METAMEMORY FOR WORDS IN NOISE

by © Sarah Oates

A Thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science in Experimental Psychology

Memorial University of Newfoundland

July 2014

St. John's Newfoundland and Labrador

ABSTRACT

This study used ease of learning (EOL) judgments to examine whether people recognize the effect that noise has on their recall performance. In four experiments, participants studied lists of visually-presented words while ignoring different types of irrelevant background noise. In Experiment 1, EOLs and recall were measured in three conditions: quiet, white noise, and irrelevant speech (cafeteria babble). Surprisingly, the typical irrelevant speech effect (ISE) was not observed. Experiment 2 determined that the lack of an ISE was not due to the additional EOL judgment. In Experiment 3, a different irrelevant speech sound file (a single talker speaking German) elicited a typical ISE. Finally, in Experiment 4, a robust ISE was found. The metamemory data showed that participants can recognize the effect of irrelevant speech prior to being tested but incorrectly believe white noise has a disadvantageous effect on their memory. The results are discussed in terms of Koriat's (1997) cue utilization theory.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Aimee Surprenant for giving me the opportunity to achieve this master's thesis. Her expertise, guidance, and continued support have been instrumental in the success of this project. She has been an inspiration and a mentor throughout this process. I would like to thank my committee member Dr. Ian Neath for his technical aid in programming my experiments, as well as his helpful insight and suggestions into the conceptualization and execution of this project. I would also like to thank my committee member Dr. Kathleen Hourihan for lending her advice, expertise and continued support throughout this process. Thank you to both of you for graciously agreeing to act as members of my committee and taking the time and effort to read and review this thesis. I would like to extend my appreciation to the staff at MUN, particularly Brenda Crewe for keeping everything running smoothly. Finally, I would like to acknowledge all of the participants who volunteered to take part in my study, without whom none of this research would have been possible.

Table of Contents

ABSTRACT ii

ACKNOWLEDGEMENTS iii

Table of Contents iv

List of Tables v

List of Figures vi

Introduction 1

Experiment 1 15

Methods 15

Results 17

Discussion 22

Experiment 2 23

Methods 23

Results 24

Discussion 25

Experiment 3 26

Methods 26

Results 27

Discussion 28

Experiment 4 28

Methods 29

Results 29

Discussion 34

General Discussion 35

Conclusion 41

References 43

List of Tables

Table 1. Mean Proportion Correct Serial Recall of Words in Quiet, White Noise, and Irrelevant Speech by Experiment (with Standard Deviation in Parentheses) 22

List of Figures

Figure 1. Mean proportion correct by noise condition (error bars are standard error of the mean) 18

Figure 2. Mean global EOL judgment rating by noise condition (error bars are standard error of the mean). Note: higher values indicate more predicted disruption 19

Figure 3. Mean frequency estimates by noise condition (error bars are standard error of the mean) 19

Figure 4. Calibration by trial and noise condition 21

Figure 5. Calibration by block and noise condition 22

Figure 6. Mean proportion correct by noise condition (error bars are standard error of the mean) 24

Figure 7. Mean proportion correct by noise condition (error bars are standard error of the mean) 28

Figure 8. Mean proportion correct by noise condition (error bars are standard error of the mean) 30

Figure 9. Mean global EOL judgment rating by noise condition (error bars are standard error of the mean). Note: higher values indicate more predicted disruption 31

Figure 10. Mean frequency estimates by noise condition (error bars are standard error of the mean) 32

Figure 11. Calibration scores by trial and noise condition 33

Figure 12. Calibration scores by block and noise condition 34

Metamemory For Words In Noise

In today's society, there are a variety of noises that surround us. Whether it is the noise of traffic, people talking, our car radio, or mechanical hums, these noises are present as we perform our daily tasks. The current body of literature has offered many important insights on how noise can directly affect our memories. Previous research has shown that there are overall decrements in memory when a task is performed in the presence of irrelevant background speech stimuli: This is called the irrelevant speech/sound effect, or ISE (Baddeley & Salamé, 1986; Colle & Welsh, 1976, Salamé & Baddeley, 1982). This effect, although not completely understood, is replicable and reliable. Even though approximately 85% of individuals show an effect of irrelevant speech, self-report measures of disruption are not predictive of the effect (Ellermeier & Zimmer, 1997). Thus, people's conscious knowledge (or "metamemory") of how much background noise affects their memory seems to be quite poor. The purpose of the current project is to further explore metamemory for the effect of irrelevant speech using on-line measures such as ease of learning judgments (EOLs; Nelson & Narens, 1990). EOLs are judgments that are made in advance of acquisition, and they are a prediction of what information will be the easiest to learn. In the following, the literature on the irrelevant sound and irrelevant speech effects will be reviewed. Next, a brief outline of research on metamemory will be provided. Finally, a description of the experiments will be given.

The Irrelevant Speech Effect

The irrelevant speech effect refers to the disruption in recall when one attempts to memorize stimuli in the presence of irrelevant speech stimuli; people tend to remember significantly fewer items when there is irrelevant speech playing in the background compared to when it is quiet. This effect is of particular interest as it occurs across modalities: the effect is observed when stimuli are presented visually, even though the distraction is auditory. In studies investigating this effect, participants are asked to attend to a set of visually presented items, which usually consist of a short list of 7 to 12 items such as words, numbers, or letters. The stimuli may be presented under various speech noise conditions, such as unintelligible speech babble, speech in a language the participants do not know, steady-state speech (unvarying or repetitive speech sounds), changing-state speech (irrelevant speech sounds that change over time), etc. Silence is the universal control group to which the noise conditions are compared. Participants are then typically asked to serially recall the words (recall the words in order).

Colle and Welsh (1976) presented participants with lists of eight written consonants either in silence or with German speech playing in the background. German was an unfamiliar language to all participants in the study, and participants were instructed to ignore any noise in the background. Results showed that participants had lower rates of recall when the German speech was playing in the background compared with when there was only silence. Colle and Welsh's (1976) study was followed up by Salamé and Baddeley (Baddeley & Salamé, 1986; Salamé & Baddeley, 1982, 1987) who explored the effect in detail. They ultimately concluded that the effect arose from a conflict in the phonological loop aspect of working memory (Baddeley, 1986). In particular, Baddeley and colleagues (see Baddeley, 2000) argued that the irrelevant speech interfered with the encoding of phonological information in the phonological store, displacing some of the to-be-remembered items. However, this explanation of the ISE is controversial, thus a more comprehensive look at theoretical considerations of this effect will be discussed below.

The Irrelevant Sound Effect

The phonological store explanation of the ISE was seriously compromised by findings that it is not just speech sounds that can impair memory (Jones, 1999). In some cases, there can be a decrease in performance even if the auditory sound has no phonological content. Similar to the irrelevant speech effect, this irrelevant sound effect is more general in that nonspeech, sounds such as tones, can also impair memory in some cases. Various noise conditions may be used to demonstrate this effect, such as pink noise (a variant of white noise), broadband noise, sine-wave speech (synthesized speech), tones, and many others. However, the research shows that in order to have a detrimental effect, this sound must be in a continually changing state, such as sounds with different consonants or tones, as little to no disruption is observed with steady state sounds such as continuous broadband noise (Jones, 1999; Neath, 2000). Additionally, some noise conditions elicit a much larger effect than others, with irrelevant speech yielding the most robust finding (Neath, 2000). One must also note that some research has shown an irrelevant sound effect regardless of the modality of item presentation. For example, it is still present when auditory items are serially recalled (LeCompte, 1996; Salamé & Baddeley, 1982).

Serial Recall

Studies have also investigated whether memory tasks other than serial recall are affected by irrelevant speech. Neath (2000) concluded that the effect can be observed on at least some tasks that did not use serial recall. Salamé and Baddeley (1990) first looked at whether serial recall was necessary to observe the ISE. They used free recall instead of serial recall, and did not observe an ISE. However, in a series of experiments, LeCompte (1994) demonstrated that maintenance of order was not a requirement for the ISE. An ISE was found in tasks of recall and recognition. For example, in a test of free recall, participants viewed lists of 12 letters under conditions of irrelevant speech, white noise, and silence. Immediately afterward, participants were asked to recall all of the letters they could remember, in any order. Results showed that the irrelevant speech significantly impaired recall performance compared to the quiet control condition.

However, with such short lists, participants could have been engaging in serial recall without explicit instruction, rendering the inferences made from this study unclear. Beaman and Jones (1997) replicated this study with one important difference: participants were given the additional task of articulatory suppression (repeating irrelevant sounds out loud that prevent rehearsal of the to-be-recalled items). Participants were asked to articulate a short series of letters during the stimulus-presentation stage to suppress serial rehearsal. With this addition, the ISE disappeared, which casts doubt on the LeCompte (1994) study, and makes it unclear whether the ISE is present in conditions of free recall.

In order to determine whether this ISE during free recall was replicable, Stokes and Arnell (2012) performed a series of experiments that tested whether the ISE could be observed using a surprise recognition task in which order did not matter and rehearsal was unlikely. Experiment 1 replicated the findings of LeCompte (1994), except using a very large list of words (100 words) compared with LeCompte's short list of 12 words. Participants' memory was tested after the presentation of all items, therefore it is rather unlikely that participants engaged in serial rehearsal. To further examine the possibility that participants were engaging in serial rehearsal, Experiment 1 was replicated, except participants were not told that they would be asked to recall the items. Participants were told that the purpose of the experiment was to examine categorization speed and accuracy under various distracting conditions. Therefore this was a 'surprise' recall task, and the idea of item rehearsal was eliminated. Results showed a significant ISE effect on this surprise nonserial recognition task. Experiment 3 replicated Experiment 1 and 2, except foreign speech was used as the irrelevant sound condition. In a further experiment, Experiment 4 compared the surprise recall task for changing-state and steady-state sounds. Results showed equally large ISE for both conditions, demonstrating that any form of irrelevant sound can potentially disrupt memory without serial recall. This argues that it is possible to obtain an ISE with recall methods other than serial recall. Additionally, this finding suggests that the ISE may result from a conflict in meaning of the stimuli as opposed to order information, but requires further investigation and replication of these results, as they are in direct contrast with Salamé and Baddeley (1990).

In addition to this research, Jones, Miles, and Page (1990) have also shown that there can be a deleterious effect of irrelevant speech without serial recall. In a series of experiments, it was shown that irrelevant speech can negatively affect a proofreading task, whereby participants detect fewer non-contextual errors than contextual errors. This series of experiments also showed that in order to elicit the effect, the speech stimuli had to be meaningful. Thus it is possible that the ISE may involve semantic confusion and not order information as it has been previously attributed. These results are interesting, though, as they provide new considerations for the cognitive locus of impairment of the ISE.

Theoretical Considerations

Phonological Store Hypothesis

Many theories have attempted to account for the ISE, the most prominent of which is the phonological store hypothesis. In working memory, the phonological loop, which is responsible for storing and rehearsing speech-based information (Baddeley, 1986), is comprised of the phonological store and the articulatory control process. The phonological store holds traces of acoustic or speech-based material. It has been suggested that irrelevant speech can interfere with the items in the phonological store in some unspecified way, which causes the decrement in memory observed in the ISE. Although not stated, it has been assumed that the interference is based on the overlap of phonological information, such as content interference.

There are various issues with this account, as outlined by Neath (2000). Firstly, the lack of specificity with regard to the explanation of the irrelevant speech effect is a concerning weakness of this account. This hypothesis does not specify how the irrelevant speech interferes with the phonological store; for example, one might logically propose that the decrement in memory observed with the ISE will be more pronounced when the phonemes in the irrelevant speech stimuli are similar to those in the stimuli to be remembered. However, this interaction has not been supported in the literature (LeCompte & Shaibe, 1997). Another issue with the phonological store hypothesis relates to the word length effect (the finding that lists of short words are better recalled than lists of long words). The word length effect is believed to arise through sub-vocal rehearsal, controlled by the articulatory control process, not the phonological store, and yet accompanying irrelevant speech abolishes the word length effect (Neath, Surprenant, & LeCompte, 1998). It is unclear how the phonological store would account for this, considering that irrelevant speech does not prevent rehearsal. This hypothesis also does not address many other factors that pertain to the irrelevant speech effect, such as serial position effects. Finally, finding an ISE with non-speech irrelevant information, which cannot enter the phonological store because it is not phonological, further adds to the aspects of the ISE that cannot be explained by this hypothesis. It proves quite difficult to account for the irrelevant speech effect using this hypothesis, and thus a more complete explanation of the ISE is necessary. Other hypotheses with a more comprehensive explanation are needed to fully explain this effect.

Changing-State Hypothesis

One theory that does offer a more complete explanation of the ISE is the changing-state hypothesis, a part of Jones's Object-Oriented Episodic Record (O-OER) model (Jones, 1993; Jones & Macken, 1993). In the past, it has been shown that irrelevant speech, strictly speaking, is not enough to produce the ISE. Miles, Jones and Madden (1991) demonstrated that there was no decrease in memory when the irrelevant speech was simple, unvarying or repetitive speech sounds, for example the letter 'C' repeatedly presented. This is type of stimulus is referred to as being steady-state, as it is unchanging. However, the typical disruption of the ISE is found when a sequence of as small as four syllables is repeated in the same order, for example "C, H, J, U". This type of stimulus is referred to as changing-state, as it is varied and more complex than steady-state stimuli. The explanation for these subtle differences that produce markedly different effects may be attributed to the likelihood of an attentional switch away from the to-be-remembered items to the irrelevant stimuli. Changing-state material regularly provides a novel auditory stimulus that may involuntary grab an individual's attention, whereas steady-state material involves a regular presentation of the same auditory stimulus, which can be quickly habituated to and thus there is no further disruptive effect. (Jones, Beaman & Macken, 1996).

To explain more specifically how irrelevant speech may disrupt recall, the O-OER model proposes that both visual and auditory stimuli are represented using amodal, abstract representations called objects (Jones, 1993). The inputs (regardless of modality) are represented using streams, whereby both the irrelevant sounds as well as the to-be-remembered items each form a different stream of objects, which have pointers. A pointer is used to help encode order information. With multiple items and pointers, a stream is created. For example, in list of seven items that are to be serially recalled, each word is represented by an object, and each word has a pointer that indicates to the next word in the sequence. The stream of objects and pointers is followed in order to accurately recall the list of words. Similarly, changing-state sounds such as irrelevant speech also create their own set of objects and pointers. Errors in recall occur when pointers from one stream of objects, such as the irrelevant speech, interfere with another set of pointers, such as the list to be recalled. Because irrelevant speech consists of different and varying items, it produces multiple objects and pointers, which can interfere with the to-be-remembered stream and thus impair recall. Conversely, if the same auditory item is presented repeatedly, such as the letter ‘C’, it creates only one object, which points back to itself and thus does not impair recall as it will not interfere with any other objects or pointers. This is different from the phonological store hypothesis in that interference occurs by crossing pointers rather than phonologically-based.

Auditory Streaming

Auditory streaming is the process by which a listener decomposes patterns of acoustic stimuli in terms of both their environmental origin, as well as aspects such as frequency and timbre (Bregman, 1990). It has been argued that the impairment in recall due to irrelevant stimuli could be a result of focused attention to said stimuli, since auditory streaming requires attention (Carlyon, Cusack, Foxton, & Robertson, 2001). However, others have argued that auditory streaming does not require attention, and that the ISE demonstrates preattentive auditory streaming (Macken et al., 2003). More specifically, several critical aspects of the methodology used in a typical ISE study suggest that acoustic processing of irrelevant stimuli is preattentive. Firstly, the noise stimuli are auditory but the list to be remembered is typically visual. Participants are instructed to ignore the irrelevant stimuli, as they are not tested on this material. Secondly, it is not common for there to be cross-modal interference of auditory stimuli on a visual task, and not all auditory stimuli elicit this effect. Only particular types of sound, such as those that are changing-state, disrupt recall. Thirdly, the primary task of serially encoding a list of words is difficult enough that processing resources must be highly allocated to this task, and thus resources required for attention to other stimuli are unavailable. Finally, the effect is also observed when the irrelevant speech occurs during a retention interval, after the list has been presented, indicating that the ISE does not necessarily occur during encoding, and the ISE cannot be a result of divided or switched attention at encoding (Norris, Baddeley, & Page, 2004). Taken in combination, this evidence suggests that the disruptive effect of irrelevant sound is not the result of deliberate processing and attention. These findings lend further support to the O-OER model's explanation of the ISE.

Individual differences in ISE

Ellermeier and Zimmer (1997) looked at the range of individual variability with respect to the irrelevant speech/sound effect. Participants were asked to memorize a visually presented sequence of nine digits, while listening to foreign speech, pink noise, or silence in the background. Participants were instructed to simply ignore the background noise. Results showed that individuals experienced a wide range of effects from the sound conditions ranging from slight facilitation to severe disruption. The size of the effect was normally distributed over a considerable range, but most participants showed decrements in memory due to noise, particularly speech noise, which is consistent with the current and previous literature. Interestingly, participants were found to be quite inaccurate when asked to assess how they thought noise would affect their memory. While participants failed to accurately predict their own susceptibility, they were able to accurately predict overall effects of the noise conditions. Additionally, after gaining experience with the task, participants' estimates of performance were somewhat improved, indicating that participants may recognize the effect of noise in-situ. Overall, the results demonstrated that the ISE was reliably observed and measured, and the effect was found to be normally distributed with speech noise causing the largest impairment in memory.

Regardless of the origin of the ISE, it is a robust and replicable effect. The data reported above show that, for the majority of individuals, background noise disrupts processing in some way that affects memory performance. Because many students report listening to background music while studying (Kotsopoulou & Hallam, 2010), it is important to determine what knowledge individuals have about the impact of noise on their memory. The section below provides a brief summary of theory and data in the area of metamemory before considering how we might examine the ISE in the context of metamemory.

Metamemory

Metamemory is the process by which individuals are able to assess the contents in their memory; it is an individual's awareness or knowledge about said contents. This monitoring process can take place either prospectively (i.e., before recalling information), or retrospectively (i.e., after recalling information). It is this ability to examine our memory that allows us to make judgments, assessments and predictions about it. Researchers use subjective judgments to explore individuals' monitoring of their memory.

The issues that these types of measures address include, but are not limited to, the process by which metacognitive judgments are formed, the implications these judgments have for self-guided learning (e.g., Metcalfe & Dunlosky, 2008), how accurate these judgments are in determining actual knowledge and memory content and what their effects are on performance (e.g., Koriat, 1997). The analyses of these issues have provided substantial insight into the relationship between our consciousness and our behavior, but there is a significant gap in the literature in terms of how individuals perceive noise to affect memory.

Metamemory and Ease of Learning Judgments (EOLs)

Researchers have distinguished between different types of judgments that can be used to measure information about people's metamemory. Judgments of learning, or JOLs (Arbuckle & Cuddy, 1969), are measures used post-study (but pre-test) to determine how much information individuals thinks they will be able to remember on a subsequent memory test. Ease-of-learning judgments, or EOLs (Richardson & Erlenbacher, 1958), are judgments that occur prior to study. When people prepare to study for an exam, they might first assess the difficulty of the material prior to study, which would be an EOL. The accuracy of these judgments is crucial for assessing study allocation time. These EOLs are important in determining how easy or difficult particular information will be to learn before the encoding has actually taken place. Nelson and Narens (1994) stated that EOLs "occur in advance of acquisition, are largely inferential, and pertain to items that have not yet been learned. These judgments are predictions about what will be easy/difficult to learn, either in terms of which items will be easiest […] or in terms of which strategies will make learning easiest" (p.16). However, the means by which individuals classify some words as more memorable or easy to learn than others must be considered. As explained below, many metamemory researchers assert that people make EOL judgments based on one or more cues, such as internal and external cues, and that these judgments are ultimately inferential in nature.

Many of these conclusions derive from Koriat's framework on cue utilization (Koriat, 1997). The cue utilization approach to judgments of learning (JOLs) asserts that JOLs are inferential in nature, thus based on past experience and knowledge. Koriat distinguished between three classes of cues used to make JOLs: intrinsic, extrinsic and mnemonic. Intrinsic cues are defined as characteristics of study items that may offer insight into an item's *a priori* ease or difficulty of learning, such as imageability, which can be an effective indicator of memorability. Extrinsic cues refer primarily to the conditions of learning, such as retention interval and stimulus duration. Mnemonic cues are internal cues that indicate the extent to which an item has been learned, such as the ease with which information comes to mind or cue familiarity. Koriat argued that EOLs are heavily based on a participant's *a priori* beliefs, and that they are inferential in nature. In terms of the ISE, it is possible that participants may recognize the differences in difficulty between memorizing words in noise conditions compared with quiet, and thus give lower EOLs to a condition that contains speech babble. If participants do recognize the possible differences between the noise conditions, this would render the judgment influencing EOLs to be a predominantly extrinsic cue by definition. Extrinsic cues encompass aspects about the learning conditions and hence the stimulus, which is noise in the ISE paradigm. Because extrinsic cues are based on the conditions of learning and this is all the information participants are offered, participants should use information from this stimulus to determine its effect on memory. Therefore, determining how noise will affect memory is most likely based on an externally-driven cue.

The ISE seems to be generalizable in that it disrupts most people, but there is a substantial subset of the population who perform cognitive tasks while watching TV, listening to music, or in other various noise-polluted environments. Is it possible that these people are less susceptible to the impairment that background noise causes? In the past, studies have indicated improvement in cognitive performance in the presence of liked music (Hallam, Price & Katsarou, 2002). Some may explain this improvement as an increase in arousal while preferred music is playing, resulting in an increase in performance (Schellenberg, 2005), and many students report often listening to music while studying (Kotsopoulou & Hallam, 2010). Perham and Vizard (2011) sought to investigate whether a preference for listening to music while engaging in cognitive tasks could actually mediate the ISE. Serial recall in quiet conditions was compared with liked and disliked music sound conditions, as well as steady-state and changing-state speech conditions. However, results revealed that performance was poorer in both music conditions and the changing-state speech condition compared to the quiet and steady-state speech conditions. Thus, not only did this study replicate previous findings with regard to steady and changing-speech conditions, it demonstrated that regardless of whether people usually listen to music while engaging in cognitively demanding tasks, performance is hindered by background noise. This finding suggests serious limitations on the beneficial effects of listening to music at the same time as task performance, and further expands the ISE to encompass music as well as other types of nonspeech noise.

In terms of metamemory, Perham and Vizard (2011) also investigated whether or not people recognized the effect of noise on memory. Participants were asked to rate different noise conditions on distractibility, among other features, and interestingly, while they rated their preferred noise condition as the most likeable condition, they did not rate their preferred noise condition as less distracting than the other noise conditions. Participants did, however, rate the quiet condition as the least distracting condition. These results indicate that participants may recognize the effect that noise has on their memories; however, these results call for further investigation.

One study by Kotsopoulou and Hallam (2010) surveyed multiple age groups of younger adults ranging in ages of 12 to 21 regarding the extent to which they played music while studying and their perception of the effects of background music on studying abilities. Interestingly, many students reported that they did not play music extensively when studying for examinations. University students reported that music had a more relaxing effect, but also reported that it interfered with their concentration, which was not reported as often in the younger age groups. Overall, there was general agreement that music could interfere with concentration. This finding indicates that people may recognize that playing music while studying can be detrimental and that they may be aware of the effect noise has on memory. However, there are a number of issues with retrospective measures and thus one cannot be sure from this study alone what meta-knowledge people possess regarding the effects of noise on memory. This uncertainty coupled with the limited research in this area calls for an investigation into what people know about how noise affects their memory.

Aims and Hypotheses

The ISE has been explored in detail. However, there is at least one practical application of this effect that has not been explored, and this is with regard to people's metamemory for words in noise; more specifically, whether the effect of noise is recognized and taken into account. The aim of this study is to examine how well people can judge future memory performance of words presented in noise and thus determine whether or not they can recognize the effect that noise has on their memory. The current literature has not produced any research to satisfactorily address this issue. It is hypothesized that some participants will be able to recognize this effect, but the overall trends are unknown. Given the lack of research in this area, the investigation into the meta-knowledge possessed by participants will be exploratory.

Experiment 1

Methods

Participants

Participants in this study were 30 students from the Memorial University community who ranged in age from 17 to 29 years, *M* = 21.73, *SD* = 3.60. There were 12 males and 18 females. All participants reported normal hearing. All participants were given either $10 or course credit for their participation. Participants who were given course credit were a part of the Memorial Psychology Research Experience Program (PREP).

Stimulus Materials

The influence of noise distortion on memory for words was manipulated within subjects. As is typical in ISE experiments, each participant received 45 lists of 7 words; 15 lists were presented in each of three noise conditions: white noise, speech babble or quiet. Each list was randomly generated using words chosen from the MRC Psycholinguistic Database and no words were repeated between or within lists. All words were concrete nouns of 1-2 syllables with medium frequency and imagibility. Each word was generated using Macbook Pro software and presented visually on the computer screen. The noise was played through headphones at a comfortable listening level. The white noise was continuously generated using a Macbook Pro. The speech babble noise was taken from a file containing digitally recorded speech babble was downloaded from the Internet (http://spib.rice.edu/spib/data/signals/noise/babble.html; copyrights TNO, Soesterberg, The Netherlands). The source of this babble was 100 people speaking in a canteen. The presentation of each of the three noise conditions was randomized.

Procedure

The experiment was divided into two phases: 1) orientation to the noise, and 2) study. In phase one, participants were told that they were about to listen to two different types of noise, after which they would then be prompted by the computer to estimate how much they thought the noise would affect their recall for lists of words if the noise were playing simultaneously in the background while they were studying the lists. These are typically called global ease of learning judgments, or EOLs. Participants were asked to give this rating on a scale from one to ten, with one indicating that they did not think the noise would affect them at all, and ten indicating that it would severely disrupt them and they would not be able to remember anything. These instructions were given verbally by the experimenter and then reiterated on the computer screen. When participants were clear on the instructions and had no further questions, they initiated the study by pressing a button on the screen with the mouse ("Next Trial").

After participants evaluated the two noise conditions, they began phase two, which was the study phase. Participants were told that for the next 45 trials they would see lists of seven words presented one at a time on the computer screen. Before each trial (list), they saw the noise condition (quiet, white noise or babble) and were asked to indicate how many words they thought they would be able to remember, in order, in that noise condition. Participants were asked to select a number between one and seven, which represented the number of words they thought they would be able to remember. These were trial-by-trial frequency estimates. Participants then had to press 'next trial' to begin the list. When this was pressed, participants were presented with each word on the computer screen, one at a time. Each word was presented for 500ms with no interstimulus interval. Immediately after the presentation of the list, participants were asked to verbally recall all the words they could remember in the order that they saw them. The experimenter recorded these responses on a laptop. In case a participant spoke too fast, or the experimenter missed a word they said, all sessions were recorded on a simple audio recorder and could be reviewed afterward.

Results

A 3 (noise: quiet versus white noise versus babble) x 7 (serial position: 1 to 7) repeated measures ANOVA was carried out on the mean recall scores for each participant with respect to serial position. There was a main effect of noise, $F(2, 58) = 7.901$, $MSe = .018$, $p = .001$. However, this main effect did not reflect an ISE considering the means of the noise conditions. In terms of the average proportion correct (see Table 1 for mean proportion correct in all four experiments), post-hoc tests showed that the quiet condition ($M = .393$, $SD = .133$) and babble condition ($M = .372$, $SD = .125$), were both significantly worse than the white noise condition ($M = .424$, $SD = .100$), $p = .009$ and $p = .001$ respectively, but performance in the quiet and babble conditions were not significantly different from each other ($p = .162$). Therefore, an ISE was not observed (see Figure 1), even though the two noise conditions were significantly different from each other. There was a main effect of serial position, $F(6, 174) = 79.308$, $MSe = .063$, $p < .001$, with the best recall performance being on serial position one and two after which there is a gradual decline in performance, and a slight recovery on serial position seven. Therefore there was a classic primacy and recency effect observed, as expected. The serial position by noise interaction was not significant, $F(12, 348) = 1.453$, $MSe = .011$, $p = .140$.

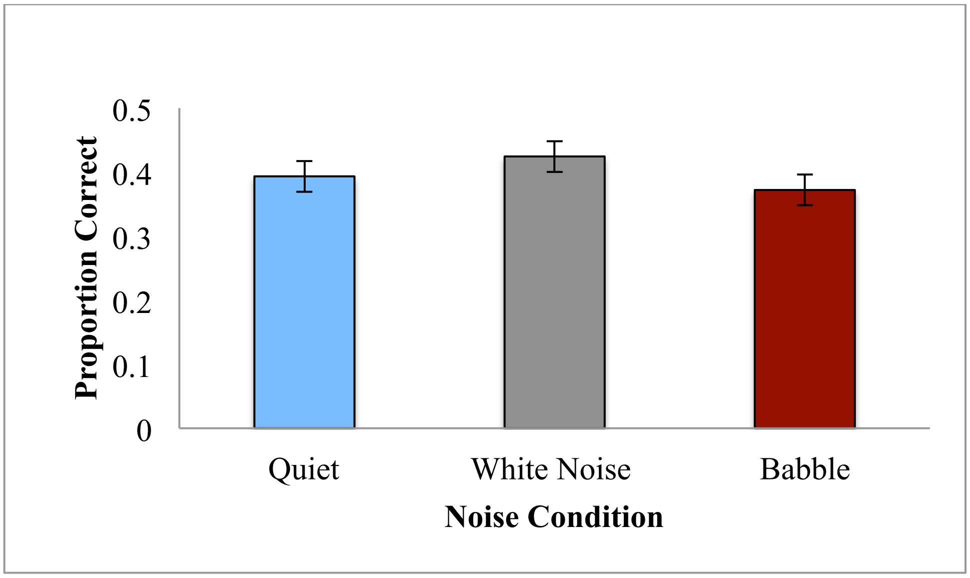

Figure 1. Mean proportion correct by noise condition (error bars are standard error of the mean).

With respect to the global EOL judgments, participants rated the babble condition as disruptive, $M = 5.07$, $SD = 2.391$. Participants also rated the white noise as being disruptive as well, $M = 4.57$, $SD = 2.402$ (see Figure 2). A paired-samples t-test showed that these judgments were not significantly different , $t(29) = -0.923$, $p = .364$. For the trial-by-trial frequency data, a mean frequency estimate was computed per condition by averaging across all 15 trials (see Figure 3). These data indicated that participants thought they would remember the most information in the quiet condition, $M = 4.05$, $SD = .846$, and the least in the white noise, $M = 3.55$, $SD = .770$, and babble condition, $M = 3.562$, $SD = .691$. A one-way ANOVA showed that the difference between these predictions was significant, $F(1, 29) = 23.849$, $MSe = .151$, $p < .001$. The average judgments for the white noise and babble conditions were not significantly different from each other ($p = .868$) but they were both significantly different from the quiet ($p < .001$)

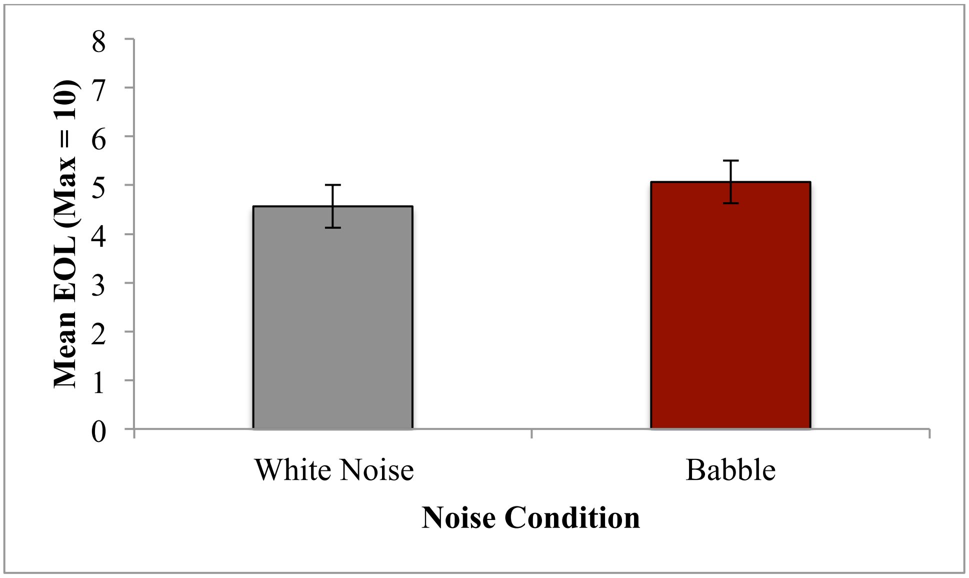

Figure 2. Mean global EOL judgment rating by noise condition (error bars are standard error of the mean). Note: higher values indicate more predicted disruption.

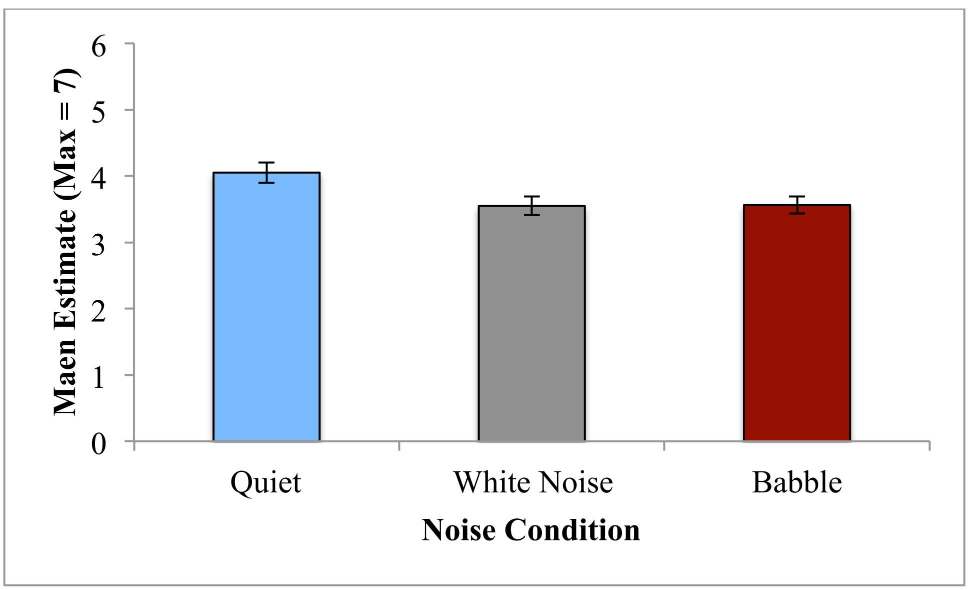

Figure 3. Mean frequency estimates by noise condition (error bars are standard error of the mean).

In order to determine how these predictions relate to actual performance, calibration, or absolute accuracy scores, were computed for each participant (see Figure 4). In terms of trial and noise condition, actual recall was subtracted from predicted recall to determine how much predictions actually deviated from performance. A score of zero would mean participants were perfectly calibrated: their predicted score was the same as their actual score. Otherwise, participants are either over or under-confident. These calculations allowed us to determine any differences in metaknowledge between the different noise conditions. A repeated-measures one-way ANOVA on the calibration data revealed a main effect of noise, $F(2, 58) = 13.814$, $MSe = 3.941$, $p < .001$ and a main effect of trial, $F(14, 406) = 3.718$, $MSe = 2.289$, $p < .001$. There was no significant interaction between the two, $F(28, 812) = 1.190$, $MSe = 2.215$, $p = .230$. Pairwise comparisons showed that the quiet and white noise conditions were significantly different from each other in terms of calibration ($p < .001$), the quiet and babble conditions were significantly different from each other ($p = .031$), and the white noise and the babble were also significantly different from each other ($p = .009$).

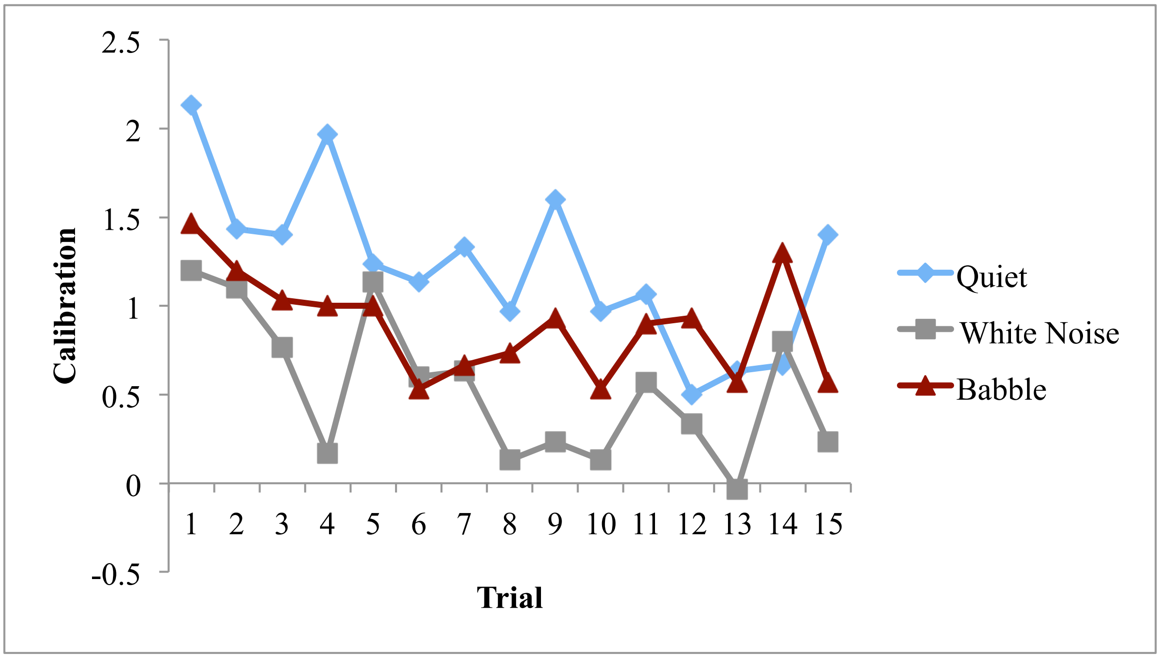

Figure 4. Calibration by trial and noise condition.

In order to further examine the main effect of trial, the trials for each noise condition were divided into three blocks: the first five trials (block one), the second five trials (block two) and the last five trials (block three). There was a main effect of noise, $F(2, 58) = 13.814$, $MSe = .788$, $p < .001$ and a main effect of trial block, $F(2, 58) = 15.818$, $MSe = .471$, $p < .001$, however the interaction between the two was not significant, $F(4, 116) = 1.123$, $MSe = .606$, $p = .349$. In terms of trial block, block 1 and 2 were significantly different from each other, $p < .000$, block 1 and 3 were significantly different, $p = .033$, and block 2 and 3 were significantly different, $p = .009$. These findings show that calibration is the best in the white noise condition, as it is the closest to approaching perfect calibration. Participants' calibration is the worst in the quiet condition, as they are the most overconfident in this condition. Overall, calibration generally improved over the course of the trial blocks for all conditions.

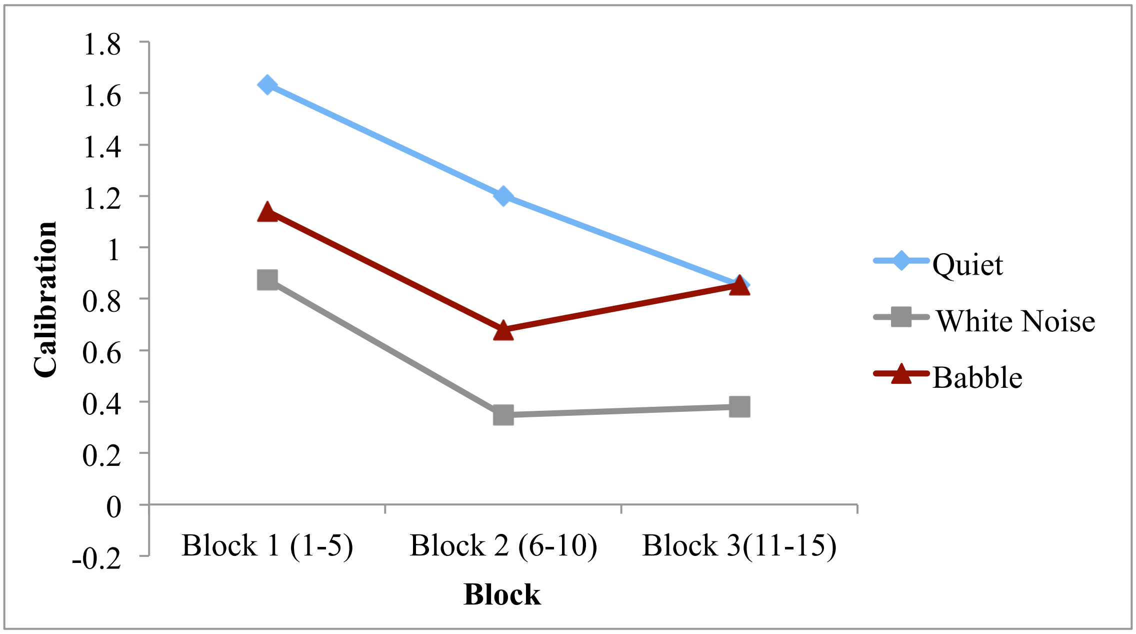

Figure 5. Calibration by block and noise condition.

Table 1

Mean Proportion Correct Serial Recall of Words in Quiet, White Noise, and Irrelevant Speech by Experiment (with Standard Deviation in Parentheses).

Noise Condition	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Quiet	.393 (.133)	.346 (.139)	.341 (.086)	.393 (.100)
White Noise	.424 (.100)	.331 (.127)	.352 (.103)	.399 (.115)
Irrelevant Speech	.372 (.125)	.344 (.138)	.270 (.100)	.330 (.092)

Discussion

Surprisingly, the results of this study showed no ISE. There are multiple explanations that could account for this. Firstly, this result could be a consequence of including trial-by-trial frequency estimates, which are atypical of an ISE study. Changing the typical procedure could have somehow disrupted the effect. It is possible that by asking people to assess their own memories that we could have changed their behavior in some way by drawing such marked attention to the noise. Secondly, it is conceivable that the speech babble sound file used was not sufficient to produce an ISE. It is possible that the sound file used was not an appropriate changing-state sound file to elicit this effect. Before conducting another metamemory study with a different sound file, it was necessary to ensure that the metamemory aspect of the study was not influencing the results, accounting for the diminished irrelevant speech effect. Therefore, a control study was implemented. This study was the same as the first experiment except the metamemory aspects were eliminated, making it a standard ISE experiment.

Experiment 2

Methods

Participants

The sample of participants in this study were 19 students from the Memorial University community who ranged in age from 17 to 26 years, $M = 19.52$, $SD = 2.50$. There were nine males and 10 females. All participants were of normal hearing. All participants were given either $10 or course credit for their participation. Participants who were given course credit were a part of the Memorial Psychology Research Experience Program (PREP).

Stimulus Materials

The stimulus materials used in this study were the same as those used in Experiment 1.

Procedure

The procedure for this study was the same as the first experiment, except that participants were not asked to give global EOL judgments for the two noise conditions in advance, and there were no frequency estimates at the beginning of each trial. Participants in this study were simply told that they were about to see 45 lists of seven words on the computer screen. After each list, participants were asked to serially recall all of the words they could remember out loud to the experimenter who recorded their responses.

Results

A 3 (noise: silent versus white noise versus babble) x 7 (serial position: 1 to 7) repeated measures ANOVA was carried out on the mean recall scores for each participant with respect to serial position. There was no significant effect of noise (see Figure 2), $F(2, 36) = .589$, $MSe = .014$, $p = .560$, as participants seemed to perform equally in the quiet ($M = .346$, $SD = .139$), white noise ($M = .331$, $SD = .127$) and babble condition ($M = .344$, $SD = .138$). There was a main effect of serial position, $F(6, 108) = 91.274$, $MSe = .037$, $p < .001$. As in Experiment 1, we observed typical serial position effects. The serial position by noise interaction was not significant, $F(12, 216) = .396$, $MSe = .010$, $p = .964$.

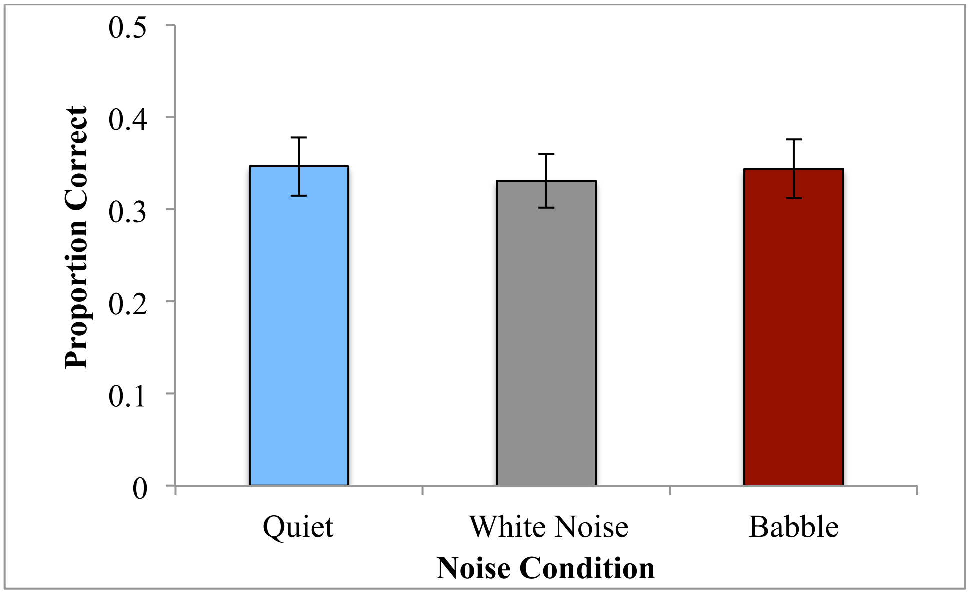

Figure 6. Mean proportion correct by noise condition (error bars are standard error of the mean).

Discussion

This control experiment yielded similar findings to the first experiment in that an ISE was not observed. This is important as it allows for the conclusion that it was not the implementation of the metamemory questions that caused the disruption of the effect. The issue with these two experiments may have to do with the type of irrelevant sound used. It is possible that the speech babble was not distinct enough to orient participants' attention compared with other babble sound files previously used in other ISE studies. In other words, the speech babble sound used in these studies may have functioned more like steady-state noise as opposed to changing-state noise as it was intended. Jones, Beaman, and Macken (1996) argued that the nature of the irrelevant speech effect derives from the sharp changes in amplitude and frequency characteristic of speech. When the number of speakers increases beyond a certain level, the boundaries between words of one speaker may be masked by the simultaneous words of the other speakers. In this situation, the varied amplitude and frequency that causes the ISE may be diminished, depending on the number and locus of the speakers. Jones and Macken (1995a, 1995b) have shown that when the babble contains more than a few voices, the disruption that causes the ISE is weakened. The nature of changing-state stimuli is that the material regularly provides a novel stimulus, which elicits an involuntary orienting response (Jones, Madden & Miles, 1992). Because there were 100 voices speaking simultaneously in the previous babble condition, there may not have been a novel stimulus embedded in this condition, and thus the sound was not a changing-state stimulus; rather, it was more akin to steady-state noise. Therefore, it was hypothesized that a new sound file using just one speaker could elicit the typical ISE observed in the literature.

Experiment 3

The results of the previous two studies showed no ISE. Considering the literature outlined above, is reasonable to conclude that the sound file used was not sufficiently changing state to cause an ISE. Therefore, a new sound, one that has been shown to cause an ISE in the past (Neath, Farley, & Surprenant, 2003), was used in the present study. Before implementing the metamemory aspect of this research, a control experiment replicating Experiment 2 was used to test whether the ISE could be elicited with a different sound file. In this study, a recording of one woman speaking in German was used instead of the speech babble sound.

Methods

Participants

Participants in this study were 20 students from the Memorial University community who ranged in age from 18 to 27 years, $M = 21$, $SD = 1.92$. There were six males and 14 females. All participants were of normal hearing. All participants were given either $10 or course credit for their participation. Participants who were given course credit were a part of the Memorial Psychology Research Experience Program (PREP). Participants who spoke German were not eligible to participate in this study.

Stimulus Materials

The stimulus materials used in this study were the same as those used in Experiments 1 and 2, except the speech babble file was replaced by a passage (in German) from Die Wilden by Franz Kafka spoken by a female used by Neath, et al. (2003). The white noise file remained the same.

Procedure

The procedure for this study was exactly the same as Experiment 2.

Results

A 3 (noise: silent versus white noise versus German) x 7 (serial position: 1 to 7) repeated measures ANOVA was carried out on the mean recall scores for each participant with respect to serial position. A main effect of serial position was observed, $F(6, 114) = 107.072$, $MSe = .066$, $p < .001$ and there was also a main effect of noise (see Figure 3), $F(2, 38) = 18.022$, $MSe = .017$, $p < .001$. The serial position by noise interaction was also significant, $F(12, 228) = 8.488$, $MSe = .017$, $p < .001$. Participants initially recalled the most items in the quiet and white noise condition, with a decline in recall performance across serial positions. Participants remembered the least initially in the German condition, and a steady recline in recall performance across serial positions. As in Experiments 1 and 2, we observed typical serial position effects. Post-hoc tests showed that recall in the quiet condition ($M = .341$, $SD = .086$) and white noise condition ($M = .352$, $SD = .103$) was significantly better than recall in the German speech condition (M

= .270, *SD* = .100), $p < .001$, but recall in the quiet and white noise conditions were not significantly different from each other, $p = .388$.

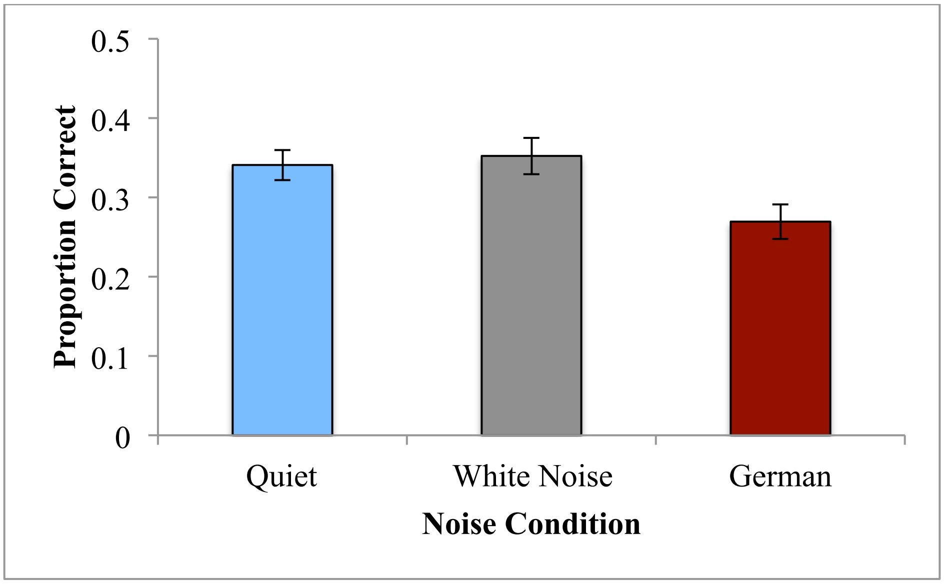

Figure 7. Mean proportion correct by noise condition (error bars are standard error of the mean).

Discussion

The results of this study demonstrated a typical, robust ISE. Participants were able to recall significantly more words in the quiet and white noise condition compared with the German condition, which is what we expected and should be observed in a typical ISE study. This finding confirms the hypothesis that the babble used in the first two experiments was not sufficient to elicit the ISE, as it likely falls into the category of steady-state noise.

Experiment 4

Experiment 3 obtained the typical ISE desired, and thus for Experiment 4, the metamemory aspect of the study originally proposed was implemented. That is, Experiment 4 was a replication of Experiment 1, but using the German speech sound file in the irrelevant noise condition instead of the cafeteria babble sound file.

Methods

Participants

Participants in this study were 36 students from the Memorial University community who ranged in age from 18 to 32 years, $M = 22$, $SD = 3.60$. There were 11 males and 25 females. All participants were of normal hearing. All participants were given either $10 or course credit for their participation. Participants who were given course credit were a part of the Memorial Psychology Research Experience Program (PREP). Participants who spoke German were not eligible to participate in this study.

Stimulus Materials

The stimulus materials used in this study were the same as those used in Experiment 3.

Procedure

The procedure for this study was exactly the same as Experiment 1.

Results

A 3 (noise: silent versus white noise versus German) x 7 (serial position: 1 to 7) repeated measures ANOVA was carried out on the mean recall scores for each participant with respect to serial position. A main effect of noise was observed (see Figure 4), $F(2, 70) = 19.191$, $MSe = .019$, $p < .001$ and there was also a main effect of serial position, $F(6, 210) = 163.313$, $MSe = .042$, $p < .001$. As in the first three Experiments, we observed typical serial position effects. The noise by serial position interaction was also significant, $F(12, 420) = 4.341$, $MSe = .011$, $p < .001$. Participants recalled the most initially in the quiet and white noise condition, with a decline in recall performance across serial positions. Participants remembered the least initially in the German condition, and a steady recline in recall performance across serial positions. Post-hoc tests showed that the quiet condition (*M* = .393, *SD* = .100) and white noise condition (*M* = .399, *SD* = .115) were both significantly different from the German speech condition (*M* = .330, *SD* = .092), $p < .001$, but were not significantly different, $p = .659$.

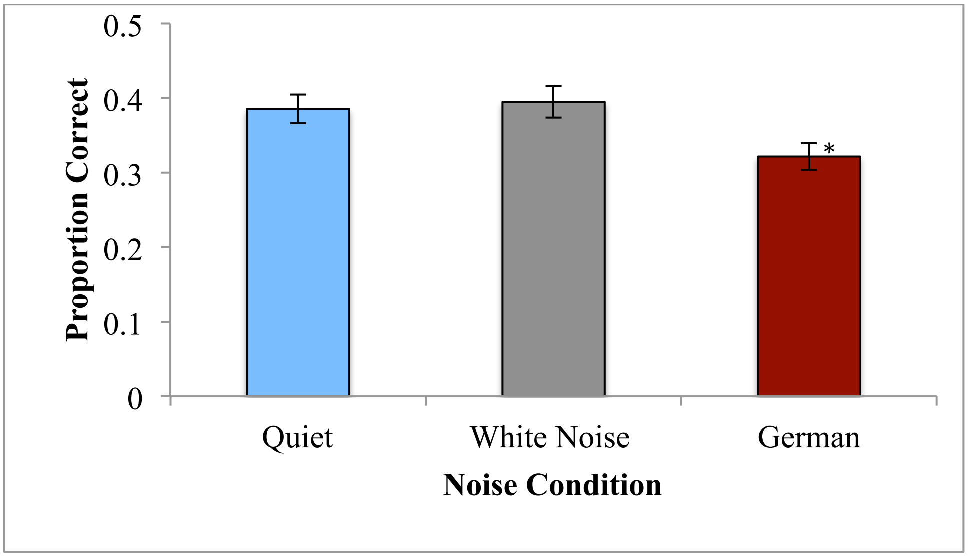

Figure 8. Mean proportion correct by noise condition (error bars are standard error of the mean).

With respect to the global EOL judgments, participants rated the German condition as disruptive (see Figure 5), *M* = 6.25, *SD* = 2.371. Participants also rated the white noise as being disruptive as well, *M* = 4.83, *SD* = 2.667. A paired-samples t-test showed that these judgments were significantly different from each other, t(35) = -2.626, $p = .013$. The trial-by-trial frequency data indicated that participants thought they would remember the most information in the quiet condition, *M* = 3.591, *SD* = .736, slightly less in the white noise condition, *M* = 3.098, *SD* = .603, and the least in the German condition, *M* = 2.828, *SD* = .704 (see Figure 6). A one-way ANOVA showed that the difference between these predictions was significant, $F(2, 70) = 28.789$, $MSe = .188$, $p < .001$. The average judgments for the quiet and white noise were significantly different from each other, $p < .001$, the average judgments for the quiet and German were significantly different from each other, $p < .001$, and the average judgments for the white noise and German were also significantly different from each other, $p = .015$.

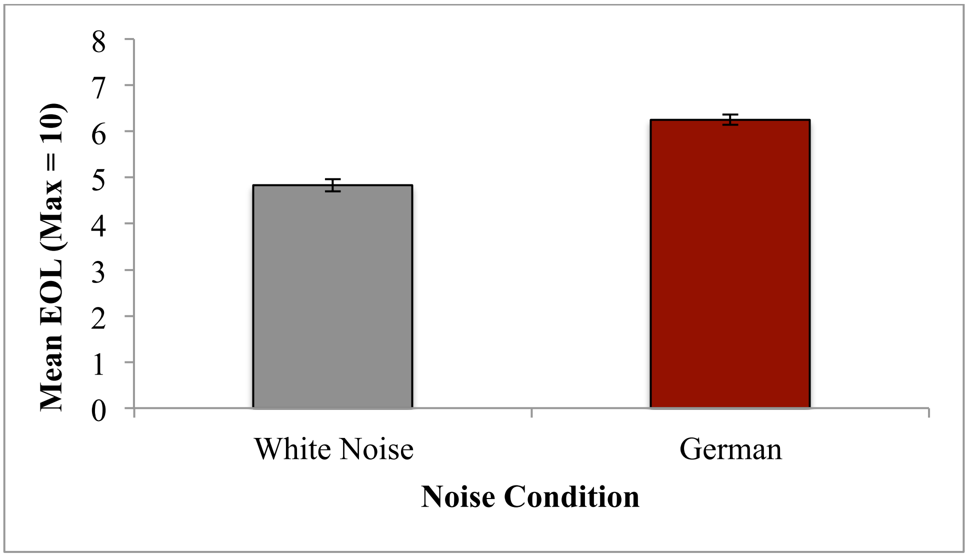

Figure 9. Mean global EOL judgment rating by noise condition (error bars are standard error of the mean). Note: higher values indicate more predicted disruption.

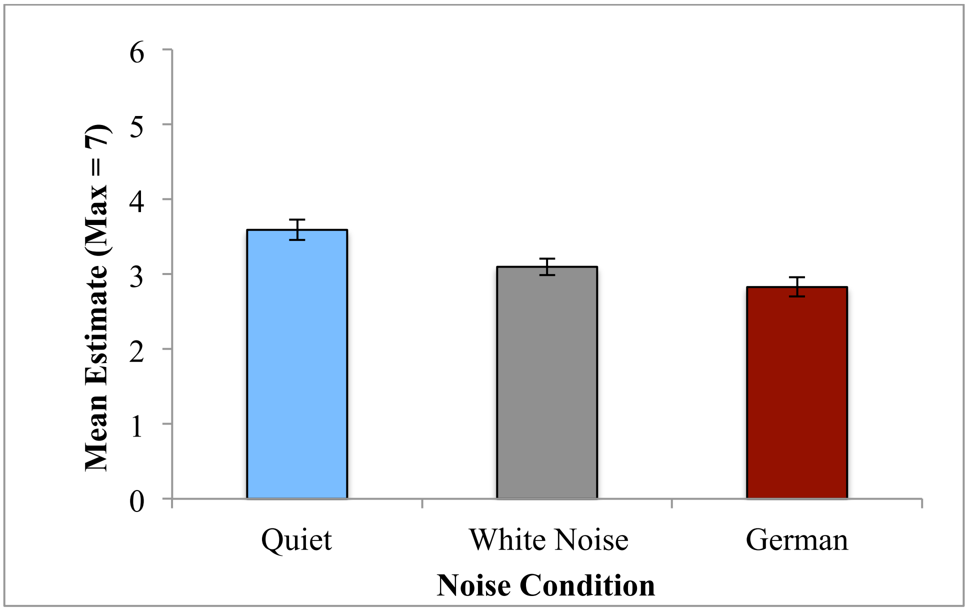

Figure 10. Mean frequency estimates by noise condition (error bars are standard error of the mean).

In order to determine how these predictions related to actual performance, calibration, or absolute accuracy scores, were computed for each participant (see Figure 11). In terms of trial and noise condition, actual recall was subtracted from predicted recall to determine how much predictions actually deviated from performance. This allowed us to determine any differences in metaknowledge across the noise conditions. A 3 (noise: silent versus white noise versus German) x 7 (serial position: 1 to 7) repeated measures ANOVA revealed a main effect of noise, $F(2, 70) = 12.001$, $MSe = 3.730$, $p < .001$ and a main effect of trial, $F(14, 490) = 18.811$, $MSe = 2.083$, $p < .001$. There was also a significant interaction between the two, $F(28, 980) = 1.921$, $MSe = 1.965$, $p = .003$. Pairwise comparisons showed that the quiet and white noise conditions were significantly different from each other in terms of calibration ($p < .001$), the quiet and German were significantly different from each other ($p = .005$), but the white noise and the German were not significantly different, $p = .076$. Calibration was better in the noise conditions than the quiet, with participants being the most overconfident initially in the quiet condition, although calibration for all conditions improved over time.

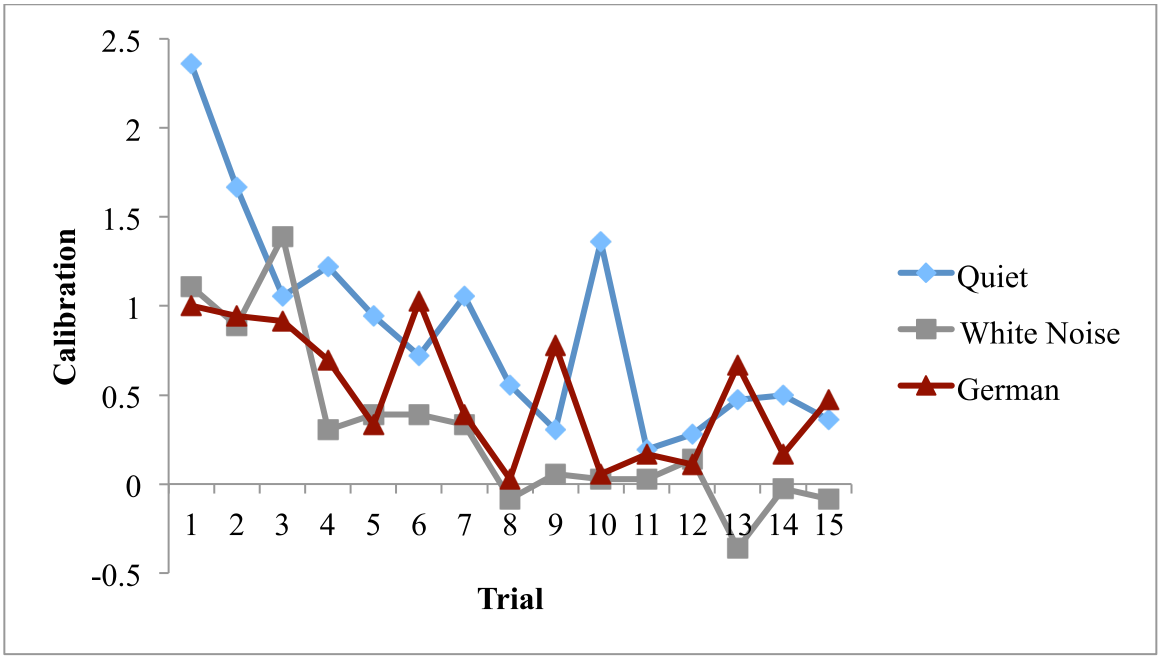

Figure 11. Calibration scores by trial and noise condition.

To more easily examine the noise x trial interaction, the trials for each noise condition were divided into three blocks (see Figure 12): the first five trials (block one), the second five trials (block two) and the last five trials (block three). A repeated-measures ANOVA showed a main effect of noise, $F(2, 70) = 12.001$, $MSe = .746$, $p < .001$ and a main effect of trial block, $F(2, 70) = 27.733$, $MSe = .664$, $p < .001$, however the interaction between the two was not significant, $F(4, 140) = 2.043$ $MSe = .485$, $p = .092$. In terms of trial block, block 1 and 2 were significantly different from each other, $p < .000$, block 1 and 3 were significantly different, $p < .001$, and block 2 and 3 were significantly different, $p = .015$. These findings show that calibration is the best in the white noise condition, as it is the closest to approaching perfect calibration. Participants are the worst calibrated in the quiet condition, as they are the most overconfident in this condition. Overall, calibration improved with experience for all three noise conditions.

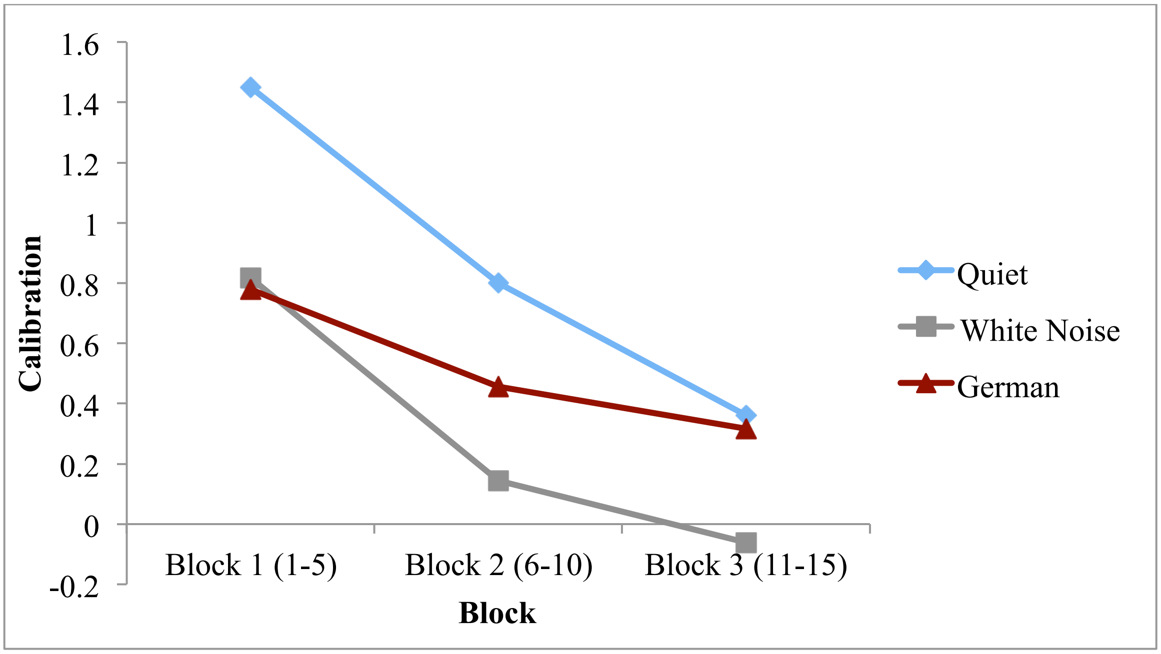

Figure 12. Calibration scores by block and noise condition.

Discussion

In terms of recall performance, the typical ISE was observed again. This confirms that the null ISE result observed in Experiment 1 was not a result of implementing metamemory questions, but was a result of the babble sound used.

The global EOLs showed that participants rated the German condition as rather disruptive, indicating that participants may hold prior beliefs about the effect of this type of noise on memory, prior to experiencing it. Interestingly, participants also rated the white noise as disruptive, even though the white noise had no effect on recall in actuality. Therefore, participants were quite inaccurate with predicting the effects of this type of noise, indicating that they also held prior (false) beliefs about the effect of white noise on memory.

The trial-by-trial frequency estimates, which were the average judgments across trials, showed that participants were generally able to recognize the effect of the German noise, giving significantly lower estimates of predicted recall than in the quiet condition. However, once again, participants were inaccurate with the white noise. Participants' estimates of performance in the white noise condition were significantly lower than those in the quiet condition, even though actual recall was similar in the white noise and quiet conditions. So the question then is: Do people remember what they think they are going to remember? The calibration data showed that in all noise conditions, participants were initially overconfident, but their predictions became more realistic over time, particularly in the white noise condition. Initially, people seemed to think that the white noise and the German conditions would affect them similarly, but with experience they became better calibrated in the white noise condition than the German condition.

General Discussion

The goal of this study was to determine whether or not people could recognize the effect that noise has on their memory. In order to address this question, four experiments were conducted within an ISE framework. The first two experiments failed to find a typical ISE due to the speech sound file used. Experiment 1 yielded unexpected findings, with participants performing significantly better in the white noise condition compared with the irrelevant sound and quiet condition. With a change in sound file, however, we were able to elicit a typical ISE as shown in Experiment 3. In Experiment 4 we found a robust ISE, indicating that EOLs do not influence this effect. From the metamemory data, we were able to determine that participants can recognize the effect of irrelevant speech sounds prior to being tested in that condition, and participants' judgments became more realistic with experience with the task. Similar to the findings of

Ellermeier and Zimmer (1997), participants were able to accurately predict overall effects of some noise conditions. Additionally, after gaining experience with the task, participants' estimates of performance improved, indicating that participants may recognize the effect of noise in-situ. Overall, the results demonstrated that the ISE was reliably observed and measured, with speech noise causing the largest impairment in memory.

In light of these findings, people may be able to recognize the effect of irrelevant speech stimuli on their memory prior to experiencing it in-situ. The metamemory assessments in this study were prospective, which enabled us to determine how participants thought the noise would affect them prior to experience with the task, as well as how these judgments changed over time with experience. These assessments were achieved using EOLs. In terms of the global EOL judgments made prior to experience with the task, participants rated the white noise, babble and German noise conditions as disruptive. The judgments between the white noise and babble conditions in Experiment 1 were not significantly different from each other, and the actual performance in these two conditions was not significantly different. However, interestingly, the judgments between the white noise and German conditions in Experiment 4 were significantly different from each other, with the German being rated as significantly more disruptive than the white noise. Actual performance shows that participants perform significantly worse in the German condition than the white noise condition. Therefore participants may recognize more subtle differences between the noise conditions than anticipated.

Within Koriat's framework on cue-utilization (Koriat, 1997), it is most likely that participants relied on extrinsic cues, which are based on the conditions of learning and hence the stimulus, which in this case were different noise conditions. By using these cues, participants were able to assess the irrelevant speech sounds in Experiment 4 and determine that it would be detrimental to their memory. Intrinsic cues, on the other hand, have been defined as “the characteristics of the study items that are perceived to disclose the items’ a priori ease or difficulty of learning” (Koriat, 1997, p. 350). Koriat has stated that intrinsic cues are often weighted more heavily by individuals making JOLs or EOLs than are extrinsic cues: “…in general, the effects of extrinsic factors should be discounted in predictions of recall relative to those of intrinsic factors” (Koriat, 1997, p. 352). Intrinsic cues are more salient to a naïve participant who may not readily recognize the influence of extrinsic cues. People often need extensive practice to account for extrinsic factors. However, in the present research there is no systematic variation of intrinsic cues for participants to allot more weight to, thus they must be using beliefs about the noise conditions to make their judgments. Thus the extrinsic cue used here (noise condition) gives participants an indication of the item’s a priori difficulty, not because of the items themselves that are about to be studied, but because of the condition of learning; therefore, extrinsic cues can be accounted for by individuals making predictions about their future memory performance.

The current research has found that participants were sensitive to extrinsic cues, depending on the type of cue. Some research on this framework suggesting that intrinsic cues are more heavily relied on than extrinsic cues for JOLs has focused on the extrinsic cue (condition of learning) of serial position (Dunlosky & Matvey, 2001). Although people typically discount extrinsic cues, there are data showing that individuals can account for extrinsic cues appropriately, particularly when they have experience. Castel (2008) showed that participants could account for extrinsic cues appropriately when they have experience with the task. In this study, Castel showed that when making JOLs after studying each item, participants tended to discount the extrinsic cues of primacy and recency effects. However, when participants were asked to make these judgments prior to studying each item, and multiple study-test sessions were employed, participants' judgments became more accurate with experience with the task, and participants were able to incorporate this extrinsic information into their judgments. The findings of the present study are also consistent with this framework. Participants can only use what is salient to make their judgments. The first thing participants were asked to do in Experiment 1 and 4 was to judge how the noise conditions would affect their memories, making the noise conditions quite salient. Participants have pre-existing theories about how noise affects them, and likely gave EOL judgments based on the theory that noise disrupts one's ability to concentrate. Over time, however, calibration improved and demonstrated that participants were able to adjust their judgments and were able to learn about these extrinsic cues with experience with the task. This is in line with Castel's research, which also showed how people learn with experience with these types of extrinsic cues.

The effect of noise condition on people's judgments warrants further investigation into the circumstances in which participants do account for extrinsic cues in making memory predictions, as observed in the present study. Extrinsic cues may affect all judgments, both JOLs and EOLs, especially if the judgment is based on prior knowledge and/or experience, as it was in this study. Overall, "…JOLs are inferential in nature. They are based on a variety of beliefs and cues that are more or less predictive of future memory performance. Thus, in making JOLs, people may rely on theories or beliefs about […] the memorial consequences of different encoding operations and of different study and test conditions…" (Koriat, 1997, pp. 363-4). The EOLs demonstrated that individuals had a theory that noise would affect their memory, but this was not completely accurate as they thought the white noise would be disruptive when it was not.

Calibration data were collected for Experiment 1 and 4. Both of these experiments had similar calibration with respect to the noise conditions. In both studies, participants were most overconfident in the quiet condition. Participants were best calibrated in the white noise conditions, although more so in Experiment 4. With respect to the irrelevant speech conditions, participants rated both the babble and German conditions as disruptive, even though the babble did not actually affect them. Participants became better calibrated over time in the German condition, but they were overconfident in both the babble and German conditions. In terms of the white noise, participants did learn over time that this type of noise does not actually affect them. Particularly in Experiment 4, participants started out overconfident, but over time became better calibrated and hence more accurate with their judgments.

Overall, participants were overconfident in all of the conditions, but became better calibrated in the quiet and white noise conditions more so than the irrelevant speech. Participants did not achieve perfect calibration, but with experience they seemed to learn more and become better calibrated. However, what do participants do when they have done either well or poorly on the previous trial? It is hard to answer this question accurately in this study because the conditions were presented in a random order, therefore participants did not know what noise condition would be coming up next. In a future study it would be worthwhile to block the different noise conditions so calibration could be assessed as participants get more familiar with only one particular type of noise condition at a time. It would be interesting to see how these judgments and thus learning would change over time. Accuracy and changes in a blocked design could then be compared to accuracy in a random-trial design in a future study.

While participants were able to recognize the effect of irrelevant speech sounds with respect to both the global EOL task and the trial-by-trial EOLs, they were rather inaccurate with respect to the effect of white noise; participants thought that the white noise would have a disadvantageous effect on their memory. Actual recall scores showed that there are no differences in performance in the white noise condition compared with the quiet condition in Experiments 2, 3 and 4. It is possible that participants may make more accurate judgments with changing-state noise than steady-state noise. It is also possible that participants think that all types of noise, regardless of the specific nature, will have an effect on their memory. Participants' EOLs were rather high for the white noise, indicating that they thought it would be disruptive. Participants' frequency estimates were also rather high, indicating that they were overconfident with their recall abilities. In the actual recall data, there is a clear pattern in which the white noise has no effect on recall accuracy. Participants seem to consistently think that this type of noise will have an effect, which is consistent with the prior belief that noise makes it more difficult to concentrate, or more difficult to remember. Even though this does not actually apply to all types of noise, it is a reasonable theory for participants to hold.

Overall, these findings support the changing-state hypothesis and O-OER model that Jones has proposed to account for the ISE. There was no semantic meaning in the German speech condition that elicited the effect in Experiments 3 and 4, therefore the decreased performance is most likely result of the nature of the stimuli in that it is changing-state. This type of noise appears to be the most disruptive, as it is more likely to elicit an involuntary attentional switch away from the to-be-remembered stimuli towards the irrelevant stimuli. This type of noise makes it difficult to focus on the intended stimuli, and thus performance is hindered.

The unique result from Experiment 1, whereby participants performed better in the white noise condition than the quiet and babble conditions, was rather unexpected. Previous literature simply shows that there should be limited to no effect observed in the white noise condition. The changing-state hypothesis posits that there should be no effect as well, as this type of noise is steady-state. Ellermeier and Zimmer (1997) demonstrated that there can be a range of individual differences with respect to memory performance in background noise, including slight facilitation. Given the number of previous studies showing no effect of white noise we will assume that this is a Type I error, pending replication. Thus, participants did not make accurate judgments with all types of noise. Due to how the white noise was assessed, as well as the variability of performance in this condition specifically, these findings call for further investigation in terms of what types of noise people can accurately assess with respect to memory performance.

Finally, these findings are also consistent with the previous literature on the ISE, save for the effect of white noise in Experiment 1. Overall, there was decreased memory performance when there was a speech stimulus in the background, the to-be-remembered stimuli were visual, and participants engaged in serial recall. These results were also observed even though the metamemory questions were implemented, which were atypical of an ISE procedure. This shows that the ISE is quite robust, and may continue to be explored further in a metamemory framework.

Conclusion

In everyday life we are often required to flexibly allocate and re-allocate our cognitive resources in order to most efficiently and satisfactorily complete the task at hand. In order to intelligently distribute a pool of limited resources, we must have some insight into the effort involved in each task. In addition, we must have some idea of how our individual cognitive systems function. It is important to determine how well individuals can accomplish this assessment of their cognitive capabilities when they are confronted by a situation in which they must remember information presented in background noise. It is of practical use and importance to investigate what people know about how noise affects memory because if their intuitions are wrong it will lead to suboptimal performance. This, in turn, could lead to failure at such everyday tasks as studying for a test or remembering instructions. Knowing whether there are systematic biases in this area will help us develop interventions to correct the errors. Therefore further research is required to explore whether or not people recognize the effect that different types of noise have on their memory, expanding upon the current research demonstrating that people may have better metaknowledge for some types of noise than others.

References

Arbuckle, T.Y., & Cuddy, L.L. (1969). Discrimination of item strength at the time of presentation. *Journal of Experimental Psychology, 81*, 126-131.

Baddeley, A. D. (1986) *Working memory.* New York: Oxford University Press.

Baddeley, A.D. (2000) The phonological loop and the irrelevant speech effect: some comments on Neath. *Psychonomic Bulletin and Review, 7,* 544-549.

Baddeley, A. D. & Salame, P. (1986) The unattended speech effect: Perception or memory? *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 525-529

Beaman, C., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 459-471. doi: 10.1037/0278-7393.23.2.459.

Bregman, A.S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, Mass.: Bradford Books, MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 115-127.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, *36*, 429-437.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*, 17-31. doi: 10.1016/S0022-5371(76)90003-7.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs). *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*, 1180-1191

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the 'irrelevant speech effect. *Journal of The Acoustical Society of America*, *102*, 2191-2199.

Hallam, S., J. Price, and G. Katsarou. 2002. The effects of background music on primary school pupils' task performance. *Educational Studies, 28,* 111–22.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A.D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control,* (pp. 87-104). Oxford: Clarendon Press.

Jones, D. M. (1999). The cognitive psychology of auditory distraction: The 1997 BPS Broadbent Lecture. *British Journal of Psychology*, 90, 167–187. doi: 10.1348/000712699161314

Jones, D. M., Beaman, C. P. & Macken, W. J. (1996). The object-oriented episodic record model. In: S. Gathercole (Ed), *Models of short-term memory*. Hove: Psychology Press, pp. 209-237.

Jones, D. M. & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 369-381

Jones, D. M. & Macken, W. J. (1995a). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition*, *23*, 192-200. doi: 10.3758/BF03197221.

Jones, D. M. & Macken, W. J. (1995b). Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied*, *1*, 216-226. doi:10.1037/1076-898X.1.3.216.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *44A*, 645-669. doi:10.1080/14640749208401304.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*, 89-108.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349-370.

Kotsopoulou, A., & Hallam, S. (2010). The perceived impact of playing music while studying: Age and cultural differences. *Educational Studies, 36*, 431-440.

LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1396-1408. doi:10.1037/0278-7393.20.6.1396.

LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: A new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *22*, 1154-1165.

LeCompte, D. C., & Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology, 50,* 100-118.

Macken, W. J., Tremblay, S., Houghton, R. H., Nicholls, A. P., & Jones, D. M. (2003). Does Auditory Streaming Require Attention? Evidence From Attentional Selectivity in Short-Term Memory. *Journal Of Experimental Psychology. Human Perception & Performance*, *29*, 43-51.

Metcalfe, J. & Dunlosky, J. (2008). Metamemory. In H.L. Roediger, III (Ed.), *Learning and memory: A comprehensive reference* (pp. 349-362). Oxford: Elsevier.

Miles, C., Jones, D. M., & Madden, C. A. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 578-584. doi:10.1037/0278-7393.17.3.578

Neath, I. (2000) Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7,* 403–423.

Neath, I., Farley, L. A., & Surprenant, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology, 56A*, 1269-1278. doi:10.1080/02724980244000756.

Neath, I., Surprenant, A. M., & LeCompte, D. C. (1998) Irrelevant speech eliminates the word length effect. *Memory & Cognition, 26*, 343-354.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-173). New York: Academic Press.

Nelson, T. O. & Narens, L. I. (1994) Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition* (pp. 1– 26). Cambridge, MA: The MIT Press.

Norris, D., Baddeley, A. & Page, M. P. A. (2004). Retroactive effects of irrelevant speech on serial recall from short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*, 1093-1105.

Perham, N., & Vizard, J. (2011). Can preference for background music mediate the irrelevant sound effect? *Applied Cognitive Psychology*, *25*, 625-631. doi:10.1002/acp.1731.

Richardson, J., & Erlenbacher, A. (1958). Associative connection between paired verbal items. *Journal of Experimental Psychology, 56,* 62-69.

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150-164. doi:10.1016/S0022-5371(82)90521-7.

Salamé, P., & Baddeley, A. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, *30*, 1185-1194. doi:10.1080/00140138708966007.

Salamé, P., & Baddeley, A. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of The Psychonomic Society*, *28*, 540-542.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science, 14*, 322–325.

Stokes, K., & Arnell, K. (2012). New considerations for the cognitive locus of impairment in the irrelevant-sound effect. *Memory & Cognition*, *40*, 918-931. doi:10.3758/s13421-012-0194-4.