Voice as a Memory Cue

by

Sandra Campeanu

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Psychology
University of Toronto

© Copyright by Sandra Campeanu 2014
Voice as a Memory Cue

Sandra Campeanu
Doctor of Philosophy
Department of Psychology
University of Toronto
2014

Abstract

The study of human memory often focuses on the benefit of context congruency. In the auditory domain, perhaps the most important context cue is speaker’s voice. The present set of auditory experiments was designed to further our understanding of voice as a memory cue. Words were spoken in voices that were congruent on either accent or pitch, both or neither. Voice reinstatement signified maximum congruency and occurred when the same voice spoke the word initially and in a subsequent memory test; voice congruency referred to either accent or pitch being comparable between presentations of a word. These stimuli were chosen to validate and extend earlier models of voice congruency based primarily on variations of pitch between speakers; we investigated multiple salient voice parameters. Experiment 1 found that voice reinstatement aids word and speaker recognition, and that voice congruency partially benefits word recognition. Experiment 2 used a different paradigm and found that voice reinstatement is required to boost word recognition when participants are not given instructions to attend to voice. Experiment 3 provided preliminary evidence that voice reinstatement benefits implicit memory. Experiment 4 examined whether voice effects vary with linguistic background; the benefit of accent congruency to word recognition appears to be augmented in bilinguals of the same language as the accented speaker voice. The first two experiments also utilized event-related potentials (ERPs) to discern the neural correlates underlying the use of voice in memory.
Taken together, results indicate that the extent of voice congruency effects, as well as the characterization of ERP modulations reflective of memory components, depends on the task/paradigm used. For example, attention (or some extended processing) is required to observe partial voice congruency effects, while a voice reinstatement benefit appears to be more robust. Finally, this work serves as a starting point for future studies to understand voice congruency in implicit memory and in people with specific language abilities.
Acknowledgments

I have always been fascinated by the human brain. My graduate work has been an effort to understand a little piece of this most complex and sophisticated organ; I am grateful to have had the opportunity to work in such an exciting area of research and with such amazing thinkers.

First and foremost, I would like to thank my supervisor, Dr. Claude Alain, for his continued guidance and support. When I started looking into graduate schools, I was often told that the most important ingredient for success was compatibility with your supervisor. I found this to be absolutely true, and I could not have asked for a more dedicated mentor and friend. Thank you so much for your patience and direction over the past several years.

I would also like to thank Dr. Gus Craik for his collaboration in many of my studies. Your leadership has been invaluable in helping me understand and navigate the vastness of memory research. My favorite moments in graduate school were brainstorming sessions with Claude and Gus, where we would bounce and develop ideas for understanding auditory memory.

I want to express my gratitude to Dr. Cheryl Grady for her thoughtful suggestions and guidance in my work, and to Drs. Lynn Hasher, Kathy Pichora-Fuller and Pascal Belin for their valuable input. Thank you to all the other researchers who have helped me along the way. Most notably, thank you to Dr. Kristina Backer for her help with programming and data analysis, and thank you to Yu He, who has always been extremely patient in her explanations and who taught me so much.

The road to this dissertation has not always been a straight and simple one, so I would like to thank all of the people who provided me with support over the years. Thank you to my family for your continued faith in me, your encouragement, and in the belief that I would succeed. I especially would like to thank my mother, who is always there for me and continues to actively search for ways to help me sort through my frustrations. Your love and dedication has given me much strength in the past few years. P.S. I would not have finished this dissertation without your help. Thank you infinitely!

Lastly, this work has been funded by lab grants from NSERC and CIHR, and by an individual Ontario Graduate Scholarship.
Table of Contents

Contents

Acknowledgments .............................................................................................................. iv
Table of Contents ............................................................................................................... v
List of Figures ...................................................................................................................... ix
Chapter 1 Background Literature ...................................................................................... 1
  1 Word Recognition ........................................................................................................... 3
  2 Source Memory ............................................................................................................... 7
  3 Event-Related Potentials and Recognition Memory ......................................................... 8
    3.1 Left Parietal Old/New Effect ....................................................................................... 9
    3.2 Early Frontal Effect .................................................................................................... 10
    3.3 Late Right Frontal Effect .......................................................................................... 13
    3.4 Three Functionally Distinct ERP Modulations ......................................................... 15
  4 Primary vs. Secondary Memory ....................................................................................... 16
  5 Implicit Memory ............................................................................................................. 17
  6 ERPs and Implicit Memory ............................................................................................. 21
  7 Language Skills ............................................................................................................. 22
  8 Summary of Circumstances in which the Voice Congruency effect on Word Memory will
    occur ............................................................................................................................. 23
Chapter 2 Experiments ....................................................................................................... 25
  9 General Hypotheses ....................................................................................................... 26
  10 General Methods .......................................................................................................... 26
    10.1 Participants .............................................................................................................. 26
    10.2 Stimuli ..................................................................................................................... 26
    10.3 Electrophysiological Recording and Analysis ......................................................... 27
Chapter 3 Experiment 1: Voice Congruency in Word and Source Memory using a Block Design

11 Method

11.1 Participants

11.2 Design

11.3 Behavioral Analysis

11.4 Electrophysiological Analysis

12 Results

12.1 Behavioral Results

12.2 ERP Results

12.2.1 Word Recognition Test

12.2.1.1 Left Parietal Old/New Effect

12.2.1.2 Frontal Familiarity Effect

12.2.1.3 Right Frontal Effect

12.2.2 Source Recognition Test

12.2.2.1 Left Parietal Old/New Effect

12.2.2.2 Frontal Familiarity Effect

12.2.2.3 Right Frontal Effect

13 Discussion

13.1 Behavioral Findings

13.2 ERP findings

14 Interim Conclusion

Chapter 4 Experiment 1a: Control Experiment

15 Methods

15.1 Participants

15.2 Design
23.3 Design ............................................................................................................................... 62
23.4 Behavioral Analysis ......................................................................................................... 63
24 Results .................................................................................................................................. 63
25 Discussion .......................................................................................................................... 65
26 Interim Conclusion .............................................................................................................. 69
Chapter 7 Experiment 4: Voice Congruency across Language Groups .......................... 70
27 Method .................................................................................................................................. 70
  27.1 Participants ..................................................................................................................... 70
  27.2 Design ............................................................................................................................ 71
  27.3 Analysis ........................................................................................................................... 71
28 Results .................................................................................................................................. 72
29 Discussion .......................................................................................................................... 73
30 Interim Conclusion .............................................................................................................. 76
Chapter 8 General Discussion and Conclusions ............................................................. 77
References ............................................................................................................................... 82
Copyright Acknowledgements .............................................................................................. 90
List of Figures

Figure 1. Behavioral Results from Experiment 1. A) Group Mean Accuracy for the Word Recognition Test. B) Group Mean Accuracy for the Source recognition Test. C) Group Mean False Alarm Rates for the Source recognition Test.

Figure 2. Word Recognition Test in Experiment 1. Panel of Electrodes.

Figure 3. Left Parietal Old/New Effect in Experiment 1. Recorded over the left parieto-occipital site (i.e., PO3) during the Word Recognition Test, showing all five conditions.

Figure 4. Source Recognition Test in Experiment 1. Panel of Electrodes.

Figure 5. Right Frontal Effect in the Source Recognition Test in Experiment 1. A) Waveform recorded over AF8, showing the same speaker condition versus the collapsed different speaker condition, B) Scalp topography showing the significant cluster over right frontal regions.

Figure 6. Group Mean Accuracy in the Word Recognition Test in Experiment 1a.

Figure 7. Group Mean Accuracy in the Source Recognition Test in Experiment 1a.

Figure 8. Behavioral Results from Experiment 2. A) Group Mean Accuracy and B) Group Mean Reaction Time. Note that accuracy is represented by hits for each lag and speaker condition (i.e., correct recognition on “old” trials).

Figure 9. Old/New Effect in Experiment 2. A) Scalp topography showing the significant cluster using BESA Statistics 1.0. B) Old vs. New conditions at a single electrode.

Figure 10. Voice Effect in Experiment 2. A) Scalp topography showing the significant cluster using BESA Statistics 1.0. B) Same speaker (SS) vs. different gender/different accent (DD) conditions at a single electrode. C) Bar graph representing the mean amplitude for all four voice conditions.

Figure 11. Effect of lag on the Old/New Parietal Effect in Experiment 2.
Figure 12. Short Lag in Experiment 2. A) Scalp topography showing the significant cluster using BESA Statistics 1.0. B) Same speaker (SS) vs. different gender/different accent (DD) conditions at a single electrode. C) Bar graph representing the mean amplitude for all four voice conditions.

Figure 13. Long Lag in Experiment 2. A) Scalp topography showing the significant cluster using BESA Statistics 1.0. B) Same speaker (SS) vs. different gender/different accent (DD) conditions at a single electrode.

Figure 14. Source Analysis in Experiment 2 using CLARA distributed source modeling algorithm. A) Old/New Effect. B) Voice Effect.

Figure 15. Voice Congruency in the Explicit Group, in Experiment 3.

Figure 16. Voice Congruency in the Implicit Group, in Experiment 3.

Figure 17. Voice Congruency in the Word Recognition Test, in Experiment 4.

Figure 18. Voice Congruency in the Source Recognition Test, in Experiment 4. A) Group Mean Accuracy in the Source Recognition Test. B) Group Mean False Alarm Rates in the Source Recognition Test.
Voices are arguably the most common and most important sounds to which humans are exposed. They are the cornerstone of auditory communication and are essential to social interaction. The importance of voice in auditory processing has made it the subject of much research in recent years. In terms of brain regions, functional magnetic resonance imaging (fMRI) research has shown that the right superior temporal sulcus (STS) responds specifically to voices (Belin et al., 2000; Belin et al., 2002). Belin and Zatorre (2003) found that regions of the right anterior STS are activated for representation of individual voices. These findings suggest that voice plays an important role in audition and, therefore, might represent an important context cue in memory. Throughout this dissertation, the terms “voice” and “speaker” will be used interchangeably, since the studies discussed and presented all utilize verbal material.

Spoken memory has been studied for several decades. One early question that emerged was whether voice is automatically encoded with memory for the word itself. Geiselman and Bellezza (1976) suggested that speaker voice is automatically encoded. In their study, Geiselman and Bellezza investigated the effect of adding an explicit voice recognition task to a sentence memory test. After free recall of sentences that had previously been heard, participants judged which voice (male or female) had originally spoken the sentence. Critically, some participants were told about the subsequent source test at the time of initial presentation (intentional condition), while some were not (incidental condition). It was hypothesized that if speaker voice was encoded automatically, then participants in the incidental condition would demonstrate speaker voice recognition without a corresponding decrease in sentence recall. In fact, participants in both incidental and intentional conditions were able to recognize the voice attribute (i.e., gender) above chance levels, and did so without suffering a decrease in sentence recall performance. The authors suggested that this automaticity of voice information might be due to an integral incorporation of the voice into the code, rather than a mere attachment to the memory trace.

If voice is encoded into the memory trace itself, it follows that reinstatement of voice would aid in retrieval of that memory trace. In fact, memory researchers have often sought to identify factors that facilitate retrieval of information. Voice represents an integral context cue in
auditory memory, which presumably might be manipulated as a parameter of interest to affect retrieval in memory experiments. Specifically, voice reinstatement refers to a situation in which the same voice speaks a word initially and in a subsequent memory test, whereas voice congruency describes a graded similarity among voice parameters between study and test, such that maximum congruency is equivalent to reinstatement. Experiments designed to understand auditory memory might, therefore, utilize voice reinstatement and voice congruency in an effort to investigate the nature of voice representation, as well as the neural correlates that underlie it. Several preliminary questions emerge, such as whether people remember words better if they are repeated in the same voice. Researchers began investigating such questions four decades ago.

In 1974, Craik and Kirsner looked at the role of voice reinstatement on memory for words. Using a continuous recognition paradigm, the study found that participants were better and faster at recognizing words spoken in the same voice at test as at study. Though the effect was relatively small, the conclusion was that voice information persists in memory. This suggests that context reinstatement (i.e., same speaker between study and test) facilitates verbal memory, a notion that is consistent with the encoding specificity principle (Tulving & Thomson, 1973). The encoding specificity principle states that the details of encoding dictate what is stored, which in turn dictates the cues that are effective in retrieval. In the case of word recognition, items are better remembered if the test episode is as close as possible to the study episode. That is, context congruency between study and test facilitates word memory.

This line of research entreats several questions. First, does partial voice congruency provide a partial facilitation effect, or is full voice reinstatement required to see a memory benefit? Second, does attention modulate this facilitation effect? The literature addressing this latter question has provided mixed results (Palmeri et al., 1993 vs. Goldinger, 1996). Next, how long does facilitation last? Several studies have employed a continuous recognition paradigm in an attempt to address this question. Do language skills (i.e., bilingualism) modulate the effect? Is facilitation present for implicit memory as well, or only for explicit recognition?

The present set of experiments was designed to add to our understanding of voice as a memory cue and the various factors that affect it, bearing in mind one potential confound in past studies. Though several subsequent studies confirmed and extended Craik and Kirsner’s findings (Palmeri et al., 1993; Bradlow et al., 1999; Goldinger, 1996; Goh, 2005), these studies have
traditionally used voices that varied primarily by pitch. It is imperative to validate that “voice” congruency has not been confused with “pitch” congruency. Therefore, all of the studies in the current set utilize the same stimuli – words recorded in four voices (one male accented, one female accented, one male unaccented, one female unaccented), so that pitch and/or accent may or may not be congruent between study and test. In our globalized society, it seems reasonable that accent is a very salient voice parameter of substantial ecological importance. In adding accent, it is possible to extend previous work into understanding voice congruency when multiple voice parameters are substantially varied between study and test. By using these stimuli, it is also possible to verify that suggested theoretical models of voice as a memory cue (e.g., Pisoni, 1993) are indeed supported by studies such as the Craik and Kirsner (1974) investigation, rather than being specific to pitch variations in voice. Pisoni (1993), for example, posited a parallel episodic memory system, perceptual and implicit, which encodes information about speaker voice with memory for the item itself. Information such as gender and dialect of the speaker are retained. This suggestion implies a general benefit for context congruency in word memory.

1 Word Recognition

Pisoni’s review (1993) concluded that information such as gender and dialect of the speaker is encoded automatically into long term memory, in parallel with the episodic memory trace of the word itself. Pisoni’s inference was based, in part, on a study done by Palmeri et al. (1993) as a follow-up to Craik and Kirsner’s (1974) investigation. Since only two voices were used in that initial experiment - one of each gender - a subsequent study by Palmeri et al. (1993) sought to clarify a potential gender confound and to extend the investigation into multiple talker conditions. Also using a continuous recognition paradigm, Palmeri et al. found that words presented in the same voice at study and at test were better recognized than words presented in different voices, regardless of whether those different voices were of the same gender or not. This finding showed that more detailed physical information, not only gender of the speaker, is encoded with the word. Interestingly, words were not better remembered if the voice at test was of the same gender as at study. Therefore, the study by Palmeri et al. indicated no partial benefit (of reinstating gender) in word recognition for similar voices.

The investigators (Palmeri et al., 1993) also manipulated the number of voices used in each continuous stream and found that talker variability did not affect performance. This implies that
the voice reinstatement benefit in word recognition utilized specific voice characteristics rather than being a function of load. If voice reinstatement was simply a quantitative benefit, one would expect greater performance when fewer talkers are used (because there would be less interference from competing voices). The lack of such an effect suggests some degree of voice specificity in memory.

Several subsequent studies have similarly found a benefit of voice reinstatement in word recognition. Goh (2005) concluded that voice information of individual speakers is encoded in long-term memory. Goh investigated the effect of repeating studied words in the same, different or new voices at test. Word recognition, again, was superior for the same voice condition. Bradlow et al. (1999) also found a same-voice facilitation effect for word memory, in that same-talker words were better remembered than different-talker words across lags in the continuous recognition paradigm used (which varied voice, speaking rate and amplitude – all perceptual attributes). In a series of experiments, Goldinger (1996) explored voice congruency as a function of levels of processing and type of memory investigated. In a first experiment, he devised a similarity matrix of voices based on participant responses – which classified voices on an ordinal scale based on pitch. This matrix was then used to classify congruency between study and test in the subsequent experiments. Using surprise recognition tasks, Goldinger found that participants were sensitive to detailed voice congruency between study and test; again, it was found that reinstating the same voice at test as at study improved word recognition. It should also be noted that recognition was better when words were re-presented in a different voice of the same gender than when they were re-presented in a different voice of a different gender, implying that the amount of voice congruency may impact memory for words. That is, the identical voice condition yielded the best word recognition, followed by the similar voice condition (same gender), and then, lastly, the dissimilar voice condition (different gender). This finding appears to contradict Palmeri et al.’s (1993) null partial congruency finding, but one must consider that Goldinger used a different (block) study design, with specific attentional focus at study. With respect to explicit memory, Goldinger found that deeper levels of encoding led to better overall recognition but also to a smaller voice effect. Given that the most shallow level of processing drew participants’ attention to voice characteristics (gender classification) while the deepest level of processing focused on syntax, it is not surprising that the latter yielded a smaller voice effect at test. In fact, the finding that attending to voice characteristics makes participants more sensitive to voice changes falls in line with the encoding specificity principle.
(Tulving & Thomson, 1973) and with the transfer-appropriate processing (TAP) framework put forth by Morris, Bransford and Franks (1977). The TAP framework suggests that congruency between encoding features and subsequent retrieval task yields better memory. That is, when semantic features are encoded, for example, performance on a semantic memory task will be enhanced. Appropriately, it appears that voice effects are more prominent when shallower levels of encoding are emphasized. In the Goldinger (1996) study, the benefits of voice congruency on recognition memory lasted at least a full day but less than a week.

Although there is evidence that voice congruency facilitates recognition, there are several studies that failed to observe a same-voice benefit for word memory. Naveh-Benjamin and Craik (1995; 1996) showed negative findings in this respect. Naveh-Benjamin and Craik (1995) used a study-test design that they suggest may have resulted in no same-voice advantage because of longer retention intervals (and a distractor task which was employed). Naveh-Benjamin and Craik (1996) investigated perceptual and conceptual encoding and the consequent effects on memory in aging. Considering only the young adult group for the present purposes, no same-voice benefit was found for item recognition. Schacter and Church (1992) also found no significant same-voice benefit for recognition. The authors here suggest that the lack of memory facilitation provided by voice congruency between study and test may reflect participants’ increased reliance on conceptual rather than perceptual encoding. There is some evidence (e.g., Goldinger 1996, above) of a difference between perceptual and conceptual information use in memory, such that increased reliance on perceptual (conceptual) encoding would yield increased perceptual (conceptual) sensitivity at retrieval. In Schacter and Church’s first explicit memory test, they utilized pitch discrimination as an encoding task and found no voice effect in word recognition. This finding is surprising given that a pitch discrimination task (i.e., a perceptual encoding task) was used. Furthermore, these results contrast Goldinger’s (1996) voice effect finding following a gender classification task, though it should be noted that Schacter and Church used far fewer trials than Goldinger did in his more powerful experiment. To account for their findings, Schacter and Church suggested that perceptual emphasis might need to be in relation to a particular word, rather than a non-specific acoustic judgment, in order to see voice effects on recognition memory. Their second manipulation involved a clarity-rating encoding task in which participants had to judge how well the speaker’s voices enunciated each word. Unfortunately, when investigating explicit memory in this design, a cued-recall test was
employed and it followed a distractor task and an implicit memory task, making a comparison of these findings to the current review inappropriate.

Church and Schacter (1994) subsequently looked at acoustic variation in implicit memory and in recognition. In their first experiment, there was a lack of a same-voice benefit for item recognition. Since a distractor task and the implicit memory task both occurred between study and the recognition test, this may have produced a confounding disruption in recognition memory. Pilotti and Beyer (2002) also found no same-voice benefit in a word recognition memory task. However, participants were familiarized with the voices used prior to completing the task, which adds a confound to the straight comparison of voice effects as we intend to study them.

A key consideration that emerges from a review of these investigations is participants’ allocation of attention during the experiment. In particular, must attention be allocated to the voice at study in order to procure voice effects? Secondly, might a shift in attention affect partial congruency conditions differently from the same speaker condition?

It seems crucial to underline the importance of instructions at study to subsequent memory performance. In the investigation by Palmeri et al., (1993) participants did not expressly attend to voice and results indicated a same speaker advantage but no partial same gender benefit. This would suggest that the exact voice reinstatement benefit in word recognition is automatic (see also Geiselman & Bellezza, 1976; Geiselman & Bellezza, 1977), while partial benefits associated with lesser voice congruency depend on attention allocation.

Goldinger (1996) suggested that the episodic trace established for each word depends on how attention is focused at study. In his study, Goldinger did find partial congruency effects when voice was emphasized at encoding. Similarly, Fisher and Cuervo (1983) suggested that the amount of information about speaker voice that is generally retained is not great (i.e., small effect found in Craik and Kirsner’s study, 1974) because perceptual attributes are not particularly meaningful. These authors suggested that semantic, conceptual properties are better encoded because they are more meaningful (also discussed by Goldinger, 1996). These suggestions are consistent with Craik and Lockhart’s (1972) levels-of-processing framework, which proposes that better memory is associated with deeper encoding.
Consideration of the levels of processing framework and the transfer appropriate processing framework (Morris et al., 1977) appear to be essential to determining expected results based on study design. Taken together, findings with respect to voice congruency effects in word recognition suggest that the presence and nature of such effects are highly dependent on the paradigm used. In particular, when using unfamiliar voices and minimal distractions between study and test, a same-voice reinstatement benefit can be found even when attention is not expressly allocated to voice, so long as encoding does not emphasize semantic processing. Furthermore, we expect to find voice effects on word recognition in studies where perceptual attributes of speaker voice are emphasized and/or where processing is shallow. There is some evidence that perceptual encoding should link the voice to the specific study word in order to produce voice effects (Schacter & Church, 1992), but other evidence suggests that such an explicit link is not required (e.g., Palmeri et al., 1993; Goldinger, 1996). Lastly, it appears likely that partial voice congruency effects in word recognition occur only when attention is paid to voice attributes.

2 Source Memory

In addition to word recognition, memory for the speaker of a word is often of interest. Recognition of the voice that speaks a word constitutes source memory, which has been assessed both concurrently and sequentially with word memory. That is, participants have been asked to identify whether a test word is new, old with the same voice as at study, or old with a different voice than at study (Palmeri et al, 1993). Alternatively, participants have been asked to first identify a test word as old or new and then subsequently determine whether the speaker voice for old test words is the same or different than at study (Craik & Kirsner, 1974; Naveh-Benjamin & Craik, 1996). Craik and Kirsner (1974) as well as Naveh-Benjamin and Craik (1996) both found no same-voice benefit for speaker recognition, though participants did not expect a voice test in the latter experiment despite being told to attend to voice at study. Palmeri and colleagues (1993), on the other hand, found that a benefit in source memory occurred when words were presented in the same voice at study and at test. Interestingly, the study also concluded that voice recognition was more accurate when words were re-presented in different voices of different genders rather than in different voices of the same gender. Speculatively, it may be that different voice trials in which voice characteristics are substantially varied are quite perceptually salient. Taken together, these findings indicate that performance on a source
recognition test depends greatly on explicit instructions to the participants, though not necessarily on attention to voice at study. Specifically, when participants are told that they will be tested on memory for voices, the same-voice source recognition facilitation, as previously found by Palmeri and colleagues, seems more likely.

On the other hand, Dodson, Holland and Shimamura (1998) investigated memory for source information (voice) during a divided attention task. The group found that memory for the voice (so-called specific source information) was compromised when attention was divided, while retrieval of the gender of the voice (so-called partial source information) remained intact. However, this study manipulated attention at retrieval rather than at encoding; results suggested that partial source information (about the most salient voice characteristic) can still be retrieved when sufficient resources required to retrieve specific voice information are not available.

3 Event-Related Potentials and Recognition Memory

Event-related potentials (ERPs) have often been utilized in order to determine the neural correlates involved in recognition memory. Other forms of imaging have also been employed to understand the role of source information in memory, namely functional magnetic resonance imaging (fMRI). Henson (2005) reviewed fMRI studies linking recognition memory to various parts of the medial temporal lobes, and described two important trends: the perirhinal cortex appears to subserve familiarity, and the hippocampus/parahippocampal cortex appears linked to source information and associations between distinct items. Furthermore, Henson concluded a left lateralization for verbal material. Diana, Yonelinas and Ranganath (2010) also found support for the idea that different regions of the medial temporal lobes handle differing kinds of episodic information, with the perirhinal cortex subserving item details and the hippocampus/parahippocampal cortex subserving recollection-based source retrieval. Such findings suggest more activation in the medial temporal lobes for retrieval of memories involving the binding of word with voice (e.g., in the same speaker condition utilized in the current set of studies). In addition, functional MRI research has implicated the left prefrontal cortex in source memory (see Mitchell & Johnson, 2009 for a review) and specifically in the retrieval of perceptual details (Ranganath, Johnson, & D’Esposito, 2000). Overall, it appears the medial temporal lobes and the prefrontal cortex are very important areas when considering memory for item and source.
Correspondingly, three ERP signatures have been observed in memory tasks utilizing electroencephalogram imaging: a left parietal old/new effect, an early frontal familiarity effect and a later right frontal modulation indexing post-retrieval work or decision-making processes.

3.1 Left Parietal Old/New Effect

Prior ERP studies have revealed a late positivity at parietal sites that is thought to index recollective processes (e.g., Wilding et al., 1995; Wilding & Rugg, 1996; Tendolkar et al., 2004; Nielsen-Bohlman & Knight, 1994). That is, the positivity is associated with correctly recognized old stimuli at test. Wilding, Doyle and Rugg (1995) investigated ERP correlates of word recognition and context (modality) recognition. Participants were told to remember words, but were not instructed to pay attention to study modality. Words correctly recognized as being initially presented in either the visual or auditory modality produced a more positive ERP deflection in the left parietal region compared with words that were correctly classified as new. This so-called parietal old/new effect occurred between 400-800ms and supposedly marked hippocampal activity (Wilding et al., 1995). Modality incongruency between study and test did not influence the appearance of this parietal effect, suggesting that it is amodal in nature. Presence of this effect seemed purely based on accuracy of recollection and was irrespective of familiarity. It could not index the latter because familiarity supposedly depends on fluency, which is greater when study and test modalities are constant. Furthermore, false alarms did not produce the old/new effect, further implicating recollection, not familiarity, as the function being illustrated. Importantly, an incorrect modality assessment still produced the parietal old/new effect as long as the word was correctly recognized as old. As long as participants correctly indicated that the test word was old, a parietal effect was seen. Specifically, this effect was larger when the subsequent modality judgment was also correct and smaller when the subsequent modality judgment was incorrect. Taken together, the findings imply that this ERP correlate of recollection is present for all correctly recognized old trials but also that its amplitude and duration can vary as a quantitative, graded effect, depending on how well the word is recollected.

In a subsequent study, Wilding and Rugg (1996) emphasized the importance of source memory in their instructions. Participants studied words in the auditory domain, in one of two voices. They were later tested visually, and source judgments referred to the voice in which the word
was originally presented (at study). The parietal old/new effect was again found, as described above.

A later study by Tendolkar et al (2004) suggested that the parietal old/new effect may actually be two superimposed processes. The first is a left inferior temporal old/new effect that is postulated to depend on retrieval demands and is only present for context retrieval in addition to item retrieval. In other words, it only seems to appear when items are correctly recognized as old and context is also correctly recognized as old. The second is a central parietal effect which is similar to the previously-described parietal recollection effect. This distinction might explain the graded effect seen in the left parietal old/new effect in the earlier studies discussed (Wilding et al., 1995; Wilding & Rugg, 1996). The quantitative variation may be a reflection of change in the left inferior temporal effect as a result of degree of recollection. By this logic, a gradient that distinguishes between remembered and non-remembered source characteristics can be seen in the parietal old/new effect. Questions emerge as to whether the need for attention to source, in observing this gradient, depends on the source in question. It is unclear if a gradient can be observed when voices are not particularly attended at study. Though Wilding et al.’s (1995) study suggests that attention is not required, it could be that modality is a much more salient source cue and that the attention requirement varies with the specific source being investigated.

In fact, there is evidence that attention plays a key role when considering this ERP modulation. Kuper et al. (2012) summarize recent findings and suggest that the parietal old/new effect is sensitive to perceptual changes (i.e., voice), and therefore might produce differences based on same vs. different voice congruency parameters, but only when such characteristics are attended/used in the task being employed. This implies that gradients reflective of voice congruency might be seen in the parietal old/new effect when participants are told to attend to voice at study.

3.2 Early Frontal Effect

Recollection, as indexed by the parietal old/new effect, is one of two processes known to affect recognition memory. Yonelinas (1994) reviewed a dual-process model for recognition memory, which suggests that familiarity and recollection both play a role during retrieval. Familiarity can be dissociated from recollection and studies have generally found distinct modulations for each of the two effects (e.g., Curran, 2000). While familiarity indicates the general strength of a
memory, recollection usually involves retrieval of contextual details of an episode. In addition to the left parietal old/new effect, then, a frontal positivity between 300-500ms, has often been shown to reflect familiarity (Curran, 2000; Curran & Dien, 2003). Curran (2000) found a left frontal positivity at 300-500ms post-stimulus in a visual experiment that used words of opposite plurality. After participants studied a group of words, a similar (familiar) group was tested (this consisted of the study words, made plural in the test phase by adding an “s”), with results corresponding to a positivity over left frontal scalp regions for similar as opposed to new words.

In a recognition study by Curran and Dien (2003), words were presented in either the visual or the auditory domain for study but were always tested visually. There was a mid-frontal effect that occurred between 300 and 500ms irrespective of study modality. The effect was indeed slightly larger for visually-studied items, thereby suggesting that perceptual congruency produces greater familiarity. However, the authors proposed that this FN400 is actually a more complex phenomenon and may reflect multiple processes. Specifically, there may be some spatial overlap with the N400, an effect known to index semantic processing. This overlap may explain the lack of significant modality sensitivity observed in this experiment. In addition, the FN400 and the parietal effect were not functionally dissociated here, though they were separable based on PCA analysis. Furthermore, the parietal effect, which is traditionally a left-lateralized effect, was somewhat bilateral in this study, though the left side did produce a larger positivity.

Schloerscheidt and Rugg (2004) used a slightly different procedure and obtained opposite results. They found that changing perceptual characteristics did diminish the FN400. Words and pictures were visually presented at study and either words or pictures were presented at test. Therefore, some trials resulted in stimulus consistency at study and at test and some trials switched words for pictures or vice versa. The early bilateral frontal effect was found to vary as a function of stimulus congruency and was quantified at 300-500ms and at 400-600ms post-stimulus, depending on the visual materials used. Two possible conclusions were suggested, and a different approach to considering the FN400 was outlined. Either the FN400 represents a familiarity positivity, as suggested by Curran (2000), or it represents a novelty negativity, wherein old words would produce a less negative novelty assessment. Since this early frontal effect is present for false positives (Nessler, Mecklinger & Penny, 2001), which obviously lack perceptual congruency between study and test, the characterization in terms of novelty appears to be more appropriate. More recently, it has been suggested that divergent results and
interpretations about the FN400 familiarity effect are due to differences in task instructions and stimulus properties (see Kuper et al., 2012). Presumably, both conceptual and perceptual information affect familiarity, but perceptual specificity effects become important when perceptual processing is highlighted at study (see Ecker et al., 2009) and/or at test. This suggests that characterization of the frontal familiarity modulation depends on study design. Specifically, observing an effect of voice congruency on this modulation might depend upon perceptual focus in task instructions. Regardless, the early frontal effect appears to be a strength marker for explicit memory that is likely modulated (when considering old words) by perceptual fluency.

One other study bears mention, with regards to the ERP modulation reflective of familiarity. Opitz (2010) observed a more negative deflection (at 300-450ms) for correctly recognized visual items where the context was congruent between study repetitions. It is crucial to note, here, that Opitz was looking at congruency between multiple study phases, and that all stimuli were decontextualized at test. The critical factor was whether the stimulus had been learned (i.e., studied) in one or across multiple contexts. Congruency was manipulated at study, when recognition was not probed, but not at test, when recognition was probed. Therefore, though one might expect the within context condition to produce the most positive deflection as a result of greater familiarity, the fact that stimuli were always decontextualized at test provides a pivotal aspect of the design. If something is studied in only one context, the memory will become quite specific. As such, seeing the stimuli in a different context at test will represent a relatively novel experience. On the other hand, if stimuli are studied across multiple contexts, then the memory trace will be less specific, actually leading to relatively greater familiarity when presented with the test stimulus. Indeed, Opitz suggested that familiarity (induced by repetition) plays a bigger role in recognition memory when context is not consistent between study repetitions. Using a remember/know paradigm, his hypothesis was based on a greater number of “know” responses and a more positive familiarity ERP modulation at test, for stimuli that had incongruent context conditions at study. Therefore, though Opitz also found familiarity to be associated with a frontal positivity, familiarity played a bigger role when participants recognized stimuli that had been repeated across different contexts at study. This is likely because repetitions across contexts promotes semantic rather than episodic memory, and familiarity involves a general feeling of knowing you have previously encountered something
(rather than retrieval of episodic details). Furthermore, since all stimuli were decontextualized at test, perceptual fluency at test became irrelevant to familiarity in this case.

To summarize, then, earlier research found a frontal positivity associated with familiarity (Curran, 2000; Schloerscheidt & Rugg, 2004), and Opitz found that this positive familiarity modulation emerged more so when stimuli were studied in decontextualized rather than congruent conditions. It is crucial to note, though, that Opitz’s design is fundamentally different than the others, in that context cues were never reinstated during the recognition test phase itself. The present set of experiments, however, all look at the importance of context reinstatement at test.

### 3.3 Late Right Frontal Effect

In 1996, Wilding and Rugg reported a right frontal positivity effect associated with voice judgments, which was posited to index context, or source, attributions. The authors proposed that this modulation represented operations done on what is retrieved in memory. Specifically, the authors suggested that it indexed recovery of contextual details of an episode (in this case, voice). Interestingly, this right frontal effect was still present when the study voice was incorrectly attributed. It was, however, much smaller in such cases. This was similar to their finding regarding the parietal old/new effect (see above) and suggested that recognition is a graded phenomenon. The accuracy seems to correlate with the amplitude of the old/new waves, such that greater accuracy is associated with larger old/new effects at parietal sites. This study did, however, differ from the previous work by Wilding et al. (1995) in their qualification of sources. Modality was used in the earlier study, whereas voice was used in the subsequent investigation. This likely contributed to the ERP differences between the two studies. Wilding and Rugg (1996) proposed that the same neural generators underlie item (word) and source (voice) memory and that source memory simply produces a bigger effect (greater recollection).

Johnson et al. (1998) later suggested that it was possible that the right frontal effect indexed either context retrieval or post-retrieval work, since the time interval quantified for the right frontal effect was post-response and also after the parietal effect had ceased. Either way, Johnson and colleagues concluded that the actual function underlying this effect is a frontal lobe function.
Senkfor and Van Petten (1998) claimed that, in their experiment, the late frontal effect represented a search for voice information. They investigated memory for words and speakers, specifically testing only speakers for old words in the source task (experiment 2). Since no difference was observed between correct and incorrect source trials, it was suggested that the modulation reflects the search process itself rather than the product of such a search. Based on observed ERP latencies, including a source-related positivity over the frontal scalp region starting at 800ms after word onset (time windows were arbitrarily chosen and the latest time window partition began at 800ms, but it should be noted that visual inspection suggested that the positivity actually began slightly earlier), they concluded that voice information was retrieved after word information, suggesting a hierarchical system. This contrasts with Pisoni’s parallel system (Pisoni, 1993), which implies the possibility of simultaneous retrieval. However, it should be noted that automatic use of voice congruency could precede actual retrieval of voice information. For example, same-voice facilitation might improve memory without necessitating conscious awareness of the similarity; subsequent retrieval of voice characteristics might involve additional mechanisms. It should also be noted that the study by Senkfor and Van Petten (1998) emphasized semantic encoding before the source recognition task. It might be that placing the emphasis on perceptual encoding could facilitate an earlier, perhaps more automatic retrieval of speaker in the source task (where, since participants knew all words were old at test, source retrieval need not follow item retrieval). Indeed, research has shown that voice congruency effects are more pronounced when perceptual encoding is emphasized (Goldinger, 1996).

Though the right frontal effect has typically been associated with post-retrieval, recent work by Hayama, Johnson and Rugg (2008) has indicated that it might actually index decision/judgment processes. Specifically, Hayama et al. found the right frontal effect whenever participants made a semantic judgment about an item, regardless of whether it had previously been studied or it was a new item. That is, the effect could be elucidated during the first exposure to an item, rather than as a result of processing of a memory. If so, then item recognition may not be a necessary precursor for the effect to be observed. Instead, the effect may index processes involved in source judgments, as long as such judgments are the task at hand. If this analysis is correct, we would not expect a right frontal effect in word recognition tests that do not require a source judgment, but would expect one in all source recognition tests (regardless of whether word recognition is tested or not).
3.4 Three Functionally Distinct ERP Modulations

Allan, Wilding and Rugg (1998) suggested a dissociation between the left parietal and the right frontal effects. The first appears to be a task-independent retrieval effect and seems to be present regardless of the explicit memory task used. The left parietal old/new effect is described by Tendolkar et al. (2004) as a parietal positivity that onsets at 400ms and lasts 400-600ms, and has been shown to overlap the N2/P3 transition for auditory stimuli (Kayser et al., 2003). The right frontal post-retrieval (though see Senkfor & Van Petten, 1998) effect appears to be task-dependent, wherein the nature and amount of post-retrieval depends on the explicit memory task employed. Specifically, the right frontal effect indexes post-retrieval processing in recognition tasks (Wilding & Rugg 1996 and 1997a, as cited by Allan et al., 1998). On the other hand, no effect is observed in associative recall tasks (Rugg 1996b, as cited by Allan et al., 1998) and a symmetrical deflection indexes post-retrieval processing in cued recall tasks (Allan et al 1996, as cited by Allan et al., 1998). There is some indication that the post-retrieval effect is also prolonged for auditory stimuli (Kayser et al, 2003).

Schloerscheidt and Rugg (2004) compared the nature of the three aforementioned ERP modulations, using both pictures and names as stimuli, and thus making both within-format (e.g., name at study and name at test) and across-format (e.g., name at study and picture at test) trials possible. When using verbal test stimuli, the old/new parietal effect was the same for both within- and across-format hits. This indicated the irrelevance of surface form of the encoding stimuli, for the parietal effect, when names are tested. This also provided evidence for a functional dissociation between the parietal effect and the earlier frontal effect, the latter showing larger amplitudes when study and test stimuli were congruent (i.e., perceptually fluent). The later right frontal old/new effect, on the other hand, seemed to be sensitive to test stimulus, only appearing when words were the test stimuli. It may be that this reflected the extra work required when words were tested because verbal cues are weaker than pictorial ones. Schloerscheidt and Rugg concluded, like other authors, that memory retrieval is a multi-component process.

In 2012, Evans and Wilding used a remember/know judgment to investigate the potential independence of recollection and familiarity in recognition tests. Using magnetoencephalogram (MEG) imaging, this study found that the recollection index was larger for remember judgments while the familiarity index was larger for know judgments. These findings provide evidence
that recollection and familiarity make separate contributions to recognition tests (see also Eichenbaum et al., 2007; Diana et al., 2007).

It appears that the ERP modulations expected in explicit memory tests can be dissociated from one another, which indicates that they should be analyzed independently of each other, in such experiments. No dependence between these imaging modulations should be automatically assumed without clear justification for it.

4 Primary vs. Secondary Memory

Since many of the studies that look at the role of voice in recognition memory employ continuous recognition paradigms, a parameter of interest is duration of potential congruency effects. Are potential voice congruency facilitation effects time-sensitive? How long do they last? Are the processes involved in recognition different at various timeframes and can this be examined using ERPs?

In general, recognition memory decays over time (e.g., Goldinger, 1996). In terms of the ERP trace, the amplitude of the late positive parietal effect has been found to decrease with increasing lag (Nielsen-Bohlman & Knight, 1994). One question that emerges is whether there is a cut-off time, so to speak, after which the neural processes involved in recognition change? Prior research suggested that a distinction occurs within 15 seconds of sensory encoding (see Tulving & Schacter, 1990), when working memory and long-term memory interact. Poon and Fozard (1980) used a continuous recognition paradigm and suggested that a transitional phase took place between lags corresponding to 12 and 24 seconds between study and test. Poon and Fozard studied word repetition effects in order to extend previous work by Waugh and Norman (1965), which had suggested a distinction between so-called primary and secondary memory. Primary memory is thought to be a limited capacity, fast-decaying system, while secondary memory is more permanent and involves slower decay. Primary memory was originally defined by James (1890) to correspond to the store of items within one’s consciousness. Today, we refer to the system that holds and manipulates these items in active awareness as the processes of working memory, thereby closely linking the terms primary memory and working memory. Poon and Fozard’s work pointed to the transition between primary and secondary memory occurring after 12 seconds. Therefore, recognition memory at very short lags likely taps into mostly primary memory processes, while longer lags are likely dependent on secondary memory
processes. Furthermore, prior research has found that primary memory utilizes mainly shallow encoding while secondary memory is more linked with semantic encoding (Craik & Levy, 1970). In experiments employing voice manipulations, then, it seems plausible that voice congruency and voice reinstatement effects would be more pronounced in primary memory processes.

5 Implicit Memory

When looking at voice congruency effects, a related but separate question is whether voice is used differently or to a different extent by implicit memory. Implicit memory tests, for example, have been found to be more dependent on perceptual characteristics changing between study and test (see Craik et al., 1994). Since voice information is perceptual in nature, it follows that voice congruency probably has a bigger effect on implicit memory than explicit memory. Questions that arise about voice congruency in implicit memory thus include: Is voice congruency utilized in the same way as in recognition memory? Do potential congruency benefits last longer? Does partial congruency provide a greater facilitation?

In 1984, Jackson and Morton looked at the potential effect of voice congruency on auditory priming. In the study phase, participants either read words or heard them spoken in one of two voices, while making a living/non-living judgment about each one. At test, words were embedded in white noise. Results indicated a priming effect but no benefit of voice congruency between study and test. However, a semantic judgment was used at encoding in this study. As noted later by Schacter and Church (1992), and similar to explicit memory, it seems reasonable that voice effects might be observed in implicit memory if participants pay attention to perceptual (i.e., acoustic) features at study.

In 1992, Schacter and Church conducted a series of experiments looking at both implicit and explicit memory as a function of voice congruency. It had been suggested that implicit memory effects are the result of a pre-semantic perceptual representation system (PRS; see Schacter & Church 1992), which drives priming via storage of superficial information about encoded items. Results of these experiments indicated that voice congruency had no effect on priming when white noise was used (independent of encoding task), but that an effect was observed in a word–stem completion task where words were spoken clearly. It was suggested that the white noise disrupted processing of the PRS in the right hemisphere. Schacter and Church did find voice
reinstatement effects using the auditory stem completion task in an implicit test, but under somewhat unexpected conditions: The same-voice advantage at test was significant when participants rated pleasantness at encoding but not when rating pitch. Perceptual sensitivity at test seems less likely following a pleasantness task meant to promote semantic encoding than following a perceptual encoding task (such as pitch rating). However, the authors suggested a possible confound in that participants had trouble dissociating pleasantness of the word from pleasantness of the voice. In addition, the authors reasoned that a perceptual task would yield voice effects if voice attributes are stressed in relation to the word itself, at encoding. To test this, in their second experiment utilizing auditory word stem completion in the clear (i.e., not in noise), Schacter and Church sought to more finely separate semantic and perceptual properties in the encoding tasks used. Participants performed a meaning rating task as the semantic encoding task and a clarity rating task (judging how clearly voices enunciated each word) as the perceptual encoding task. Results indicated that voice reinstatement improved performance on the implicit test following both encoding tasks, but performance was not greater following the clarity rating encoding task. In their discussion, the authors speculated that voice effects might not be affected by encoding task because encoding of voice is a compulsory component of speech perception (Goldinger, Pisoni, & Logan, 1991, as cited by Schacter & Church, 1992).

Church & Schacter (1994) then used the same clarity rating task in an implicit task that required identification of low-pass filtered words at test. Instead of being embedded in white noise, words were degraded at test using a low-pass filter in an effort to determine whether voice effects on priming can be found in an auditory identification task. In fact, voice reinstatement from study to test resulted in enhanced priming. Additionally, the experiments revealed that voice effects were larger for implicit memory tasks than for explicit word recognition memory tasks.

In his series of experiments, Goldinger (1996) also looked at implicit memory. In order to address the issue of white noise producing null results, Goldinger used words presented in noise both at study and at test (as opposed to words clearly studied that were tested in noise, as in Jackson & Morton, 1984; Schacter & Church, 1992). It was hypothesized that an increase in perceptual fluency between study and test should presumably bolster voice effects in implicit memory. In his second experiment, words presented in white noise were equally likely to be spoken in the same or a different voice between study and test, and participants always had up
to 20 seconds to type the word presented (in both study and test phases). Results indicated that the same-voice advantage was prominent in the perceptual identification task for at least a week.

In a subsequent study, Sheffert (1998) directly tested Goldinger’s (1996) suggestion that voice effects can be found in noise when the same stimulus/mask combination is used at both study and test. Indeed, by looking at words in noise as well as words degraded by a filter, Sheffert found evidence for Goldinger’s supposition. The finding of voice congruency effects using white noise contradicted Schacter and Church’s (1992) notion that the PRS system is disrupted by white noise. Sheffert (1996) had previously investigated implicit and explicit memory for words and voices in an attempt to elucidate whether voice and word information are stored together or separately. Not surprisingly, voice congruency between study and test resulted in increased priming on word identification tasks. This effect was boosted when the type of processing engaged at study mirrored that engaged at test. Sheffert suggested that word and voice information are represented together in an episodic system, and that the primary determinant of voice change effects is the similarity of processing at study and at test (i.e., transfer-appropriate processing), irrespective of which type of memory the task primarily taps into. As such, perceptual fluency and similar processing at study and at test appear to be the key factors contributing to voice congruency effects in word identification tasks.

Meehan and Pilotti (1996) looked at the importance of voice familiarity, talker variability and voice congruency on the speed of response to old words (i.e., priming). Importantly, talker variability and voice congruency were investigated in a between-subjects design. The task used here was quite different than in the other studies described. Participants were asked to identify if the first phoneme of each word was “b”, both at study and at test. Results indicated that increased talker variability (i.e., a greater number of voices used) had a detrimental effect on priming, which was partially mitigated by equating for familiarity with all voices. That is, if participants were given extra trials to familiarize themselves with the larger number of voices in the multi-talker scenarios, priming occurred in all but the most acoustically variable scenario. When familiarity was not taken into account, priming occurred when only one or two voices were used but not when three or four were employed. In terms of voice congruency, trials of same voice conditions were compared with trials of different gender voice conditions. Voice changes between study and test had a detrimental effect on priming, again indicating that voice information is essential to priming. The authors suggested that auditory priming was dependent
on memory for sub-word components of stimuli. In sum, the findings supported the idea that both talker variability and voice changes are potentially detrimental to priming when a phoneme monitoring task is used. It was suggested that this task oriented participants to sub-word processes, while a task such as word identification focuses on the whole word.

Pilotti et al. (2000) highlighted the importance of the type of task used, on whether voice effects are present or not. Here, we limit our discussion to the conditions referring to voice changes and auditory encoding only. In a direct comparison, Pilotti et al. investigated potential voice effects in four implicit memory tasks and two explicit memory tasks so that the only variable was task employed at test. All participants performed a semantic encoding task while hearing words in the clear, and then performed one of the tests. Results indicated that voice changes between study and test were found to lower performance on auditory identification tasks (in noise and when low-pass filtered) and word recognition tasks, but not on completion tasks such as stem and fragment completion or cued recall. The results support Church and Schacter (1994) but contradict Schacter and Church (1992) and Jackson and Morton (1984). While it should be noted that Pilotti et al. have a more powerful experiment, informed by a greater number of trials, differences might also be the result of speakers used to record the stimuli. Specifically, Pilotti et al. used only two voices – one male and one female – as opposed to Schacter and Church, who used six voices to record stimuli. As such, it may be that voice information is easier to discern and therefore less salient in the experiment by Pilotti et al., and/or that familiarity became a factor in Pilotti et al.’s investigation, given the few voices and the many trials that made up the experimental procedure. Taken together, it becomes apparent that the effect of voice congruency on auditory implicit tasks is largely dependent on experimental design.

One major point of interest in auditory identification tasks is the type of noise or filtering used. When considering multi-talker babble, for example, a problem occurs. Pichora-Fuller, Schneider and Daneman (1994) suggested that the extra resources allocated to deciphering words in babble result in working memory deficiencies. Similarly, in her dissertation, Heinrich (2006) posited that encoding is limited in babble scenarios due to additional required processing of the words. Interestingly, the allocation of extra resources might also provide greater processing of words embedded in babble. As such, it seems possible that priming effects might be observed for words embedded in babble, given the right combination of study and test
tasks/processing. Therefore, the paradigm used and the type of noise or degradation employed plays a crucial role in investigating voice effects in implicit memory.

6 ERPs and Implicit Memory

In terms of the ERP trace, Rugg et al. (1998) reported a dissociation between the neural correlates that underlie implicit and explicit memory systems. Using visual stimuli, a marker of implicit memory was found by comparing new words with old words incorrectly judged to be new. Theoretically, a difference between these two conditions was thought to index memory without awareness (i.e., implicit memory). This marker was found over parietal sites, at 300-500ms, where old words always produced a more positive deflection than new words, regardless of memory performance. This was differentiated from the frontal familiarity effect at the same latency, which was found here to distinguish between old words correctly recognized and old words mistakenly thought to be new. However, it should be noted that a recent review has called into question the validity of inferring an electrophysiological component of implicit memory when a behavioral implicit memory measure is lacking (Voss & Paller, 2008).

Nevertheless, Henson et al. (2005) used functional magnetic resonance imaging (fMRI) to isolate hemodynamic counterparts to the modulations found using Rugg et al.'s (1998) ERP method. Left inferior parietal activation was associated with recollection while left anterior medial-temporal deactivation corresponded to familiarity. Unfortunately, correlates of implicit memory could not be determined.

Later research suggested that the ERP component that indexes implicit perceptual memory is an N350 modulation (see Schendan & Kutas, 2007). Kuper et al. (2012) used an indirect visual memory task whereby participants made a semantic classification judgment both at study and at test. Findings indicated a repetition priming effect for exact same exemplars, which elicited centro-parietal positivities between 300-700ms post-stimulus. It should be noted, however, that behavioral results indicated near-ceiling performance on the task. Nevertheless, this modulation appeared to coincide with the N400 effect commonly found over central sites that is thought to index implicit semantic priming. The latter has been dissociated from the FN400 familiarity effect (Bridger et al., 2012).
Taken together, these findings provide a departure point for ERP studies of implicit memory. The sensitivity (if any) to voice congruency of this relatively early parietal positivity remains unknown.

7 Language Skills

When considering the potential benefits of voice congruency in memory for words, an obvious issue that arises is the importance of voice familiarity. While the scope of this work does not extend to specific voice familiarity, the question of accent familiarity (via language skills) was a point of interest in the present set of experiments. Does bilingualism confer a greater or different type of benefit, if any at all, when voices are congruent between study and test?

A few studies have looked at language effects on auditory processing. Thompson (1987) examined the effect of language bilingualism on voice identification. Recordings by bilingual speakers resulted in stimuli that were in the native language of the monolingual participants, in an accented voice in the monolingual’s native language, and in another language. Voices were better identified if they were speaking the participant’s native language and less so if an accented voice was used. Thompson proposed that language is represented schematically, and that schemas for specific words involve the sounds of words spoken in the person’s own accent. By this account, words presented in foreign accents are farther from a monolingual’s schema for that word, making voice recognition more difficult.

Goggin et al. (1991) similarly found that language familiarity is important to voice identification. This work compared scenarios using several different languages, and showed that voice recognition suffers when language familiarity is reduced. It was suggested that, since words and speakers have a reciprocal relationship (Nearey, 1989, as cited by Goggin et al., 1991), disruption caused by language unfamiliarity might have a negative impact on voice recognition. The authors also speculate that attention might play a role, in that resources that might otherwise be allotted to voice recognition are depleted when an utterance is spoken in a non-standard way.

Lagrou et al. (2011) looked at performance on a lexical decision task (participants decided if they heard a non-word or a word) as a function of bilingualism. Results indicated that, for bilinguals, recognition that a word is in fact a word was faster for non-accented speakers. This
implied that there was a better match to the lexical representation of that word when non-accented speakers said it.

To our knowledge, however, studies directly assessing the impact of accent congruency on word and voice memory in monolinguals vs. bilinguals are lacking. Nevertheless, considerable research concerning bilingualism has been undertaken by Bialystok and colleagues (Bialystok, 2009; Wodniecka et al., 2010). Evidence indicates that bilinguals generally perform better on memory tasks that utilize executive control (see Bialystok, 2009). Not surprisingly, recollection appears to be affected by bilingualism, though familiarity is not (Wodniecka et al., 2010). Therefore, it may be that bilinguals perform better when accented voices are used because they are better able to focus attention to pronunciation. If so, one would expect greater performance on recognition tasks across all voice congruency conditions, for all bilinguals compared to monolinguals. On the other hand, or perhaps in addition, when considering bilinguals of the same language as the accented voices utilized in a given stimulus set, less attention may be required on their part due to incorporation of that accent into their schemas in that language.

8 Summary of Circumstances in which the Voice Congruency effect on Word Memory will occur

Clearly, a voice congruency facilitation effect on word memory occurs only under specific conditions. With respect to recognition memory, it appears that while specific voice reinstatement (i.e., same speaker) facilitates recognition memory even without attention to voice at study, a partial benefit of similar voices between study and test is less clear-cut. In terms of explicit memory experiments utilizing unfamiliar voices, encoding methods appear to play a crucial role. Voice congruency effects have been found when voice is specifically attended at study (i.e., when relatively shallow, perceptual encoding takes place), though some evidence (Schacter & Church, 1992) suggests that such a benefit specifically requires a focus on voice attributes in relation to the words being spoken.

Electrophysiological studies have found three modulations of interest when considering the effect of voice congruency on recognition memory. First, a parietal old/new recollection effect that distinguishes between correctly identified old words and correctly identified new words whose amplitude may be affected by allocation of attention to source attributes such as voice. Second, there is an early frontal familiarity effect that may be sensitive to perceptual fluency for
old words. Third, a late right frontal modulation appears to index either post-retrieval work (potentially in the form of voice search) or decision-making processes, but a characterization of this effect based on voice congruency in correct old trials only is currently lacking. To our knowledge, the only investigation that looked at even some of these modulations in purely auditory experiments was done by Senkfor and Van Petten (1998), but they emphasized semantic encoding and only used a male and a female voice. Thus, further purely auditory research is required to look at the more fine-grained effects of voice congruency on the three aforementioned ERP modulations, especially in situations where they might be bolstered by highlighting perceptual attributes at study.

With respect to implicit memory, voice congruency effects appear to depend on the task employed. Using a word identification task, perceptual similarity between study and test conditions confers a benefit of voice congruency. Emphasis on similar processing at study and at test might also improve voice congruency effects. Though it seems even more likely to find a similarity gradient reflective of voice congruency (rather than just voice reinstatement) in implicit memory tests rather than explicit memory tests, due to a greater reliance on perceptual attributes in the former, a direct comparison between implicit and explicit memory tests using more than two voices is currently lacking. Such an investigation would provide a more fine-grained analysis of voice congruency effects, reflective of more than just pitch differences, across multiple memory systems.
Chapter 2
Experiments

The present set of experiments is designed to further our understanding of voice as a memory cue. First, we investigate the effect of voice congruency on word recognition and on voice recognition when voices are easily distinguishable by gender and/or by accent, and participants are told to attend to both word and voice at study. The neural markers underlying potential voice effects are studied using ERPs. In addition, one possible caveat to this work is that the same-speaker conditions employ the exact same acoustic-word combination rather than a word said twice by the same speaker on separate occasions. Since this introduces the possibility that superior performance in the same-speaker condition involves an extra acoustic identification that is beneficial for recognition, a behavioral study using new stimuli is used to rule out or confirm this confounder.

Then, in a continuous-paradigm word recognition test, we seek to confirm and extend the findings of the first experiment to a different paradigm. Importantly, the design of the second experiment allows us to extend characterization of potential voice effects by looking at lag and explicit attentional focus as potential factors. Specifically, in this second experiment, participants are not expressly told to attend to voice, and are not tested on voice recognition. Once again, neural correlates are investigated using ERPs.

A (pilot) experiment next explores the role of voice congruency in multiple forms of memory. In order to compare the role of voice in explicit vs. implicit memory, a between-groups experiment that uses the same study phase followed by either a word recognition task or a word-in-noise identification task is reported.

Lastly, a behavioral study then examines whether voice facilitation effects vary as a function of linguistic background. Using the same task as in the first experiment, we compare a group of monolinguals, bilinguals of the same accent as the recorded stimuli and bilinguals of a different accent than the recorded stimuli. Since this study is the first, to our knowledge, to directly assess potential benefits of voice congruency in word and source recognition for different language groups, our investigation is exploratory and we can make few a priori hypotheses.
9 General Hypotheses

We expect behavioral same-speaker facilitation effects in all memory tests, regardless of type or of attentional focus at study. In recognition memory, we predict a gradient reflective of voice congruency only when attention is allocated to voice at study, though this may not be a requirement to observe a similar gradient in an implicit memory test. In terms of imaging, presence and characterization of the three potential recognition memory modulations would depend on task and attention. Namely, the parietal old/new effect will be present for all correctly recognized old words and a potential voice congruency gradient might also be attention-dependent. The frontal familiarity deflection would reflect voice congruency when considering only correctly recognized old words and that the later right frontal modulation would be present only for voice recognition tests (when a decision about context would have to be made). We cannot make specific a priori predictions regarding the difference between gender congruency and accent congruency as partial voice congruency conditions.

10 General Methods

10.1 Participants

Each experiment involves healthy volunteers aged between 18 and 35 years (except in Experiment 4, where participants were students in a first-year undergraduate psychology course and could, therefore, be 17). Participants were recruited from the subject pools at the University of Toronto and the Rotman Research Institute. All participants had normal hearing – either with audiograms indicating pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000Hz (both ears), or as self-reported. Only participants whose primary language was English were included in the studies. Further information concerning musical training and language expertise was asked of each participant. All participants signed a consent form approved by Baycrest Hospital.

10.2 Stimuli

The stimuli were the same across all experiments, except in the control investigation (Experiment 1a). The word list consisted of 336 high-frequency (30+ from Kucera & Francis, 1967), two-syllable nouns, taken from the MRC psycholinguistic database (Wilson, 1988). All
words were recorded by four speakers – one native-English female, one native-English male, one Chinese-accented female and one Chinese-accented male – in continuous streams, at 32000 sampling rate, mono, with 16-bit resolution. The native English-speaking female was 31 years old at the time of recording and the native English-speaking male was 27 years old. The Chinese-accented female was 57 years old and had learned English at 27 years of age. The Chinese-accented male was 44 years old and had learned English at 19 years of age. With a Shure KSM44 microphone and a USBPre preamplifier with digitizer, speakers recorded the words using Adobe Audition 1.5 on a Dell laptop. All words were then spliced into individual files, using a Matlab (version 5.3) script. The Matlab script used to splice the words specified thresholds, duration and pre-stimulus intervals, which could be varied as necessary for different groups of words. Using a batch process in Adobe Audition 1.5, all wavefiles (words) were analyzed for loudness and then normalized to the average level of all the words. The normalization used an equal loudness contour, which places emphasis on mid-frequencies rather than high/low frequencies to which the human ear is less sensitive. No clipping was permissible. Words were inspected for background noise, clipping, duration (1 second) and clarity. Editing was done as appropriate, either manually or using batch files. To check for intelligibility, two young adults with English as a first language each listened to three blocks of words and indicated where word intelligibility would be an issue. Reshuffling and processing of new words occurred as necessary to ensure that potential participants (young adults with English as a first language) would have no trouble understanding the words spoken. Once all words were judged adequate, they were again normalized to their average loudness, at -32.44dB.

### 10.3 Electrophysiological Recording and Analysis

For Experiments 1 and 2, electroencephalogram (EEG) recordings were obtained. The EEG was digitized continuously (sampling rate 500 Hz; bandpass of 0.05–100 Hz) during the study phase and the test phases using NeuroScan Synamps2 (Compumedics, El Paso, TX, USA). The ERPs were sampled at 64 scalp locations that include electrodes placed at the outer canthi and at the inferior orbits to monitor eye movements. During recording, all electrodes were referenced to the Cz electrode; for off-line data analysis, they were re-referenced to an average reference. The analysis epoch consisted of 200 ms of pre-stimulus activity and 1300 ms of post-stimulus activity.
For each participant, a set of ocular movements was obtained prior to and after the experiment (Picton et al. 2000). A Matlab program was used to calculate averaged eye movements for both lateral and vertical eye movements as well as for eye-blinks. A principal component analysis of these averaged recordings provided a set of components that best explained the eye movements. The scalp projections of these components were then subtracted from the experimental ERPs to minimize ocular contamination, using Brain Electrical Source Analysis (BESA 5.2).

After correcting for eye movements, all experimental files for each participant were then scanned for artifacts; epochs including deflections exceeding 130 µV were marked and excluded from the analysis. The remaining epochs were averaged according to electrode position and trial type, using BESA 5.2. Each average was baseline-corrected with respect to the pre-stimulus interval and digitally low-pass filtered at 20 Hz (zero phase, 24 dB/oct), using BESA software.
Chapter 3
Experiment 1: Voice Congruency in Word and Source Memory using a Block Design

In our first experiment [Campeanu, S., Craik, F.I.M., & Alain, C. (2013). Voice congruency facilitates word recognition. *PLoS ONE, 8*(3), e58778.], we sought to extend previous work by assessing how two voice parameters (two aspects of a source) interact in both word and speaker memory. We investigated the effects of context congruency on word and source recognition, wherein context might be reinstated at test, might vary in either gender or accent while remaining congruent on the other voice parameter or might largely vary on both voice parameters. Based on previous work, we predicted a same-speaker (reinstatement) advantage in both the word recognition test and the source (speaker) memory test, indicating either specificity for voice-word binding or a quantitative benefit of context congruency at test. Given that attention was allocated to voice in the study phase of the experiment, we did expect a partial congruency effect to manifest behaviorally. Though a partial congruency facilitation effect would fall in line with the hypothesis of a graded voice congruency benefit, we could make no prediction about pitch versus accent as a congruency cue.

Three ERP modulations were also expected. First, we expected a parietal old/new effect for all correctly recognized old words in the word recognition test, and for all words in the source recognition test. Since attention to voice was emphasized at study, we anticipated that the old/new effect would be larger when test words were spoken by the same voice as at study (reinstatement) than when spoken by a different voice and that the level of congruency among the different voice conditions would be reflected as partial voice congruency at this modulation. In addition, since the present study used the same voices for old and new words, the analysis of familiarity across old and new words was not straightforward. The primary scope of this investigation was the distinction between varying amounts of voice congruency in memory. As such, we restricted the “familiarity” analysis to old words only, because there can be no context congruency for new words. We predicted that item, or word, familiarity would be modulated by voice congruency and that this would be shown as a modulation in the early frontal deflection, which would be sensitive to perceptual fluency in old words only. Lastly, we predicted a late (right) frontal deflection in the source recognition task only and at a somewhat earlier latency than the one found by Senkfor and Van Petten (1998), since attention was allocated to voice at
study. The following description of Experiment 1 and its results are largely taken from our published work (Campeanu et al., 2013) with the exception of the frontal familiarity results and discussion, which are taken from unpublished work.

11 Method

11.1 Participants

Twenty-two participants provided written informed consent according to the guidelines set out by the Baycrest Centre and the University of Toronto. One participant did not complete the experiment because their pure-tone threshold fell out of the normal range in the left ear. EEG data from two participants were excluded because of excessive muscle artifacts and/or eye movements during recording. Lastly, three participants were excluded because of insufficient number of trials per condition. The final sample of 16 participants comprised four males and twelve females aged between 19 and 32 (M = 25 ± 4.3 years); they all had English as their first language, were right-handed and had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000Hz (both ears).

11.2 Design

The experiment was programmed using Presentation software (version 11.0). Words (as *wav files) were converted to analogue using a computer soundcard (16 bit, stereo, with a sampling rate of 44100 Hz). The analogue output was fed into a 10 kHz filter (Tucker Davis Technologies (TDT, Alachua, FL), FT6-2), and then to a GSI 61 audiometer. Stimuli were presented binaurally through insert earphones (EAR-TONE 3a), at 70dB SPL. The words were divided into six independent blocks and balanced such that words beginning with each letter were equally distributed into the six blocks. The six blocks, which were identical in structure, consisted of a study phase of 32 words, plus four buffer words on each end, and two recognition tests. There were eight study words in each of the four voices. The word recognition task used 16 studied words and 16 new words and asked participants to give an old/new judgment for each word. In terms of voices, four words that were originally spoken in each voice at study were presented during the word recognition task. Participants were asked to judge a word as “old” if they had heard it earlier, irrespective of whether the test word was presented in the same or a different voice as at study. However, the number of words in each voice was also balanced.
in each word recognition test. The second task (source recognition) used the other 16 studied words only, and asked participants for a yes/no answer as to whether the speaker for that word at test was the same as during study. Again, each source recognition test was balanced for the number of words presented in each voice. In each block, there were four same speaker conditions for each test, one for each voice. This gave a total of 24 same speaker trials per participant. There were also four different gender/same accent and four same gender/different accent conditions, as well as four different gender/different accent conditions, in each test of each block. This made each voice congruency condition possible, with each voice, in each test and each block. Participants were given mandatory two-minute breaks between blocks.

Each scenario began with a warning sound (1.5 seconds duration) to alert participants that the word list was about to begin. The first word was presented two seconds after the cue, and subsequently the stimulus onset asynchrony (SOA) between words (in the study phase) was three seconds. In the encoding phase, participants were instructed to pay attention to both the words and the speakers of each word. They were told that there would be two subsequent tests, one for word recognition and another for source recognition. Since the word and source recognition tasks used different words, counterbalancing order of the tests was not needed. Instead, we felt that it was important to maintain a consistent test order so that duration between study and each test was the same for all participants. Therefore, we presented the word recognition test first during each block. All participants received the same test lists. During the test phases, participants responded to the old/new and same speaker/different speaker judgments in a self-paced manner. There was no visual instruction screen in either the study or the test phases; stimuli were strictly auditory. Time between the end of the study phase and the beginning of the word recognition test was approximately 30s – 1min. Time between the end of the word recognition test and the beginning of the source recognition test was only a few seconds, but again, studied words did not overlap between tests. Responses were made on a keyboard using the right hand (the dominant hand since all participants were right-handed), but we did not control for which fingers were used.

11.3 Behavioral Analysis

For the word recognition task, a sensitivity measure was calculated for the four possible voice conditions – same speaker, same gender/different accent, different gender/same accent, different gender/different accent – by subtracting false alarms from hit rates. The false alarm rate used
was a common rate based on the proportion of new words incorrectly judged as old, for each participant.

For the source memory test, a hit rate was used for the same speaker condition. This hit rate was compared to the chance level for the same speaker condition (0.25) in a one-sample t-test. For the three different speaker conditions in the source memory test correct responses are represented by correct rejections, given that participants were asked to simply respond “same” or “different” rather than distinguishing between the different speaker conditions. For this reason false alarm rates were compared for these three conditions.

11.4 Electrophysiological Analysis

As previously mentioned, files were collapsed across the six blocks for each participant; weighted combinations for each condition per participant were made. Group average files were then made by combining all participants’ blocks in an unweighted fashion. The parietal old/new effect has been consistently found in the literature to index recollection when old/new judgments are required, as is the case in the present study. Tendolkar et al. (2004) described that “[t]he first old/new effect identified during tests of recognition memory onsets approximately 400 ms post-stimulus, typically lasts around 400–600 ms, and is largest in amplitude over left temporo-parietal scalp electrodes” (p.236). To both correspond with that expected latency range and to surround the observed peak of the deflection in the present study, the modulation was quantified here at the 700-900ms interval, over P5, P3, PO3 and P1 on the left side and P6, P4, PO4 and P2 on the right side. The early frontal familiarity deflection was quantified between 300-500ms, bilaterally, over F5, F3 and F1 electrodes, as well as over F6, F4 and F2 electrodes, in the word test. Electrodes were chosen to correspond to scalp areas analyzed in previous, relevant studies (Wilding et al., 1995; Wilding & Rugg, 1996; Tendolkar et al., 2004; Johnson et al., 1998; Curran, 2000; Curran & Dien, 2003). In all cases, mean amplitude measurements were exported and analyzed using repeated measure ANOVAs with trial type as the within-subject factor.

The right frontal effect is traditionally described for studies where a source judgment follows an old/new judgment. Since this study used a different methodology, we could not predict a clear time window based on the literature. As such, the right frontal deflection predicted for the source test was quantified using more objective pairwise permutation tests, using BESA
Statistics 1.0. Comparisons were made between the same speaker voice condition and the different speaker voice condition, regardless of response success, from 100ms to 1300ms post-stimulus over the entire scalp. This two-stage analysis first computes a series of t-tests that compare the ERP amplitude between the two conditions at every time point. This identifies clusters (in space and in time) when the ERPs differ between the conditions. In the second stage of this analysis, permutation tests are performed on these clusters. The permutation test uses a bootstrapping technique to determine the probability values for differences between conditions in each cluster. The final probability value computed is based on the proportion of permutations that are significant for each cluster, and corrects for multiple comparisons. In the current analysis, we used an overall cluster alpha of 0.05, one thousand permutations and electrode clusters defined using a channel distance of 4 cm, which resulted in an average of 3.125 neighbors per channel. The literature points to a frontal modulation before 800ms (see Senkfor & Van Petten, 1998). Based on this, we examined significance in the (right) frontal area between 100ms-900ms.

In the word recognition test, the parietal old/new effect and the frontal familiarity effect were measured using correct trials only; hence, the mean number of correct trials used in each condition was 17.8 (range 11-22, s.d. 2.63). In the source recognition test, the mean number of correct trials used in the same speaker condition was 17.1 (range 14-20, s.d. 2.2). Since the different gender/same accent, same gender/different accent, and different gender/different accent conditions all resulted in the same response, “different voice”, in the source recognition test, they were combined together (mean number of trials per condition = 13.5; range 5-21, s.d. 3.8).

12 Results

12.1 Behavioral Results

Figure 1a shows the group mean accuracy for all four conditions in the word recognition test. The ANOVA yielded a main effect of voice congruency ($F(3,45) = 11.41, p < 0.001, \eta^2 = 0.43$; linear trend $F(1,15) = 24.69, p < 0.001, \eta^2 = 0.62$), with pairwise comparisons indicating that performance in the same speaker condition was superior to all three different speaker conditions ($p < 0.05$). In addition, performance in the different gender/same accent condition was significantly better than in the same gender/different accent condition and the different
gender/different accent condition ($p < 0.05$). Consistent with this, a preliminary analysis using gender and accent as independent factors found only a main effect of accent in the word test ($F(1,60) = 11.74, p = 0.001, \eta^2 = 0.164$).

For the source recognition test, the same speaker performance measure is shown in Figure 1b along with the false alarm rates taken to measure performance in the other three conditions (Figure 1c). Since participants were asked to make a “same” vs. “different” speaker judgment, the analysis of hit rates for the three different speaker conditions would not take into account any bias associated with simply responding “different” without having to clarify which type of different speaker condition was present. Therefore, we restrict our analyses to the hit rate for the same speaker condition and to the false alarm rates for the three different speaker conditions.

Two analyses were performed on the accuracy data from the source memory test. For the same speaker condition a one-sample t-test was conducted to see if the group mean was significantly above chance (25%, since \( \frac{1}{4} \) of the trials were same speaker trials). Accuracy for the same speaker condition was significantly above chance ($t(15) = 21.07, p < 0.001, r^2 = 0.97$). Next, we compared the false alarm rates for the other three (different) speaker conditions to see if there was a bias. These were analyzed in a repeated-measures ANOVA. Even though the different gender/different accent false alarm rate was somewhat smaller than the other two, this effect was not significant, ($F(2,30) = 1.32, p = 0.28$). Therefore, there appeared to be no bias between the three different speaker conditions in the source memory test.

### 12.2 ERP Results

#### 12.2.1 Word Recognition Test

Figure 2 shows the ERP results for the word recognition test at various electrode sites over the entire scalp.

##### 12.2.1.1 Left Parietal Old/New Effect

In the word recognition test there was an increased positivity for correctly identified old words as compared to correctly identified new words, which peaked at around 750ms post-stimulus. The ANOVA on the mean amplitude for the 700-900ms interval over bilateral parietal regions yielded a significant main effect of trial type, $F(1,15) = 11.70, p = 0.004, \eta^2 = 0.44$, with old
words producing a significantly more positive deflection than new words. There was also a main effect of hemisphere, $F(1,15) = 4.96, p = 0.042, \eta^2 = 0.25$, with a significantly more positive deflection over the left hemisphere. Lastly, there was also a significant trial type by hemisphere interaction, $F(1,15) = 24.54, p < 0.001, \eta^2 = 0.62$. Due to the hemispheric effect, which coincides with the literature in naming this deflection as the left parietal old/new effect, we restricted further analyses to the left hemisphere only.

We then analyzed the left parietal old/new deflection over P5, P3, PO3 and P1 at 700-900ms for all five trial types – same speaker, different gender/same accent, same gender/different accent, different gender/different accent and new words. There was a main effect of trial type, $F(4,60) = 4.55, p = 0.011, \eta^2 = 0.23$, with pairwise comparisons indicating that the same speaker, the different gender/same accent and the same gender/different accent conditions were each more positive than the new word condition ($p < 0.05$). The correctly identified new words had the smallest positivity (see Figure 3). There was also a significant linear trend, $F(1,15) = 8.10, p = 0.012, \eta^2 = 0.35$, and a significant quadratic trend, $F(1,15) = 6.81, p = 0.02, \eta^2 = 0.31$.

12.2.1.2 Frontal Familiarity Effect

Considering only correctly identified old words in the word test, ERPs showed a bilateral deflection at 300-500ms that was most negative for words spoken in the same voice, followed by the two partial congruency conditions and lastly by the different gender/different accent condition; there was a significant main effect across these conditions, $F(3,45) = 3.59, p = 0.027, \eta^2 = 0.19$ (linear trend: $F(1,15) = 10.06, p = 0.006, \eta^2 = 0.40$). Pairwise comparisons indicated that the same speaker condition was different from the different gender/different accent condition ($p < 0.05$). There was no hemispheric effect or interaction.

12.2.1.3 Right Frontal Effect

Research by Hayama et al. (2008) indicated that right frontal deflections traditionally associated with post-retrieval work are actually indicative of decision/judgment processes rather than post-episodic source retrieval processing. To that end, we analyzed the item recognition data and found no late right frontal effect, indicating support for the decision/judgment hypothesis over the source retrieval hypothesis. Since no source judgment was required in the word recognition test, it is not surprising that we did not find a late right frontal effect here.
12.2.2 Source Recognition Test

Figure 4 shows the ERP results for the source recognition test at various electrode sites over the entire scalp.

12.2.2.1 Left Parietal Old/New Effect

Since all words presented in the source recognition test were old words, and participants were aware of this, it is not surprising that there was a consistent positivity over parietal regions for all words. Interestingly, no gradient reflective of voice congruency was evident in the voice recognition test, perhaps indicating that the amplitude of the parietal old/new effect varies only based on item-related judgments.

12.2.2.2 Frontal Familiarity Effect

Since the early frontal familiarity effect is an index of item (or lexical) familiarity, it is not surprising that no difference between voice conditions was observed in the source recognition test. Again, an analysis of the effects of voice congruency on parameters indicative of item recognition is not indicated for the source test.

12.2.2.3 Right Frontal Effect

The right frontal effect was quantified in the source recognition test. Since the hypothesized right frontal effect was not predicted at a specific time window, we conducted an analysis using BESA Statistics 1.0, over all scalp regions for 100-1300ms post-stimulus, and were particularly interested in significant modulations found over the right frontal area between 100-900ms. Previous research has shown that this frontal effect is found for both correct and incorrect source judgment trials (Senkfor & Van Petten, 1998); therefore we compared all same speaker trials vs. all different speaker trials, regardless of response success. The results of this cluster analysis produced a significant difference between same speaker and different speaker conditions over right frontal electrodes (FP2, AF8 and F8), which peaked at 722 ms ($p = 0.016$). The same speaker condition produced a significantly more positive deflection than the collapsed different speaker condition (Figure 5).
We also conducted ANOVAs at 100ms intervals, beginning at 100ms post-stimulus, bilaterally over AF7, FP1 and AF8, FP2. The windows that produced a significant main effect of voice condition were 700-800ms, $F(1,15) = 6.13, p < 0.05$ and 800-900ms, $F(1,15) = 5.54, p < 0.05$. This appeared to be a bilateral modulation, showing no hemispheric effect.

13 Discussion

The purpose of the first experiment was to assess the impact of reinstating speaker voice context at test on both word and source memory, and to identify memory-related ERPs that reflect a potential voice congruency benefit.

13.1 Behavioral Findings

In the present study, participants were most likely to remember a word if it was presented with the same speaker voice at study and at test. This finding is consistent with Pisoni’s (1993) notion of a parallel memory system that encodes speaker voice with word memory, and with prior behavioral research showing that reinstating the study voice at test provides a performance benefit both in word (e.g., Craik & Kirsner, 1974) and source (Palmeri et al., 1993) memory. In addition, we found facilitation in word recognition when the accent of the speaker was congruent with the original presentation, compared to reinstating gender or neither characteristic of voice. Though this provides some support to prior research indicating a context congruency benefit in word recognition (see Goldinger, 1996), further research is needed to clarify why the effect was restricted to accent. Our findings may be related to the fact that the words used in the present experiment were recorded individually, and therefore were not identical on any one parameter (accent or gender) in the conditions when that parameter was said to be maintained. That is, the voice conditions represented a categorical distinction rather than a quantitative gradient. Alternatively, it may be that certain voice characteristics are more salient than others.

13.2 ERP findings

There were three expected ERP deflections. The left parietal effect in the word recognition test produced a positivity gradient. Correctly identified old words where voice was also reinstated provided the most positive deflection. As the voice at test became less similar to the study voice, the correctly identified old words produced a weaker parietal positivity. Lastly, the
correctly identified new words were the least positive of the conditions. This finding is somewhat surprising in light of work done by Schloerscheidt and Rugg (2004), which indicated that the old/new parietal effect was the same for within and across format hits when verbal test stimuli were used, thus showing the irrelevance of surface form for this effect. However, their results do not speak to the effect of manipulating parameters within one of the formats, on this modulation. A possible explanation for the observed gradient is that, since only correct trials were used in the present comparison, the positivity gradient might reflect an implicit benefit of reinstating the speaker’s voice at test. This would then fall in line with findings by Wilding et al. (1995) and Wilding and Rugg (1996), who found that recollection was associated with greater positivity when context was recollected as well as the word itself. Moreover, our result is also in line with the description by Tendolkar (2004), that this well-known effect is actually two separate effects that overlap – one that is constant for recollection and another that varies based on strength of the memory.

The second predicted ERP modulation was an early frontal familiarity deflection, which corresponded to voice congruency in the word test. Since new words were presented in the same four voices as the old words, a comparison across hits and correct rejections would not be straightforward in terms of familiarity. Therefore, we investigated the frontal familiarity positivity for old words only, in an effort to determine the level at which the voice effect on word memory occurs. The familiarity gradient produced the most positive deflection for the most decontextualized situation (that is, different accent and gender), and the most negative deflection for the same speaker condition. This was somewhat surprising considering that a prior study that investigated context congruency at test revealed a frontal positivity for more familiar trials (Schloerscheidt & Rugg, 2004). However, this was the first experiment, to our knowledge to investigate the familiarity modulation in a purely auditory experiment. It may be that this ERP modulation is reversed in polarity for auditory familiarity, due to different sources that reflect an interaction between item familiarity and auditory processing. Nevertheless, this result should be treated with caution since it was based on a relatively small number of trials, which resulted in a less-than-optimal signal-to-noise ratio. Further investigation is required to confirm this finding.

The third modulation was a late right frontal effect traditionally associated with source processing. Though this modulation has typically been thought to investigate post-retrieval
processing, recent work by Hayama et al. (2008) has indicated that the deflection might actually index decision/judgment processes. Indeed, the word recognition test indicated no late right frontal ERP effect, lending support to Hayama et al.’s theory. Therefore, we investigated the right frontal effect in the source test only. We compared trials across the two voice conditions, same speaker and different speaker, regardless of response success. Trials were therefore classified by voice congruency type rather than by response type. This falls in line with evidence from previous studies that show a late frontal effect for correct as well as incorrect source judgments (e.g., Senkfor & Van Petten, 1998; see Hayama et al., 2008 for a brief summary). The finding that the same speaker condition produced a more positive deflection than the different speaker collapsed condition, peaking at 722 ms over right frontal sites, indicated a difference between voice congruency conditions regardless of response success. Given that the same speaker condition involved specificity in the binding of word and voice while the different speaker conditions implied a less demanding (and specific) judgment, we suggest that the more positive deflection associated with same speaker trials might reflect greater monitoring demands in this more stringent voice condition. However, this suggested explanation requires future research to confirm our findings and directly test this theory.

14 Interim Conclusion

In this first experiment, behavioral results indicated that voice congruency at test facilitates both word memory and source memory. These behavioral effects were paralleled by three expected ERP modulations. In the word recognition test, the same speaker conditions corresponded to the most positive left parietal old/new deflection and the most negative early frontal familiarity deflection. Since this condition had the most context congruency between study and test (i.e., reinstatement), it follows that recollection was the strongest in recognition for same speaker trials. In the source recognition test, a data-driven analysis indicated a right frontal positive deflection for the same speaker trials compared with the different speaker trials, regardless of response success. This deflection likely indexes some aspect of decision/judgment processes.
Chapter 4
Experiment 1a: Control Experiment

A potential confound was identified in our first experiment. If facilitation in word recognition depends on perceptual fluency (i.e., context congruency), then is the same-speaker benefit specific to the same exact acoustic example or is voice encoded? In order to address this potential confound in our results, we conducted a control experiment using new stimuli. We recorded new two-syllable frequent nouns in two voices – one male, one female (both unaccented). Participants performed the same tasks as in Experiment 1, but, in this case, half of the same speaker trials for each test were the exact same recording while the other half were different exemplars recorded by the same voice.

15 Methods

15.1 Participants

The sample comprised 21 participants who provided written informed consent according to the guidelines set out by the Baycrest Centre and the University of Toronto. One participant did not follow instructions for the speaker recognition test, so the final sample comprised 21 participants for the word recognition test and 20 for the speaker recognition test. Of the 21 total participants, there were eight males and thirteen females aged between 18 and 34 (M = 23 ± 4.6 years); they all had English as their first language and had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000Hz (both ears).

15.2 Design

The design and procedure used was identical to Experiment 1 except that only three blocks were used here. This control experiment used new voices and new words, and wavefiles were processed similarly to the method described in the General Methods (Stimuli) section. The words were again normalized to -32.44dB. Even though the average decibel level for these stimuli was -22.5dB, preliminary testing indicated that the words were quite loud. The shift to a quieter decibel level and transduction through the GSI at 60dB SPL (rather than 70dB SPL) were utilized in order to afford participants a more comfortable volume level. We did not
perform acoustical analysis of the stimuli and, so, we cannot quantify the difference between same-speaker trials that utilized different exemplars recorded by the same voice. However, the variation between such cases reflected normal ecological variation, which was the potential confound of interest in this experiment.

15.3 Behavioral Analysis

For both the word and the source recognition tests, we calculated hit rates for the three conditions – the same speaker with the same recording, the same speaker with a different recording, and a different speaker – and compared them using a repeated measures ANOVA.

16 Results

In the word recognition test, there was a main effect of voice condition, $F(2,40) = 5.12, p = 0.011, \eta^2 = 0.20$ (linear trend: $F(1,20) = 8.77, p = 0.008, \eta^2 = 0.31$; see Figure 6). Pairwise comparisons indicated that the two same speaker conditions did not differ from each other ($p = 0.76$) but that both differed from the different speaker condition ($p = 0.008$ for the same recording vs. the different speaker condition; $p = 0.024$ for the different recording vs. the different speaker condition). In the speaker recognition test, there was a main effect of voice condition, $F(2,38) = 22.62, p < 0.001, \eta^2 = 0.54$ (linear trend: $F(1,19) = 64.49, p < 0.001, \eta^2 = 0.77$; see Figure 7). Pairwise comparisons indicated that the two same speaker conditions did not differ from each other ($p = 0.32$) but that both differed from the different speaker condition ($p < 0.001$ in both cases).

17 Discussion

Results indicated no significant differences between same-speaker trials in which the same recording was used compared to trials in which a different recording by the same speaker was used. This was true for both the word and the speaker recognition tests. These results provide evidence that the benefit of voice reinstatement is not limited to the exact acoustic replica at test, but that the benefit results from a more general encoding of voice attributes, which might be normalized across exposures for individual voices. As such, it appears reasonable to suggest that auditory memory is sensitive to individual voices, and that the benefit of reinstating encoded voices is not specific to each acoustic exposure.
18 Interim Conclusion

Our findings of a same speaker voice reinstatement benefit do not simply reflect a benefit of acoustic repetition. Rather, individual voices are encoded with the word and act as context cues so that voice reinstatement facilitates memory across exposures.
Chapter 5
Experiment 2: Voice Congruency in Word Recognition using a Continuous Recognition Design

As a follow-up to Experiment 1, we used a continuous recognition paradigm to assess the impact of voice congruency on spoken word recognition [see Campeanu, S., Craik, F.I.M., Backer, K., & Alain, C. (2014). Voice reinstatement modulates neural indices of continuous word recognition. *Neuropsychologia, 62*, 233-244; Chapter 5 is largely taken from this work]. In Experiment 2, we sought to extend the characterization of a voice congruency effect by using a different paradigm than that used in Experiment 1 (Campeanu et al., 2013), but one that also allowed a direct comparison with previous behavioral research (Craik & Kirsner, 1974; Palmeri et al., 1993). Specifically, the second ERP investigation allowed an examination of the importance of focused attention to voice in word recognition. While our previous investigation (Experiment 1; Campeanu et al., 2013) used a block design and emphasized attention to words and to voice at study (since both were directly tested), the current experiment asked participants to make word judgments alone, with no attention expressly allocated to voice. Consequently, this design should significantly reduce attention to voice and we expect little or no gradient reflective of voice congruency, at least at longer lags when the study trace has been dropped from active awareness.

Since we employed a continuous recognition paradigm, a further important objective of Experiment 2 was to investigate the effect of lag on recognition memory, and more specifically on potential voice congruency effects in word recognition. In the current experiment, we manipulated lag within each voice congruency condition, so that old words were re-presented after 1, 7 or 15 intervening words. This allowed us to extend our investigation to determine how long the voice congruency benefit lasts and if/how it decays over time. Most importantly, it may be that the elapsed time between study and test in the different lag conditions corresponds to functionally different processes at different lags. Since prior research has found that working memory and long-term memory both contribute to memory performance within 15 seconds of sensory encoding (see Tulving & Schacter, 1990) and that the working memory component drops out between 12 and 24 seconds after the study episode (when using a continuous recognition paradigm; Poon & Fozard, 1980), we utilized lags that allowed us to compare retrieval before and after this critical time window. Specifically, we tested words after 1
intervening word (i.e., lag 2), where stimulus-onset asynchrony (SOA) was 8 seconds between study and test, and after 7 and 15 intervening words (i.e., lag 8 and 16, respectively), where SOA was 32 seconds and 64 seconds, respectively. It should be noted, therefore, that the longer lags corresponded to more intervening items, not only an extended delay. As such, recognition memory at lag 2 may have tapped into working memory (primary memory) processes, while lags 8 and 16 were more dependent on long-term memory (secondary memory) processes. To that end, we compared lag 2 with lags 8 and 16 (collapsed), aiming to determine whether behavioral and ERP signatures of voice congruency would differ for primary and secondary memory processes. In terms of primary memory, past research has shown that auditory perceptual information persists longer than visual information (Craik, 1969). In his review, Shulman (1971) proposed that encoding in short term memory (essentially the same concept as primary memory) is primarily phonemic, unless the task specifically engages other processes. Taken together, this suggests a dominance of auditory information in primary memory. For Experiment 2 (which uses auditory stimuli), this implied two things: first, that performance at lag 2 should be superior to performance at later lags; second, primary memory (lag 2) was likely to be more sensitive to voice congruency effects.

With respect to voice congruency, the second presentation of each word may have been spoken by the same voice, a voice of the same gender but different accent, a voice of the same accent but different gender, or a voice of different gender and different accent (as in Experiment 1). Since participants were not explicitly instructed to pay attention to voice, we anticipated overall superior performance for old words spoken in the same voice and no difference between old words spoken in any of the three different voice conditions. In addition, we predicted overall memory performance to be inversely correlated with lag and that, while more specific voice congruency effects might be stronger in primary memory, same-voice reinstatement facilitation would be present in both primary and secondary memory. Though voice effects might be expected to decrease over time, complete decay of such effects would be expected only at delays exceeding one day (Goldinger, 1996), a far more extended time course than we investigated in Experiment 2.

In terms of ERP modulations, we expected the presence of a parietal positivity for all correctly recognized old words. Though we predicted that a same-voice reinstatement benefit might correspond to the most positive deflection, we expected no gradient based on voice congruency
in the parietal old/new effect for this experiment. We also anticipated the presence of an early familiarity modulation that would be sensitive to voice congruency effects. The polarity of such a deflection was less clear, but based on the results of Experiment 1 (which is the only other one, to our knowledge, that investigated familiarity as a function of voice congruency in a purely auditory experiment), we expected the modulation to be most negative for same-speaker trials. As previously mentioned, it seemed reasonable that the sources underlying any neural activity observed should be different than those previously found (Curran, 2000; Schloerscheidt & Rugg, 2004; Curran & Dien, 2003), thereby resulting in a potentially different orientation and/or location of ERP modulations. Lastly, since no source judgment was required, we anticipated no late right frontal effect.

19 Method

19.1 Participants

Nineteen participants provided written informed consent according to the guidelines set out by the Baycrest Centre and the University of Toronto. EEG data from three participants were excluded because of excessive muscle artifacts and/or eye movements during recording and data from one participant was excluded because he did not follow task instructions. The final sample of 15 participants comprised six males and nine females aged between 22 and 33 (M = 24, SD = 3.4 years). All participants learned English as their first language, were right-handed and had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000 Hz (both ears).

19.2 Experimental Design and Procedure

The experiment was programmed using Presentation software (Neurobehavioral Systems, http://www.neurobs.com/, version 14.5) and custom Matlab code (version 7.1). On each trial, participants heard one word, presented binaurally through insert earphones (EAR-TONE 3a), at 70 dB SPL. Words (digitized as .wav files) were converted to analog using a computer soundcard (16 bit, stereo, with a sampling rate of 44100 Hz). The analog output was fed into a 10 kHz filter (Tucker Davis Technologies (TDT, Alachua, FL), FT6-2), and then to a GSI 61 audiometer. The ISI was jittered between 2.8 and 3.2 seconds (33 or 34 ms steps, rectangular distribution). Participants indicated whether or not they had previously heard the current word,
by making a button press corresponding to “old” or “new” with their left or right forefinger, respectively. Participants were asked to judge a word as “old” if they had heard it earlier, irrespective of whether the test word was presented in the same or a different voice as its previous occurrence. They were instructed to respond as accurately as possible on each trial. The experiment consisted of two 23-minute blocks, each comprising 348 trials (one word per trial). No words were repeated across blocks. Participants were given a mandatory break, lasting at least two minutes, between the two blocks.

In this experiment, we manipulated two factors: voice condition (i.e., the gender/accent of the voice repetition) and lag (i.e., number of intervening words separating the first and second presentation of the same word). There were four voice conditions: same speaker, same gender/different accent, different gender/same accent, and different gender/different accent. The three lag conditions included pairs, in which the first word was presented 2- (i.e., 7.6-8.4 seconds delays), 8- (30.4-33.6 seconds delay), or 16-back (60.8-67.2 seconds delay). Since these various lag conditions were randomly interleaved within each block, we created a custom Matlab program to determine the fewest number of extra “filler” trials that would allow for the 144 word pairs of interest (48 per lag, within each block) to be presented at each lag. This process generated ten possible lag templates. The 348 trials per block consisted of 288 “new”/“old” words of interest (144 pairs), 24 early “filler” trials (12 pairs) used to build up a memory store, and 36 “filler” trials throughout (14 pairs separated by lengthy lags and 8 unpaired words). The probabilities that a word was old or new were therefore roughly equal.

Custom Matlab code was used to create unique stimulus lists for each participant and block, based on two of the ten lag templates (one per block), which were chosen randomly for each participant. The word pairs were randomly ordered within each lag for each participant. The number of trials per lag and voice condition was counterbalanced as maximally as possible. In each block, within the 144 word pairs of interest, there were exactly 36 pairs within each voice condition per block (72 total pairs for each subject) collapsed across all three lags, and approximately 48 pairs (range: 47 to 49) within each lag per block (approximately 96 total pairs [ranging from 94 to 98 pairs across participants], collapsed across the four voice conditions). Within each voice condition, there were 23 to 25 pairs per lag, averaging to approximately 24 pairs per lag across participants. We initially collapsed over lag, since the number of observations per lag was small (at each lag, the number of trials per speaker condition was a
maximum of 25 trials). However, visual inspection of ERP traces allowed us to notice that lag 2 often behaved differently than lags 8 and 16, which were more similar. In addition, behavioral analysis of hits at separate lags (see Results section) indicated that performance at lag 2 followed a different voice congruency pattern than at lags 8 and 16. As such, we conducted further analyses on ERP data to see if a difference between short (2) and long (8, 16) lags emerged for different voice conditions. Moreover, since performance was superior at lag 2, we were able to get a reasonable number of trials to examine the interaction between voice condition and lag, in this way. At lag 2, the mean number of hits for the same speaker condition was 23.47 (SD = 0.83). It was 23.33 (SD = 1.05) for the different gender/same accent condition, 22.87 (SD = 0.92) for the same gender/different accent condition and 22.60 (SD = 1.24) for the different gender/different accent condition. At lag 8, the mean number of hits for the same speaker condition was 23.33 (SD = 1.18). It was 20.53 (SD = 2.36) for the different gender/same accent condition, 20.40 (SD = 2.59) for the same gender/different accent condition and 20.40 (SD = 2.23) for the different gender/different accent condition. At lag 16, the mean number of hits for the same speaker condition was 21.40 (SD = 2.53). It was 19.27 (SD = 3.15) for the different gender/same accent condition, 19.53 (SD = 2.61) for the same gender/different accent condition and 19.60 (SD = 1.50) for the different gender/different accent condition.

19.3 Behavioral Analysis

Hit rates were calculated for the four possible voice conditions – same speaker, different gender/same accent, same gender/different accent, different gender/different accent – within each lag (2-, 8-, or 16-back). The false alarm rate reported is a group mean rate, averaged from individual common rates (i.e., across all new trials) - based on the proportion of new words incorrectly judged as old - for each participant. Since voice conditions represented levels of congruency rather than individual voices, it was not possible to calculate false alarm rates per condition; instead, a common rate was calculated for each individual. As such, using hits minus false alarm rates in the analysis of voice congruency effects would yield no additional information compared to using hit rates alone. Therefore, accuracy analyses were calculated based on hit rates in the current experiment. In addition to hit rates, we computed RTs relative to word onset, using only correct old trials, for each voice condition within each lag.

The effect of context congruency on continuous word recognition was assessed using repeated measures ANOVAs with voice condition and lag as the within-subject factors, using IBM SPSS
Statistics (version 20). The Greenhouse-Geisser p-values are reported if the sphericity assumption was violated. Post-hoc pairwise least significant difference (LSD) comparisons were done following significant main effects.

19.4 Electrophysiological Analysis

BESA Statistics 1.0 was the primary means of ERP analysis. Two conditions could be compared at a time, over all scalp regions and up to 1300ms post-stimulus. A two-stage analysis first computed a series of t-tests that compared the ERP amplitude between the two conditions at every time point. This identified clusters in time (adjacent time points) and space (adjacent electrodes) where the ERPs differed between the conditions. In the second stage of this analysis, permutation tests were performed on these clusters. The permutation test used a bootstrapping technique to determine the probability values for differences between conditions in each cluster. The final probability value computed was based on the proportion of permutations that were significant for each cluster, and implicitly corrected for multiple comparisons. In each of the current analyses, we used a cluster alpha of 0.05, one thousand permutations and clusters defined using a channel distance of 4 cm, which resulted in an average of about 5.08 neighbors per channel. Using this technique, we computed 4 comparisons. First, we compared all correctly identified “old” trials with all correctly identified “new” trials. Then we sought to investigate the role of voice reinstatement on item familiarity in word recognition. To that end, we compared the same speaker condition (voice reinstatement) with the different gender/different accent condition (most decontextualized of the voice conditions), always using words that were correctly identified as old. Lastly, we investigated the role of voice reinstatement at short (lag 2) and then at long lags (lags 8 and 16), both times comparing the same speaker and the different gender/different accent conditions.

When the results of the data-driven analysis indicated potential voice reinstatement effects, we then exported mean amplitude measurements and analyzed them using repeated measures ANOVAs with voice condition as the within-subject factor. This was done to identify potential pairwise comparisons reflective of more fine-grained voice congruency effects (rather than just reinstatement). In addition, when looking at the effect of lag, trials were collapsed across voice conditions and then peak amplitude and latency measurements of the parietal old/new effect were exported and analyzed using repeated measures ANOVA with lag as the within-subject factor. Again, all deflections were measured using correct trials only.
20 Results

20.1 Behavioral Results

Overall, participants correctly identified 88.0% (SD = 5.4%) new words and 89.3% (SD = 5.0%) old words. For analysis, a false alarm rate was calculated for each participant; the group mean false alarm rate was 12%. Figure 8a shows the group mean accuracy (hit rates) for all four voice conditions at each of the three lags in the word recognition test; all accuracy analyses were conducted using hit rates. The ANOVA yielded a main effect of voice congruency, $F(3,42) = 15.63, p < 0.001, \eta^2 = 0.53$, with pairwise comparisons indicating that performance in the same speaker condition was superior to all three different speaker conditions ($p < 0.001$). Performance did not significantly differ among the three different speaker conditions ($p = 0.80$ for different gender/same accent vs. same gender/different accent; $p = 0.81$ for different gender/same accent vs. different gender/different accent; $p = 0.98$ for same gender/different accent vs. different gender/different accent). There was a significant linear trend, $F(1,14) = 47.15, p < 0.001, \eta^2 = 0.77$, and a significant quadratic trend, $F(1,14) = 11.20, p = 0.005, \eta^2 = 0.44$. There was also a main effect of lag, $F(2,28) = 45.57, p < 0.001, \eta^2 = 0.77$, with pairwise comparisons indicating that performance at all three lags were significantly different from each other ($p < 0.001$). The analysis of lag also produced a significant linear trend, $F(1,14) = 60.59, p < 0.001, \eta^2 = 0.81$. There was an interaction that approached significance between voice condition and lag, $F(6,84) = 2.42, p = 0.058, \eta^2 = 0.15$. This interaction indicates that voice effects were largely equivalent for lag 2, but the same-voice condition was superior to the others at lags 8 and 16 (Figure 8a).

Since the interaction between voice condition and lag closely approached significance, separate ANOVAs were then calculated at each lag, with voice condition as the within-subject factor. At lag 2, the main effect of voice condition was not significant, $F(3,42) = 1.87, p = 0.16, \eta^2 = 0.12$, though there was a significant linear trend, $F(1,14) = 9.31, p = 0.009, \eta^2 = 0.40$. It should be noted that the lack of significance at lag 2 probably reflected a ceiling effect, as hit rates were greater than 90% in all four voice conditions at that short lag. At lag 8, voice condition produced a significant effect, $F(3,42) = 10.61, p < 0.001, \eta^2 = 0.43$ (linear trend: $F(1,14) = 20.33, p < 0.001, \eta^2 = 0.59$; quadratic trend: $F(1,14) = 11.32, p = 0.005, \eta^2 = 0.45$). Pairwise comparisons indicated that performance in the same speaker condition was greater than in the
other three voice conditions (all $p \leq 0.001$) but that performance among the three different speaker conditions did not vary significantly ($p > 0.35$ in all cases). At lag 16, voice condition again produced a significant main effect, $F(3,42) = 5.54, p = 0.004, \eta^2 = 0.28$ (linear trend: $F(1,14) = 12.22, p = 0.004, \eta^2 = 0.47$). Once again, pairwise comparisons indicated that performance in the same speaker condition was greater than in the other three voice conditions ($p = 0.006$ for same speaker vs. different gender/same accent; $p = 0.009$ for same speaker vs. same gender/different accent; $p = 0.001$ same speaker vs. different gender/different accent), but that performance among the three different speaker conditions did not vary significantly ($p > 0.70$ in all cases).

Figure 8b shows RT for correct old trials in all four voice conditions at each of the three lags. There was a main effect of voice condition, $F(3,42) = 11.44, p < 0.001, \eta^2 = 0.45$ (linear trend: $F(1,14) = 20.87, p < 0.001, \eta^2 = 0.60$; quadratic trend: $F(1,14) = 9.33, p = 0.009, \eta^2 = 0.40$). Pairwise comparisons revealed shorter RTs for the same speaker condition relative to the other three voice conditions ($p = 0.019$ for same speaker vs. different gender/same accent; $p < 0.001$ for same speaker vs. same gender/different accent; $p = 0.002$ for same speaker vs. different gender/different accent). Also, RTs on the different gender/same accent trials trended toward being faster than those on the same gender/different accent ($p = 0.013$) and the different accent/different gender trials ($p = 0.049$). In addition, the ANOVA indicated a main effect of lag, $F(2,28) = 113.93, p < 0.001, \eta^2 = 0.89$ (linear trend: $F(1,14) = 140.19, p < 0.001, \eta^2 = 0.91$; quadratic trend: $F(1,14) = 68.80, p < 0.001, \eta^2 = 0.83$), with pairwise comparisons showing that participants responded faster at lag 2 than at lags 8 and 16 (both $p < 0.001$). There was no significant difference in RT between lags 8 and 16, $p = 0.51$. Finally, there was no significant interaction between lag and voice condition, $F(6,84) = 0.29, p = 0.86, \eta^2 = 0.02$.

In summary, behavioral results indicated that the same speaker condition generally produced the best accuracy and the shortest RT among the voice conditions. However, the benefit to accuracy of reinstating voice was significant only at longer lags (8 and 16), but not at the short lag (2). RTs also indicated faster recognition when accent was congruent between study and test. Overall, accuracy decreased and RT increased with increasing lag.
20.2 ERP Results

The ERPs comprised N1-P2 modulations and a late positive complex over parietal sites. The exogenous ERPs (i.e., P1, N1, P2) were not affected by voice condition or lag.

20.2.1 Effect of Voice Condition

The first analysis using BESA Statistics 1.0 focused on old and new words. It revealed a significant difference at parietal and frontal sites ($p < 0.001$; see Figure 9). At parietal and occipital sites, this modulation was characterized by an increased positivity for old words. This old/new effect began at about 380 and peaked at about 700 ms after word onset. The polarity of this modulation was inverted at frontal sites.

In a second analysis, we examined the effect of voice reinstatement on correctly identified old trials. Here, we collapsed across all lags and used BESA Statistics 1.0 to compare only the two most extreme voice conditions of correct old trials (i.e., same speaker vs. different gender/different accent), since we were looking for an indicator of how decontextualization affects recognition. The analysis revealed a significant difference between the same speaker condition and the different gender/different accent condition. The same speaker condition was consistently more negative over left frontal sites (F7, F5, F3, AF7, AF3, FC5) between approximately 200 and 600ms post-stimulus ($p < 0.001$; Figures 10a and 10b). This deflection appeared to inflect over central and right parietal sites between approximately 400-550ms post-stimulus ($p < 0.001$).

Since the data-driven analysis produced a significant difference between same speaker and different gender/different accent conditions as early as 200ms over the anticipated frontal area (see Curran, 2000; Curran & Dien, 2003), we then conducted an ANOVA to look for variations between all four voice conditions using the mean amplitude for the 200-600ms interval, over F7, F5, F3, AF7, AF3 and FC5. There was, as expected, a significant main effect of voice condition, $F(3,42) = 4.75$, $p = 0.009$, $\eta^2 = 0.25$ (linear trend: $F(1,14) = 7.98$, $p = 0.013$, $\eta^2 = 0.36$). Of particular interest, pairwise comparisons indicated that the deflection corresponding to the same speaker was significantly more negative than that corresponding to the different gender/different accent condition ($p = 0.007$) and that the same speaker deflection was more
negative than the deflection corresponding to the same gender/different accent condition ($p = 0.017$; see Figure 10c).

### 20.2.2 Effect of Lag

Since behavioral results indicated a main effect of lag, specifically that there was a significant difference between lags 2, 8 and 16, we collapsed across voice conditions to look at where and when this main effect would manifest over the scalp. Prior research has indicated that short lag trials produce a decreased-latency positive deflection over central and parietal sites at 550-650ms (Nielsen-Bohlman & Knight, 1994); therefore, we were particularly interested in lag effects as a function of peak latency of the old/new parietal effect.

We measured peak latency of the old/new parietal effect, at 600-800ms over bilateral parietal electrodes PO3, P5, P3, P1 and PO4, P6, P4, P2. The main effect of lag was significant, $F(2,28) = 8.29$, $p = 0.003$, $\eta^2 = 0.37$ (linear trend: $F(1,14) = 7.04$, $p = 0.019$, $\eta^2 = 0.34$; quadratic trend: $F(1,14) = 11.24$, $p = 0.005$, $\eta^2 = 0.45$), with pairwise comparisons indicating that lag 2 peaked earlier than lags 8 ($p = 0.002$) and 16 ($p = 0.019$; see Figure 11). The peak amplitude also produced a main effect of lag, $F(2,28) = 14.09$, $p < 0.001$, $\eta^2 = 0.50$ (linear trend: $F(1,14) = 20.56$, $p < 0.001$, $\eta^2 = 0.60$), with pairwise comparisons indicating that lag 2 produced a more positive deflection than both lags 8 and 16 ($p = 0.002$ and $p < 0.001$, respectively).

#### 20.2.2.1 Short vs. Long Lag

By collapsing trials at lags 8 and 16, we were able to compare old trials on short (lag 2) vs. long (lags 8/16) lag conditions. This analysis was conducted because behavioral findings indicated that voice congruency effects differed between short and long lags. Specifically, a main effect of voice congruency was found for accuracy at long lags, but not at short lags. This led us to believe that voice congruency effects might manifest differently in the ERP trace, based on grouping of correctly recognized words at short vs. long lags. Indeed, a preliminary analysis using BESA Statistics 1.0 indicated that short lag trials produced a more positive modulation over parietal sites from about 370-770ms post-stimulus, peaking at approximately 650ms ($p < 0.001$), which was inflected over frontal sites around 400-700ms ($p < 0.001$). As such, we conducted separate analyses on short and on long lags using BESA Statistics 1.0, looking at the same speaker vs. different gender/different accent voice conditions each time.
For the short lags, there was a significant difference between the two voice conditions at approximately 300-700ms, primarily over F7, F5, AF7 and FC5 ($p < 0.001$; see Figures 12a and 12b), which inflected over central and slightly right-lateralized parietal sites at 440-650ms ($p < 0.001$). Once again, in an effort to discern pairwise comparisons that would indicate more fine-grained voice congruency benefits, we followed the data-driven analysis with an ANOVA comparing mean amplitudes for all four voice conditions, at 300-700ms over F7, F5, AF7 and FC5 (as indicated from the BESA Statistics 1.0 results). As expected, there was a main effect of voice, $F(3,42) = 8.13, p = 0.001$, $\eta^2 = 0.37$ (linear trend: $F(1,14) = 5.92, p = 0.029$, $\eta^2 = 0.30$; see Figure 12c). Of particular interest, pairwise comparisons indicated that the same speaker condition was more negative than the same gender/different accent condition ($p = 0.001$) and the different gender/different accent condition ($p = 0.004$). Moreover, the different gender/same accent condition was more negative than the same gender/different accent condition ($p = 0.006$).

For the long lags, the data-driven analysis showed a significant difference between the same speaker and different gender/different accent conditions at 400-550ms over F7, F5, F3, AF7 and AF3 ($p = 0.027$; see Figure 13). A subsequent ANOVA comparing all four voice conditions at 400-550ms over F7, F5, F3, AF7 and AF3 (again, as indicated from the BESA Statistics 1.0 results) was not significant, $F(3,42) = 2.40, p = 0.093$, $\eta^2 = 0.15$, though there was a significant linear trend, $F(1,14) = 8.55, p = 0.011$, $\eta^2 = 0.38$. This indicates that there was a same-voice advantage and perhaps the hint of a more fine-grained effect based on voice congruency.

### 20.2.3 Source Analysis

A source analysis using BESA 5.3 was conducted on the group mean ERPs in order to localize the origin of the activity underlying the old/new and voice effects in the present experiment. For the old/new effect, we modeled the difference wave between correctly recognized old words and correct new words using a CLARA (Classical LORETA Analysis Recursively Applied) distributed model approach. Three stable sources emerged and are all displayed at 740ms post-stimulus, in Figure 14a. The first was stabilized around 500ms in the left temporal area and moved slightly posterior around 700ms. The source remained stable in the left temporo-parietal area for the remainder of the 1300ms trace, though it did diminish slightly around 800ms. Second, starting at approximately 600ms and lasting until the end of the 1300ms post-stimulus trace, stable activity was apparent in frontal areas, slightly right-lateralized and increasing in strength as time elapsed. Lastly, a third source was stable in right temporo-parietal areas,
starting around 520ms, and then beginning to move more toward central parietal areas starting around 830ms, where it remained (though steadily diminished) until approximately 1200ms. Dipoles were then placed at each source location and rotated to minimize the residual variable. The dipoles provide the time course of source activity at each source location identified by the CLARA algorithm.

For source localization of voice effects, we utilized the same approach. Using CLARA, we analyzed the difference wave between correctly recognized, same-speaker old words and correctly recognized, different gender/different accent old words (thus comparing voice reinstatement to the most de-contextualized condition). There was one source that emerged from 430-510ms in the left frontal area (see Figure 14b).

21 Discussion

21.1 Behavioral Results

For old words, the same speaker condition produced a higher accuracy score and a shorter reaction time than the other three voice conditions. While accuracy between the three different speaker conditions differed in Experiment 1, there were no significant differences in accuracy of the different speaker conditions in the present experiment. There are two possible explanations for this. First, there were differences in task instructions; in Experiment 1, participants were told to attend to words as well as to which speaker uttered each word, since they would be tested on both aspects. In the present study, participants were told to judge each word as old or new regardless of whether the voice was the same as the first time they had heard the word. Hence, the task instructions in the present study placed less emphasis on remembering the speaker voice. This suggests that attention to speaker voice may be important for the gradient in voice congruency to emerge.

Indeed, past research has looked at the role of attention in auditory recognition. According to Goldinger (1996) “it appears that the content of study words’ episodic trace is influenced by the focus of attention at study” (p.1177). By this account, one would expect voice congruency effects when participants are told to pay attention to both word and voice at study, but perhaps to a lesser extent, or not at all, when they are not instructed to pay attention to the voice. Palmeri et al. (1993) has previously used a continuous recognition paradigm in which voice was
not attended and found a same speaker advantage but no benefit when old words were presented in voices of the same gender as at study. Goldinger (1996), on the other hand, did find a partial same gender voice congruency benefit, both when voice was specifically attended at study and even in some conditions when participants were not asked explicitly to attend to voice. However, this difference might be explained by different designs, whereby participants in the latter experiment typed out the entire word during a study phase prior to a surprise recognition test. The more thorough encoding of each word (i.e., typing it out versus simply listening) may have facilitated a stronger association between voice and word at study in the Goldinger experiment, resulting in fine-grained voice congruency effects even when voice was not explicitly attended at study. The procedure involved in the present study was more similar to the Palmeri et al. study. Based on our results and previous research, it seems reasonable to suggest that a voice reinstatement benefit occurs even in the absence of attention being allocated to voice, while a more fine-grained voice congruency benefit requires attention to voice or some other form of emphasizing voice at study.

It is important to note that study instructions did not indicate a need to respond as quickly as possible, yet there were still significant effects of lag and voice condition on RT. Participants generally responded faster in the same speaker condition, and there was also some indication of a speed benefit when accent was congruent between study and test. It should, however, be noted that since overall accuracy was much higher at lag 2, the potential differences among voice conditions at this lag were very likely masked by a ceiling effect. In addition, the relatively small sample size in the current experiment might have limited potential voice effects, especially since there is a graphic trend to a behavioural voice effect at lag 2 (see Figure 8a).

Therefore, though behavioral results provide indications of same-voice facilitation at all lags, further investigations are needed to better characterize potential voice congruency benefits at short lags in particular. Further research is also needed to assess whether directing attention to an aspect of voice (i.e., gender or accent), or familiarity training with a voice, influences this facilitation pattern.

21.2 ERP Results

The parietal old/new effect showed more positive waves for the old word conditions and a most negative wave for the new word condition, as expected. However, no voice congruency effects
were found (and no systematic gradient between the different voice conditions of the old word trials). This parietal old/new effect has previously been interpreted as a combination of two effects, reflective of recollection and of retrieval demands (Tendolkar et al., 2004). Given that task instructions did not stress attention to voice, it is perhaps a reasonable finding that the strength or specificity of recollection was not reflected in voice congruency in the parietal old/new effect. However, since we did not manipulate attention allocation in the present study, further research is needed to clarify whether attention is, indeed, a pivotal factor. Alternatively, it may be that the parietal modulation deals with amodal, lexical information in recognition memory and is largely insensitive to perceptual attributes.

Conversely, some separation between voice congruency conditions did exist over frontal regions. Since frontal deflections are associated with familiarity in memory studies (e.g., Curran, 2000), we suggest that voice congruency might affect item familiarity even when it is not expressly attended. This modulation began as early as 200ms post-stimulus, which might reflect the fact that even the longest lag in the current paradigm is shorter than the typical lag in block paradigms. It seems likely that familiarity is more sensitive to perceptual fluency than recollection. Overall, these findings fall in line with imaging studies that have found recollection and familiarity to independently support recognition memory (e.g., Evans and Wilding, 2012).

It is interesting to note that our findings over frontal sites indicate a most negative deflection for the same speaker condition. This is perhaps unexpected in light of other findings (Curran, 2000; Curran & Dien, 2003). However, although the frontal familiarity effect presents as a positivity for more familiar items (Curran, 2000; Schloerscheidt & Rugg, 2004; Curran & Dien, 2003), the use of familiarity across various voice congruency conditions is less clear. The only previous study that has found a frontal negativity for context congruency (Opitz, 2010) employed a paradigm that utilized visual stimuli and very different context cues. Furthermore, though Opitz (2010) found that studying stimuli across various contexts resulted in greater familiarity (resulting in a greater number of “know” responses and a more positive wave for test stimuli that had incongruent context conditions at study), the test phase in that experiment did not reinstate potential context cues. Therefore, congruency was manipulated at study, when recognition was not probed, but not at test, when recognition was probed. Opitz posited that familiarity plays a bigger role in recognition memory when context is not consistent between
study repetitions. This is likely because repetitions across contexts promotes semantic rather than episodic memory, and familiarity involves a general feeling of knowing you have previously encountered something (rather than retrieval of episodic details). Furthermore, since all stimuli were decontextualized at test, perceptual fluency at test became irrelevant to familiarity in this case. In the present study, however, we examined episodic memory for words spoken once in one of four voices, rather than semantic knowledge. In summary, though Opitz did investigate context congruency in relation to the familiarity ERP deflection, that experiment was fundamentally different from the current investigation.

There is the possibility of another explanation for the finding of a familiarity modulation that is most negative for the voice reinstatement condition. Since the present experiment utilized only auditory stimuli, it is possible that the sources underlying any neural activity observed are different than those previously found, thereby resulting in a potentially different orientation and/or location of ERP modulations. It seems plausible that, since previous studies investigating this effect were qualitatively very different from the current investigation, e.g., in terms of modality, task itself (see Curran, 2000; Schloerscheidt & Rugg, 2004; Curran & Dien, 2003), familiarity might simply generate different sources of activity for this task in the auditory domain. Given that our findings are roughly in the same time window and over approximately the same regions as previous work, we do still suggest that it is a marker of familiarity. A negative polarity for the same speaker condition might indicate agreement with Opitz’s (2010) conjecture – that decontextualization results in greater use of familiarity. However, Opitz was examining subsequent memory following decontextualized repetitions. In that experiment, participants who learned the stimuli across various contexts probably developed a richer constellation of traces that could signal familiarity. Since the present experiment involves context reinstatement at test following one exposure, episodic memory is at play and perceptual fluency is likely a factor. Therefore, it might also be that the negativity associated with the same speaker condition simply reflects the novel paradigm utilized in the current investigation. The positioning and orientation of the sources underlying this purely auditory effect require further study.

This continuous paradigm meant that test words were successfully encoded 2-, 8- or 16-back (that is, with 1, 7 or 15 words intervening between first and second presentations). There was a main effect of lag both behaviorally and in the ERP trace. The significant behavioral results
showed that participants were better and faster with old/new judgments at lag 2, and this was mirrored in peak latency measurements of the parietal old/new effect, where lag 2 peaked higher and significantly earlier. Comparing the short and long lag conditions, ERP results indicated that the use of recollection in recognition memory is stronger (i.e., more positive parietal deflection) at short lags.

Since behavioral analysis indicated that voice congruency behaved differently at lag 2 as opposed to lags 8 and 16, we conducted further analyses on the ERP data for short (lag 2) vs. long (lags 8 and 16) lag conditions. Results indicated that a voice reinstatement benefit was present at both short and long lags, but that more fine-grained voice congruency benefits were present only at the short lag. Since only 8 seconds elapsed and only one word intervened between presentation and test for lag 2, it is likely that subjects still retained the word in conscious awareness. This factor would not operate at longer lags. Thus the recognition decision at lag 2 was likely done on a different basis – recognition of a trace still present in auditory awareness. This distinction between items in awareness and items that must be brought back into awareness is referred to as the distinction between primary memory and secondary memory in the cognitive literature (see Waugh & Norman, 1965; Poon & Fozard, 1980; Nielsen-Bohlman & Knight, 1994). Similarly, working memory refers to memory that is held and manipulated in active awareness while long-term memory is analogous to secondary memory. The current design involves test words at lag 2, where stimulus-onset asynchrony (SOA) is 8 seconds between study and test, and test words at lags 8 and 16, where SOA is 32 seconds and 64 seconds, respectively. Therefore, recognition memory at lag 2 may tap into primary memory processes, while lags 8 and 16 are likely more dependent on secondary memory processes.

Moreover, in terms of primary memory, past research has shown that auditory perceptual information persists longer than visual information (Craik, 1969). In his review, Shulman (1971) proposed that encoding in short term memory is primarily phonemic, unless the task specifically engages other processes. Taken together, this suggests a dominance of auditory information in primary memory. For the present study (that uses auditory stimuli), this might explain why performance at lag 2 was superior to performance at later lags. In addition, this framework would also imply that primary memory (lag 2) is likely to be more sensitive to voice congruency effects, an effect that we did not observe behaviorally (most likely due to ceiling
effects) but that was supported by early fine-grained voice congruency effects over frontal sites in the ERP trace.

This framework can therefore help us understand the ERP findings at short and long lags. No effect of voice reinstatement emerged when comparing recollection at short or long lags; this implies that while recollection is overall stronger in primary memory, there appears to be no differential use of voice congruency in this case. With respect to the frontal familiarity modulation, however, there was a same speaker facilitation effect at both short and long lags, and the nature of the facilitation differed between the two lag conditions. At the short lag, there appeared to be some separation between voice congruency conditions, while such fine-grained distinctions were absent at longer lags. In particular, the voice conditions that corresponded to congruent accent between study and test produced more negative deflections at the short lag. This could suggest that accent congruency plays a significant role in primary auditory memory.

It may be that accent is an important component of voice, which is more resistant as the auditory trace decays. However, it is also important to note that the speakers who recorded the accented stimuli were older than the speakers who recorded the non-accented stimuli. Therefore, we must note that the accent congruency condition also involved age congruency of the speaker’s voice, making it possible that the benefit in performance is the result of quantitatively richer context congruency (accent and age of the speaker) rather than a benefit specific to accent congruency itself. Further research is needed to clarify this potential confound.

Nevertheless, our findings of a more fine-grained voice congruency effect at primary memory fall in line with the idea that this type of memory involves a still-active trace that places emphasis on phonemic/auditory properties (Craik, 1969; Shulman, 1971). In addition, it is perhaps more appropriate to discuss the familiarity deflection as a marker for perceptual fluency with respect to primary memory, given that the initial auditory trace is still in active awareness at this point. Taken together, this suggests that perceptual fluency in the form of voice congruency between study and test likely facilitates recognition in primary memory. At longer lags, there was a significant voice reinstatement benefit but no more fine-grained voice congruency effect. This suggests that familiarity picks up on same-voice reinstatement, which appears to meet the threshold required for this strength marker, but is less sensitive to more subtle voice differences. Therefore, though there were some pairwise comparisons that indicated a voice similarity gradient beyond just a same-voice advantage at lag 2, the frontal
modulation at longer lags only produced a same-voice advantage. Recall that the overall frontal modulation collapsed across all three lags also produced a same-voice advantage, though with added significant pairwise comparisons. This implies that lag 2 trials, which likely have a strong primary memory component, utilize still-active auditory traces and are therefore more sensitive to perceptual detail. On the other hand, trials at longer lags, which correspond to secondary memory, indicate a same-voice effect whereby context reinstatement results in familiarity in recognition memory.

Overall then, in using the primary vs. secondary memory framework to explain our findings, we suggest that both recollection and perceptual fluency/familiarity are used differently in primary memory as opposed to secondary memory. Recollection appears to be used more at lag 2, irrespective of voice congruency. The so-called familiarity deflection, on the other hand, appears to be sensitive to voice congruency in both primary and secondary memory, though more so in the former (when the initial trace is still active).

Lastly, we must point out a confound in our experiment. Although we have discussed lag effects in terms of trace decay, it should be pointed out that longer lags necessarily involve more intervening items between a word’s study and test. That is, the effects we observed may be due to interference rather than decay, and this confound should be addressed in future studies.

22 Interim Conclusion

There is a reliable and robust same-voice facilitation effect in word recognition. This and the existence of a significant parietal old/new effect replicate previous findings from Experiment 1. The current work also extends our understanding of voice as a memory cue. First, a same speaker facilitation effect is apparent for both accuracy and reaction time measurements, while a more general voice congruency benefit is not found in this paradigm. ERPs demonstrate a lack of a recollection effect for voice congruency. ERPs do, however, show a frontal pattern reflective of voice congruency, which greatly precedes response times. This suggests that familiarity uses more detailed perceptual characteristics and may therefore be more involved in implicit memory. Future studies should therefore extend this line of speculation into voice congruency as a cue in implicit memory. Is this familiarity gradient only present when explicit memory is tested, or does it represent an automatic linking of word and voice that precedes and underlies the use of voice in multiple forms of memory?
Chapter 6
Experiment 3: Pilot Study

An intriguing finding from Experiment 2 was that the familiarity modulation was sensitive to voice congruency in the absence of explicitly directing participants’ attention to voice, while the parietal modulation was not. This result suggested that voice congruency would have a bigger effect on implicit memory than on explicit memory (see also Craik et al., 1994). Experiment 3 was designed to directly test this suggestion.

23 Method

23.1 Participants

The sample comprised 23 participants who provided written informed consent according to the guidelines set out by the Baycrest Centre and the University of Toronto. There were nine males and fourteen females aged between 18 and 34 (M = 24 ± 4.8 years); all participants had English as their first language and had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000Hz (both ears). In the explicit group, we tested 11 participants (four males and seven females aged between 20 and 34, M = 24 ± 5.2 years). In the implicit group, we tested 12 participants (five males and seven females aged between 18 and 32, M = 24 ± 4.6 years).

23.2 Stimuli

The words used were the same as for Experiments 1 and 2 (see General Methods). Specifically, words were high frequency nouns and they were spoken in one of four voices: male/non accented, male/accented, female/non accented, female/accented. In all phases of this experiment words were embedded in noise. The noise bits came from a continuous stream of cafeteria babble, which was cut into 2-second pieces using an Adobe Audition 1.5 script. Since each word file was 1 second in duration, noise was presented 500ms before word onset so that it lasted until 500ms after offset of the word file.

In order to increase perceptual congruity between study and test, words were always embedded in babble noise (both at study and at test). When using words embedded in noise, loudness variations between stimuli became evident. As such, additional processing became necessary.
Using another Matlab (version 7.4) script, all words were normalized to a maximum root-mean-square (RMS) rather than an average RMS (as for our previous studies), skipping points at which the decibel level was zero. Each trial was then checked manually to make sure loudness was consistent between word-in-noise trials.

23.3 Design

Experiment 3 had all participants perform the same encoding task, and then gave half of the participants an incidental explicit recognition (old/new) test and the other half an implicit word identification test. This between-subjects design was chosen so that all participants presumably processed/encoded the stimuli in the same way; analysis at test would therefore focus on differences in retrieval (conscious or subconscious). Participants who performed the explicit test realized they were being tested for memory as soon as the test phase began; conversely, participants who performed the implicit test were never told that they were being tested for memory. Regardless, all participants were told at study that they were participating in an experiment that looked at auditory processing in noise, and therefore Experiment 3 utilized incidental encoding for both groups. Participants were meant to be balanced between the explicit and implicit groups, based on vocabulary, musical and language expertise, to minimize potential confounding variables.

Participants first underwent a subjective noise identification task, wherein they had to identify practice words presented in various levels of noise. Relative volume levels between word and noise were adjusted by attenuation of the noise stimulus. The noise levels chosen varied based on the participant’s response to the previous noise level – i.e., if participants had an easy time at a given noise level, the noise was less attenuated in the next pass at the words. Each participant performed this preliminary task so that individual word-in-noise comprehension levels could be gauged for later use in the experiment. Ideally, the study phase was run at a noise attenuation for which the participant heard almost 100% of words embedded in noise. At test, we wanted participants to be able to easily identify about 70-80% of words embedded in noise.

Once appropriate noise levels were chosen, participants performed the study task. Participants were asked to identify, by pressing the corresponding button on the keyboard, the last letter of each word they heard. At test, participants were asked to type out the word (implicit group) or to make an old/new judgment and then type out the word (explicit group). Responses were
checked manually, so that minor spelling mistakes could be overlooked in analysis. At study, 144 words-in-noise were presented. Those words were then interspersed with 144 new words, and the resulting 288 words were split into 4 parts, at test.

23.4 Behavioral Analysis

In the explicit group, participants performed a word-in-noise recognition (and identification) task in the test phase. An old/new analysis first compared words that were correctly recognized as old and correctly identified at test with words that were correctly said to be new and correctly identified at test. Then, voice congruency effects were investigated: hit rates were calculated for the four possible voice conditions – same speaker, same gender/different accent, different gender/same accent, different gender/different accent. Hits could each be analyzed in two ways. In Analysis 1, hits were given by words that were correct at study and then also correctly recognized as old as well as correctly identified at test. In Analysis 2, hits were given by words that were correctly recognized as old and correctly identified at test, regardless of whether or not they had been correct at study. Correct trials at study were those where the last letter of the word was correctly identified.

In the implicit group, participants performed a word-in-noise identification task in the test phase. Again, correct trials at study were those where the last letter of the word was correctly identified. Hits, for the implicit group, were those old trials where the entire word was correctly identified. In Analysis 3, priming was calculated by considering only old words that were correct at both study and test, in comparison with new words. In Analysis 4, priming was calculated by considering old words that were correctly identified at test, irrespective of whether they had been correct at study or not, in comparison with new words. In addition to the priming analysis, potential voice congruency effects were investigated by separating correct old trials into the four voice conditions – same speaker, different gender/same accent, same gender/different accent, different gender/different accent.

24 Results

In terms of a potential old/new effect, there was no difference in performance between old words that were both correctly recognized as old and correctly identified at test (M = 47.7%, SD = 11.9%) vs. new words that were correctly said to be new and correctly identified at test (M =
47.3%, SD = 11.7%; shown as a dotted line in Figure 15), \( t(10) = 0.064 \), two-tailed, \( p = 0.95 \).

(Note that since the analysis using all correctly recognized old words that had been correctly identified yielded insignificant results, limiting the analysis to only those correct old words that had also been correct at study could not yield results indicative of priming. As such, the analysis of this subset was omitted. Subsequently, we also looked at all words correctly recognized as old – both dependent on correct study trials and independent of study performance – vs. words correctly said to be new, regardless of whether the word had been correctly identified at test. Whether comparing hits minus false alarm rates for both old words and new words, or performing the same comparison using hit rates only, the mean number of correct trials was always larger for new words.)

Voice congruency results for the explicit group (Figure 15) varied based on the method of analysis used. Interestingly, there were significant voice effects in spite of the null old/new effects. For Analysis 1 with only words that were correct at study, results indicated the presence of a significant overall voice congruency effect, \( F(3,30) = 5.35, p = 0.009, \eta^2 = 0.35 \) (linear trend: \( F(1,10) = 32.30, p < 0.001, \eta^2 = 0.76 \)). Pairwise comparisons revealed that the same speaker condition produced better word recognition than all three different voice conditions (all \( p < 0.05 \)). For Analysis 2 with all correct test words, regardless of whether or not they were correct at study, results also indicated a main effect of voice congruency, \( F(3,30) = 3.37, p = 0.047, \eta^2 = 0.25 \) (linear trend: \( F(1,10) = 27.99, p < 0.001, \eta^2 = 0.74 \)). Pairwise comparisons revealed that performance was weaker in the same gender/different accent condition than in the same speaker and the different gender/different accent conditions (both \( p < 0.05 \)). In terms of attention allocation, participants were not expressly told to attend to voice but they did perform an auditory perceptual task at study. As such, voice effects were expected in this investigation. Interestingly, overall performance was much lower for this group, in comparison with Experiment 1. This difference may be explained by the fact that the present investigation utilized incidental encoding and/or it may be that the additional task of typing out the whole word resulted in a test that was simply more difficult.

For the implicit group, priming results depended on the approach used in the analysis. Using Analysis 3, we compared old words that were correct at both study and at test (\( M = 54.3\%, SD = 11.9\% \)) with correctly identified new words (\( M = 64.8\%, SD = 12.3\% \)). Results indicated a main effect of old vs. new, \( t(11) = -4.90 \), two-tailed, \( p < 0.001, r^2 = 0.69 \), where participants
were better at identifying new words. This implied no priming effect at all. On the other hand, using Analysis 4, we compared all old words that were correct at test (M = 66.4%, SD = 11.5%) with correctly identified new words (M = 64.8%, SD = 12.3%), and did find a small but insignificant benefit of old vs. new, $t(11) = 1.08$, two-tailed, $p = 0.30$, $r^2 = 0.10$.

With respect to the voice conditions, results indicated that correctly identified words at test might also be susceptible to voice congruency effects. The main effect was significant for Analysis 3, $F(3,30) = 13.60$, $p < 0.001$, $\eta^2 = 0.58$ (linear trend: $F(1,10) = 12.49$, $p = 0.005$, $\eta^2 = 0.56$; quadratic trend: $F(1,10) = 31.05$, $p < 0.001$, $\eta^2 = 0.76$), and for Analysis 4, $F(3,30) = 5.88$, $p = 0.012$, $\eta^2 = 0.37$ (quadratic trend: $F(1,10) = 30.61$, $p < 0.001$, $\eta^2 = 0.75$). Word identification was significantly better for the same speaker condition compared to all three different speaker conditions, when considering trials that were correct at study and at test. When looking at all trials that were correct at test (regardless of whether the last letter of the words had been correctly identified at study), pairwise comparisons only showed superiority of the same speaker condition in comparison to the different gender/same accent and the same gender/different accent conditions, as well as better performance for the different gender/different accent condition than the different gender/same accent condition (see Figure 16).

25 Discussion

Results for the explicit group provided converging evidence for the benefit of voice reinstatement in explicit memory. Reinstating the speaker’s voice at test bolstered recognition performance, even when the memory test was incidental and even when participants had the added task of identifying the words at test. Pairwise comparisons also indicated that there might be some more fine-grained voice congruency effects, but these depended on the method of analysis used. Interestingly, voice effects existed despite there being no significant old/new effect for the explicit group. This was unexpected and might perhaps be due to low study power ($n = 11$). For example, when considering words that are correctly classified as old or new and that are also correctly identified, a small benefit exists for old words. A larger sample might yield significant effects corresponding to this benefit. (However, when considering only words that are correctly classified as old or new, regardless of whether or not they are correctly identified, new word performance is superior. It is possible that when words cannot be clearly
identified, their initial presentation acts as a form of interference which actually decreases performance.)

With respect to the implicit group, this pilot study raised a major point of theoretical consideration when analyzing priming results. Namely, it is not obvious whether calculations involving correctly identified old words should only include trials that are correct at study as well, or if analysis should be independent of performance at study. This calls into question the mechanism of priming. If, for example, priming is a subconscious mechanism that is mainly acoustic in nature, then mere exposure might be enough to improve subsequent memory. By extension, Analysis 4 is the appropriate approach and we observed a small but insignificant priming effect in the current experiment. It follows from the logic underlying this approach, then, that the benefit of voice reinstatement would be more pronounced here, than for Analysis 3. If priming works based on acoustic exposure, then reinstatement of the same speaker between study and test would bolster memory at test much more than for old words in which voice was not reinstated between study and test. Our results do not support this, however, since the magnitude of the voice effect was larger and the pairwise comparisons separating the same speaker trials from all the different speaker trials were more significant for Analysis 3. If, on the other hand, the mechanism of priming (at least in the present circumstances) requires participants to understand the words at study (as signaled by the correct identification of the last letter of the word at study), then Analysis 3 is the appropriate method. However we did not observe a priming effect using this approach.

In addition, it is possible and perhaps most likely that priming works by both acoustic and semantic mechanisms. If so, then the difference between correct and incorrect study trials would be a quantitative one. Speculatively, an incorrect study trial could signify sufficient acoustic priming without additional semantic priming. The question is whether this perceptual priming would be sufficient to improve memory at test. If it is, then we expect the effect of voice reinstatement (i.e., same speaker trials vs. different speaker trials) to differ based on accuracy in the study task. Specifically, higher accuracy at test for same speaker trials (compared with different speaker trials) subsequent to incorrect study judgments would provide evidence that enough perceptual information was being processed to facilitate priming even when trials were incorrect at study. To that end, we did look at words that had been correct at study and at test, and separated them based on voice reinstatement (same speaker; group mean
average proportion was 0.64) or not (different speaker; group mean average proportion was 0.51) at test. We did the same for words that had been incorrect at study but correct at test (same speaker group mean average was 0.065 and different speaker group mean average was 0.14). Comparison of these group mean proportions indicated no benefit of voice reinstatement when trials were incorrect at study. We, therefore, cannot conclude that enough perceptual information was processed at study to facilitate priming at test; this did not provide justification for using Analysis 4 rather than Analysis 3 in the implicit group. Actually, there seemed to be a comparative boost in performance for words heard in a different voice at test after being incorrect at study (which was largely driven by performance in the different accent conditions), compared to the same voice. It seems reasonable that words heard in babble in the accented voices were often too difficult to discern, and a second presentation in a non-accented voice was required for identification. As such, the “old” condition, when conditionalized on a correct study response, is deflated in Analysis 3 but not in Analysis 4. This may explain why performance on new trials exceeded performance on old trials in Analysis 3.

We should also note, however, that the mechanism of priming might depend on the type of noise employed. Though words were checked for comprehension, the use of multi-talker babble may have interfered with processing, since this type of noise is made up of sounds that are quite similar to the stimuli themselves, resulting in potential interference/competition at the level of auditory perception and processing. In addition, a lack of comprehension is not necessarily the only reason for participants’ incorrect study trials. Participants were not tested for spelling proficiencies, so incorrect study trials might have resulted simply from spelling errors. We manually inspected trials for these errors at test, but no such check could be performed when participants’ responses were a single letter, as was the case at study.

Another concern relates to the method of Analysis 3. In considering only words that were correct at both study and at test, an alternate method would be to do a conditional analysis. Namely, we could have calculated a proportion by dividing correct old test words by correct study words, rather than deflating our old word group by placing two conditions on hit trials. However, such an approach could not be directly compared with correct new trials, since the parameters would be different between the two groups. Furthermore, the difficulty of hearing certain words in noise would likely vary based on the voice speaking it at study and at test, thereby introducing a bias. Sheffert (1998) noted the latter bias and reported conditionalized
results using a modified baseline. That is, performance for each word was affected by how many other subjects had correctly identified that word. This would suggest that perhaps a better approach to the current experiment would have been to use a between-subject design wherein the designated study words and the designated new words are switched for half of the participants, in an effort to minimize bias. Goldinger (1996) used a different way of analyzing word identification trials. Specifically, he looked for improvement in identification across sessions – an approach that required participants to engage in the same task at study and test, which made it inappropriate as a method to analyze our current experiment.

Regardless of whether priming occurred or not, the more intriguing finding in this pilot study was the presence of voice congruency effects. As was the case in Experiments 1, 1a and 2, voice reinstatement provided a benefit for memory. Moreover, the pattern observed between the voice congruency conditions appeared to graphically trend in the same direction across both groups. More fine-grained voice congruency effects were not found in the implicit group of Experiment 3 (but were suggested in the explicit group), perhaps due to the small sample size and/or perhaps due to a lack of attentional focus to voice at study. Further research should address the question of more fine-grained voice congruency effects in implicit memory using a larger sample size; indeed, we would expect perceptual attributes such as voice congruency to have a bigger effect in implicit memory paradigms than in explicit memory paradigms.

Theoretically, the issue of whether or not priming can be said to have occurred has potentially significant consequences for the interpretation of voice effects in this experiment. If we judge that priming has occurred, then voice reinstatement effects appear to follow a similar pattern in this implicit memory paradigm as we previously found using various explicit memory tasks. A lack of more robust and specific voice congruency effects might simply reflect the power limitation of the sample size used in this pilot study. If, on the other hand, we judge that priming has not occurred, then the theoretical implications might be different. In our previous experiments, we found evidence to support a binding of word and voice, so that reinstatement of voice at test represents an integral context cue. If we have found evidence of voice effects in the absence of priming, then the current pilot study provides preliminary evidence to suggest that the acoustic and semantic traces of an episode in auditory memory run parallel. This suggestion is speculative at this point and conditional on appropriate analysis methods.
26 Interim Conclusion

An investigation to determine whether sufficient priming exists when trials are incorrect at study is required as a first step to discern the implications of Experiment 3. Nevertheless, results of this pilot study suggest that voice congruency is utilized in the same way across various forms of memory.
Chapter 7
Experiment 4: Voice Congruency across Language Groups

We investigated the potential role of language skills in conferring the voice congruency effect we had observed in Experiment 1. Though some prior research had investigated the role of bilingualism in auditory processing (Thompson, 1987; Goggin et al., 1991; Lagrou et al., 2011), the effect of bilingualism on word and voice recognition had not yet been studied, to the best of our knowledge. As such, we employed the behavioral paradigm from Experiment 1 and tested participants in three groups – monolinguals, bilinguals in the same language as the voices that recorded accented stimuli and bilinguals in other languages. In all groups, participants had English as their primary language. Bilingualism was determined by self-assessment, wherein participants were deemed bilingual if they reported speaking a second language on a daily basis.

We predicted a main effect of voice congruency for all groups, and that bilinguals would generally outperform monolinguals. Since participants were told to attend to both word and voice and since memory function involves executive control at several stages, we anticipated this trend based on earlier research, which suggested that memory function is superior in bilinguals when executive control is required (Bialystok, 2009; Wodniecka et al., 2010). However, we were unclear as to whether a benefit for bilinguals would be uniform across voice congruency conditions (or perhaps a greater benefit would result specifically when accent was congruent between study and test) and whether potentials trends would be similar between the two different bilingual groups (i.e., is a potential benefit of bilingualism specific to the language or is there a more general benefit related to executive functioning?).

27 Method

27.1 Participants

One hundred and twelve participants provided written informed consent according to the guidelines set out by the Baycrest Centre and the University of Toronto. They received course credit for participating in the one-hour session. Forty-eight participants were excluded from analysis based on exclusion criteria (i.e., English was not their primary language) or if their level of bilingualism was ambiguous (i.e., monolinguals who spoke a second language at a
rudimentary level or bilinguals who only spoke a second language weekly, monthly or occasionally). The final sample of 64 participants comprised fourteen males and fifty females aged between 17 and 34 (M = 19 ± 2.3 years). All participants had English as their primary language and self-reported as having normal hearing. Twenty-one participants were monolingual (seven males and fourteen females aged between 17 and 34, M = 20 ± 3.8 years), 20 were Chinese-speaking bilinguals (spoke Chinese daily; four males and sixteen females aged between 17 and 21, M = 18 ± 1.0 years) and 23 were bilinguals who spoke a different second language (daily; three males and twenty females aged between 17 and 21, M = 18 ± 0.8 years).

27.2 Design

Participants first completed Shipley’s Institute of Living Scale - Vocabulary Test (Shipley, 1940) and the Language Background Questionnaire - Short Form (Craik, Bialystok, & Freedman, 2010). Then, participants performed five blocks of the design used in Experiment 1, each consisting of a study phase followed by a word recognition test and a voice recognition test. Procedure was exactly the same as for Experiment 1, except that participants proceeded at their own pace. The importance of taking breaks was emphasized but not enforced. Participants performed the auditory behavioral experiment at a computer, which played the words through headphones. The volume was automatically set at 90% of the computer’s maximum, which was deemed to be a comfortable volume.

27.3 Analysis

In the word recognition test, performance (hits) was analyzed using a mixed ANOVA. Language group was the between-subjects factor and voice condition was the within-subjects factor.

In the source recognition test, two analyses were conducted. First, an independent-measures ANOVA compared same speaker performance (hits) among the three language groups. Second, false alarms were compared for the three different speaker conditions (see Experiment 1 for explanation) using a mixed ANOVA with language group as the between-subjects factor and voice condition as the within-subjects factor.
Results

Participants in the three groups were balanced; there were no significant group differences between vocabulary score ($p = 0.24$), age ($p = 0.17$) or gender distribution ($p = 0.27$).

Figure 17 shows the results of the word recognition test. In this test, the main effect of language group was not significant, $F < 1$. There was a significant main effect of voice congruency, $F(3,183) = 26.54$, $p < 0.001$, $\eta^2 = 0.30$ (linear trend: $F(1,61) = 17.82$, $p < 0.001$, $\eta^2 = 0.23$; quadratic trend: $F(1,61) = 65.07$, $p < 0.001$, $\eta^2 = 0.52$). Pairwise comparisons indicated that performance in the same speaker condition was better than in each of the three different speaker conditions, that performance in the different gender/same accent condition was significantly better than in the same gender/different accent condition and that performance in same gender/different accent condition was significantly worse than in the different gender/different accent condition (all $p < 0.05$). There was also a significant interaction between voice condition and language group in the word recognition test, $F(6,183) = 2.79$, $p = 0.014$, $\eta^2 = 0.08$ (quadratic trend: $F(2,61) = 7.79$, $p = 0.001$, $\eta^2 = 0.20$). This interaction appeared to be driven by two observations. First, Chinese-speaking bilinguals performed better in conditions when only accent was reinstated between study and test. Second, the other bilingual group showed decreased performance in the same gender/different accent condition.

In light of the significant interaction of voice condition and language group, I then tested for simple effects. I ran individual repeated measures ANOVAs for the three language groups, in an effort to discern the effects of voice congruency in each group. In the monolingual group, there was a significant effect of voice congruency, $F(3,60) = 10.24$, $p < 0.001$, $\eta^2 = 0.34$ (quadratic trend: $F(1,20) = 53.05$, $p < 0.001$, $\eta^2 = 0.73$). Pairwise comparisons indicated that performance in the same speaker condition was significantly better than in the different gender/same accent and the same gender/different accent conditions (both $p < 0.001$). In addition, performance in the different gender/different accent condition was better than in the same gender/different accent condition ($p = 0.009$). For the Chinese-speaking bilingual group, voice congruency produced a significant main effect, $F(3,57) = 4.33$, $p = 0.009$, $\eta^2 = 0.19$ (linear trend: $F(1,19) = 7.21$, $p = 0.015$, $\eta^2 = 0.28$). Pairwise comparisons revealed that performance in the same speaker condition was superior to performance in the same gender/different accent condition ($p = 0.002$) and that performance in the different gender/same accent condition was
better than in the same gender/different accent condition ($p = 0.006$). Lastly, there was a main effect of voice congruency for the other bilingual group as well, $F(3,66) = 17.22, p < 0.001, \eta^2 = 0.44$ (linear trend: $F(1,22) = 7.89, p = 0.01, \eta^2 = 0.26$; quadratic trend: $F(1,22) = 45.83, p < 0.001, \eta^2 = 0.68$). Pairwise comparisons indicated that performance in the same speaker condition was better than in the different gender/same accent condition ($p = 0.004$) and the same gender/different accent condition ($p < 0.001$). Performance was also better in the different gender/same accent condition than in the same gender/different accent condition ($p = 0.001$) and performance in the different gender/different accent condition was better than performance in the same gender/different accent condition ($p < 0.001$).

In the source recognition test (see Figure 18), results indicated no significant difference among the language groups for same speaker performance, $F < 1$. When comparing the three different speaker conditions, no significant difference between language groups was found, $F(2,61) = 2.08, p = 0.13, \eta^2 = 0.064$. However, and unlike in Experiment 1, there was a significant main effect of voice condition among false alarms of the different speaker trials, $F(2,122) = 9.79, p < 0.001, \eta^2 = 0.14$ (linear trend: $F(1,61) = 12.81, p = 0.001, \eta^2 = 0.17$; quadratic trend: $F(1,61) = 6.64, p < 0.012, \eta^2 = 0.10$). Specifically, there were significantly more false alarms for the different gender/same accent condition than for the same gender/different accent and the different gender/different accent conditions (both $p < 0.05$). There was no significant interaction between language group and voice condition, $F < 1$.

29 Discussion

Experiment 4 sought to extend the characterization of voice congruency effects by looking at potential influence of bilingualism. Using three language groups – monolinguals, bilinguals who had specific knowledge of the same language as the speakers who recorded stimuli, bilinguals of any other language (i.e., non-specific to the language of the speakers who recorded stimuli) – we were able to investigate three points of interest. First, we investigated whether there were overall differences in performance on the two recognition tests between monolinguals and bilinguals. Second, and of more interest, we were able to ascertain whether levels of voice congruency had a differential effect on monolinguals vs. bilinguals. Third, by comparing the two bilingual groups, we could investigate whether potential voice congruency
differences were the result of specific language sensitivity or of more general executive benefits associated with bilingualism.

Results indicated no overall main effect of language group in either recognition test. Since prior work had suggested that bilingualism would impact recollection but not familiarity (Wodniecka et al., 2010), it may be that the present design lent itself to greater use of familiarity in recognition. This seemed less likely in the source recognition test, but it should be noted that participants knew that all test words were “old” in the source task. Additionally, participants simply gave a same/different voice judgment (i.e., they did not have to specifically identify the voice speaking each word). This may have lessened the need for specific recollection in the source task. However, since this experiment was behavioral, we could not investigate the neural correlates underlying memory function in this case. Future investigations should compare monolinguals and bilinguals using ERPs, in order to observe potential modulations of the early frontal familiarity effect and the parietal old/new recollection effect.

With respect to the second point of interest, we were interested in whether voice congruency effects represent general acoustic sensitivity or whether language skills confer sensitivity. In the word recognition test, voice congruency did modulate performance in all three language groups. Results indicated more fine-grained voice congruency effects than in Experiment 1, which may have been the result of increased power achieved by using a larger sample. Interestingly, voice congruency effects manifested differently between the three language groups. Specifically, accent congruency appeared to preferentially benefit bilinguals of the same language as the speakers who recorded the stimuli. This was indicated by the similarity of performance in the same speaker and the different gender/same accent conditions, for the Chinese-speaking bilinguals only. For the other two language groups (monolinguals and other bilinguals), pairwise comparisons did find a significant difference between the same speaker and different gender/same accent conditions. Skills in a particular language conferred a word memory benefit when that accent was congruent between study and test, which suggests some specific benefit of language training.

There were two findings reported here that were not expected. First, in the word recognition test, performance in same gender/different accent condition was significantly worse than in the different gender/different accent condition. We suggest that increased performance in the different gender/different accent condition represents a kind of novelty effect (Tulving & Kroll,
1995). If so, then the false alarm rate for the different gender/different accent condition would be relatively low, in the word recognition test, compared to the other three voice conditions. However, the present design does not lend itself to such an analysis and therefore the reason for increased performance in the different gender/different accent condition cannot be convincingly ascertained.

The second unexpected finding was a main effect of voice condition on the false alarm analysis in the source recognition test. We had not previously found such a distinction in Experiment 1, and suggest that the finding might be the result of increased power in this experiment (i.e., a much larger sample size). A significant increase in false alarms in the different gender/same accent condition relative to the other two different speaker conditions, in the source recognition test, provides converging evidence that suggests that accent congruency might result in greater perceptual fluency than gender congruency between study and test. Recall that results from Experiment 1 also suggested that accent plays an especially important role in voice congruency effects. However, as for Experiment 2, it is worth noting that the accented voices used to create stimuli were also older than the non-accented voices. It is, therefore, unknown at this point whether a greater benefit of accent congruency (as compared to gender congruency) is the result of accent being an especially salient acoustic cue or whether this effect is simply quantitative and reflects congruency on multiple vocal parameters (i.e., accent and age).

Experiment 4 did also have a limitation that should be pointed out. Participants were deemed monolingual or bilingual based on their language background, and participants with intermediary levels of bilingualism were omitted from analysis in order to minimize this potential confound, yet the language environment was not taken into account in this experiment. That is, all participants were undergraduate students at the University of Toronto, in downtown Toronto. In such a cosmopolitan society, it seems reasonable to assume that participants in all language groups have been regularly exposed to various accents. Therefore, monolinguals were likely not so naïve in terms of multiple language schemas, for example. A future study should address this potential confound by studying monolinguals with limited exposure to accented voices.
30 Interim Conclusion

In conclusion, bilingualism did not result in better overall performance in this experiment. However, this study did provide an indication for a differential effect of bilingualism on voice congruency in explicit word memory, such that accent congruency conferred a benefit in word recognition for bilinguals of the same language as the speakers who had recorded the stimuli.
Chapter 8
General Discussion and Conclusions

The purpose of this work was to investigate how voice is used as an auditory memory cue. A series of experiments were designed to build on our understanding of voice as a context cue. The experiments utilized the same set of stimuli (except Experiment 1a, which served as a control study) to investigate various types of memory, using different paradigms and different types of participants. Experiments were behavioral and/or utilized EEG recordings in an effort to understand the neural correlates of voice congruency in memory. Crucially, past studies investigating the effects of voice congruency on word memory utilized stimuli recorded by non-accented voices that varied primarily based on pitch. The stimuli used in this set of experiments were created using voices that varied on several parameters, most notably accent and gender. As such, one major purpose was to confirm that models of voice representation in memory (e.g., Pisoni, 1993) do, in fact, apply to voices rather than to just one parameter of voice, namely pitch.

In Experiment 1, participants were tested for word and voice recognition using a block design. Voice reinstatement (i.e., same speaker) was found to benefit performance on both types of memory tests. In addition, congruency of the accent parameter of voice appeared to boost performance on the word recognition test. These findings confirmed that same-voice facilitation, which is expected based on previous literature, occurs when multiple, substantial voice attributes are varied between study and test (i.e., accent).

Furthermore, analysis of ERP modulations reflective of recollection (i.e., the left parietal old/new effect) and familiarity (i.e., the frontal familiarity effect) showed sensitivity to voice congruency between study and test. The parietal old/new effect showed a non-significant gradient reflective of voice congruency, where the same speaker condition resulted in the most positive wave. The frontal familiarity deflection showed a significant voice effect. In addition, the voice recognition test revealed that the late frontal positivity, which has been known to index post-retrieval work and/or decision/judgment processes, was sensitive to voice reinstatement at test.
A potential caveat of Experiment 1 was that the same speaker trials always utilized the same exact recording, making it possible that the observed voice reinstatement benefit had merely been the result of acoustic repetition rather than voice representation. Experiment 1a had participants do the same tasks as in Experiment 1, but with half of the same speaker trials being the same acoustic recording and the other half being a second recording of the word by the same voice. Results indicated no difference between the two same speaker versions in either test, ruling out this potential confound.

Experiment 2 was designed to confirm and extend our findings from Experiment 1. Participants performed a continuous word recognition test, thereby extending the investigation of voice congruency in word recognition to another paradigm; results indicated that voice reinstatement facilitated explicit word memory in this paradigm as well. In addition, Experiment 2 allowed an investigation into the importance of focusing participants’ attention to voice at study, on subsequent voice congruency effects at test. Though participants had been told to pay attention to the speaker voice at study in Experiment 1, they were not told to attend to voice in Experiment 2. Results indicated a benefit of voice reinstatement (i.e., same speaker) in both experiments, but only Experiment 1 provided evidence for more fine-grained voice congruency effects. While the power of Experiment 1 was somewhat limited, pairwise comparisons indicated some differentiation among voice congruency effects both behaviorally and in the frontal ERP trace. Visual inspection also suggested fine-grained voice congruency effects in the parietal old/new modulation, but pairwise comparisons within this gradient failed to reach significance. In Experiment 2, behavioral results indicated only voice reinstatement benefits at longer lags, while potentially broader voice congruency facilitation was masked at lag 2 (when the trial was still in participants’ active awareness) by ceiling effects. In terms of the ERP trace, there was no separation at the parietal old/new effect based on voice, though congruency did appear to affect the frontal familiarity deflection. Taken together, these results suggest that voice congruency effects can be obtained using various recognition memory paradigms and that the specificity of voice congruency effects depends on processing at study. More detailed voice congruency benefits require either attentional focus to voice and/or deeper perceptual processing at study.

Experiment 3 attempted to extend the characterization of voice effects to another type of memory. Since implicit memory is known to be more sensitive to perceptual features of the
stimuli, we expected more detailed voice congruency effects in the word identification task than in our explicit memory investigations. However, since the nature and difficulty of the task was quite variable between Experiment 3 and Experiments 1 and 2, it is inappropriate to make a direct comparison across investigations. To that end, Experiment 3 included an explicit (recognition) memory task and utilized a between-subjects design. The purpose was to have participants perform a common encoding task and then to perform either an implicit or an explicit memory task at test, making possible a direct comparison of retrieval processes. Unfortunately, the appropriate method of analysis of priming in Experiment 3 was ambiguous. Nevertheless, a voice reinstatement benefit was observed for both implicit and explicit groups.

Lastly, in a between-subjects investigation, we extended our understanding of the mechanism of voice effects by comparing groups of participants that differed based on language skills. The most intriguing finding was that accent congruency at test conferred a similar benefit to word recognition as voice reinstatement, in the Chinese-speaking bilingual group but not in the other two groups. This provided evidence of accent sensitivity in memory for people who are fluent in that particular language. As such, language skills appear to confer a benefit that is specific to particular voice congruency conditions rather than a more general context congruency benefit that might have been related to executive control.

Overall, this set of experiments points to several key developments in understanding voice congruency in memory. We can conclude that the extent and specificity of voice congruency effects are dependent on the task/paradigm used. Correspondingly, ERP modulations reflective of various memory components can be found in auditory studies looking at voice congruency, but their characterization is highly dependent on task/paradigm. An important factor in such auditory investigations is the allocation of participants’ attention, particularly at study. The general finding that attention (or some extended processing) is required to observe specific voice effects might reflect the presence of a quantitative memory effect. For example, attention might not be explicitly needed to observe such effects if other components of the design make perceptual characteristics salient enough. That is, allocating attention may just be the easiest/strongest way to engender the processing required to obtain voice effects. This can be inferred from the present set of experiments, for recognition tests.

In terms of automaticity of voice encoding and its use as a retrieval cue, the current findings support the notion that voice is an integral context cue. All of the current experiments found a
voice reinstatement effect, suggesting that the benefit of reinstating the same voice at test appears to be robust at least when using these stimuli. On the other hand, a more general benefit of voice congruency appeared to be paradigm-specific: Experiment 2, which focused less on perceptual attributes, produced a robust same-speaker benefit but not more generalized voice congruency effects. Speculatively, it may be that voice is automatically encoded and retrieved with a memory trace, so that same-speaker trials reflect this automaticity. Partial congruency trials, however, do not benefit from automaticity because they are not an exact match; instead, they appear to involve more effortful processing (e.g., attempting to match the test stimulus to the memory trace). The ERP finding of an early voice effect (i.e., before recollection) further supports the notion of voice automaticity in auditory memory. The present electrophysiological findings suggest that partial congruency might also be involved in pre-recollective processing of voice, but more powerful studies are needed to confirm this possibility. Hypothetically, it may be that partial congruency is a quantitative component which must meet a certain threshold in order to facilitate memory. Extended perceptual processing – for example, attention allocation – might provide a sufficiently strong sense of familiarity so as to promote recognition memory.

Unfortunately, analysis and interpretation of implicit memory results are less straight-forward, and therefore a suggestion as to the nature of voice congruency effects in implicit memory cannot be inferred based on these results. Nevertheless, it is interesting to note that the effect of voice congruency was more robust than the (possible but insignificant) priming effect in Experiment 3, suggesting that perceptual congruency is very salient in word identification tasks. More research is needed to investigate this further, but the current experiment serves as a point for relevant discussion around two critical factors involved in such experiments: Type of analysis employed in priming studies, and type of noise used in word identification tasks investigating voice congruency.

A common finding in this set of experiments was that fine-grained voice congruency benefits, where they did exist, were driven by a benefit of accent rather than gender congruency between study and test. This could suggest that accent is a more salient vocal attribute than pitch, in memory. However, such a conclusion is premature in light of the potential confound of age differences between accented and non-accented voices. Alternatively, it may have been that voice is a quantitative context cue, wherein reinstatement or congruency of an increasing number of parameters leads to increased facilitation at test. This clarification is especially
important for the interpretation of a potential accent benefit in Experiment 4. Since the accent benefit was specific to bilinguals of the same language as the voice speaking the accented words, it seems more likely that the benefit involves some sort of specificity, rather than simply a general sensitivity to detailed vocal information. The latter would have resulted in a similar benefit for bilinguals of other languages, unless this sensitivity exists in parallel to an accent specificity, both of which sum to some sort of quantitative voice effect. Clearly, understanding the subcomponents of human voice and how they are processed in memory is an avenue for future research, with important implications for language learning and memory.

To that end, future studies should address our accent/age confound by utilizing voices that vary orthogonally on multiple vocal parameters. Such an investigation would allow for clarification as to the extent to which voice is the quantitative amalgamation of various paralinguistic attributes and if/which of these attributes are stronger cues than others (qualitative distinctions between the attributes). Overall, voice appears to be encoded in multiple attributes that are either encoded together into one representation or separately into many. The classification of voice as a holistic or composite cue is largely irrelevant from a functional standpoint; it seems likely that the number and/or salience of voice characteristics determine its benefit as a reinstated or congruent context cue.
References


doi:http://dx.doi.org/10.1017/S1366728908003477


doi:http://dx.doi.org/10.1016/j.neuroimage.2012.07.047


Copyright Acknowledgements

A significant portion of Chapter 1 was taken from a paper that has been accepted for publication at the International Journal of Psychophysiology:


A significant portion of Chapter 3 was taken from:


A significant portion of Chapter 5 was taken from:

Figure 1. Group Mean Accuracy for Experiment 1. Note that DgSa = different gender/same accent condition, SgDa = same gender/different accent condition and DgDa = different gender/different accent condition. Error bars represent standard error. A) Accuracy in the Word Recognition Test. Results indicated that performance in the same speaker condition was significantly better than the other three voice conditions, and that performance in the DgSa condition was significantly better than the SgDa and DgDa conditions. B) In the Source (Speaker) Recognition Test, performance in the same speaker condition was significantly above chance level. C) In the Source (Speaker) Recognition test, there was no significant difference between the three different voice conditions.
Figure 2. Group Mean Event-Related Potentials for the Word Recognition Test in Experiment 1.
Figure 3. Left Parietal Old/New Effect in the Word Recognition Test in Experiment 1. Note that SS = same speaker; DD = different gender/different accent; NW = new word. Results indicated that the Same Speaker, DgSa and SgDa conditions each produced a more positive deflection than the new word condition. In addition, there were significant linear and quadratic trends.
Figure 4. Group Mean Event-Related Potentials for the Source (Speaker) Recognition Test in Experiment 1. Note that the three different speaker conditions are collapsed into the “Different Speakers” condition shown here.
Figure 5. Right Frontal Effect in the Source Recognition Test in Experiment 1. Results indicated a significantly more positive deflection for the same speaker trials (compared to the different speaker trials) over right frontal sites, with this effect peaking at 722ms. Note: significant effects are demonstrated using star(s) enclosed in a box over the corresponding electrode.
Figure 6. Group Mean Accuracy in the Word Recognition Test in Experiment 1a. Error bars represent standard error. Results indicated no significant difference between the two same-voice conditions.
Figure 7. Group Mean Accuracy in the Source Recognition Test in Experiment 1a. Error bars represent standard error. Results indicated no significant difference between the two same-voice conditions.
Figure 8. Group Mean Accuracy and RT in Experiment 2. Error bars represent within-subjects standard error. A) Accuracy in the Continuous Word Recognition Test. There was a significant main effect of voice condition, reflecting a superiority of performance in the same speaker condition compared with the other three voice conditions. There was a significant main effect of lag, with all three lags being significantly different from each other. When analyzing simple effects, different patterns of voice effects were seen as a function of lag. Voice condition produced significant linear trends and significant main effects at lags 8 and 16, with the same speaker condition showing better performance than the other three conditions (which were not significantly different from each other). At lag 2, there was no significant voice effect – perhaps due to ceiling effects. B) RT in the Continuous Word Recognition Test. There was a significant main effect of voice condition, with performance being faster for the same speaker condition relative to the other three voice conditions. In addition, RTs were shorter for the different gender/same accent condition than the same gender/different accent and different gender/different accent conditions. There was also a main effect of lag, with performance being faster at lag 2 than at lags 8 and 16. There was no significant interaction between voice condition and lag.
Figure 9. Old/New Effect in Experiment 2. A) Scalp topography showed a significant positivity over (left) parietal sites. Note: significant effects are demonstrated using star(s) enclosed in a box over the corresponding electrode. B) Shown at P1, the old/new effect began at around 380ms and peaked at around 700ms post-stimulus.
Figure 10. Voice Effect in Experiment 2. A) Scalp topography reflected a significant effect of voice reinstatement. Note: significant effects are demonstrated using star(s) enclosed in a box over the corresponding electrode. B) Voice reinstatement effect shown over F5. C) Voice congruency resulted in a significant linear trend. Mean amplitudes were averaged over F7, F5, F3, AF7, AF3 and FC5 at 200-600ms. Note that SS = same speaker; DS = different gender/same accent; SD = same gender/different accent; DD = different gender/different accent.
Figure 11. Lag Effect at the Parietal Old/New Modulation, in Experiment 2. Recorded over bilateral parietal sites (i.e., PO3 and PO4) showing all three lags, with all voice conditions collapsed at each lag. Results showed a significantly earlier and larger peak for lag 2, as compared with lags 8 and 16.
Figure 12. Voice Effects at Lag 2, in Experiment 2. A) Scalp topography reflected a significant effect of voice reinstatement. Note: significant effects are demonstrated using star(s) enclosed in a box over the corresponding electrode. B) Voice reinstatement effect shown over F7. C) Voice congruency resulted in a significant linear trend. Mean amplitudes were averaged over F7, F5, AF7 and FC5 at 300-700ms. Note that SS = same speaker; DS = different gender/same accent; SD = same gender/different accent; DD = different gender/different accent.
Figure 13. Voice Reinstatement at Lags 8 and 16, in Experiment 2.  A) Scalp topography showed a significant negativity over left frontal sites when voice was reinstated. Note: significant effects are demonstrated using star(s) enclosed in a box over the corresponding electrode. B) Shown at F7, the voice reinstatement effect was largest at around 450ms.
A) Old/New Effect.

B) Voice Effect.

Figure 14. Source Analysis for Experiment 2. The letters L and R refer to the left and right hemisphere, respectively.
Figure 15. Voice Congruency in the Explicit Group, in Experiment 3. The blue bars correspond to Analysis 1 and the red bars correspond to Analysis 2. Note that “corr at test” refers to correct test trials, regardless of whether the words were correct at study, or not. DgSa = different gender/same accent condition, SgDa = same gender/different accent condition and DgDa = different gender/different accent condition. Error bars represent standard error. The dotted line across the graph shows the group mean average proportion of new words correctly identified and correctly said to be new at test, for the explicit group.

Using Analysis 1, there was a significant linear trend as well as a significant main effect of voice condition, with performance being significantly better in the same speaker condition than in each of the other three voice conditions. Analysis 2 also produced a significant linear trend and a significant main effect of voice condition, but pairwise comparisons were different using this approach. Specifically, performance was weaker in the same gender/different accent condition than in the same speaker and the different gender/different accent conditions.
Using Analysis 3, there was a significant linear trend and a significant main effect of voice condition, with pairwise comparisons indicating that word identification was better in the same speaker condition compared with the three different speaker conditions. Using Analysis 4, there was a significant quadratic trend and a significant main effect of voice condition. Pairwise comparisons showed superiority of the same speaker condition in comparison to the different gender/same accent and the same gender/different accent conditions, as well as better performance for the different gender/different accent condition than the different gender/same accent condition.
Figure 17. Voice Congruency in the Word Recognition Test, in Experiment 4. SgSa = same speaker, DgSa = different gender/same accent condition, SgDa = same gender/different accent condition and DgDa = different gender/different accent condition. Error bars represent standard error. There was a significant main effect of voice condition as well as an interaction between voice condition and language group. For each language group, there was a significant main effect of voice condition, but pairwise comparisons suggested that the Chinese-speaking group obtained a greater benefit from accent congruency. Specifically, performance in the same speaker condition was significantly better than the different gender/same accent condition for the monolingual and other bilingual group, but not for the Chinese-speaking group.
Figure 18. Voice Congruency in the Source Recognition Test, in Experiment 4. SgSa = same speaker, DgSa = different gender/same accent condition, SgDa = same gender/different accent condition and DgDa = different gender/different accent condition. Error bars represent standard error. A) Results indicated no significant difference in same speaker performance between the three language groups. B) Results indicated significantly more false alarms for the DgSa condition than the other two, with no interaction between language group and voice condition.