Incidental auditory learning and memory-guided attention: Examining the role of attention at the behavioural and neural level using EEG

Manda Fischer a,b,⁎, Morris Moscovitch a,b, Claude Alain a,b

a Rotman Research Institute, Baycrest Hospital, Toronto, Canada
b University of Toronto, Department of Psychology, Toronto, Canada

ARTICLE INFO

Keywords:
Auditory Attention Memory Implicit EEG

ABSTRACT

The current study addressed the relation between awareness, attention, and memory, by examining whether merely presenting a tone and audio-clip, without deliberately associating one with the other, was sufficient to bias attention to a given side. Participants were exposed to 80 different audio-clips (half included a lateralized pure tone) and told to classify audio-clips as natural (e.g., waterfall) or manmade (e.g., airplane engine). A surprise memory test followed, in which participants pressed a button to a lateralized faint tone (target) embedded in each audio-clip. They also indicated if the clip was (i) old/new; (ii) recollected/familiar; and (iii) if the tone was on left/right/not present when they heard the clip at exposure. The results demonstrate good explicit memory for the clip, but not for tone location. Response times were 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 and a difference between old clips that were previously associated with the target tone and those that were not. These results are consistent with the attention-dependent learning hypothesis and suggest that associations were formed incidentally at a neural level (silent memory trace or engram), but these associations did not guide attention at a level that influenced behaviour either explicitly or implicitly.

1. Introduction

Experimental Psychology is shifting focus from studying cognitive processes in isolation to considering them as components in a complex, interactive system (Moscovitch et al., 2016; Romero and 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 (Alain and Bernstein, 2015; 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. Attention and memory, two important and tightly related phenomena, play a vital role in disambiguating complex acoustic signals in everyday listening situations (Backer and Alain, 2013; Backer et al., 2015). Most research has focussed on the effects of attention on memory; few studies have investigated the converse, that memory may guide attention. The proposed study seeks to examine the relation among learning, attention, and awareness that mediate memory and attention when a person listens to audio-clips of everyday real-world environments.

1.1. Attention as a mediating factor for the effect of implicit long-term memory (LTM) on perceptual organization

Previous research suggests that attention is the primary mediating factor that enhances sensory processing when familiarity of stimuli is manipulated. Chun and Jiang (1998) were among the first to demonstrate that memory may influence the deployment of attention and facilitate target localization. This study, and the majority of subsequent studies examining contextual cueing, presented participants with randomly generated configurations. Each configuration had a target that always appeared in the same location. Old invariant conditions were those that repeated throughout the entire experiment. Participants were not told to memorize the spatial arrangement of the distractors or the target location. Their response times, however, were faster when the target was embedded in old invariant configurations compared to new
variant conditions. A recognition test at the end of the testing session was used to determine the nature of the participants’ knowledge of the repeated configurations. The results indicated that explicit memory for the configurations themselves were at chance level (52%). Based on these results, the authors concluded that implicit processes facilitated the encoding of context which, in turn, influenced and constrained expectation and visual search behaviour (e.g., Chun and Jiang, 1998; Chun and Jiang, 2005). Interestingly, however, three out of 14 participants (Experiment 1) and seven out of 18 participants (Experiment 5) reported being aware that certain configurations were repeated. Even though awareness of repetition was not associated with explicit recall of the configurations, one cannot rule out the possibility that memory representations for configurations might have been contaminated by explicit processes.

Zimmermann et al. (2017) examined the effect of memory-guided attention in audition. They showed that deliberately forming associations between audio-clips and tones enables memory to bias attention to auditory stimuli. 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 of a subsequent test on participants’ memory for the learned associations exhibited facilitation of target localization. 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 of a separate implicit system at play, independent from any explicit knowledge.

It is unclear, however, whether incidental learning of the association between context and target and independent implicit processes are truly implicated in these studies. A number of criteria have been established to assess awareness (Reingold and Merikle, 1999; Shanks and St John, 1994). Chun and Jiang (1998) violated the information criterion criteria by not asking participants to report the nature of their memory for the target-context association. Although participants’ memory for the context may be implicit, it is unknown whether or not they can explicitly express context-to-target associations. This information regarding the association is presumably what guides attention to afford a behavioural benefit. Therefore, in order to truly test for implicit learning, the test used must measure the same stored knowledge that is influencing behaviour (Dulany, 1961; Shanks and St John, 1994).

Zimmermann et al. (2017) trained participants to explicitly pair the context with the target location. At test, however, they used “absence of explicit recall” of these associations as a metric for implicit memory. This is problematic, as previous studies have demonstrated above chance performance on implicit tasks may still occur without a clear dissociation between recognition and priming. Statistical anomalies and noise introduced into a model for explicit learning can yield these dissociative results (Ostergaard, 1992; Plaut, 1995; Poldrack et al., 1999; Shanks, 2005). Therefore, these tests do not rule out contamination from explicit memory processes.

The fact that participants were instructed to do the same task both at learning and at test may be additionally problematic. Using different tasks at learning and at test can minimize contamination of implicit memory by explicit memory systems and allow for differential “transfer-appropriate processing” across conditions (Paller et al., 1998; Schacter et al., 2007). If attention at learning is not manipulated, contamination between explicit and implicit memory systems cannot be ruled out (Wolters and Prinsen, 1997). Therefore, whether implicit learning can indeed facilitate memory-guided attention to auditory objects remains an open question.

Together, the literature on memory-guided attention suggests that when attention is drawn to either (i) the target but not to the context (e.g., Chun and Jiang, 1998; Chun and Jiang, 2005) or (ii) the association between scene (context) and tone (target) (Zimmermann et al., 2017), memory-guided attention is facilitated. However, what has not been examined is the extent to which incidental associations between context and target can also facilitate memory-guided attention. The attention-dependent learning hypothesis posits that expression of learning, but not learning of associations themselves, in terms of a behavioural benefit, depend on task relevance and attention. Therefore, in order to confirm whether incidental learning and implicit processes can guide attention at the behavioural level, a paradigm that tests for the incidental learning of the associations themselves is needed.

1.2. Current study

The current experiment aims to address these limitations, by examining the role of attention at encoding on memory-guided attention at retrieval. Here, we investigated whether the effect of memory-guided attention holds in natural listening situations in which incidental associations are formed through implicit learning (exposure). By not having participants deliberately associate the target and the clip at encoding, we provided the conditions at learning necessary to implicate implicit processes. Additionally, we included a forced-choice question about the association between clip and target in order to test objectively for implicit learning and awareness of the associations formed (Shanks, 2005).

The paradigm used by Chun and Jiang (1998) has an unaware element with repeated meaningless geometric configurations. Other studies by Patai et al. (2012) use meaningful stimuli but rely on explicit knowledge of the target location in relation to the context (e.g., target to the left of the beach umbrella). Our study combines elements of both of these paradigms to address implicit processing within real-world contexts.

The validity of this effect was tested behaviourally, and the neural correlates underlying this memory-retrieval process were examined using electroencephalogram (EEG). Specifically, we aimed (1) to examine further the link between auditory memory and attention, by testing whether incidental associations between tone and audio-clip facilitated localization of a lateralized pure tone (i.e., target); and (2) to use EEG to index neural (implicit) processes underlying this effect.

Rugg et al. (1998) used event-related potentials (ERPs) to reveal separate neural patterns of memory-related activity. Contrasts between old recognized and old unrecognized words and new and old unrecognized words in a recognition task over parietal areas of the scalp revealed the neural dissociation between implicit and explicit memory. Furthermore, to address the issue of simply studying the difference between recognized and unrecognized items in the standard recognition paradigms, Paller et al. (1998) conducted a study to manipulate the amount of priming, independent of influences on recollection. This paradigm allowed for the identification of a unique ERP correlate over occipital sites, indexing implicit memory. Although we used different procedures and materials, in the current study, we predicted that ERPs would show a difference during the sustained potential over parietal and auditory temporal sites between clips that had been associated with the target compared to those that were not, indexing the implicit encoding of audio-clip-to-target associations. Evidence of priming in the auditory cortex would complement the findings of Paller et al. (1998), who found an electrophysiological measure of priming over occipital sites for visual stimuli.

The results of the current study revealed neural differences indicative of the formation of a neural trace/engram of the learned associations between tone presence and audio-clip (i.e., auditory scene), but no differences at the level of behaviour (i.e., reaction time for localizing the pure tone). For the remainder of the paper, we will refer to effects observed with EEG as reflective of a neural trace/engram and behavioural effects as reflective of implicit memory. The dissociation between a neural memory trace and an implicit memory (the behavioural consequence) are discussed in detail in the Discussion section.
2. Methods

2.1. Participants

Twenty-six healthy young adults participated in the experiment. Data from one participant was removed due to issues with task compliance. The remaining 25 participants (M = 26.1 years; SD = 4.3; 10F) had normal hearing, as assessed using pure tone audiometry, and had no history of psychiatric, neurological, or other major illnesses. Effect sizes pertaining to the main effect of memory cue in three experiments described in Zimmermann et al. (2017) (Cohen’s $d_z = 1.53$, 0.58, and 0.64, respectively) and in one experiment in Zimmermann et al. (2020) (Cohen’s $d_z = 0.53$) were calculated. We then used the most conservative effect size to compute the estimated power of our study for a paired t-test between dependent variables, using G’Power 3.1.9.4 (Faul et al., 2007). The estimated power achieved was 0.84. Participants were recruited from the Rotman Research Institute participant database and received monetary compensation for their participation. None of the participants in the current study took part in previous experiments of a similar nature. The study is certified for ethical compliance by the Research Ethics Board at Baycrest and all participants provided written consent.

2.2. Stimuli

A total of 108 everyday audio-clips, selected from “http://www.freesounds.org/,” were included in the experiment. An initial pilot test confirmed that each clip in the set was nameable and that there was considerable consistency among participants. Each clip could be given a semantic label (e.g., coughing), which increased 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 and Gregg, 2011). 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 65 decibel (dB) sound pressure level (SPL) on average across stimulus duration, with brief peak amplitudes of 80 dB SPL.

Eighty clips were used during the exposure phases, in which 20 were paired with a left ear auditory target and 20 with a right ear auditory target. These clips (referred to as old items) were also presented at the surprise memory-test phase. 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) was embedded in 50% of the clips at exposure (80 dBA SPL) and in all clips at test (55 dBA SPL). 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 was constant across participants. The screen was black with a white fixation cross at the center during the presentation of auditory stimuli. Acoustic stimuli and visual cues were presented, using Presentation software (version 13, Neurobehavioral Systems, Albany, CA). Sound levels were measured, using a Larson Davis System 824 and a 2CC coupler.

2.3. Procedure

Participants were first fitted with an electrode cap and prepared for EEG recording. Then, they were presented with the exposure and the test phases (see Fig. 1).

2.4. Exposure phase (encoding)

In the exposure phase, participants were presented with four blocks of 80 audio-clips; half of them comprised of a lateralized (left or right ear) pure tone stimulus. Prior to the study, we ensured that the pure tone was clearly audible (suprathreshold 80 dBA SPL), allowing implicit incidental learning, via mere exposure, to proceed. During the presentation of the sound clip, participants were instructed to fixate on the cross at the center of the black screen. For each trial, participants indicated whether the audio-clip depicted a natural (nature) scene (e.g., waterfall) or a manmade object (e.g., airplane engine). In line with the incidental nature of encoding, we made no reference to the embedded tone or to any memory test. To respond, participants used the “left” and “right” arrow keys on a computer keyboard. 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-cue clip pairings with clip were randomized across participants. This ensured that any effects uncovered would not be attributable to key-response, order-of-clips, or specific pairing between clip and spatial location.

2.5. Test phase (retrieval)

After the exposure phase, participants were given a surprise memory test (retrieval) on the learned associations. In each trial, participants were presented with a cue, followed by a brief delay and a probe audio-clip that included a faint lateralized target tone (i.e., 55 dBA SPL). An initial pilot test was conducted by Zimmermann (2018), demonstrating that a single probe was not sufficient to allow participants to activate
auditory memory for the target location. We conducted a pilot study, with a sample of six participants, in which participants detected the target. The duration and intensity of the target tone was chosen based on their performance, such that the target was not entirely masked by the clip (approximately 80–90% correct localization), thereby remaining audible. This was done to ensure that all participants engaged in effortful listening (Alain et al., 2018; Pichora-Fuller et al., 2017).

The test was comprised of old and new audio-clips. The old audio-clips were divided into memory and neutral-cue clips. The memory-cue clips included a lateralized tone during the exposure phase. The neutral-cue clips did not include a lateralized tone during exposure. In addition, we included catch trials in which the pure tone target was presented in the opposite ear to the one used during the exposure phase. The ratio between memory-cue and catch trials was 80:20. During the presentation of the sound clips, participants were instructed to fixate on the cross at the center of the black screen.

Participants were required to indicate 1) if, at test, the tone was presented on the left or on the right (localize faint pure tone as quickly as possible with key press); 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 exposure; and 4) if, at exposure, the tone was either on the left, right, or not present. Participants used the “left” and “right” arrow keys for questions 1–3 and “left”, “right”, and “down” arrow keys for question 4. If participants responded “New” to question 2, the experiment proceeded to the next trial and did not present questions 3 and 4. 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-clips, respectively.

Question 3 (Recollect/Familiar?), common to traditional memory paradigms, was included in order to disentangle processes underlying recollection of the clip. Recollection was defined as a sound that participants “can remember, from [their] own perspective […] because [they] can remember additional details associated with it.” Familiar was defined as a sound that participants “can recognize […] but can’t remember any additional details associated with it. Nonetheless, [they] have the strong impression, or are even sure, that it occurred earlier in the experiment.”

2.6. EEG recording and analysis

The EEG was recorded continuously during each exposure phase and during the memory-guided attention test phase, 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 (two symmetrical electrodes on the mastoid, pre-auricular points, outer canthus of each eye, inferior orbit of each eye, and 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).

Data pertaining to five participants were not used because of excessive eye movements and/or muscle artifacts. As a result, 20 participants were included in the EEG analysis. The EEG data were first 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 accounted 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-blinks. 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 two sets of epochs: Cue audio-clip and probe audio-clip. The cue audio-clip epochs started 200 ms before clip onset (0 ms) 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 audio-clip).

3. Statistical analyses

Catch trials were not analyzed, as there were too few correctly-localized 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 clips to old catch trials, was not possible.

3.1. Behavioural

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

3.1.1. Old-new effect

In order to assess the contextual component of memory-guided attention, old clips (memory-cue and neutral-cue) were compared to new clips, using a paired two-sample t-test. The 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.

3.1.2. Memory-guided attention effect

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

3.1.3. Recollect and familiar effect

To investigate the potential memory system that underlies explicit retrieval of the sound-clip, we examined whether sound-clips that were correctly categorized as old were recollected or familiar. Although recognition of the sound-clip itself was not the main focus of the paper, we nevertheless wanted to better understand how explicit processing of the clips might have affected incidental encoding and processing of the associations between sound-clip and target tone. It is possible that RT for localizing the pure tone at test differs for recollected and for familiar
sound-clips. Despite the fact that the memory-guided attention effect was not observed behaviourally, we compared RT for recollected and familiar clips to first see if there were any marked differences between the two dependent measures.

Furthermore, clips that were deemed “missed” were those that were old but were judged by the participant as new (i.e., implicit memory of the sound-clip). We had hoped to analyze missed trials, but the number of trials for this condition was too low due to above chance accuracy at correctly classifying the sound-clips as old or new.

3.1.4. ERP statistical analyses

For statistical analyses, the ERP waveforms were exported into BESA Statistics 2.0. BESA Statistics 2.0 software automatically identifies clusters in time and space, using a series of t-tests that compared the ERP amplitude 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 roughly 95% (one-sided t-test) of all clusters derived by random permutation of the data. The number of permutations was set at 1000.

Cluster-based statistics were performed on two different time windows: cue audio-clip and probe audio-clip. For old and new audio-clips, it was expected that ERPs would show a difference during cue audio-clip, indicating that a neural trace reflecting the formation of a memory representation, specifically of auditory clips, was not observed behaviourally, we compared RT for recollected and familiar clips to first see if there were any marked differences between the two dependent measures.

Furthermore, clips that were deemed “missed” were those that were old but were judged by the participant as new (i.e., implicit memory of the sound-clip). We had hoped to analyze missed trials, but the number of trials for this condition was too low due to above chance accuracy at correctly classifying the sound-clips as old or new.

We 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 1000.

Cluster-based statistics were performed on two different time windows: cue audio-clip and probe audio-clip. For old and new audio-clips, it was expected that ERPs would show a difference during cue audio-clip, indicating that a neural trace reflecting the formation of a memory representation, specifically of auditory clips, was not observed behaviourally, we compared RT for recollected and familiar clips to first see if there were any marked differences between the two dependent measures.

4. Results and comment

4.1. Behavioural results

Overall memory accuracy for old and new audio-clips (question 2: Old versus New) and the sound-tone association (question 4: Left, Right, None) were assessed. The results revealed that old (M = 77.3%, SE = 2.40) and new (M = 78.2%, SE = 2.30) clips were correctly categorized as old and new above chance level (50%), t(24) = 11.13, p < .001, d = 2.22 and t(24) = 12.04, p < .001, d = 2.41, respectively (Fig. 2 left panel). There was no significant difference between memory accuracy for recollected clips (M = 88.57% SE = 0.03) and for familiar clips (M = 89.22%, SE = 0.04), t(24) = −0.157, p = .88, d = 0.03. Hits minus false alarms for clip accuracy (M = 55.5%, SE = 2.95) was also significantly above chance, t(24) = 18.84, p < .0001, d = 3.77, suggesting that participants were capable of distinguishing between old and new items.

4.1.1. Sound-tone association

Accuracy for reporting whether, for memory-cue and neutral-cue clips, the target tone was on the left, right, or not present during exposure, was compared to chance accuracy (33%), using a one-sample t-test. The results revealed that participants performed at chance level when asked to report the tone’s location relative to the particular clip, t(24) = −1.62, p = .12, d = −0.35, BF10 = 2.04e + 15 (Fig. 2 right panel). Using a one-sample t-test, sound-tone accuracy was analyzed separately for memory-cue clips (left/right vs. no tone) and for neutral-cue clips (no tone vs. left/right). Participants still performed at chance level (50%), t(24) = −0.04, p = .97, d = −0.008, BF10 = 0.21 and t(24) = 0.89, p = .38, d = 0.18, BF10 = 0.30, respectively, suggesting that they could explicitly report neither the tone’s location (memory-cue) nor its presence or absence (neutral-cue). All comparisons were Bonferroni corrected.

4.1.2. Old-new effect

The average RT for localizing the pure tone at test for correctly categorized old clips (M = 690.82 ms, SE = 34.07) was compared to the average RT for correctly categorized new clips (M = 734.62 ms, SE = 42.40). The contrast yielded a significant difference between the two conditions, t(24) = 2.56, p = .009, d = 0.52, suggesting that memory for the clip facilitates target localization (Fig. 3 left panel).

4.1.3. Memory-guided attention effect

Contrasts between average RT in localizing the pure tone at test for correctly categorized memory- and neutral-cue conditions did not yield a significant difference, t(24) = 1.14, p = .26, d = 0.23, BF10 = .375 (Fig. 3 right panel).

4.1.4. Recollect and familiar effect

Audio-clips that were categorized as old were further categorized as either recollected (M_{recollected old} = 55.38%, SE = 3.81) or familiar (M_{familiar old} = 44.62%, SE = 3.81). A two-way repeated measures analysis of variance (ANOVA) was conducted to compare RTs for memory-cue and neutral-cue conditions in terms of the recognition memory involved. The analysis showed a significant interaction, F(1, 24) = 5.62, p = .03, η²_p = .19 (Fig. 4). Post-hoc Bonferroni corrected t-tests revealed no significant differences between the pairwise comparisons, p > .05.

As there were no differences in accuracy, and the interaction observed in RT was not predicted, we have no ready interpretation for these results. One possibility is that in the memory condition, recollection brings to mind a tone, but not its location, thereby distracting the participant at test and slowing down responses compared to the familiarity condition when no contextual cue (tone) is recalled. Because no tone was presented at exposure in the neutral-cue condition, recollection, compared to familiarity, led to more rapid processing of neutral-cue clips.

4.1.5. Electrophysiological results at encoding phase (encoding)

In order to rule out the possibility of attentional capture/distraction by the tone at exposure, we examined group mean ERP at exposure. Fig. 5 pertains to group averaged ERPs for trials that contained the left or right tone (memory-cue) and for trials that did not contain a tone (neutral-cue), time-locked and baseline corrected, using 100 ms prior to the onset of the tone. Both left and right ear tones embedded in the audio clip generated an acoustic change complex (ACC), which was expected. There was no evidence of a P3a. The presence of ACC and lack of P3a suggest that the tones were clearly perceptible, but that they did not capture nor yield a major reorienting of attention.

4.1.6. Electrophysiological results at test phase (retrieval)

Fig. 6 shows group mean ERPs elicited by the cue and probe audio-clips. The onset of both was associated with large transient evoked responses.

The contrast assessing whether ERPs differed between old and new and memory-cue and neutral-cue clips was performed from 0 to 2000 ms relative to stimulus onset (i.e., cue or probe onset). We did not assess whether ERPs differed between correctly recollected and familiar clips due to an insufficient number of trials.

4.1.7. Cue audio-clip (0 ms–2000 ms)

Generally, the cue audio-clips generated transient onset responses that were largest at midline central electrodes, and which were followed by a sustained potential over the left central, temporal and temporal-parietal sites.

\[ d = \frac{z}{\sqrt{N}} \]

1 All Cohen’s d calculations were computed using the following formula: \( d = \frac{z}{\sqrt{N}} \).
Old vs. New. In this analysis, we examined whether there were neural indices underlying the behavioural difference in new and old clips. ERP averages for old clips included both memory-cue and neutral-cue clips. Cluster permutations yielded five significant clusters (Table 1, Fig. 7). The first cluster was associated with enhanced positivity for old over new clips over parietal areas of the scalp at 777–1072 ms, which reversed polarity at frontal sites (Table 1, Cluster 3). The second cluster followed the first one in time and revealed enhanced negativity at 1307–1998 ms for old over new clips over lateral frontal scalp area. The third cluster overlapped with the first cluster in time and showed

Fig. 2. Group mean accuracy at test phase. Left panel: Memory accuracy for old and new clips (Q2: Old or New?). Right panel: Memory accuracy for sound-tone association for memory-cue clips. (Q4: At exposure, was target on left, right or no target?). Chance level of 33% was determined based on the three choices provided in question 4 (left, right, none). Error bars represent standard error of the mean.

Fig. 3. Reaction times (RT) for localizing the faint pure tone at test. Q1: At test, localize tone on the left or right. Left panel: RT for localizing the pure tone at test for correctly categorized old and new clips. Old clips were comprised of both memory-cue and neutral-cue clips. Right panel: Average tone-localization RT for correctly “categorized as old” memory-cue and neutral-cue clips. Error bars represent standard error of the mean.

Fig. 4. Interaction plot depicting RTs for old clips in terms of the recognition type. Q3: At test, is the audio-clip recollected or familiar? Group mean RT for localizing the faint pure tone at test for correctly remembered old recollected and familiar clips. Error bars represent standard error of the mean.

Fig. 5. Group mean event-related potentials (ERPs) at exposure. ACC = acoustic change complex. Fz = midline frontal electrode.

Fig. 6. Butterfly plot of group mean 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.

Old vs. New. In this analysis, we examined whether there were neural indices underlying the behavioural difference in new and old clips. ERP averages for old clips included both memory-cue and neutral-cue clips. Cluster permutations yielded five significant clusters (Table 1, Fig. 7). The first cluster was associated with enhanced positivity for old over new clips over parietal areas of the scalp at 777–1072 ms, which reversed polarity at frontal sites (Table 1, Cluster 3). The second cluster followed the first one in time and revealed enhanced negativity at 1307–1998 ms for old over new clips over lateral frontal scalp area. The third cluster overlapped with the first cluster in time and showed
modulations over the frontal scalp area. The fourth and fifth cluster overlapped one another, preceded all other clusters in time and showed modulations in the central-parietal and parietal areas of the scalp. Correlations between memory accuracy for old clips and the absolute difference in ERP amplitude between old and new clips were computed for each cluster latency. These correlations were not significant, \( p > .05 \).

**Memory-cue vs. Neutral-cue.** Next, we compared memory- and neutral-cue clips. The ERP analyses yielded two significant clusters. The first significant cluster was located over frontal regions, with memory-cue clips eliciting greater sustained negativity than neutral-cue clips (Table 4 and Fig. 10). The second cluster followed the first cluster in time and occurred over right parietal areas of the scalp, with greater sustained negativity for neutral-cue clips than for memory-cue clips due to a reversal in polarity from anterior to posterior sites (Table 4 and Fig. 10).

### 5. Discussion

The current study further investigated the role of attention at encoding on memory-guided attention at retrieval. We tested whether mere exposure can facilitate memory-guided attention to real-world sounds. The results show that when conditions at learning are conducive to the incidental encoding of associations between sound-clip (context) and tone (target), explicit memory for the clip is good, but memory for the tone’s presence and location is at chance. Target localization was also faster for old clips with or without an embedded

![Fig. 7. Left panel: Group mean ERPs for the cue audio-clip: Old vs. New. The grey boxes depict significant time windows. Dark grey represents overlapping time windows. The time window for Cluster 1 and 2 were 900–1000 ms and 1300–1400 ms, respectively. P1 = left parietal electrode; FC4 = right fronto-central electrode.](image)
target than for new clips, suggesting that memory facilitates processing of the tone by speeding responses to it. These behavioural effects support previous research that report processing efficiency due to decreased RTs for old relative to new stimuli. These behavioural effects support previous research and as with explicit behavioural tests, neuro-electric activity revealed an old-new effect at midline and frontal sites.

RTs did not vary as a function of the associations between target presence and audio-clip. Importantly, neuroelectric activity yielded a significant difference between clips that had been associated with the target compared to those that were not associated with it, yet ERPs were not lateralized. The difference in ERPs between memory-cue and neutral-cue clips may reflect the formation of a neural trace/engram of the association. Research has shown that silent engrams are memory traces in the brain that do not control behaviour and are a product of normal mnemonic processing (for a review see Joselyn and Tonegawa, 2020). In other words, the memory of the association exists orthogonal to the recovering of the memory (Moscovitch, 2007). The results suggest that a trace of the associations, not specific to the tone’s location per se, but rather to the presence or absence of the tone (i.e. memory-cue associated with ‘target present’ and neutral-cue associated with ‘target absent’) was formed. The EEG results pertaining to the difference between memory-cue and neutral-cue clips may index the formation of this memory trace, while the behavioural results pertaining to the lack of difference between memory-cue and neutral-cue clips suggest that this learned association was not expressed at the behavioural level. Together, these results support the latent learning and attention-dependent learning hypothesis that learning of the association itself does not require directed attention (Chun and Jiang, 2005), but expression of learning in the form of behavioural enhancement does require attention (Treisman and Gelade, 1980), respectively. Specifically, in relation to memory-guided attention, when the association between context and target is truly incidentally encoded, we do not find evidence for implicit processes playing a role in guiding auditory attention at the level that influences behaviour.

In the current study, our memory to attention paradigm can also be construed as auditory priming for non-verbal information. Church and Schacter (1994) found that, for words, if acoustic features, such as speaker voice, intonation, and fundamental frequency are altered, performance is worsened, or priming effects are completely eliminated. In our paradigm, the tone, when localized at test, can be thought of as a measure of priming of the sound-tone association. Based on Church and Schacter (1994), one might have expected that detection (priming) would be better in the memory condition, with the tone lateralized on the same side at study and test, rather than in the neutral-cue condition because the acoustic features would remain unchanged. This outcome was not evident in behaviour, suggesting either that non-verbal stimuli, the embedded tone, the task itself (localization) or all three are not sufficiently powerful conditions to lead to behavioural priming. As no-one, to our knowledge, has looked at nonverbal auditory priming, it is interesting to consider these findings in light of Church and Schacter (1994), and determine, in future experiments, whether they provide some boundary conditions for obtaining auditory priming. It is important to note, however, that tone-presence vs tone-absence does lead to neural priming, indicating that the stimuli can leave a neural trace that is not realized in behaviour.

5.1. Old-new effect

The fact that old clips 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 and Jiang, 1998) and tactile (Assumpção et al., 2015; Nabeta et al., 2003) search tasks. In our paradigm, endogenous control, which was comprised of habits developed over time as a result of repeated exposure, may have facilitated use of the context for subsequent target localization. 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 Vo and Wolfe, 2015; Thompson and Sabik, 2018; Trick et al., 2004).

The behavioural old-new effect is supported by the ERP analyses, which showed a number of significant modulations at cue audio-clip and probe audio-clip. The old-new 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

Table 3

<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>$p = .0007$</td>
<td>AF6, F8, FC5, C6, T8, TP8, CP6, CP4, P8, P10, PO8, Cz2, TP10, FT10, F10, LO2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1311–1645</td>
<td>$p = .04$</td>
<td>F1, F3, FC1, C1, CP3, CP1, Fz, FC2, Cz</td>
</tr>
</tbody>
</table>
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 and Rugg, 1998; Friedman and Johnson, 2000; Rugg and Curran, 2007; Vilberg et al., 2006; Vilberg; Rugg, 2007; Wilding, 2000). In addition, the later sustained activity may index maintenance of the audio-clip in working memory, as participants were not instructed to make a response during the cue audio-clip.

5.2. Memory-guided attention: memory-cue clips vs. neutral-cue clips

The behavioural results revealed that participants’ memory for the clip was significantly above chance, but their memory for the sound-tone association was not significantly different from chance. There was no significant difference in RT to lateralized targets when comparing old clips that contained a memory-cue to those that did not (neutral-cue). Therefore, memory for the clip facilitated processing of the target; this effect, however, was not specific to target location or presence. This suggests that when attention is drawn only to the clip (background), incidental encoding of the sound-tone association is not strong enough to yield a congruency benefit in RT for localizing the target at test.

The absence of a behavioural effect may be due to conditions during exposure that may have affected encoding of the sound-tone association. 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, and that only objects that are grouped in the foreground are learned over repeated exposure. Information that is perceptually grouped and foregrounded is thought to be prioritized for attention (Mazza et al., 2005; Nakayama et al., 1989) and contextual learning (Jiang and Leung, 2005; Pollmann and Merginelli, 2009). Jiang and Leung (2005) concluded that learning of context occurs independently of attention and that the effect of memory-guided attention is influenced by where attention is directed at exposure. In the current study, attention was guided towards the clip at exposure, rather than the target. At test, however, 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 clip (foreground) and poor memory for the target’s location at exposure (background). Therefore, the incidental learning of the associations between clip and target and its behavioural manifestation may be particularly sensitive to the degree to which attention is manipulated at test and learning may consequently strongly depend on which auditory object is attended at encoding.

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 and on the conditions at learning that affect the encoding of the sound-tone association. When attention is drawn to the clip and the association between clip (context) and tone (target) is formed incidentally, we do not see a behavioural contextual-cueing benefit. This finding highlights 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, 2020).

By contrast, we did find a significant difference in ERP amplitudes comparing the memory- and neutral-cue conditions, suggesting that participants differentially processed these two conditions. It was expected that a difference in the sustained potential between cue

Table 4

<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-cue</td>
<td>1</td>
<td>21–201</td>
<td>$p = .01$</td>
<td>FP1, AF7, P7, 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>$p = .04$</td>
<td>P1, P3, O1, Iz, Oz, POx, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2, C2</td>
</tr>
</tbody>
</table>

Fig. 9. Left panel: Group mean ERPs for the probe audio-clip: Old vs. New. The grey shaded box denotes significant differences between old and new clips. Right panel: The iso-contour maps show the group mean amplitude for cluster 1 and 2, with 1900–2000 ms and 1500–1600 ms intervals, respectively. F1 = left frontal electrode; C6 = right central electrode.
ERP amplitude for neutral-cue clips than for memory-cue ones may afford the maintenance of a single representation. Furthermore, greater memory for neutral-cue clips compared to memory-cue clips may support the notion that there is neural evidence of a memory trace, indexed by the ERP differences between memory-cue and neutral-cue conditions. The fact that there are differences in the neural signature between these conditions suggests that there is residue of prior experience. 

The lack of ERP lateralization suggests that while there may be an indication of retention of an association (a neural trace/engram), the tone might have blended in with the clip as a Gestalt, leading to the lack of independent encoding of spatial information. Memory-cue clips may contain predictive cues that neutral-cue clips do not, but only in the unitized case (clip + tone). The memory effect distinguishing between memory-cue and neutral-cue clips may reflect anticipation of an item that is not specific to its location. At test retrieval, there may be a pattern completion where the cue either activates the entire Gestalt (clip + tone) or that activates the tone alone. The current study cannot distinguish between these two possibilities. However, the possibility of encoding an entire Gestalt representation of “clip + tone” may explain why the information was not used for guiding attention, as measured with behavioural memory-cue vs. neutral-cue contrast.

The discrepancy between a behavioural effect and a brain effect brings to light the intricate relationship between attention and learning in relation to memory. Our findings are consistent with previous work that has shown that learning the association, in and of itself, as demonstrated by the neural difference we find between memory- and neutral-cue conditions, does not require directed attention (Chun and Jiang, 2005; Tolman, 1948); expression of this learning, in terms of behavioural performance enhancement, however, does require attention (Treisman and Gelade, 1980).

We show that depending on the task, attention at encoding may be differentially manipulated and that this can lead to very different behavioural outcomes. This finding highlights the importance of task choice. We have provided evidence for the notion that the object to which one attends and the degree to which the associations are made explicit at encoding strongly affect the nature of the association at retrieval and how this association may be used for future processing. When task conditions are controlled, such that the context-to-target association is learned incidentally, and contamination from explicit memory processes minimized, we find that (i) memory-cue and neutral-cue conditions are neurally distinguishable and form a neural memory, but only in the unitized case (clip + tone). The memory effect distinguishing between memory-cue and neutral-cue clips may reflect anticipation of an item that is not specific to its location. At test retrieval, there may be a pattern completion where the cue either activates the entire Gestalt (clip + tone) or that activates the tone alone. The current study cannot distinguish between these two possibilities. However, the possibility of encoding an entire Gestalt representation of “clip + tone” may explain why the information was not used for guiding attention, as measured with behavioural memory-cue vs. neutral-cue contrast.

The discrepancy between a behavioural effect and a brain effect brings to light the intricate relationship between attention and learning in relation to memory. Our findings are consistent with previous work that has shown that learning the association, in and of itself, as demonstrated by the neural difference we find between memory- and neutral-cue conditions, does not require directed attention (Chun and Jiang, 2005; Tolman, 1948); expression of this learning, in terms of behavioural performance enhancement, however, does require attention (Treisman and Gelade, 1980).

We show that depending on the task, attention at encoding may be differentially manipulated and that this can lead to very different behavioural outcomes. This finding highlights the importance of task choice. We have provided evidence for the notion that the object to which one attends and the degree to which the associations are made explicit at encoding strongly affect the nature of the association at retrieval and how this association may be used for future processing. When task conditions are controlled, such that the context-to-target association is learned incidentally, and contamination from explicit memory processes minimized, we find that (i) memory-cue and neutral-cue conditions are neurally distinguishable and form a neural memory, but that (ii) these traces require directed attention at encoding in order to retrieve them and form an implicit or explicit memory that guides auditory attention at the level that influences behaviour.

6. Concluding remarks

In sum, the current study contributes to our understanding of the nature of implicit and explicit memory processes that guide attention. In the context of memory-guided attention literature, it has been shown that when a direct association is made at encoding (explicit) between context and target (Patai et al., 2012; Zimmermann et al., 2017), one observes a behavioural effect. The results from the current study add to these findings and help to elucidate the role that attention at encoding performs.
plays in the effect of memory-guided attention. We demonstrate that in the absence of the formation of an explicit association between target and clip at encoding (incidental encoding), no effect is observed behaviourally, though we do see evidence for a neural trace effect reflected in the neural signature. Therefore, learning conditions that lead to explicit knowledge of the context and implicit knowledge of the sound-tone association are not sufficient to guide attention at the behavioural level. We acknowledge, however, that some other tests of attention may be more effective in eliciting behavioural evidence associated with the priming we found. As well, some behaviour may have been influenced by the formed associations but may not have been detected by our measures.

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 (Moscovitch, 1992; Moscovitch et al., 2016; Schacter et al., 2004) that can benefit from, and contribute to, attentional processes. This study, also, contributes to the field of auditory perception, by further defining the internal mechanism underlying schema-based scene analysis. 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.

CRediT authorship contribution statement

Manda Fischer: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Project administration. Morris Moscovitch: Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. Claude Alain: Conceptualization, Methodology, Resources, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

None.

Acknowledgements

We thank Vanessa Chan, Karishma Ramdeo, and Shahier Paracha for their assistance. We would also like to thank the editor and anonymous reviewers for their very thoughtful and helpful input.

References


