finding was surprising to us; however, previous EEG research has shown that electrophysiological measures are, in some cases, more sensitive than behavioural measures (e.g., Cameron et al., 2019; Francois & Schön 2011; Peretz et al. 2009).

Regarding the behavioural data, when comparing old scenes that contained a memory-cue to those that did not (neutral), there was no significant difference in response time to lateralized targets. Memory for the scene facilitates processing of the target; this effect, however, was not lateralized. The absence of lateralization may be due to conditions during exposure that may have affected encoding of the scene-tone association.

Firstly, interference from the suprathreshold tone at exposure may have disrupted the encoding of the scene-tone association. This interpretation is supported by the fact that memory accuracy
of target detection for memory-cue scenes that contained a suprathreshold target were significantly worse than for neutral scenes that did not contain a tone embedded in them.

Alternatively, the behavioural results suggest that when attention is drawn to the scene (background), the encoding of the scene-tone association is not strong enough to elicit a congruency benefit in reaction time for detecting the target at test. This interpretation is supported by the fact that participants’ memory for the scene was significantly above chance, but their memory for the scene-tone association was not significantly different from chance. Previous research has shown that visual foreground-background segmentation is an important aspect of guiding attention (Wolfe, 2003). Furthermore, Zang et al. (2016) showed that the perceptual segmentation of foreground and background precedes, and may even mediate, contextual learning. Information that is perceptually grouped and foregrounded is thought to be prioritized then for attention (Mazza et al., 2005; Nakayama et al., 1989) and contextual learning (Jiang & Leung, 2005; Pollmann & Manginelli, 2009). In the case of the current study, selective attention was guided towards the scene at exposure, rather than the target. However, at test, attention was guided towards the target. It is possible that at test, the target may have already been established as the background. This possibility may account for the strong memory for the scene (foreground) and poor memory for the target’s location at exposure (background).

We conducted a follow-up study to see if changing where attention is oriented at exposure could account for the null behavioural effects. The follow-up study methods were very similar to the original study, with the exception of the target stimulus that could vary in pitch: High (1000 Hz) or low (500 Hz). At exposure, participants were required to indicate if the target was high or low. No reference was made to the scenes or to the spatial location of the target. The test phase remained identical to the original experiment. Preliminary analyses of the data reveal that memory-cue clips had faster RTs than did neutral clips. These results suggest that memory-guided attention in the case of mere exposure is possible if the listener’s attention is guided towards the target at exposure.

Taken together, the behavioural results suggest that the effect of memory-guided attention in the case of mere exposure is dependent on how the attended object is processed in the auditory scene (i.e. foreground and background relations). In addition, the behavioural results highlight the difference between associations that are formed via mere exposure and those that are formed in a
context in which there is conscious effort, but no conscious memory, of the associations (Zimmermann et al., 2017). Shanks (2004) argues that there is a need to obtain evidence for an independent implicit learning process. Therefore, an interesting avenue for future research could be centered around quantifying the learning conditions necessary to implicate true implicit memory systems.

5.2.2 EEG results

It was expected that the neural indices would reflect both memory and attentional components. Previous studies have shown that neural indices of memory-guided attention are expressed in terms of processes that enable preparation that is driven by reinstatement of the context for the memory, and, also, target selection (Chun & Jiang, 2003; Stokes et al. 2012; Summerfield et al., 2011). The neural data showed significant modulations, corresponding to two separate effects at cue audio-clip, probe audio-clip, and target.

The first effect revealed that memory-cue stimuli elicited greater sustained negativity than did neutral stimuli. This difference suggests that although differences in behaviour are not observed, clips that have a memory-cue have a significantly different neural signature than scenes that do not. Greater ERP amplitude for neutral clips than for memory-cue ones may reflect the fact that neutral scenes elicited more prediction violations compared to memory-cue scenes. For example, neutral clips may be less informative of the target’s location compared to memory-cue clips. Without any a priori expectations regarding the spatial location of the tone, participants may split their attention equally between the left and right auditory fields. Consistent with this hypothesis are studies that have shown that activity related to the attention controlling networks is modulated by the predictability of a certain stimulus event to occur, in which predictable events have been shown to elicit lower activity than do less predictable ones (Doricchi et al., 2010; Shulman et al., 2007).

The second effect revealed that neutral clips elicited greater sustained negativity than for memory-cue clips over frontal sites. This finding was surprising, as we did not have any a priori hypotheses pertaining to this effect.
5.3 Future directions

5.3.1 Behavioural

Research examining individual differences in attentional style is growing and is becoming more sophisticated (van Calster et al., 2018). Attentional style is thought to impact highly on perceptual abilities (Scolari & Serences, 2009). In the future, it would be interesting to have participants complete a questionnaire on their attentional style. Individual differences in allocation of attention have been shown to be affected by working memory capacity (Unsworth & Robison, 2015), Type A behaviour pattern (Ishizaka, et al., 2001), global task difficulty (Valérya et al., 2019), need for closure (Szumowska & Kossowska, 2017), and the ability to resist distractors (Fukuda & Vogel, 2008).

A preliminary analysis of the data set, using hierarchical clustering on the behavioural reaction time data across conditions, revealed two groups, each with high internal consistency (Group 1: Cronbach’s $\alpha = 0.913$ $N = 14$ and Group 2: Cronbach’s $\alpha = 0.873$ $N = 11$, respectively). This finding suggests that individual differences may play an important role in auditory memory-guided attention, but the number of participants in each group were too few to enable meaningful statistical analyses.

In future experiments, examining individual differences may be a useful way to capture variability in performance in memory-guided attention tasks that may be particularly sensitive to cognitive style.

There are also important cultural factors to consider. Interestingly, but not surprisingly, vision studies have shown that Westerners differ from East Asians in their cognitive style. Westerners have been shown to pay attention to salient objects and spend a longer time processing focal objects, while East Asians have been shown to take longer to process central items and to focus on the relations between foreground and background (Boduroglu et al., 2009; Chua, et al., 2005; Masuda & Nisbett, 2001; 2006; Nisbett & Masuda, 2003; Nisbett & Miyamoto, 2005).

5.3.2 Neural

Time-frequency analysis could be used to complement the ERP analyses. Time frequency analysis is a good tool for examining cognitive processes that are not precisely time-locked, such
as memory and attention (Roach & Mathalon, 2008). Change in oscillatory activity in the inferior parietal lobe has been shown to be indicative of automatic attention (Corbetta & Shulman, 2002). Furthermore, alpha oscillations have been associated with spatial attention (Worden et al., 2000). Assessing these neural indices would allow us to examine further how spatial information about target location is stored and attention guided laterally.

Dynamic coherence techniques would allow us to bridge the gap between acoustic material and higher-level factors, such as memory and attention. Research has shown that memory-guided attention is, in part, driven by the reinstatement of the context for the memory (Chun & Jiang, 2003; Stokes et al. 2012; Summerfield et al., 2011). If the pattern of activity at encoding is similar to the pattern at retrieval of old scenes, it will indicate that there is a reinstatement of context for the audio-clip. This finding may, additionally, serve as evidence that our capacity to form successful associations is supported by implicit associative learning that aids in preparatory memory-guided gains. Additionally, if the degree of similarity between neural patterns at encoding and at retrieval varies according to reaction time and to recollection or familiarity, it may reflect a reinstatement of context for the memory (Xue, 2018).

Connectivity analyses would also be of benefit. They would allow us to examine further the specific memory system that underlies memory-guided attention in natural listening. In vision, contextual cueing has been shown to be dependent on the medial temporal lobes (Chun & Phelps, 1999; Manns & Squire, 2001; Negash et al., 2007). The hippocampus has been identified as an important structure for quickly guiding attention (Goldfarb et al., 2016). There is some evidence that contextual cueing relies on associative learning that is supported by medial temporal lobes, regardless of whether or not the information is accessible to conscious awareness (Park et al., 2004; Ryan et al., 2000).
6 Conclusion

Taken together, the results show that familiar contexts elicit greater processing efficiency. The old-new effect is supported by both the behavioural and neural data. Although memory facilitates efficient allocation of auditory resources, it does not facilitate the spatial allocation of attention. The neural data, however, reveal a significant difference between memory-cue clips and neutral ones. It is possible that memory-cue clips contain predictive cues for orientation of spatial attention that neutral clips do not, thus eliciting more prediction-violations.

Memory-guided auditory attention is a largely unexplored area. Further elucidating the link between memory and attention will help us better understand memory as a dynamic system (Romero & Moscovitch, 2015) that can benefit from, and contribute to, attentional processes. In addition, this study will contribute to the field of auditory perception, by further defining the internal mechanism underlying schema-based scene analysis. Finally, the ecological validity of the proposed paradigm will enable us to adapt it to the world outside the laboratory, and aid people who have memory or hearing impairments, such as older adults and those with memory disorders.
References


## Appendices

### Appendix A: List of audio-clips used

<table>
<thead>
<tr>
<th>OLD SCENES:</th>
<th>CHICKEN</th>
<th>WARGUNSHOTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTRUCTION</td>
<td>CHILDREN</td>
<td>WASHINGDISHES</td>
</tr>
<tr>
<td>COOKING</td>
<td>CHIMES</td>
<td>WINDSHUTTERS</td>
</tr>
<tr>
<td>COWSMOOING</td>
<td>CHURCH</td>
<td>CAROUSEL</td>
</tr>
<tr>
<td>DOGS</td>
<td>CLASSROOM</td>
<td>SANDPAPER</td>
</tr>
<tr>
<td>SMALLROCKSFALLING</td>
<td>NOISYCHEWING</td>
<td>ELECTRICITY</td>
</tr>
<tr>
<td>BOXING</td>
<td>LIONSROARING</td>
<td>PLANEOVERHEAD</td>
</tr>
<tr>
<td>EMERGENCY</td>
<td>MARINA</td>
<td>LEAVES</td>
</tr>
<tr>
<td>FACTORY</td>
<td>MUSICBOXES</td>
<td>LIGHTNING</td>
</tr>
<tr>
<td>FIREBURNING</td>
<td>SNAKES</td>
<td>MARCHINGBAND</td>
</tr>
<tr>
<td>FROGPOND</td>
<td>SNORING</td>
<td>MEDIEVALBATTLE</td>
</tr>
<tr>
<td>GLASSBREAKING</td>
<td>TYPING</td>
<td>MONKEYSZOO</td>
</tr>
<tr>
<td>HAIRDRYING</td>
<td>WHISPERINGVOICES</td>
<td>MONKSPRAYING</td>
</tr>
<tr>
<td>HIGHWAY</td>
<td>BOWLING</td>
<td>NOISYOFFICE</td>
</tr>
<tr>
<td>HORSES</td>
<td>SEAGULLS</td>
<td>ORCHESTRATUNING</td>
</tr>
<tr>
<td>LAUGHING</td>
<td>SHEEP</td>
<td>OWLS</td>
</tr>
<tr>
<td>ALARMS</td>
<td>SICK</td>
<td>PARADE</td>
</tr>
<tr>
<td>ANGRYMOB</td>
<td>STREETHOCKEY</td>
<td>PEOPLETALKING</td>
</tr>
<tr>
<td>APPLAUSE</td>
<td>SUMMERNIGHT</td>
<td>RUNNINGDOWNSTAIRS</td>
</tr>
<tr>
<td>ARCADES</td>
<td>SUPERMARKET</td>
<td>PRESSCONFERENCE</td>
</tr>
<tr>
<td>BABIES</td>
<td>TABLETENNIS</td>
<td>RAINING</td>
</tr>
<tr>
<td>BALLDRUMMING</td>
<td>TOILET</td>
<td>RIVER</td>
</tr>
<tr>
<td>BASEKETBALLGAME</td>
<td>TRAFFIC</td>
<td>SCREAMING</td>
</tr>
<tr>
<td>BEESANDINSECTS</td>
<td>TRAIN</td>
<td>FLIPPINGBOOKPAGES</td>
</tr>
<tr>
<td>BUBBLING</td>
<td>TRIBALAFRICAN</td>
<td>FORESTWOODPECKER</td>
</tr>
<tr>
<td>CATS</td>
<td>UNDERWATER</td>
<td>CHRISTMAS</td>
</tr>
<tr>
<td>CHEERINGCROWD</td>
<td>WALKINGINWOODS</td>
<td>COINS</td>
</tr>
</tbody>
</table>
NEW SCENES
AMUSEMENTPARK
BUS
COINS
DRUNKENSINGING
METRO
PARACHUTEOPENING
ROCKETLAUNCH
BLACKSMITH
BOMBING
CHRISTIANPRAYER
COWBELLS
DEMOLITION
DRAGCARRACE
EERIEFOREST
FIREWORKS
GEESE
HORROR
SOUNDSCAPE
MACHINERY
PARROTS
PIGS
SCARYPLAYGROUND
SMASHINGGUITARS
THROWINGGLASS
WOLVES