Implicit and explicit effects of context on episodic auditory-verbal memory:

A hybrid repetition-learning recognition paradigm

by

Jessica S. Arsenault

A thesis submitted in conformity with the requirements for the degree of

Master of Arts

Graduate Department of Psychology

University of Toronto

© Copyright by Jessica S. Arsenault (2012)
Abstract
The objective of this research was to investigate the extent to which context contributes to the learning and recognition of episodic auditory-verbal memories (EAM). By combining the Hebb repetition paradigm (HRP) and continuous recognition paradigm (CRP), I capitalized on the advantages of both while manipulating the context in which EAM were retrieved. Through repetition, participants learned sequences of pseudowords in which word order and speaker were varied. A recognition test of either a pseudoword (Experiment I) or the speaker of a pseudoword (Experiment II) revealed temporal and sensory context effects. Results showed that the encoding manipulation did not impact short-term memory but did have an effect on long-term learning. This research helped to clarify the role of context in EAM in both short- and long-term memory, as well as added to the current literature of HRP and CRP. Future directions are discussed.
# Table of Contents

Abstract.............................................................................................................. ii

Table of Contents............................................................................................... iii

List of Tables........................................................................................................ iv

List of Figures....................................................................................................... v

List of Appendices............................................................................................. vi

Chapter 1: Background Literature................................................................. 1
  1.1 Introduction................................................................................................. 1
  1.2 Literature Review...................................................................................... 2
  1.3 The Present Experiments.......................................................................... 8

Chapter 2: Experiment I................................................................................ 13
  2.1 Methods..................................................................................................... 13
  2.2 Analysis.................................................................................................... 17
  2.3 Results..................................................................................................... 17
  2.4 Discussion................................................................................................. 19

Chapter 3: Experiment II.............................................................................. 21
  3.1 Methods..................................................................................................... 21
  3.2 Analysis.................................................................................................... 22
  3.3 Results..................................................................................................... 22
  3.4 Discussion................................................................................................. 25

Chapter 4: General Discussion.................................................................... 29
  4.1 Future Directions...................................................................................... 39
  4.2 Conclusion................................................................................................. 40

References......................................................................................................... 41

Tables................................................................................................................. 43

Figures............................................................................................................... 51

Appendices......................................................................................................... 59
List of Tables

Table 1. Descriptions and examples of each condition as represented at the first and second trial…………………………………………………………………………………………………………………………..44

Table 2. A 6 x 5 (Block x Condition) within-subjects ANOVA on accuracy scores for Experiment I……………………………………………………………………………………………………………………………45

Table 3. A 6 x 5 (Block x Condition) within-subjects ANOVA on RT scores for Experiment I……………………………………………………………………………………………………………………………………………46

Table 4. Independent samples t-test (2-tailed) of confidence ratings of participants in Experiment I who were and were not explicitly aware of repeating sequences………………47

Table 5. Repeated measures ANOVA of accuracy scores for Experiment II………………..48

Table 6. Repeated measures ANOVA of reaction times for Experiment II…………………..49

Table 7. Independent samples t-test (2-tailed) of confidence ratings of participants in Experiment II who were and were not explicitly aware of repeating sequences………………..50
List of Figures

Figure 1. Graphical depiction of design of Experiment I………………………………52

Figure 2. Main effect of block averaged across conditions for all participants in Experiment I…………………………………………………………………………………53

Figure 3. Line graphs of Block x Condition interaction effects from Experiment I……54

Figure 4. Serial position effects for Experiment I………………………………………………………………………………………………………………………………..55

Figure 5. Probability of responses in SIDV trials according to gender congruency at study and test………………………………………………………………………………………………56

Figure 6. Mean reaction times by block number for Experiment II averaged across participants and conditions………………………………………………………………………………………………57

Figure 7. Serial position effects for Experiment II…………………………………………………………………………………………………………………………………………58
List of Appendices

Appendix 1. List of pseudowords used in Experiments I and II..........................60
Chapter 1
Background Literature

1.1 Introduction

Although memory exists within and across multiple modalities, including visual, auditory, sensory, and olfactory, a substantial proportion of one’s episodic memories involve speech and language, which is strongly tied to the auditory modality. Episodic auditory memories (EAM) refer to memories that individuals can re-experience and relive in their mind’s ear. For example, one can recall the sound of his or her child’s first word or replay a telephone conversation from last week. Many previous studies have investigated the impact of single-trial exposure on auditory memories for items (Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1993), however, the cumulative effect of a series of learning trials on the formation of episodic memories has rarely been studied. While episodic memory has typically been thought of as relating to a single episode, perceptually rich memories can nevertheless develop over a series of experiences (Greenberg & Verfaellie, 2010). The episodic context in which a memory is encoded and retrieved is presumed to play a significant role in the strength of the memory trace (Tulving & Thomson, 1973), yet the systematic manipulation of voices has received relatively little attention. The present study was designed to address some of the gaps that still exist in the episodic auditory-verbal memory literature. Speaker and word order of aurally presented pseudowords were manipulated across a series of trials within a recognition memory paradigm in order to examine the effects of voice-context on repetition learning.

1.2 Literature Review
The Hebb repetition effect (Hebb, 1961) was originally observed in tests of immediate serial recall (ISR) and occurs when the repetition of certain “target” lists improves recall performance, even when the repetition is unknown to the subject. In Hebb’s original study, participants were presented with sequences of nine digits and were asked to recite the lists, in order, immediately following each presentation. Every third sequence that appeared was an identical repetition and by the eighth presentation recall performance for the repeated list had significantly improved. This repetition effect has been observed for lists of digits, words, and non-words (Cumming, Page, & Norris, 2003; Parmentier, Maybery, Huitson, & Jones, 2008); in children and adults (Page & Norris, 2009); across repetition gaps of twelve intervening lists (Page & Norris, 2009); and in the visual, auditory, and spatial domains (Page & Cumming, 2006; Couture & Tremblay, 2006). Overt recall in the Hebb repetition paradigm (HRP) is necessary for learning to occur. For example, Cohen and Johansson (1967) found that when participants were not asked to actively recall a list following each presentation, but were merely passive listeners, no repetition effect was observed. When the authors directed participants to note digits or positions that felt familiar (i.e., were recognized), performance was enhanced compared to controls, however, further analysis revealed that this was driven by participants basing responses on only the first or first and last digits of each list. Therefore, the role of recognition memory in Hebb repetition learning remains unclear.

In contrast to repetition learning, which typically focuses on recall performance, many recognition memory studies use single-trial exposure. For instance, in the continuous recognition paradigm (CRP), a stream of stimuli is presented and participants are asked to judge whether each item is old or new immediately following its presentation. Much of the research on the
impact of modality on auditory verbal memory has used this type of design (e.g., Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993; Goldinger, 1996) and has converged on some general conclusions. Early work on the effect of voice established that a word spoken by the same voice will be better recognized than the same word that was later presented by a different voice (Craik & Kirsner, 1974). A further observation was that indexical properties of speech, such as pitch, can persist in long-term memory (LTM). This was in direct contrast to the predominant belief at the time, that sensory aspects of speech stimuli could only be held in memory for a few seconds (Murdock, 1971). Instead, Craik and Kirsner found that participants were better able to recognize a word when it was repeated in the same voice at a lag of up to 32 items (the largest lag tested), which corresponded to a delay of approximately 2 minutes. Further, the authors noted that participants were able to identify the original voice in which a word was presented across the same interval, although recall was best before lag 4. This led the authors to conclude that specific verbal attributes of a speaker’s voice could be stored in LTM.

The above findings, however, were based on an experiment that only used two speakers – one male and one female. It could be argued, then, that simple gender representations, and not specific voice attributes, were persisting in LTM. To address this, Palmeri, Goldinger, and Pisoni (1993) conducted a CRP using 20 speakers – ten male and ten female. In line with the Craik and Kirsner (1974) study, the authors found that item (word) recognition was enhanced for same-voice compared to different-voice repetitions. This result was obtained without any explicit instruction to attend to voice, suggesting implicit learning was responsible for this effect. When participants were specifically asked to attend to voices in a follow-up experiment, however, it was found that even at lag 0 (no intervening items) participants relied on gender to classify
speakers, often judging same-gender/different-voice words as “same” when asked to recognize the speaker. Essentially, when participants were asked to explicitly attend to voices, recognition performance decreased compared to implicit judgments of just the item. By contrasting the results of the two experiments, Palmeri and colleagues thus concluded that voice attributes are encoded and stored in LTM relatively automatically. In a review of studies examining the role of LTM in speech perception, Pisoni (1993) concluded that words and voices are processed as integral dimensions and that the voice dimension was hardest to ignore, suggesting again that perceptual attributes of voice are encoded implicitly.

More recent research on voice-specific effects has further addressed the issue of voice similarity. Although participants tend to rely on gender to classify voices, it appears that a strong same-voice repetition effect still exists, especially in implicit memory tests. Goldinger (1996) found that hit rates decreased as within-gender voice distinctiveness increased during a recognition memory task with either 6 or 10 speakers (half male and half female). However, Goh (2005) found the opposite result in his recognition memory test, which included 10 speakers, all of whom were male. Goh tested voice distinctiveness between subjects by creating perceived similarity voice maps and then assigning participants to either the distinctive-voices or similar-voices condition. It was reported that overall voice discriminability was significantly higher in the distinctive condition compared to the voice-similarity group. It was concluded that the voice similarity effects observed in Goldinger (1996) were the result of response bias, as recorded by hit rate. Thus, in cases of the same word being presented by a different speaker, when voice similarity increases participants are more likely to respond with a false alarm, but when
similarity decreases participants are more likely to correctly identify that the word was the same and spoken by a different voice.

While both the HRP and the CRP provide insight into the effects of context on episodic auditory-verbal memory encoding and retrieval, many questions about how EAM are accessed still exist. According to a recent dissertation, Campeanu (2008) concluded that speech is encoded, and subsequently retrieved, as an entity rather than in acoustical parts. By varying two qualitative parameters of speech - gender and accent - Campeanu tested both item and speaker recognition using a block design. It was found that for both item and voice recognition, the same speaker condition yielded the highest performance, while the other three conditions – same gender/different accent, different gender/same accent, and different gender/different accent – did not differ significantly from each other. It was concluded that reinstating only one voice parameter (either gender or accent) at test did not significantly aid in either item or speaker recognition. Therefore, it was hypothesized that voice is encoded as a pattern rather than as a sum of its acoustical elements, and that the reinstatement of only one parameter was not sufficient to reach the threshold necessary for recognition.

These results must be approached with caution, however, as only four speakers were used in the Campeanu study – male/native English, female/native English, male/Chinese accent, and female/Chinese accent. This small number of speakers introduced a confound between speaker identity and voice characteristics. In order to investigate how source manipulation impacts EAM retrieval, an important control would include the reinstatement of same gender/same accent but with a different speaker. Otherwise, it is impossible to determine whether or not the advantage of
the same speaker condition is due to the benefit of having reinstated both parameters or simply due to the reinstatement of the *exact same speaker* (Campeanu, 2008).

This issue can be partially addressed by the findings of Goh (2005), which are summarized above. Using 10 male speakers in an incidental-learning paradigm with a surprise recognition test, Goh highlighted the importance of testing for both instance-specific matching (an exact match between the memory probe and the memory trace) as well as voice-specific familiarity (contributions from prior exposure to the attributes of a talker’s voice) within the same study. Therefore, he used both studied and unstudied talkers at test in order to examine the effects of old and new words presented by same, different (studied), or new (unstudied) talkers at test. In addition to replicating various other findings, such as an advantage for perceptual attributes of spoken word stored in LTM, Goh also found evidence of both instance-specific and voice-specific familiarity, but only on hit rate in the latter condition. In particular, instance-specific effects were evidenced by enhanced word recognition when the same voice at study was represented at test. While this is a replication of previous findings, it extends the effect to at least 5 different studied voices of the same gender. Accuracy decreased when old words were mismatched with different but previously studied voices as well as with unstudied voices compared to the same voice; however, there were no statistically significant differences between the accuracy scores for the different/studied and unstudied conditions. As mentioned earlier, this finding implies that familiarity may influence a response bias (as indicated by the increased hit rate among different/studied and unstudied contexts), but does not aid in word recognition. This would imply that instance-specific effects play a larger role in word recognition, indicating that the findings in Campeanu (2008) may have been due to the reinstatement of the same speaker.
rather than the reinstatement of qualitative voice parameters. It has, however, also been shown that male voices tend to be easier to identify than female voices (Nygaard & Pisoni, 1998), making the above findings of Campeanu (2008) and Goh (2005) even more difficult to interpret.

Some indirect support for the idea that memories are retrieved as aggregate units comes from attention and perception studies. For example, Mullennix and Pisoni (1990) asked participants to attend to only one stimulus dimension (i.e., voice) while ignoring the other (i.e., phoneme) during a speeded classification task. Substantial interference effects were observed, leading the authors to conclude that words and voices are processed interdependently. While this does not explicitly apply to episodic memory, it nonetheless suggests that if linguistic and indexical properties of speech are processed simultaneously, then subsequent retrieval may be similarly impacted. Overall, however, support for the theory that episodic auditory-verbal memories are accessed as all-or-none entities rather than as independent acoustical features is still lacking.

Nonetheless, much of our EAM depend on sequential context – that is, the sensory features that occur immediately before and after an item being attended to. Music and language, for example, both require a temporal organization in order to make sense. The information that is received both helps to interpret what has just been experienced as well as helps to predict what will come next. While the HRP asks participants to recognize patterns of sequences, it has not been used to test for the effect of voice-specific attributes on long-term learning. Conversely, the CRP emphasizes recognition of repetitions within or across modality, but neglects the impact of sequential information. Neither approach can address the impact of the sensory details that occur in the immediate temporal surroundings of an item or a list of items to be learned. According to
the encoding specificity principle (Tulving & Thompson, 1973), episodic memories rely on the specific context of the specific situation in which they were encoded. This would suggest that our memories are encoded as rich holistic episodes rather than as a combination of associated elements. Current approaches to studying EAM do not account for the impact of both temporal and sensory context effects; thus, a new paradigm was developed, in which sequential episodic context was manipulated across a series of learning trials. As such, the graded impact of either partial or total preservation of context could be assessed.

1.3 The Present Experiments

The purpose of the current research was to examine the effects of context on the learning of EAM for auditory-verbal sequences. The temporal and sensory context of each memory trace was systematically manipulated while accuracy and reaction time (RT) were recorded. In the first Experiment, only item (word) recognition was tested. In the second Experiment, both item and speaker recognition was tested explicitly.

The current research was designed to address many of the inconsistencies noted in previous studies. The design reflects a combination of the HRP and the CRP, with the overall objective being to examine repeating temporal and sensory context effects on subsequent recognition and learning. Similar to a HRP, participants were exposed to sequences of pseudowords delivered by six different speakers; similar to a CRP, participants were asked to make judgments on individual pseudowords immediately following the presentation of each sequence. The purpose of using pseudowords was to encourage episodic memory formation – as the pseudowords were
meaningless, it was more difficult for participants to adopt semantic strategies in order to remember the stimuli.

The purpose of the hybrid study design was to gain access to changes in learned sequences as episodic memories formed. In Buchsbaum, Padmanabhan, and Berman (2010), the authors highlight the importance of looking at short-term memory (STM) and LTM jointly rather than in isolation as is typically done. The proposed research allowed for the investigation of the effects of context on both STM (as evidenced by a recognition judgment following each sequence) and LTM (represented by changes in recognition memory over repeated sequences). Typically, the CRP focuses on recognition memory via single-trial exposure but does not assess the cumulative effect over a series of trials. Conversely, HRPs investigate learning effects with immediate serial recall rather than recognition. By combining the two paradigms, the goal was to determine whether recognition memory for individual items benefits from the reinstatement of the temporal and sensory context of a repeating sequence.

In the current study, speaker and word order were manipulated. Participants were asked to listen to sequences of six pseudowords that were spoken by six different speakers – three males and three females – each of whom presented one pseudoword per sequence. In some conditions, the sequences maintained a consistent word order and in others they did not; speaker was similarly manipulated. There were five different sequence conditions which refer to the manner in which serial word order and voice order were manipulated across repetitions, namely: same order/same speaker (SOSS), same order/different speaker (SODS), different order/same speaker (DOSS), different order/different speaker (DODS), and non-repeating sets (NRS; filler sequences). Table
1 displays examples of how order and voice were varied across trials and conditions (note that different sequences contained different pseudowords). In the SOSS condition, the six pseudowords within each sequence appeared in the same word order and were presented by the same speaker every time that the set was presented. In the SODS condition, the pseudowords appeared in the same order, however, the speaker of each pseudoword was shuffled across trials. In the DOSS condition, the order in which the pseudowords were presented was rearranged, but the speaker order of the set was consistent across trials. In the DODS condition, both word order and speaker was shuffled across trials, yet the sets consisted of the same six pseudowords. Finally, the NRS can be thought of as a control condition in which the order and speaker, as well as the pseudowords themselves, were presented randomly. This was the only condition in which the same six pseudowords only appeared together once, and therefore did not repeat.

If individual pseudowords within each sequence were better remembered over time when they were presented in a consistent context, this would lend support to the notion that EAM for word sequences are retrieved as episodic wholes (i.e., sequences). If, however, the same stimuli presented an equal number of times but within a different context (e.g., scrambled order/speaker) were remembered equally well, then this would suggest that EAM may be accessed as individual units (i.e., pseudowords). The main condition of interest, then, was the SOSS condition in which all of the contextual details are maintained across repetitions. The other repeating sequences served as repetition controls and also helped to assess whether there was a graded role of context on auditory memories. For example, the SODS condition served as a control for word order while DOSS served as a control for the prosody of the sequence. Finally, the DODS condition helped to assess the role of overall voice familiarity and item repetition.
Based on Pisoni’s (1993) parallel episodic memory system, which states that “perception of one dimension (i.e., phoneme) affects classification of the other dimension (i.e., voice) and vice versa, and subjects cannot selectively ignore irrelevant variation on the non-attended dimension” (p. 112), enhanced performance was expected in both experiments for the SOSS condition in which the episodic context for the set was strongest due to the consistent pairing of order and speaker. The highest accuracy and lowest RT data was expected for this condition. Based on HRP studies, improved performance was anticipated around the eighth repetition of each list (Page & Norris, 2009), however, the novelty of multiple lists and use of recognition in repetition learning required tentative predictions. Aside from the performance expected in the SOSS condition, a slight advantage was predicted for the consistent serial word order of the SODS condition compared to DOSS and DODS, which were expected to be somewhat enhanced or equal to the NRS condition.

For the simple item judgments required in the first experiment, some effects of implicit learning on accuracy were expected. As discussed, previous research on the effect of speaker variability on word recognition suggests that recognition accuracy is enhanced when the word is presented by the same voice at study and at test. An advantage in RT was also anticipated for same-voice compared to different-voice repetitions, such that RT would be shortest for old words that were presented by the same voice at study and at test. Due to the consistent pairing of words, order, and speakers, it was suggested that the episodic memory trace for the SOSS condition would be strongest and that this would aid in subsequent recognition judgments for individual items.
In general, an overall increase in RT and decrease in accuracy was predicted for Experiment II as participants were required to make an additional judgment regarding the speaker as well as the simple item judgment. Despite this, the general pattern of results outlined above was anticipated, specifically with enhanced learning in the SOSS condition. Previous research with numerous speakers has also suggested that participants have difficulty explicitly discriminating between two different same-gendered voices speaking the same word. As discussed above, participants have a tendency to judge different-voice/same-gender repetitions as ‘same’ rather than as ‘different’. Therefore, a higher proportion of false alarms were expected in the ‘same’ category when the speakers at study and at test were of the same gender.

Across both experiments, a modest serial position effect was anticipated. Specifically, recency effects often emerge during recognition paradigms (Ward, Avons, & Melling, 2005), such that a slight advantage exists for the items at the end of a list compared to those in the beginning or middle of the set. Therefore, an increase in recognition accuracy was expected for pseudowords that were presented as the fifth or sixth pseudoword in each set, especially in Experiment I when only item was being attended to.
2.1 Methods

In this experiment, a set of auditory stimuli was manipulated in order to investigate the impact of varying the strength of context on subsequent pseudoword recognition. During each trial, participants heard a set of six pseudowords and, following a short inter-stimulus-interval (ISI), heard a single test pseudoword and were asked to identify whether or not that pseudoword was present in the preceding set. There were six blocks of 40 trials, for a total of 240 trials.

Participants:
Participants consisted of 18 healthy young adults (average age 23 years; 6 males) recruited from both the Baycrest volunteer subject database and the PSY100 pool at the University of Toronto, as well as advertisements on the UofT website. All subjects were primary English speakers with no history of neurological/psychiatric illness or hearing disorders.

Stimulus Materials:
The stimulus materials in this study consisted of 89 auditory pseudowords (see Appendix 1), each spoken by six speakers. There were three male speakers and three female speakers, one each with a generally high-, medium-, and low-pitched voice. All speakers were between the ages of 21-25 and were primary English speakers. The pseudowords were monosyllabic consonant-vowel-consonant combinations (e.g. “beb” or “hup”). Each pseudoword was recorded in isolation in a soundproof room at the Rotman Research Institute. The task was programmed and delivered in Eprime 2.0, with a new randomization of the experiment being generated for
each participant (i.e., the repeating lists differed across participants). In each study, four repeating lists (one for each of the four repeating conditions) of six different pseudowords were produced, as well as multiple non-repeating sets from the overall list of 24 pseudowords for each participant.

*Design:*

Pseudowords were aurally presented in sets of six, which varied both by order and by speaker across sets. The five different conditions in which a sequence could appear (SOSS: Same Order/Same Speaker; SODS: Same Order/Different Speaker; DOSS: Different Order/Same Speaker; DODS: Different Order/Different Speaker; NRS: Non-Repeating Sets; see Table 1) occurred in a pseudo-random order such that each condition appeared 8 times per block. A single set of six items was presented repeatedly in each of the first four conditions. Each pseudoword appeared in only one repeating set and in multiple non-repeating sets. Each repeating sequence was repeated a total of 48 times throughout the course of the experiment.

*Procedure:*

Participants were tested individually at either the Rotman Research Institute at Baycrest Hospital or at U of T’s St. George campus, with the same equipment being used at both sites. Informed consent was acquired and instructions were given. Participants were asked to attend to each word in the set and were informed that there would be a brief recognition test following each set. The experiment started with a short practice session to familiarize participants with the task. The practice phase consisted of 16 trials of non-repeating sets, and contained different pseudowords than those used during the experiment. The set of speakers used in the practice phase were
identical to those used during the experiment; therefore participants had the opportunity to familiarize themselves with the speakers before the task began.

Participants were directed to sit at a desk facing a computer screen with headphones on. They were asked to listen to sets of pseudowords, followed by a probe pseudoword, and then to make a yes/no judgment as to whether or not the probed pseudoword appeared in the just-presented set. Participants were asked to respond as quickly and accurately as possible. Responses were recorded using the computer keyboard controlled by participants’ dominant hand, with the “b” key recording “yes” responses and the “n” key recording “no” responses. The computer screen displayed a speaker icon during the listening phase and displayed “yes/no?” with the corresponding key prompts during the response phase. Once a response was recorded, feedback was given (a green “Correct!” or red “Incorrect”), followed by a brief crosshair before the next study phase.

The presentation of each pseudoword persisted for about 700ms with a gap of 100ms between each pseudoword, followed by a gap of 3s and a subsequent auditory probe (target word), which lasted approximately 700ms. Participants were immediately prompted to respond as “yes/no?” appeared on the computer screen. Their response was not time-limited and was followed by 1s of feedback. The total length of each trial was around 10s, with 40 trials per block. Participants were offered a short break between each of the six blocks, with the entire task lasting about 45 minutes. See Figure 1 for a graphical depiction of the task design. In half of the trials the probe pseudoword was a repetition (requiring a “yes” response), and in the other half of the trials the probe was a pseudoword that was not in the preceding sequence. For the trials in which the probe
matched an item in the sequence, half of these probes were delivered by the same voice, and half were delivered by a different voice. Unfortunately, after data collection was completed for the first Experiment, an error in the design files was discovered, revealing that proper counterbalancing and labeling of old pseudowords being repeated by the same speaker at study and test did not occur. Therefore, word recognition was confounded by inconsistent speaker reinstatement across trials and conditions. The impact of this flaw is discussed below.

Upon completion of the task, each participant was asked to perform a final follow-up task in order to assess explicit awareness of the repetition manipulation. First they were verbally asked if they noticed anything specific about the paradigm; if no repetition was mentioned, they were asked if they noticed anything about the words and speakers. Finally, if the repeating sets were not explicitly noted, participants were asked if they thought any of the sequences were repeating. If at any point participants suggested that they may have been aware of the repeating sets, they were asked how many sequences they thought repeated and during which block they noticed the repetition.

Participants were then told that some sets had indeed been repeating and were asked to complete a final computerized task, rating six different sequences from their experiment. The sequences consisted of the SOSS set, the SODS set, and 4 randomly selected NRS sets, all of which were presented in random order. The participants were told that some of the sequences in the follow-up task were heard multiple times during the experiment, and others were heard only once. They were asked to rate how confident they were that each set repeated during the experiment by using a 1-5 rating.
2.2 Analysis

A glitch in the design files for the first Experiment was discovered after data collection was complete which rendered probe type confounded; therefore, no probe-specific analyses were performed. Instead, learning effects as well as serial position effects were investigated. To normalize the data, all RTs above 3000ms were excluded. Following this, the average and standard deviation was calculated separately for each participant. All trials 2.5 standard deviations higher than each participant’s mean RT were then excluded.

Follow-up data was coded as (1) spontaneous detection of repeating sequences (i.e., “did you notice anything in particular about the task?” / “did you notice anything about the words and the speakers?”), (2) probed recognition of repeating sequences (i.e., “do you think that any of the sequences were repeating?”), or (3) no recognition of repeating sequences. Explicit recognition was compared to the confidence ratings collected when participants were presented with non-repeating and repeating sequences.

2.3 Results

Although responses were confounded by probe type, 6 x 5 (Block x Condition) repeated measures ANOVAs were performed separately for both accuracy (percent correct) and RT data to search for possible main effects of block or condition. Table 2 displays the results for accuracy scores; while block number was significant, $F(5, 85) = 3.187, p = 0.011$, it did not follow a linear trend (Figure 2); rather, an order 5 trend was the most significant, $F(1, 17) = 8.320, p = 0.010$. There were no significant effects of condition, $F(4, 68) = 0.084, p = 0.987$, or of the interaction, $F(20, 340) = 0.551, p = 0.943$. For RT data, there was again a significant effect of block, $F(5, 85)$
= 5.105, \( p < 0.0001 \), but no main effect of condition, \( F(4, 68) = 0.709, p = 0.588 \); see Table 3. The block effect followed a linear trend, \( F(1, 17) = 7.066, p = 0.017 \), with RTs generally getting smaller as block number increased. Figure 3 shows the interaction of block and condition, which was not significant, \( F(20, 340) = 0.606, p = 0.909 \).

There was a significant effect of the serial position of the test word on both accuracy, \( F(5, 85) = 10.400, p < 0.0001 \), and RT, \( F(5, 85) = 5.229, p < 0.0001 \). Across serial positions, mean accuracy scores ranged from 77% for the first word in the sequence to 96% for the sixth word in sequence, following a generally linearly increasing trend (see Figure 4). For RTs, the data indicated a primacy/recency effect and followed an inverted-U shape.

In the follow-up task, results showed that both the SOSS and SODS trials were better recognized than the four novel trials (\( t(17) = 5.374, p < 0.0001 \)). Specifically, when asked to rate each sequence on a scale of 1-5, 5 being completely confident that the sequence was heard multiple times during the experiment, the average rating for SOSS (mean = 4.39, SD = 1.01) and SODS (mean = 4.33, SD = 0.67) trials were much higher than the average novel confidence level (mean = 2.90, SD = 0.67). There was no significant difference between the ratings for the SOSS and SODS conditions (\( t (17) = 0.212, p = 0.834 \)).

Of the 18 participants in Experiment I, five individuals indicated that they were not aware of the repetition manipulation, even when it was explicitly made reference to (i.e., “do you think that any of the sequences were repeating?”). Despite this, when presented with the follow-up task, these five participants did not differ from the rest of the group in their ability to identify the
repeating sequences \((t(16) = -1.436, p = 0.170;\) see Table 4). That is, both of the repeating sequences (SODS and SOSS) had a higher mean rating than the novel sequences regardless of whether or not the participants explicitly recognized that some sets were repeating.

2.4 Discussion

Due to the error discovered in the design files for Experiment I, few comparisons can be made. There are, however, a number of statistically significant findings. Serial position and practice effects can be addressed. Additionally, the results of the follow-up tasks provide insight into participants’ awareness of the repetition manipulation.

A recency effect was observed such that items near the end of the list were better and more quickly recognized than those near the beginning of the list. This is typical of recognition memory paradigms, especially those tested in the auditory domain (Ward, Avons, & Melling, 2005).

The increase in accuracy and decrease in RT across block number suggests simple learning effects as participants got better at the task over time. The trend analysis for accuracy, however, suggests otherwise. As can be seen in Figure 2, a dip in accuracy occurred during the fourth block across most of the conditions; therefore, evidence of a linear learning effect was not found, at least in the accuracy scores. Aside from this, accuracy generally seemed to be increasing. The drop in accuracy scores also corresponded to a decrease in RTs, which is uncommon; typically accuracy and RT scores are negatively correlated. It is possible that these results are partly attributable to fatigue effects, but this is purely speculative.
No effect of condition was found, indicating that even after 48 repetitions, pseudowords in the SOSS trials were not being remembered better (or worse) than those in the other trials. While the reasoning for this is unknown, some possibilities are suggested in the general discussion. Curiously, the results of the follow-up task strongly suggest that repetition is indeed having an effect on later recognition, despite the null result of condition.

Following the completion of the six blocks of the experiment, participants were questioned about any patterns they noticed during the task and then performed a confidence rating on selected sequences from the experiment. These results allowed for an examination of whether participants were aware of the sequence repetitions, and if they were differentially aware of the SOSS and SODS conditions. For example, when the five individuals who indicated that they did not think that any of the sequences were repeating were compared to the rest of the participants, no significant differences arose. Regardless of explicit awareness, participants were more confident that the repeating sets were heard multiple times compared to the non-repeating sets. These results demonstrate Hebbian learning and also follow previous research, which states that indexical properties of speech are encoded and stored implicitly in memory (Pisoni, 1993). The inconsistent effects of condition on recognition memory between the main and follow-up tasks suggest that the different testing procedures may be impacting the results. These findings are discussed in conjunction with those of Experiment II in the general discussion.
Chapter 3
Experiment II

The purpose of Experiment II was to investigate the impact of explicitly attending to temporal and sensory context effects, compared to the implicit effects noted in Experiment I.

3.1 Methods

The materials, design, and procedures of this experiment were nearly identical to those used in Experiment I (although the issue with probe type was corrected), with different participants in each study. The major difference between experiments was the judgment required of participants. In Experiment II, participants were asked to attend to both the item (pseudoword) and voice in order to make a recognition decision which encompassed both. Rather than a “yes/no” response, they were asked to make a “same voice/different voice/new item” response. A Same Item/Same Voice (SISV) response would indicate that the probe was a pseudoword that was present in the test set and was delivered by the same speaker at study and test. A Same Item/Different Voice (SIDV) response would be required if the probed pseudoword was the same, however, it was presented by a different speaker than at study. Finally, a Different Item (DI) response would indicate that the probed pseudoword was not present in the just-presented set. Therefore, ‘same speaker’ and ‘different speaker’ responses both represent old pseudowords and can be used as a measure of recognition for items irrespective of speaker. The ‘b’, ‘n’, and ‘m’ keyboard buttons recorded responses for ‘SISV’, ‘SIDV’, and ‘DI’, respectively. Within this Experiment, 25% of the trials were exact matches (item and speaker), 25% contained the same pseudoword presented by a different speaker, and the remaining 50% consisted of a different
pseudoword that did not appear in the preceding sequence. As such, half of the responses required ‘old’ judgments (collapsed across same/different) and half required ‘new’ judgments.

Participants:
Participants in this experiment consisted of 21 healthy young adults (average age 21.5 years; 6 males) who were primary English speakers with no history of neurological/psychiatric illnesses or hearing disorders. None had participated in the first Experiment. One participant was unable to complete the auditory follow-up task; therefore only 20 subjects contributed to those data. There were an additional 7 participants whose files were corrupted in the same way as those in Experiment I, and therefore their data was removed from analysis.

3.2 Analysis
All trials with RTs above 4500ms (2.5 standard deviations from the mean) were excluded. Following this, the mean and standard deviation (SD) for each participant was calculated, and trials above 2.5 SDs from the individual’s mean were removed. Accuracy (percent correct) and RT were both investigated across block number (1 through 6), condition (SOSS, SODS, DOSS, DODS, NRS), and probe type (SISV, SIDV, DI).

3.3 Results
A 6 x 5 x 3 (Block x Condition x Probe Type) repeated measures analysis of variance (ANOVA) was performed on the accuracy scores of the second Experiment (see Table 5). There was a significant effect of probe type \( (F(2, 40) = 69.789, p < 0.0001) \), however no other significant effects were observed \( (\text{block}, F(5, 100) = 0.253, p = 0.937; \text{condition}, F(4, 80) = 0.854, p = \)
The main effect of probe type appeared to be driven by the DI condition (mean accuracy = 0.869, standard error = 0.009) compared to SISV (mean accuracy = 0.576, standard error = 0.029) and SIDV (mean accuracy = 0.453, standard error = 0.026) trials. A further paired samples t-test (2-tailed) also revealed a significant difference between the SISV and SIDV conditions ($t(20) = 2.562, p < 0.019$), which was reflected higher accuracy scores in SISV compared to SIDV trials. A paired samples t-test (2-tailed) was performed on just the SIDV trials, which revealed no significant effect of gender congruency (i.e., same vs. different gender at study and test) on accuracy scores ($t(20) = -1.011, p = 0.324$). Response patterns, however, did show an effect of gender within the SIDV trials; see Figure 5. While participants correctly responded only about half of the time, a significant false alarm rate was detected. That is, participants were significantly more likely to respond with ‘same voice’ on SIDV trials when gender at study and test were congruent ($t(20) = 2.281, p = 0.034$).

A 6 x 5 x 3 (Block x Condition x Probe Type) within subjects ANOVA was also performed for RT data, with significant main effects of block ($F(5, 100) = 7.05, p < 0.0001$) and probe type ($F(2, 40) = 26.55, p < 0.0001$) but no other significant effects emerging, including interactions (condition, $F(4, 80) = 0.552, p = 0.698$; see Table 6). For block, there was a generally linear decrease in RT from Block 1 (mean = 1926.07, standard error = 86.27) to Block 6 (mean = 1672.47, standard error = 69.96; see Figure 6). The main effect of probe type was again driven by the DI condition (mean = 1595.9, standard error = 59.93) compared to the SISV (mean = 1879.66, standard error = 80.61) and SIDV (mean = 1951.02, standard error = 90.75) conditions. Once again, a paired samples t-test (2-tailed) revealed a significant difference between the SISV and SIDV trials ($t(20) = -2.274, p = 0.034$), showing that participants were significantly faster to
respond to SISV trials than SIDV. The RTs for the SIDV trials were not affected by reinstating the gender of the voice at study and at test ($t(20) = -1.612, p = 0.123$).

A second analysis was performed on the RT data in which block number was collapsed into either Bin 1 (blocks 1-3) or Bin 2 (blocks 4-6) in order to test the effect of repetition. In this ANOVA ($2 \times 5 \times 3$), there were no significant effects of condition, $F(4, 80) = 0.323, p = 0.862$, although there was again a significant effect of probe type ($F(2, 40) = 33.82, p < 0.0001$) and of binned block ($F(1, 20) = 8.242, p = 0.009$). The significant effect of probe type reflected quicker responses to DI trials, while binned block was the result of significantly faster RTs in the last three blocks (mean = 1759.468, standard error = 72.17) compared to the first three blocks (mean = 1883.763, standard error = 79.28). No interaction effects approached significance.

Overall, the accuracy scores for task 2 were smaller than those in task 1, while the RTs were generally longer, likely due to the increased difficulty of the task and extra judgment to be made. Participants’ average accuracy scores ranged between 63%-77%. There was an effect of serial position of the test word on both accuracy, $F(5, 100) = 30.750, p < 0.0001$, and RT, $F(5, 100) = 5.908, p < 0.0001$ (see Figure 7). Across conditions, mean accuracy scores ranged from 37% for the second serial position to 82% for the final item in the list, following a generally increasing trend. Reaction times followed a loosely inverted-U shaped curve.

The data for all 27 available participants in Experiment II was analyzed at follow-up (recall that the probe type was confounded for these participants, rendering their accuracy and RT scores unusable, but their follow-up data unaffected). The results showed that SODS and SOSS trials
were better remembered than the novel trials \((t(26) = 5.071, p < 0.0001)\). The average confidence rating for the SOSS trials was highest (mean = 4.37, SD = 0.74), followed by the SODS trials (mean = 3.78, SD = 1.25) and finally the non-repeating trials (mean = 3.02, SD = 0.65). The difference between the SOSS and SODS trials was statistically significant, \(t(26) = 2.254, p = 0.033\).

Nearly one third (30\%) of the participants in Experiment II failed to recognize that some of the sequences were repeating by the end of the study. Despite this, individuals who were and were not aware of the repetition did not differ from each other in their confidence ratings of either the repeating sets, \(t(25) = 0.766, p = 0.451\), or novel sets, \(t(25) = 0.493, p = 0.626\); see Table 7.

### 3.4 Discussion

Similar to the effects observed in Experiment I, a significant recency effect also appeared in Experiment II. While the pattern was comparable, the range of both the accuracy scores and RTs was much larger in the second Experiment than in the first due to the increased difficulty of the task.

There was no significant main effect of condition, either in accuracy or RT data, indicating that the contextual manipulation was not impacting participants’ responses. Block number significantly affected RT but not accuracy. As can be seen in Figure 6, RTs steadily decreased during each block of the experiment. This suggests a simple practice effect – participants became faster at the task over time, regardless of condition or probe type. Across both accuracy and RT
data, a significant main effect of probe type emerged, are there were no significant interaction effects between any of the main variables of interest.

The significant main effect of probe type was largely driven by the DI condition, as these trials were more quickly and accurately identified than Same Items, both Same Voice and Different Voice. This is not overly surprising – it is easier to reject a new word than correctly identify whether it is both old and if the speaker had been maintained. Further complicating the judgment of old words are the effects of gender and voice similarity; recall that word recognition is often facilitated by the reinstatement of the same speaker but source (voice) recognition is enhanced when words are re-presented in different voices of a different gender compared to different voices of the same gender (Campeanu, 2008). The results from the post hoc test investigating a significant difference between SISV and SIDV trials are consistent with this research and suggest that voice recognition is heavily dependent on similarity – if the voice is either exactly the same or very different, it is more easily recognized as such; when similarity increases, accurate voice recognition is more challenging.

This would suggest that SIDV trials in which the probe is of the same gender as the target item (and hence are more similar) would be subject to an increased false alarm rate. As discussed earlier, there have been mixed results when assessing whether or not participants rely on gender as a primary cue used to determine whether a probe item is spoken in the same voice as a target. When SIDV trials were isolated, accuracy scores and RTs did not differ when the voice at study and at test were of the same gender. When looking at response patterns, however, a clear difference arose. Participants correctly identified the probe as a SIDV trial only about 50% of the
time. An examination of Figure 5 reveals that SIDV and DI responses follow a similar pattern – when gender at study and test changes, participants are more likely to either judge a pseudoword as spoken by a different voice or as a new pseudoword altogether (but not significantly so) compared to when gender remains congruent. The pattern of responses for SISV, however, reflected the opposite trend; that is, participants were significantly more likely to judge an SIDV trial as SISV when the gender was congruent at study and test. This confirms the elevated false alarm rate when gender is reinstated at study and test.

In order to allow for learning to occur, the first three blocks and last three blocks of the experiment were collapsed and a new 2 x 5 x 3 (Binned Block x Condition x Probe Type) ANOVA was performed. It was hypothesized that because any differences between conditions would not be present in the first few blocks, this measure might be more sensitive to the effect of condition. Despite this, a significant main effect of condition was not found. A significant effect of probe type again emerged, as well as a main effect of block bin. Simple practice effects drove this, as participants produced quicker responses in the last three blocks of the experiment than in the first three.

The follow-up task in the second Experiment replicated the results seen in Experiment I; specifically, repeated sequences were rated significantly higher than non-repeated sequences. Following the same trend, SOSS trials had the highest confidence ratings. When all available data for Experiment II was tested, a significant difference between the SOSS and SODS trials emerged. Recall that the SOSS condition reflects the strongest contextual cues across trials, and that both the SOSS and SODS sequences were repeated an equal number of times. Differences
between the SOSS and SODS conditions are therefore attributable to the additional effect of speaker repetition as the same speakers presented the same words in every trial in the SOSS condition. This indicates that context (condition) does indeed affect the recognition of the entire memory trace, despite no significant effects of condition in the main analyses. Unfortunately, only two of the four repeating conditions were tested, making it impossible to determine if the differences between SOSS and SODS represent the upper end of a gradual benefit due to contextual preservation, or an all-or-none threshold effect of reinstating both parameters. Once again, those who were aware of the repeating sequences did not significantly differ from those who were unaware of the repetition in their confidence ratings of either the repeated or novel sequences, implying that repetition effects were largely implicit.
Chapter 4
General Discussion

The present research was conducted in order to assess the impact of sequential context across entire sequences. The purpose was to investigate whether episodic auditory-verbal memory traces are accessed as coherent wholes or individual elements. To achieve this, the sequential context of individual pseudowords was systematically manipulated and accuracy and RT were recorded. Theoretically, within and across experiments, access to STM and LTM, implicit and explicit learning, as well as word and voice recognition is gained. The results help to deconstruct how these types of memories are accessed and tell us about the effects of context on these various interactions of memory. After an explanation of the null effect of condition in the main experiment, a theoretical account of what the results showed and could have produced (had Experiment I not been confounded) will be presented, followed by the implications of the current research on the existing literature.

Encoding Manipulation:

It was hypothesized that over time, participants would develop voice-specific episodic LTM traces for the sets in the SOSS condition. Therefore, by analyzing the accuracy and RT of responses in this condition across the blocks of the experiment, access to the formation of memories from STM to LTM can be gained. While the immediate STM judgments were impacted by probe type, they did not appear to be differentially affected across trials by the total or partial preservation of context; that is, the immediate judgments were not facilitated by speaker and word order reinstatement even after 48 repetitions. Despite this, long-term learning as assessed by the follow-up task shows a clear impact of context. This data implies that
condition did have an impact on the recognition of auditory verbal memory traces despite the null results of the main Experiments. The overall results imply that STM was unaffected by contextual effects, while LTM was indeed impacted by contextual reinstatement.

There are many possible reasons why an improvement over time as a function of the contextual manipulation was not observed. It could simply mean that sequential memories are not retrieved as holistic entities, but rather as several unique elements that remain associated but segmented. Previous research, however, suggests that auditory memories are encoded cohesively, with linguistic and perceptual information being stored interdependently and automatically. Thus, there are a number of practical task-specific reasons why an explicit learning effect for the SOSS condition was not observed, including the inability of participants to learn voices from isolated pseudowords, the inability to learn multiple sequences, the fact that all conditions contained pseudowords from the same item set, or that the encoding of the sequences did not emphasize temporal context. Finally, the differences observed between STM and LTM could have been attributable to the different tasks used to assess the different types of memory. These possibilities are discussed below.

The possibility that the null results are due to participants’ inability to learn voices from isolated pseudowords can be explored through previous research on voice recognition. For example, Nygaard and Pisoni (1998) tested long-term recognition for participants who were trained to identify 10 different speakers using both isolated words (Experiment 1) and entire sentences (Experiments 2 and 3). In the first experiment, groups of listeners were exposed to the voices of 5 male and 5 female speakers, each of which were paired with common gender-specific names,
and performed recognition tests with feedback. The researchers found that after 9 one-hour training sessions spread over two weeks, only about half of the participants successfully learned to identify (by name) the voices of isolated words based on a criterion of 70% accuracy. Upon further investigation, it was found that “good” and “bad” learners did not differ in performance on the first day of training, but quickly diverged over the following training days. While the authors suggest that simple exposure to isolated words was not sufficient for learning to occur, there are many important differences between the paradigm used by Nygaard and Pisoni and the paradigm used in the current study. First, our study only requires participants to learn 6 speakers, rather than 10. Further, through recognition rather than serial recall, the current research largely promotes incidental learning (especially in the first Experiment), which appears to be more sensitive to talker variability than explicit learning, as was emphasized in the Nygaard and Pisoni study. Finally, the results of the Nygaard and Pisoni study imply that learning is indeed possible as nearly half of the participants showed dramatic improvement. Regardless, in the current study, STM recognition was not affected by repetition of the same sequences, even with voice repetition, which could be addressed by extending the practice phase and using stimuli with an inherent sequential structure such as sentences rather than random word sets for training.

Indeed, Nygaard and Pisoni explored the use of full sentences rather than isolated words to train participants in Experiments 2 and 3. Participants were asked to identify the ten voices over three days of training, and it was found that a greater proportion of participants (about 72% in Experiment 2 and 100% in Experiment 3) were able to identify the voices with at least 70% accuracy when tested with full sentences. When tested with isolated words, voice identification was much lower (around 63%). These results indicate firstly that learning voices from sentences
is far superior to learning voices from isolated words, and secondly that learning voices from sentences does not generalize very well to identifying those voices using isolated words. Thus, what is lacking from lists of isolated words – namely syntactic structure of some kind – is useful in helping participants to identify a particular speaker. Further, it appears that participants learned a qualitatively different set of properties to identify voices from full sentences than they did from isolated words. These results imply that surrounding contextual information, provided that it is embedded in a syntactic structure, can have a profound effect on learning.

It is also possible that participants were unable to learn the multiple sequences that were required of them using this paradigm. The vast majority of studies using the HRP train participants to learn only single lists. According to one source (Page & Norris, 2009), at least four different sequences can be learned via the HRP. Unfortunately, these results do not appear to have been published, and were likely performed using an ISR task. While ISR requires recall of an entire list (including serial order), recognition memory of an isolated item does not emphasize or reinforce the sequential structure of the word lists. Nonetheless, as it was discovered that participants were unable to improve performance with multiple lists, a between-subjects design could be employed such that each participant would be tested in only one condition with one repeating sequence. The follow-up data, however, suggest that participants in the current study were aware of the repetition manipulation and could reliably detect two different repeating sequences when presented as wholes.

Whether or not participants would experience Hebbian learning for the Different Order conditions (DOSS/DODS) is also an issue of debate. Early research on the HRP demonstrated
the importance of maintaining the first two digits of a nine-digit ISR task (Schwartz & Bryden, 1971). When just the first two digits of a sequence were changed, performance was no better than when all of the numbers were changed. Further, Cumming, Page, and Norris (2003) noted that changing the positions of half of the digits in an ISR task with nine digits eliminated the Hebb repetition effect. Thus, word order may have a substantial impact on the efficacy of learning. According to a model based on immediate serial recall and Hebb repetition effects, Page and Norris (2009) have posited that long-term sequences are encoded sequentially as chunks. Essentially, the model suggests that chunk nodes, made up of individual elements, are encoded temporally. For example, a chunk node representing the sequence DOG is made up of its’ individual components (i.e., the letters D, O, and G) which have temporally graded connections. That is, the positive connections from the DOG unit to the D unit are stronger than those from the DOG unit to the O unit, which in turn are stronger than for those to the G unit. A crucial aspect of this theory, therefore, is the temporal order of the chunk such that the memory trace for DOG would not activate for the chunk GOD. As applied to the current study, this theory would suggest that the different-order conditions (DOSS and DODS) would not be subject to Hebbian learning due to the inconsistent word associations. Unfortunately, it has not been specified how this model relates to recognition learning, however, the authors suggest that it can be applied. Regardless, because the conditions in the current research did not produce significant STM effects, this theory cannot be tested, but could have been assessed with a follow-up task that included the DODS and DOSS conditions.

There is also the possibility that the sets were not learned, or perhaps simply weakly so, due to the fact that the pseudowords in the NRS sequences came from the same item set as those in the
repeating sequences. As reviewed by Page and Norris (2009), early work suggested that Hebb repetition learning was not observed for lists with large gaps (i.e., presented every sixth rather than every third list). Page and Norris’ own group, however, showed that learning was effective for spacings as large as every twelfth list so long as the repeated and filler lists came from different item sets. In the current study, all lists were constructed from the same item set (24 pseudowords per participant), so it is possible that learning was not observed due to this reason. However, the study by the Page and Norris group again does not appear to have been published and was likely tested using an ISR task rather than recognition which would aid in memory; therefore, I believe that the paradigm was worth investigating. Also, the lists repeated at shorter gaps than every twelfth list in the current study. Regardless, a larger item set could be employed in future experiments to address this issue.

Another potential issue with the current research relates to Morris, Bransford, and Franks’ transfer appropriate processing theory (1977), which states that it is important to test individuals using a measure that adequately captures the design of the study. For example, the authors showed that deep semantic processing only results in better memory of a word if the word was being tested in a semantic way. If, on the other hand, it was being tested using a rhyming paradigm, it would be better remembered if a rhyming manipulation was also used during encoding. As applied to the current research, because the study phase did not emphasize serial order, it is possible that encoding did not stimulate a sequential memory trace. If the participants in the current study had been tested with an immediate serial recall task, perhaps condition effects may have emerged (although previous studies suggest that changing the order of even
half of the items in a repeating list removes any Hebbian learning effects; Schwartz & Bryden, 1971; Cumming, Page, & Norris, 2003).

While the above possibilities could each have impacted the learning of the sequences over time, none explain why an effect was observed in the follow-up task but not in the main experiment. The key to this result may lie in the memory systems responsible for each task. The significant impact of probe type suggests that the immediate judgment was being dominated by STM. Thus, even after 48 repetitions of a sequence, participants were using their immediate STM to make the judgment. The follow-up task, however, clearly relied on LTM. Thus, even if the episodic LTM trace was building up (as suggested by the follow-up task), participants were not utilizing this to make the immediate recognition judgment. Unfortunately, the conclusion that the effects are due to the separate memory systems is confounded by the possibility that the observed differences between tasks may be attributable to the different ways in which these memory systems were tested. The STM judgment was based on recognition of an isolated pseudoword, whereas the LTM judgment was based on the entire sequence. While this was done in order to assess the question of how EAM are accessed, in order to conclusively state that these systems are differentially affected by context per se, they must be tested using the same method. It would be possible to simply record a judgment after each sequence regarding the recognition of the entire sequence (rather than an isolated pseudoword); however, this would not allow for an investigation into whether EAM are accessed as coherent units or individual elements.

In order to further assess the differential impact of the two memory systems, another option would be to block rehearsal between when the participants hear the study sequence and when
they hear the probe. If rehearsal were blocked, for example, by a verbal distractor task, participants would be forced to rely on the LTM trace of the sequence which was just activated. In this situation, if context is having an effect on LTM, performance should improve in the immediate judgments with increasing repetitions. By testing two groups – one that is allowed to rehearse and one that is not – the differential effects of contextual reinstatement on STM and LTM could be properly assessed.

*Implicit and Explicit Effects:*

Theoretically, by looking within experiments, similarities and differences between implicit and explicit learning could be investigated. In Experiment I, implicit encoding and retrieval of the voices that present each pseudoword could be analyzed, as word recognition alone was being tested. Therefore, if word recognition was enhanced when the same speaker presented a pseudoword at study and at test, this would provide evidence of implicit voice learning consistent with previous research (Craik and Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993). This would also support the TAP framework, which suggests that congruency of features at encoding and retrieval enhances memory. Further, considering that there would be trials in which voice and order are systematically preserved or shuffled, the extent to which episodic context enhances recognition memory implicitly through repetition could be assessed. Specifically, if accuracy was improved and RT decreased for the SOSS condition compared to the others (especially SODS), it would suggest that the preservation of word- and speaker-order creates a stronger episodic context and enhances the learning effect over a series of trials within the CRP.
Experiment II, in which both item and voice judgments were made, reflects explicit learning and any differences in comparison to Experiment I would speak to the way in which EAM are explicitly encoded and retrieved. Overall, a similar pattern of results was expected for both experiments in terms of item recognition. Recall that in Experiment II, ‘same’ and ‘different’ responses can be collapsed to create an ‘old’ category that is comparable to the ‘old’ and ‘new’ judgments that would be acquired in Experiment I. It would be expected that ‘old’ responses would be influenced (enhanced) by the presence of consistent voices at study and test, regardless of condition. This was evidenced by the significant difference between the SISV and SIDV trials in Experiment II – pseudowords repeated in the same voice were more quickly and accurately identified than those repeated in a different voice. Unfortunately, this could not be assessed in Experiment I, but contrasting the differences between implicit and explicit encoding would have been interesting. If no difference between the Experiments arose, it would imply that intentionally attending to voices does not impact the probability or speed of remembering them later. Previous research, however, suggests a slight disadvantage of explicit learning on voice recognition (Palmeri, Goldinger, & Pisoni, 1993), as participants tend to rely on gender when classifying same-gender/different-voice repetitions. This would indicate that more false alarms would be present in the data from Experiment II than Experiment I.

In terms of simple auditory word recognition, a cumulative effect of repetition can be assessed over a series of trials, compared to single-trial learning. As mentioned earlier, most CRP studies utilize single-trial exposure to isolated words, so the effect of speaker and word order on sequences has not been systematically examined in the past. Our results suggest a simple repetition effect – RTs decreased as a function of block number, at least for the first Experiment.
It is somewhat curious that the accuracy scores did not reflect the learning effects observed in the RT data; the reasoning for this remains to be seen. Nonetheless, a strong same-voice bias was observed for word recognition accuracy and RT in Experiment II.

Regarding voice recognition, it has been shown that participants tend to exhibit greater accuracy in identifying ‘same’ voices as opposed to ‘different’ voices (Palmeri, Goldinger, & Pisoni, 1993). However, it is possible that this effect is driven by gender as Palmeri and colleagues observed a high rate of false alarms for words presented by different speakers of the same gender. In the present studies, participants were quicker to respond to the same voice than to a different voice presenting a pseudoword at study at test. Despite this fast recognition, there was a significantly higher rate of false alarms in SIDV trials when gender was reinstated from study to test than when gender was incongruent. Therefore, the current results suggest that gender impacts responses but does not specifically aid in correctly identifying a voice (as evidenced by non-significant differences in accuracy and RT scores among SIDV trials with gender reinstatement compared to gender incongruency).

Previously, the learning of multiple sequences using a recognition-based HRP had not been fully documented. The results from the follow-up tasks of both Experiments suggest that a recognition-based HRP (as opposed to an immediate serial recall task) still produces learning effects. As the unannounced repetition of both the SOSS and SODS lists were better remembered than the non-repeating sets, even though they were never verbally recalled, support was found for the notion that Hebbian learning can occur through simple recognition. Further, since the item set for each participant overlapped (and therefore the same pseudowords appeared in
repeating and non-repeating sets) it is not likely that relying on only the first or first and last words would be an adequate strategy, as has been suggested in previous research (Cohen & Johansson, 1967). In addition, the fact that two separate lists (SOSS and SODS) were better remembered suggests that the learning of multiple lists within an HRP is possible. The current research also represents the first known attempt to investigate the learning of sequences through a pseudo-CRP task using repetition. Rather than single-trial exposure, a series of trials was used to build an EAM, which may have created an increasingly stronger trace.

4.1 Future Directions

While the results of this study were somewhat inconclusive, some general findings were observed. Overall, it appeared that STM was dominating the immediate recognition judgment and that it was unaffected by the encoding manipulation, while long-term learning was indeed impacted by context. Some support was found for the notion that episodic auditory verbal memories are accessed as entire units rather than individual elements, and that sequential context plays an important role. In order to further investigate this, it would be helpful to take into account the potential issues with the current study. A replication of the current research utilizing a larger item set, recall instead of recognition, sequences with a syntactic structure, and/or tests of the entire memory trace may be more likely to pick up on the effects of context on episodic auditory-verbal memories. Further, a controlled manipulation of STM and LTM involving similar tests for each would clarify whether the differential results are due to the memory systems themselves or the ways in which they were tested. A between-subjects design manipulating rehearsal will also properly assess the effects on STM and LTM.
A further manipulation condition can also be added, in which the speaker of each pseudoword is maintained across trials, yet word order if shuffled. This would provide a unique control for the information being tested. In this case, when compared to the SOSS condition, the exact same information is being presented, the only difference is the order (and hence the prosody of the sequence). The new condition could replace the current DOSS condition in order to keep the number of sets to be learned to a minimum.

4.2 Conclusion

In summary, the proposed research contributes to our understanding of sequence learning and recognition memory. The results help clarify the role of recognition within the HRP, as well as whether or not multiple lists can be learned simultaneously. The current research also addresses the cumulative effect of a series of learning trials within the framework of the CRP. Finally, I was able to examine the implicit and explicit temporal and sensory context effects for spoken words in both STM and LTM. Evidence for enhancement of an episodic memory trace through the strengthening of context would support the notion that EAM for verbal sequences are accessed in an all-or-none fashion. Further research is needed to support this claim, however, preliminary results from the current research suggest that it is plausible.
References


Tables
Table 1. Descriptions and examples of each condition as represented at the first and second trial. Coloured words represent different speakers, such that the speaker order is maintained across trials in the ‘same speaker’ conditions (SOSS and DOSS). Note that sequences would contain different pseudowords across conditions, however, the same list was used for simplicity and to highlight the non-repetitive nature of the NRS condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
<th>Example Across Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOSS</td>
<td>Both word-order and speaker retained</td>
<td>1: “Kax Bim Zoy Rit Cov Meb” 2: “Kax Bim Zoy Rit Cov Meb”</td>
</tr>
<tr>
<td>SODS</td>
<td>Only word-order retained</td>
<td>1: “Kax Bim Zoy Rit Cov Meb” 2: “Kax Bim Zoy Rit Cov Meb”</td>
</tr>
<tr>
<td>DOSS</td>
<td>Only speakers retained</td>
<td>1: “Kax Bim Zoy Rit Cov Meb” 2: “Bim Zoy Kax Meb Cov Rit”</td>
</tr>
<tr>
<td>DODS</td>
<td>Neither word-order nor speakers retained, but still same 6 words</td>
<td>1: “Kax Bim Zoy Rit Cov Meb” 2: “Bim Meb Kax Zoy Cov Rit”</td>
</tr>
<tr>
<td>NRS</td>
<td>Nothing retained across trials; different words in each set</td>
<td>1: “Kax Bim Zoy Rit Cov Meb” 2: “Nud Pid Wis Hin Jaf Dop”</td>
</tr>
</tbody>
</table>
Table 2. A 6 x 5 (Block x Condition) within-subjects ANOVA on accuracy scores for Experiment 1.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>0.227</td>
<td>5</td>
<td>0.045</td>
<td>3.187</td>
<td>0.011</td>
</tr>
<tr>
<td>Condition</td>
<td>0.008</td>
<td>4</td>
<td>0.002</td>
<td>0.084</td>
<td>0.987</td>
</tr>
<tr>
<td>Block*</td>
<td>0.177</td>
<td>20</td>
<td>0.009</td>
<td>0.551</td>
<td>0.943</td>
</tr>
</tbody>
</table>
Table 3. A 6 x 5 (Block x Condition) within-subjects ANOVA on RT scores for Experiment 1.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>890960.213</td>
<td>5</td>
<td>178192.043</td>
<td>5.105</td>
<td>0.0001</td>
</tr>
<tr>
<td>Condition</td>
<td>23966.219</td>
<td>4</td>
<td>5991.555</td>
<td>0.709</td>
<td>0.588</td>
</tr>
<tr>
<td>Block*Condition</td>
<td>102908.013</td>
<td>20</td>
<td>5145.401</td>
<td>0.606</td>
<td>0.909</td>
</tr>
</tbody>
</table>
Table 4. Independent samples t-test (2-tailed) of repeating and novel sequences in Experiment I. When the responses of those aware of repeating sequences (n=13) are compared to those who were unaware of repetition (n=5), no significant differences arose.

<table>
<thead>
<tr>
<th>Explicit Recognition</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeating</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5</td>
<td>4.0000</td>
<td>.70711</td>
<td>.31623</td>
</tr>
<tr>
<td>Yes</td>
<td>13</td>
<td>4.5000</td>
<td>.64550</td>
<td>.17903</td>
</tr>
<tr>
<td>NOVEL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5</td>
<td>3.0500</td>
<td>.57009</td>
<td>.25495</td>
</tr>
<tr>
<td>Yes</td>
<td>13</td>
<td>2.8462</td>
<td>.74679</td>
<td>.20712</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeating</td>
<td>-1.436</td>
<td>16</td>
<td>.170</td>
<td>-.50000</td>
<td>.34807</td>
</tr>
<tr>
<td>NOVEL</td>
<td>.548</td>
<td>16</td>
<td>.591</td>
<td>.20385</td>
<td>.37193</td>
</tr>
</tbody>
</table>
Table 5. Repeated measures ANOVA of accuracy scores for Experiment II.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>0.121</td>
<td>5</td>
<td>0.024</td>
<td>0.253</td>
<td>0.937</td>
</tr>
<tr>
<td>Condition</td>
<td>0.319</td>
<td>4</td>
<td>0.080</td>
<td>0.854</td>
<td>0.495</td>
</tr>
<tr>
<td><strong>ProbeType</strong></td>
<td><strong>57.494</strong></td>
<td>2</td>
<td><strong>28.747</strong></td>
<td><strong>69.789</strong></td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td>Block*Condition</td>
<td>1.489</td>
<td>20</td>
<td>0.074</td>
<td>0.756</td>
<td>0.766</td>
</tr>
<tr>
<td>Block*ProbeType</td>
<td>1.071</td>
<td>10</td>
<td>0.107</td>
<td>0.879</td>
<td>0.554</td>
</tr>
<tr>
<td>Condition*ProbeType</td>
<td>0.759</td>
<td>8</td>
<td>0.095</td>
<td>1.035</td>
<td>0.412</td>
</tr>
<tr>
<td>Block<em>Condition</em>ProbeType</td>
<td>3.316</td>
<td>40</td>
<td>0.083</td>
<td>0.877</td>
<td>0.688</td>
</tr>
</tbody>
</table>
Table 6. Repeated measures ANOVA of reaction times for Experiment II.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>1.275E7</td>
<td>5</td>
<td>2.549E6</td>
<td>7.053</td>
<td>0.0001</td>
</tr>
<tr>
<td>Condition</td>
<td>442849.272</td>
<td>4</td>
<td>110712.318</td>
<td>0.552</td>
<td>0.698</td>
</tr>
<tr>
<td><strong>ProbeType</strong></td>
<td><strong>4.446E7</strong></td>
<td><strong>2</strong></td>
<td><strong>2.223E7</strong></td>
<td><strong>26.555</strong></td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td>Block*Condition</td>
<td>1.838E6</td>
<td>20</td>
<td>91888.872</td>
<td>0.512</td>
<td>0.962</td>
</tr>
<tr>
<td>Block*ProbeType</td>
<td>774734.827</td>
<td>10</td>
<td>77473.483</td>
<td>0.497</td>
<td>0.891</td>
</tr>
<tr>
<td>Condition*ProbeType</td>
<td>1.144E6</td>
<td>8</td>
<td>142986.508</td>
<td>0.747</td>
<td>0.650</td>
</tr>
<tr>
<td>Block<em>Condition</em>ProbeType</td>
<td>6.523E6</td>
<td>40</td>
<td>163068.293</td>
<td>0.819</td>
<td>0.781</td>
</tr>
</tbody>
</table>
Table 7. Independent samples *t*-test (2-tailed) of confidence ratings of participants in Experiment II who were (n=19) and were not (n=8) explicitly aware of repeating sequences.

### Group Statistics

<table>
<thead>
<tr>
<th>Explicit Recognition</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>4.2500</td>
<td>.75593</td>
<td>.26726</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19</td>
<td>4.0000</td>
<td>.78174</td>
<td>.17934</td>
<td></td>
</tr>
<tr>
<td>NOVEL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>3.1250</td>
<td>.80178</td>
<td>.28347</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19</td>
<td>2.9868</td>
<td>.60366</td>
<td>.13849</td>
<td></td>
</tr>
</tbody>
</table>

### Independent Samples Test

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeating</td>
<td>.766</td>
<td>25</td>
<td>.451</td>
<td>.25000</td>
<td>.32646</td>
</tr>
<tr>
<td></td>
<td>.777</td>
<td>13.647</td>
<td>.451</td>
<td>.25000</td>
<td>.32186</td>
</tr>
<tr>
<td>NOVEL</td>
<td>.493</td>
<td>25</td>
<td>.626</td>
<td>.13816</td>
<td>.28032</td>
</tr>
<tr>
<td></td>
<td>.438</td>
<td>10.507</td>
<td>.670</td>
<td>.13816</td>
<td>.31549</td>
</tr>
</tbody>
</table>
Figure 1. Graphical depiction of design of Experiment I. For convenience, only 15 of the 40 trials in each block are represented. Each letter denotes order and speaker, respectively (i.e., SS = Same Order/Same Speaker; DS = Different Order/Same Speaker).
Figure 2. Main effect of block averaged across conditions for all participants in Experiment I. A) Accuracy scores show a non-linear trend. B) Reaction times suggest a general learning effect.
**Figure 3.** Line graphs of Block x Condition interaction effects from Experiment I. A) Accuracy scores by block do not follow a linear trend, indicating a lack of learning effects. Note the small range of the scale, in order to see the different conditions. B) Reaction time data produce a general linear trend by block, with no significant differences between conditions.

A)

B)
Figure 4. Serial position effects for Experiment I. A) Average accuracy scores by serial position of test word; data indicates a strong recency effect. B) Average reaction time by serial position of test word.
Figure 5. Probability of responses in SIDV trials according to gender congruency at study and test.
Figure 6. Mean reaction times by block number for Experiment II averaged across participants and conditions.
Figure 7. Serial position effects for Experiment II. A) Average accuracy scores by serial position of test word shows some primacy and a strong recency effect. B) Average reaction time by serial position of test word shows an inverted-U pattern consistent with primacy/recency effects.
Appendices
### Appendix 1. List of pseudowords used in Experiments I and II.

<table>
<thead>
<tr>
<th>bap</th>
<th>jid</th>
<th>rof</th>
</tr>
</thead>
<tbody>
<tr>
<td>beb</td>
<td>jop</td>
<td>rud</td>
</tr>
<tr>
<td>bim</td>
<td>jub</td>
<td>sab</td>
</tr>
<tr>
<td>bof</td>
<td>kax</td>
<td>seb</td>
</tr>
<tr>
<td>bux</td>
<td>kem</td>
<td>sif</td>
</tr>
<tr>
<td>cas</td>
<td>kig</td>
<td>sof</td>
</tr>
<tr>
<td>cex</td>
<td>kog</td>
<td>sut</td>
</tr>
<tr>
<td>cif</td>
<td>kug</td>
<td>tam</td>
</tr>
<tr>
<td>cov</td>
<td>lar</td>
<td>teg</td>
</tr>
<tr>
<td>cug</td>
<td>lev</td>
<td>tiv</td>
</tr>
<tr>
<td>daw</td>
<td>lif</td>
<td>tox</td>
</tr>
<tr>
<td>deg</td>
<td>loz</td>
<td>tuz</td>
</tr>
<tr>
<td>div</td>
<td>lut</td>
<td>vav</td>
</tr>
<tr>
<td>dop</td>
<td>mav</td>
<td>viv</td>
</tr>
<tr>
<td>duf</td>
<td>meb</td>
<td>vur</td>
</tr>
<tr>
<td>fam</td>
<td>mip</td>
<td>wem</td>
</tr>
<tr>
<td>fip</td>
<td>mov</td>
<td>wis</td>
</tr>
<tr>
<td>fod</td>
<td>mup</td>
<td>wog</td>
</tr>
<tr>
<td>fud</td>
<td>naf</td>
<td>wug</td>
</tr>
<tr>
<td>gan</td>
<td>nen</td>
<td>yat</td>
</tr>
<tr>
<td>ged</td>
<td>nid</td>
<td>yeb</td>
</tr>
<tr>
<td>gim</td>
<td>nop</td>
<td>yig</td>
</tr>
<tr>
<td>goy</td>
<td>nud</td>
<td>yog</td>
</tr>
<tr>
<td>har</td>
<td>pab</td>
<td>yuv</td>
</tr>
<tr>
<td>hep</td>
<td>pex</td>
<td>zam</td>
</tr>
<tr>
<td>hin</td>
<td>pid</td>
<td>zeg</td>
</tr>
<tr>
<td>hod</td>
<td>pux</td>
<td>zin</td>
</tr>
<tr>
<td>hup</td>
<td>ras</td>
<td>zoy</td>
</tr>
<tr>
<td>jaf</td>
<td>rem</td>
<td>zub</td>
</tr>
<tr>
<td>jer</td>
<td>rit</td>
<td></td>
</tr>
</tbody>
</table>