MIXED FINDINGS OF THE IRRELEVANT SOUND EFFECT IN SURPRISE RECOGNITION MEMORY TASKS

by Anna MacDermid

A thesis submitted

to the School of Graduate Studies in partial fulfillment of the

requirements for the degree of

MSc Experimental Psychology

Memorial University of Newfoundland

January 2021

St. John's   Newfoundland and Labrador

Abstract

The irrelevant sound effect (ISE) is the finding that irrelevant sound impairs recognition memory performance. Traditionally explored in serial recall, the current study attempts to elicit the effect in a surprise recognition task. Of the many ISE models, only the Object-Oriented Episodic Record (O-OER) model approaches the effect from an order-centric position. As such, given reduced need for order information in this design, O-OER predicts a null effect. A successful manipulation check in Experiment 1 confirmed that the stimuli were sufficient to produce the effect under a standard serial recall design and confirmed statistical equivalency between in-person and online participants. Experiment 2, expanding on Stokes and Arnell (2012), implemented a surprise two alternative forced choice (2AFC) recognition task under quiet, steady-state, and changing-state sound. Performance in the two statistically equivalent sound conditions was impaired, a result consist with all predictions. Experiments 3 and 4 followed this same design but employed alternative cover tasks. Neither experiment reported statistical differences between sound conditions. These results are best described by the O-OER model, however implications for other models such as the Feature model are also discussed.

General Summary

The irrelevant sound effect describes the phenomenon of background sound having a negative effect on our ability to recognize items in a recognition memory task. The cause of this effect is disputed. One theory is that the order of the sounds becomes muddled with the order of the words being remembered. The current series of experiments tests this theory by removing focus from the order of the words being remembered. If order is behind the effect, this will reduce sound and word muddling and therefore have less of a negative effect on recognition memory. This design was achieved by using a surprise multiple-choice test (with 2 choices). This surprise test element stopped participants from adding order by reciting the words. Different cover tasks were used to present the words that were part of the surprise test. The results of the four experiments provided the strongest support for the order-based theory.

Acknowledgements

I would like to thank my supervisors Drs. Aimée Surprenant and Ian Neath for their guidance and support throughout my master's program. Their mentorship has been integral in my development as a scientist. I am grateful for the opportunities they have afforded me and for the atmosphere they cultivated in the Cognition lab. I would also like to thank Dr. Jonathan Fawcett for sitting on my supervisory committee, providing constructive feedback, and walking me through the finer points of Bayesian statistics.

It is incredible what can be learned and achieved in a few short years. I want to thank those friends and family who came along for the ride, offering conversation, encouragement and distraction when needed.

My graduate work was supported by a Canadian Graduate Scholarship awarded to me by the Natural Sciences and Engineering Research Council of Canada. I was also supported by NSERC Discovery Grants awarded to Drs. Aimée Surprenant and Ian Neath.

Table of Contents

Abstract	ii
General Summary	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
Mixed Findings of the Irrelevant Sound Effect in Surprise Recognition Memory Tasks	1
The Effect	2
The Models	5
Challenging the Order-Interference Account	12
The Current Study	17
Experiment 1	18
Methods	18
Results and Discussion	21
Experiment 2	25
Methods	26
Results and Discussion	27
Experiment 3	30
Methods	31
Results and Discussion	32

Experiment 4 35

Methods 35

Results and Discussion 36

General Discussion 38

References 44

List of Tables

Table 1	Proportion Correct Responses in Experiment 1 as a Function of Sound Condition and Participant Type	24
Table 2.1	Hit and False Alarm Rates as a Function of Sound Condition	28
Table 2.2	d' and c in Experiment 2 as a Function of Sound Condition	29
Table 3.1	Hit and False Alarm Rates in Experiment 3 as a Function of Sound Condition	33
Table 3.2	d' and c in Experiment 3 as a Function of Sound Condition	33
Table 4.1	Hit and False Alarm Rates in Experiment 4 as a Function of Sound Condition	36
Table 4.2	d' and c in Experiment 4as a Function of Sound Condition	36

Mixed Findings of the Irrelevant Sound Effect in Surprise Recognition Memory Tasks

In the modern world, noise is increasingly difficult to escape. We can thank industrialization for much of it: the buzzing and beeping of cellphones, the hum of lights and appliances, and the clicks and clacks of keyboards. As workplace environments evolve, becoming reminiscent of the open office layouts of the 1950s, this mechanical and digital soundscape is enhanced by the voices and gadgetry of colleagues. The advent and proliferation of noise-cancelling headphone technology accentuates the relevance and importance of determining the impact background noise has on human cognitive processing. Serial recall, for example, has been shown to be impaired under conditions of background noise, a phenomenon referred to as the irrelevant sound effect (ISE).

In the following sections, I will discuss the ISE literature, delving into the effect's history, its varied empirical directions and explanations, and its theoretical importance, with a focus on the conditions under which the ISE presents. The debated necessity of reliance on order information in a task will be reviewed and the reasoning behind the current study described in detail. A four-experiment series, the current study began with a manipulation check that simultaneously verified that the irrelevant sounds developed elicited the changing-state effect and that online and in-person participants did not differ greatly. Building upon the method developed by Stokes and Arnell (2012), the remaining three experiments investigated the impact of irrelevant sound on recognition memory using a variety of study block cover tasks (lexical-decision, pleasantness rating, and word frequency) in a two alternative forced choice (2AFC) recognition memory task. Recognition tasks are often thought to be relatively free of order information, but a traditional recognition design can allow order information to slip in through participant rehearsal of study items. The use of cover tasks and a single surprise test block circumvents this confounding factor. The results of this series will help determine whether order information is a requirement of the ISE, supporting or undermining the Object-Oriented Episodic Record (O-OER) model (Jones, 1993; Jones & Macken, 1993; Jones et al., 1993).

The Effect

The irrelevant sound effect (ISE) was first discovered accidently. In their seminal study, Colle and Welsh (1976) set out to evoke a simple sound effect in efforts of supporting Sperling's (1967) Model III. In this dual-component model of primary memory, simultaneous presentation of target and irrelevant auditory information was hypothesized to overtax auditory sensory memory, impeding memory processing, and thereby impairing serial recall performance. Colle and Welsh (1976) devised an experiment to clarify the role of filtering (attending only to relevant information) and compensation (rehearsing relevant information aloud or more intensely). Their participants completed 48 trials in which they attempted to serially recall either a phonologically similar or dissimilar set of eight visually presented consonants under noise (foreign speech; quiet) and articulation (restricted; unrestricted) conditions. The simultaneous presentation of foreign speech and target stimuli significantly impaired serial recall performance, but articulation and noise conditions did not interact, a result that conflicted with the compensation hypothesis. This marks the first report of the ISE, defined as worse serial recall performance in the irrelevant sound condition than the quiet condition (note that the O-OER model has a unique definition of the ISE which will be discussed at a later point). Within this seminal work, an intriguing interaction was discovered, the irrelevant foreign speech eliminated the phonological similarity effect (PSE). As discussed below, this finding suggests Colle and Welsh had happened across more than a simple sound or general distraction effect.

Also known as the acoustic confusion effect, the PSE is the finding that participant serial recall performance is worse when items are acoustically similar than when they are acoustically dissimilar (Baddeley, 1966; Conrad, 1964). A robust effect, the PSE has been reported when items are presented visually and even when participants are asked to read items silently. This difference in performance on acoustically similar and dissimilar items is central to the logic that supports the phonological loop as it suggests that visual items are recoded at the encoding stage into auditory ones (Baddeley, 1986). That Colle and Welsh's (1976) foreign speech, later referred to as irrelevant sound, was able to disrupt the PSE suggests that it too operates within the encoding stage of memory. Irrelevant sound is not unique in its ability to interact with the PSE. A similar pattern of results has been reported when individuals study words under concurrent articulation (e.g., while reciting the alphabet); this and other similarities to concurrent articulation will be addressed momentarily (Murray, 1968; Murray et al., 1988).

In efforts to better understand the effect, the types and components of irrelevant sound have received comprehensive study. Usually, this is conducted within a standard serial recall task. Participants study visually presented items under conditions of either quiet or irrelevant sound and are then asked to serially recall the studied items. While Gaussian and white-noise are consistently reported to have no effect on serial recall performance (e.g., Ellermeier & Zimmer, 1997), other non-speech sounds such as tones (Beaman & Jones, 1998; Jones & Macken, 1993; Jones et al., 1993; Jones et al., 1992; LeCompte, 1994; LeCompte et al., 1997; Tremblay & Jones, 1998), music (Nittono, 1997; Schlittmeier et al., 2008), environmental sounds (Buchner et al., 2008), and auditory sequences of randomly ordered consonants (Jones & Macken, 1993; Tremblay et al., 2000) have all been reported to impair serial recall and therefore elicit the ISE. A series of experiments within the same lab using 40 different sound backgrounds, including speech, tones, music, and environmental noise, conclusively reported that speech is the strongest manipulation of irrelevant sound (Schlittmeier et al., 2012). A meta-analysis conducted in 2014 corroborates this, reporting that speech sounds as a group, i.e., foreign, reversed, and laboratory transformations, consistently produce larger ISEs than their non-speech counterparts (Ellermeier & Zimmer, 2014).

While sound type is clearly central to the ISE, it is independent of many other factors. Habituation, for example, has little to no effect on the ISE. While serial recall performance has been reported to improve in a second session, further exposure provides no further benefit and despite modest overall performance gains, the ISE is relatively unchanged between practice sessions (Hellbrück et al., 1996). More theoretically meaningful, as it refutes attentional processes, is the finding that the intensity of the irrelevant sound has no impact on the magnitude of the effect (Colle, 1980; Ellermeier & Hellbrück, 1998). So long as the irrelevant sound is discernable, defined as above 30 dB, the ISE will be observed (Colle, 1980).

The timing of sound presentation has also been reported as inconsequential (Hanley & Bakopoulou, 2003; Hughes et. al, 2005; Jones et al., 1992; Miles et al., 1991; Röer et al., 2014). Irrelevant sound at encoding, rehearsal, and recall phases were investigated in a modified serial recall design (Miles et al., 1991; Experiment 1). The 10 s rehearsal phase, an addition to the standard design, was situated between the encoding and recall phases. Results showed that irrelevant sound presented during encoding and rehearsal phases equally impaired serial recall performance. Had the ISE only occurred when irrelevant sound was presented during the study phase, or even if it had more of an impact, then it would be clear that it affected the encoding process. This ISE equivalency between encoding and rehearsal phases clearly shows that the effect impacts more than just encoding. Further evidence that the ISE is unlikely to be caused by attentional processes comes in the report of a non-effect of irrelevant sound in a traditional

Stroop task (Miles et al., 1989). Given the sensitivity of the Stroop test to interference, had attention been central to the ISE significant impairment in performance would have been found.

Features of to-be-remembered (TBR) items have also been investigated. The role of semantics for the most part has been considered independent from the ISE (Buchner et al., 2004; Klatte et al., 1995; LeCompte et al., 1997). Presentation modality of TBR items has also been manipulated and the ISE has been produced in visual, auditory, and lip-reading presentation modalities (e.g., Campbell & Dodd, 1984; Hanley & Broadbent, 1987; Jones, 1994; Salamé & Baddeley, 1982; Surprenant et al., 1999) indicating cross modal interference.

The concurrent articulation effect also occurs across modalities and furthermore, interacts with changes in modality in similar ways to the ISE, negating the PSE when using visual TBR items, yet leaving it intact when using auditory TBR items (Longoni et al., 1993; Peterson & Johnson, 1971; Surprenant et al., 1999; Hanley, 1997; Hanley & Broadbent, 1987; Rouleau & Belleville, 1996; Salamé & Baddeley, 1982, 1987). The two phenomena also interact in a similar way with the word-length effect, the finding that serial recall for short words is superior to that for long words, reducing it in all presentation modalities (Baddeley et al., 1975; Mackworth, 1963; Neath et al., 1998; see Tremblay et al. (2000) for a replication refuting this finding).

The Models

Four competing theoretical models of the irrelevant sound effect (ISE) have emerged. The phonological store hypothesis, out of Baddeley's Working Memory model (1986, 1992), the changing-state hypothesis from Jones' O-OER model (1993; Jones & Macken, 1993; Jones et al., 1993), the Attentional Account (Bell et al., 2012; Cowan, 1995; Elliott, 2002), and the feature adoption hypothesis from Neath's Feature model (2000). I discuss each in turn.

The Phonological Store Hypothesis

Baddeley's Working Memory model places the locus of action for irrelevant sound interactions within the phonological loop, a system consisting of the phonological store and articulatory control process (1986; 1992). The phonological store retains language information for a short time before memory decay begins. This decay can be delayed by activating the articulatory control process (ACP). The primary undertaking of the ACP however, is to recode visual information into auditory speech-based information. It is this process, under the Working Memory model, that is assumed to be behind subvocal rehearsal (Baddeley, 1986; 1992).

Baddeley theorizes that irrelevant speech enters the phonological store alongside TBR items. It is here in the phonological store that the irrelevant speech is said to interfere with the TBR items, impairing performance. The mechanism behind this interference is not explained in detail and therefore it is difficult to derive specific predictions. The model cannot explain how between-stream phonological similarity does not consistently elicit the ISE (Bridges, 1996; Jones & Macken, 1995; LeCompte & Shaibe, 1997). It is also unable to address the similarities between irrelevant sound and concurrent articulation as it places the locus of action for the two effects in the phonological store and ACP, respectively (Baddeley et al., 1984; Baddeley et al., 1975; Neath et al., 1998; Peterson & Johnson, 1971; Surprenant et al., 1999). Similarly, the model is unable to explain how the ISE interacts with the word length effect which, like concurrent articulation, is believed to take place in the ACP (Neath et al., 1998).

The Changing-state Hypothesis

Given the shortcomings of the phonological loop hypothesis, an alternative model of the ISE soon emerged. While arguing for the equivalency of speech and non-speech sounds, Jones et al. (1992) uncovered a novel finding. A series of four experiments, in which the characteristics of the irrelevant sound stimuli in a serial recall task were manipulated, showed that a minimum of four repeating syllables were required for the ISE to be produced. Speech sounds in and of themselves (i.e., a repeated single syllable) were not sufficient. The terms steady-state and changing-state sound were developed shortly thereafter to distinguish irrelevant sounds comprised of a single repeating sound (steady-state sound) from those comprised of at least four repeating sounds (changing-state sound). These results underlie the changing-state effect – that changing-state sound impairs serial recall performance to a greater extent than steady-state sound – and form the basis of the Object-Oriented Episodic Record model (O-OER).

The O-OER model represents items as abstract amodal objects (Jones et al., 1996). Serial information is encoded within pointers that link these objects together in what is referred to as an episodic trajectory. The episodic trajectory, also called a stream, needs to be intact for accurate serial recall. As pointer linkages are vulnerable to interference and decay over time, serial recall performance is often less than perfect. Changing-state sound, due to its segmented nature, creates its own stream complete with auditory objects held together by order information laden linkages (Jones et al., 1996). Conversely, the lack of segmentation in steady-state sound means it creates a single auditory object with no linkages and no order information. Within this model serial rehearsal is the act of tracing the episodic trajectory of the encoded information (Jones et al., 1996). However, if there are irrelevant changing-state sound stimuli present during rehearsal, the two streams can intersect and break the links in the TBR stream. Thus, like encoding, serial rehearsal can be influenced by concurrent streams and pulled off the correct trajectory, impairing performance.

Within this model, speech is merely another example of a changing-state sound. This claim is supported by equivalent impairment in serial recall performance under conditions of irrelevant speech and irrelevant sequences of syllables and tones that varied together in pitch (Jones & Macken, 1993, Experiment 5). In addition to suggesting speech and non-speech equivalency these results point to the ISE having an acoustic rather than phonological cause. This distinction between steady-state and changing-state sound has led the O-OER model to the development of a unique definition of the ISE – worse serial recall performance in a changing-state sound condition than in both a steady-state sound condition and a quiet condition. In other words, within this model the ISE requires the presence of the changing-state effect. This is because according to the model, the ISE only occurs when streams intersect and break links in the TBR stream – a possibility only under changing-state sound conditions.

Unlike the Working Memory model, the O-OER model and changing-state hypothesis provide an explanation for the similarity between irrelevant sound and concurrent articulation. Both effects are described as impairing serial recall performance by interfering with the episodic trajectory of TBR information. The changing-state hypothesis is also able to support findings that the ISE does not habituate. Changing-state sound is described as evoking constant attentional orientation, while steady-state sound only orients the once. Given this constant re-orientation, habituation cannot occur under changing-state sound. Nevertheless, some of the O-OER model's assumptions and predictions have been challenged. Most damaging are reports of the ISE beyond the confines of order information and serial recall designs. Because the interference in this model is thought to be confined to the breaking of connections between adjacent items, rather than interference with memory for the item itself, the O-OER model predicts that only tasks that rely on seriation will show an ISE.

While early work trying to show an ISE in free recall, a non-seriated task, was unsuccessful (Salamé & Baddeley, 1990), LeCompte (1994, 1996) has since reported significant ISEs using free recall, paired-recall and recognition – all three tasks that are thought to have reduced reliance on order information. Additionally, disparate results of irrelevant speech and irrelevant tones on the word length effect challenge the assumed equivalency of irrelevant speech and tones (Neath et al., 1998). The O-OER model has also been challenged by reports of modality dependent effects (e.g., Peterson & Johnson, 1971; Surprenant et al., 1999) that suggest assumptions of amodality and equivalency with concurrent articulation are flawed.

The Attentional Account

The Embedded Processes model defines working memory as a collection of information stored in long-term memory that has been activated and made accessible (Cowan, 1995). The most highly activated and accessible representations of memory information are referred to as the focus of attention. According to this model, to complete a serial recall task participants must bring target items into the focus of attention and maintain them in working memory until testing. Selective Attention, the process of selecting and processing only relevant information in an environment, allows for this to occur. To be selected and processed incoming information must be either novel (i.e., changing) or relevant to the current task (Cowan, 1995).

The Attentional Account argues that, due to its novel status, irrelevant sound is often selected and processed. Because attentional processing resources are limited, the processing of irrelevant sound means that these resources are diverted from target items (Cowan, 1995). Irrelevant sound is also thought to pull the focus of attention away from target items, reducing target rehearsal in working memory (Cowan, 1995). Both arguments are supported by evidence showing that young children are more adversely affected by irrelevant sound than adults (Elliot, 2002). This is to be expected as not only do young children have less control over attention, they are also less equipped to properly select relevant visual items (Cowan, 1995).

The habituation of attentional orienting responses (Sokolov, 1963) is theorized to be a mechanism behind Selective Attention. Habituation helps filter out irrelevant information and conserve limited attentional processing resources. It also provides an alternative explanation for the O-OER model's changing-state effect as changing-state sound is harder to habituate than steady-state sound and so should cause more interference. Yet habituation does not occur using the standard ISE paradigm (Hellbrück et al., 1996). Bell et al. (2012) created optimal circumstances for habituation by incorporating a pre-trial 45 s passive listening phase to a serial recall task. This change in design did reduce the ISE, suggesting that habituation of irrelevant sound is possible and strengthening the argument of a central role of attention in the effect.

The Attentional Account is unable to explain the near non-effect of intensity of sound (Colle, 1980; Ellermeier & Hellbrück, 1998). If the ISE is caused primarily by attentional factors, increasing the intensity of the irrelevant sound would increase the size of the ISE; however, reports have shown that irrelevant sounds ranging from 40dB to 75dB have the same deleterious effect on serial recall performance (Colle, 1980).

Feature Adoption Hypothesis

Originally designed to account for immediate memory phenomena, including modality effects, the Feature model (Nairne, 1988, 1990) represents items in memory as vectors of features (Hintzman, 1991). Like pixels on a screen, features do not convey much information alone. It is only when enough are assembled that the represented item can be identified (Neath, 2000). In sharp contrast to the phonological store hypothesis, changing-state hypotheses, and Attentional Account, the Feature model addresses modality effects. The primary assumption of the model is that there are two categories of features, those that are modality-dependent and those that are modality-independent. The former are features that carry information unique to the presentation modality (e.g., letter case, accent, etc.) while the latter carry information that is semantic and speaks to the nature of the item (Neath, 2000).

The second assumption of the Feature model is that there are two types of memory: Primary memory, which maintains and processes cues, and secondary memory responsible for item storage and recall. At presentation, an item enters both primary and secondary memory. The item in primary memory becomes degraded over time due to retroactive interference, conversely, the item in secondary memory remains intact. Retrieval is achieved by matching the degraded primary memory item (the cue) to the item that best matches it from secondary memory (Neath, 2000).

Mechanistically modelled after the Ranschburg effect - that items are rarely recalled more than once, even if they were presented more than once (Crowder & Melton, 1965; Henson, 1998) - the Feature model addresses memory loss by using a process called feature overwriting. If an item in a list shares features with the item immediately preceding it the features in the earlier item are overwritten, degrading the primary memory cue. While the O-OER model stores order information in object pointers, the Feature model encodes and stores order information with the primary memory cue (Neath, 1999). Like primary memory cues can be degraded, order information can drift. A positional uncertainty gradient is calculated for each primary cue indicating the probability of the cue being in each of the possible serial positions. This means there are two ways order error can come about. Drift can occur causing the cue to get selected out of order or the cue can be matched with the wrong secondary item (Neath, 2000).

Of the four models presented, the Feature model is the only one that can explain the presence of a recency effect when auditory items are presented concurrently with irrelevant sound. The last item - by virtue of being the last item – does not experience modality-dependent retroactive interference (Neath, 2000). The Feature model also uniquely addresses suffix effects – the phenomenon where the addition of a speech item at the end of a TBR list impairs the serial recall of the last item. When irrelevant sound is presented during the retention period the suffix overwrites the modality-dependent features of the last list item. The word length effect too is explained; the longer the word, the more phonemes and segments it has (Melton, 1963). More complexity brings more opportunity for error, leading to worse serial recall performance for longer words than shorter ones. The Feature model also explains the PSE which is rooted in the model itself as the matching of primary memory cues to secondary memory stores rests on similarity. Lastly, the similarity of irrelevant sound and concurrent articulation is understood by the concept of feature adoption which is described as the mechanism of action for both effects (Neath, 2000). Features from the articulated words and irrelevant sound are both able to be adopted by the TBR items, adding noise to the TBR modality-independent features (Murray et al., 1988).

The Feature model is not devoid of problems. It does not predict that irrelevant sound will interfere with serial order, as the interference is at the feature (item) level. In addition, the concept of feature adoption does not extend well to the use of non-speech irrelevant sounds. This is because retroactive interference occurs only due to overlapping features. Non-speech sounds do not share significant features with TBR word stimuli.

Challenging the Order-Interference Account

The O-OER model is the only model to hypothesize that order information is essential to the ISE. Reports of ISEs in serial recall tasks align with this hypothesis, however, it is also important to confirm that the effect occurs outside of this unique task and inversely, does not occur in tasks that require reliance on limited order information. Farley et al. (2007) were able to address this first requirement in a novel study of irrelevant sound in sequence learning, a task for which of course serial order is essential. The O-OER model's hypothesis was supported by findings of significantly lengthened response times in irrelevant sound trials that required reliance on sequence learning. Results of studies investigating the ISE in tasks where participants do not use order have been less conclusive.

Simple reaction time, rhyming judgement, and 16-item free recall tasks have all been investigated under irrelevant sound conditions and have yielded inconsistent results (Baddeley & Salamé, 1986; Kjellberg & Sköldström, 1991; Salamé & Baddeley, 1990). Missing-item tasks have also been used to investigate the ISE and while reports state a null-effect (e.g., Beaman & Jones, 1997; Hughes et al., 2007; Jones & Macken, 1993) the implications of this are disputed. Bell et al (2013) argue that as the task relies on familiarity, it is more akin to an associative memory than an item memory task. They then theorize that context binding, rather than order information, underlies the ISE. Items, they argue, can be bound to temporal, perceptual and spatial contexts. The ability of items to bind to temporal contexts explains why tasks that require order information elicit the ISE. A 2X2 design comparing an item-order associative task and an item-colour associative task under both irrelevant sound and quiet conditions explored this further (Bell et al., 2013). The item-order associate task presented participants with seven consonants in a random order. At test participants were asked to indicate the order in which the consonants had been shown. The item-colour associative task was methodologically similar but deviated at test where participants were instead asked to indicate the background colour presented alongside each consonant. Irrelevant sound effects were found for both tasks. This result was replicated in Experiment 2 which substituted words for consonants and incorporated an item memory measure by including seven new items to the test list (bringing it to a total of 14). Not only did this new design replicate the results of the first experiment, it also provided evidence that irrelevant sound can impair recognition memory in both item-order and item-colour association tasks.

Earlier work by LeCompte (1994) investigated the ISE in free recall and recognition memory tasks using visual TBR item presentations. The first three experiments of the series used a free recall design where participants were presented with 12 consonants in either quiet, short bursts of white noise, or irrelevant sound (speech and non-speech). Results showed a small but significant difference between all conditions and replicated the findings of presentation timing independence (ISE occurred when irrelevant sound was presented during or after encoding); Experiment 2) and of semantic independence (non-words were also capable of generating the ISE; Experiment 3). Experiment 4 adopted the 16-item list length used previously by Salamé and Baddeley (1990) and similarly found no evidence of an ISE. This was determined to be due to differences in serial rehearsal strategy for short and long lists and so Experiment 5 expanded to a recognition design in which participants completed 96 trials of slightly varied tasks and methods. Word lists ranged from 8 to 12 and both rhyming and non-rhyming distractor words were used at test, which in Experiment 5A followed a yes-no probe procedure and in 5B and 5C a forced choice probe procedure. All three versions of Experiment 5 produced an ISE. In his final experiment, LeCompte designed a cued recall task that discouraged serial rehearsal by presenting items as pairs. Participants completed 48 trials each consisting of 12 word-pairs. A small but significant difference was found between the white noise condition and the speech condition.

This study has been criticized for its use of short list lengths (word lists contained 16 or fewer items) and high frequency of testing (testing occurred after each word list; Le Compte, 1994; Beaman & Jones, 1997). Proponents of the O-OER model argue it was possible participants were using serial rehearsal in LeCompte's experiments of their own volition. Supporting this argument, the addition of a concurrent articulation task - to disrupt any attempted serial rehearsal - was found to reduce the ISE in both free recall and recognition tasks (Beaman & Jones, 1997).

To address this concern of serial rehearsal, Stokes and Arnell (2012) implemented a series of five experiments. The first of these, a replication and extension of LeCompte (1994), acted as a control for those that followed. Fully informed participants began by performing a lexical-decision task in blocks of either quiet or irrelevant sound (square-wave tones that ranged in pitch, i.e., changing-state sound). In total they viewed 200 visually presented words and non-words. They were then given a checklist of words and instructed to indicate which they recognized from the lexical-decision task. Next, participants completed a serial recall task consisting of 30 trials split between blocks of quiet and irrelevant sound. Nine randomly selected consonants were used as visual stimuli for each block. Significant ISEs were found in both the recognition task and the serial recall task. As the word list consisted of 100 items it is unlikely that participants were using serial rehearsal. The presence of an ISE in this recognition task suggests that the effect does not require serial order information.

The second experiment in the series further isolated the ISE from serial rehearsal by surprising participants with the recognition task. As in Experiment 1, participants completed a lexical-decision task, only this time they were unaware of the imminent recognition task. Results again reported an ISE. Further experiments in the series maintained this surprise recognition task element and manipulated the irrelevant sound condition. Experiment 3 substituted foreign speech for the square waveform tones used in Experiments 1 and 2. Experiment 4 returned to these earlier tones and implemented a steady-state sound condition (tones did not vary in pitch) alongside the quiet and changing-state sound conditions. Irrelevant sound impaired recognition memory in all sound conditions: Foreign speech, steady-state sound (tones), and changing-state sound (tones). These results were interpreted as providing good evidence that order information is not required for the ISE to occur.

Although the results reported by Stokes and Arnell (2012) are intriguing, flaws in their design call their conclusions into question. First, and most damaging, recognition memory performance in Experiment 4's steady-state and changing-state sound conditions were statistically equivalent. This means that there was no changing-state effect and suggests that their results might be due to general attentional factors, rather than an ISE. The absence of a changing-state effect in Experiment 4 (and lack of changing-state sound in Experiments 1-3) also means that these results do not fit the definition of an ISE under the O-OER model and therefore do not challenge it. Second, given the use of an old/new recognition design, the results of the recognition task had to be analyzed using a shared false alarm rate. It is possible that this design hides a small changing-state effect. In an old/new recognition task participants are presented with individual items and must decide if they are from the study list (an old item) or not (a new item). These decisions have four possible outcomes: Hits (in which the participant correctly identifies an old item as old), Misses (in which the participant incorrectly identifies an old item as new), false alarms (in which a participant incorrectly identifies a new item as old) and Correct Rejections (in which a participant correctly identifies a new item as new). Hit rates and false alarm rates are then used to calculate sensitivity (how well the observer can detect the correct stimulus) using a measurement called *d'* (*d'* = *z*(hit rate) - *z*(false alarm rate)). Note that a larger *d'* score indicates greater sensitivity, meaning that participants are able to discriminate the signal (old item) from the noise (new item) at a rate higher than chance (*d'* = 0.00). Because old/new recognition only presents one item at a time false alarm rates cannot be calculated for unique sound conditions and so a shared false alarm rate is used in the calculation of *d'*. As such, Stokes and Arnell's *d'* analysis is not as informative as it could be and possible nuances between conditions are lost. The only clear results come from proportion correct (Hits) which introduces the possibility of response biases clouding the results. To help illustrate this consider an extreme example: if a participant were to respond old to every word under proportion correct analysis they would be scored as performing at 100%. Finally, Stokes and Arnell (2012) only used one orienting task in their design (lexical-decision) and I was curious as to whether other cover tasks led to the same result.

The Current Study

The current study extends the work of Stokes and Arnell (2012) by using a more robust manipulation of irrelevant sound (speech) and adjusting the experimental design to allow for improved calculation of *d'*. A two alternative forced choice (2AFC) recognition design was implemented as it allows for the calculation of unique false alarm rates thanks to participants having to make a choice between an old item and a new item. The proper sound condition can be attributed through the old item that was paired with the new item that the participant false alarmed to at test. The focus in Experiment 1 was on ensuring that the irrelevant auditory stimuli were successful in producing a changing-state effect in serial recall prior to exploration of the irrelevant sound effect (ISE) within recognition paradigms. Online participants recruited through Prolific Academic were also tested for participant quality. Experiments 2 through 4 employed a lexical-decision, pleasantness rating, and word frequency cover task, respectively. Participants were not told that a 2AFC recognition memory task would follow.

It was hypothesized that the procedures followed in Experiments 1-4 would elicit the changing-state effect – worse serial recall/recognition memory performance in the changing-state sound condition than the steady-state sound condition. The segmentation of sound in continuous speech (the changing-state sound condition) should be, according to the O-OER model, sufficient to disrupt the streams, thereby disrupting order information and the object pointers. Repeated speech (the steady-state sound condition) in comparison should not disrupt streams or order information. The second hypothesis was that all experiments would report an ISE– worse serial recall/recognition memory performance in both sound conditions than the quiet condition. The unanticipated nature of these surprise recognition memory tasks circumnavigates some of the criticisms of past designs. Cover tasks, infrequency of testing, and large set size (300 stimuli) makes unprompted serial rehearsal unlikely. This would lend support to models positing that the ISE does not require order information and would call the O-OER model into question.

Experiment 1

Experiment 1 was designed as a manipulation check. It was critical to first ensure that the irrelevant stimuli used were able to produce a "classic" changing-state effect in a serial recall task. Thus, participants were presented with TBR visually presented words in conditions of quiet, steady-state sound, and changing-state sound in a typical serial recall task. In addition, to assess online participant quality, a manipulation of participant pool was included. Half of the participants were tested in a traditional face-to-face laboratory setting and half using an online testing service called Prolific Academic. It was predicted that, like the many studies showing a changing-state effect with a similar design, an ISE and changing-state effect would be found and that the two participant types would yield similar results.

Methods

Participants. Thirty participants were recruited from Memorial University of Newfoundland's St John's campus for Experiment 1A. All were native speakers of English. The mean age was 22.63 (SD 4.41, range 18-36) and 26 self-identified as female and 4 as male. Most participants were undergraduate students recruited through the university's Psychology Research Experience Pool (PREP). These students received course credit for their participation. All other participants were recruited on campus through a paid recruitment process and were paid $10.00 for their time.

Experiment 1B recruited 30 volunteers from Prolific Academic. These participants were compensated £8 per hour (prorated). All participants were native speakers of English; between

19 and 39; and had an approval rating of at least 90% on prior studies. The mean age was 31.33 (SD 4.65, range 22-39) and 15 self-identified as female and 15 as male. Location information was not recorded.

The sample size for both Experiment 1A and 1B was determined by two considerations. First, a changing-state effect was reported by Jones et al. (1992, Experiment 4) using 24 participants in a similar within-subjects design. The primary difference between the current study and that of Jones et al. (1992) is that in addition to having a quiet condition, a steady-state, and a changing-state sound condition, Jones et al. also had a 4th condition that can be described as "partially changing". In this condition, four letters were repeated. Second, estimates of power made by Faul et al. (2009) indicate that a sample size of 30 would yield power in excess of 0.90 for a range of likely effect sizes. For these calculations, we estimated the changing-state effect size in studies that used similar designs.

Stimuli/Materials. For those participating in Experiment 1A, the experimental stimuli were presented using Apple Macintosh desktop computers in private testing rooms in the Cognition and Memory lab. Stimuli were presented visually in white capitalized letters in 28 point Helvetica in the centre of a black background. Visual stimuli were 434 two-syllable common English words gathered from the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm; Kučera & Francis, 1967). All words were between four and six letters in length, emotionally neutral, and of moderate to high familiarity and concreteness. Each trial consisted of the individual presentation of six randomly selected words.

The irrelevant stimuli – quiet, irrelevant repeating speech (steady-state sound), and irrelevant continuous speech (changing-state sound)– were delivered to participants through headphones. The volume was set by participants to a comfortable level. The irrelevant stimuli consisted of the digits 1-9 produced by the default voice "Tom" on macOS 10.14. We chose to use digits as the irrelevant speech because the content of the irrelevant speech has little effect (e.g., LeCompte & Shaibe, 1997) and digits are easy to generate digitally. Two digits were excluded from the steady-state sound condition, the digit 7 because it has two syllables and the digit 6 because pilot testing indicated that some participants began perceiving this as "sex" when repeated multiple times.

The participants in Experiment 1B completed the experiment online through a web browser on their personal computers with their own headphones. All stimuli were as stated above in Experiment 1A.

Design. A within-subjects design, participants performed a standard serial recall task with three sound conditions: quiet, steady-state sound, and changing-state sound.

Procedure. Participants in Experiment 1A were welcomed into the lab where they were informed of the experimental design and procedure, gave their voluntary and ongoing informed consent, and received debriefing and experimenter contact information (as per the study's ethics procedures which received ethics approval from the Interdisciplinary Committee on Ethics in Human Research Participants). Task instructions were reviewed with participants in detail and an opportunity for questions was provided before the experiment began. Participants in Experiment 1B received this information and completed the consent procedure online. All participants responded to a brief demographic questionnaire before transitioning to the experiment proper. Thirty trials were evenly divided into three counterbalanced sound blocks (quiet, steady-state sound, and changing-state sound). A trial consisted of the individual presentation of six randomly selected words. To be remembered words were each displayed in the centre of the screen for 1000 ms with the next word appearing immediately afterward.

For the two noise conditions, a digit was played every 800 ms. For the changing-state sound condition, digits were randomly chosen with the constraint that the same digit could not be played twice in a row. For the steady-state sound condition, a digit was randomly chosen and that digit was used for all steady-state sound lists for that particular participant.

In the test phase, which began immediately after the offset of the final word, participants received a series of on-screen prompts. Each prompt asked participants to enter a word from the current trial. Typing prompts were ordered sequentially, starting with a prompt for the first word presented. After submitting their response participants were unable to go back and make corrections. An option to pass was available at every serial position. Upon answering or passing on the sixth prompt participants could take a short break before pressing a key to launch the next trial. Participants were informed at the onset of the task that they would be presented with a series of six words which they would later be asked to recall in order. They were also made aware that some trials would be accompanied with sound, which they were to ignore.

Results and Discussion

Data from Experiments 1 - 4 were analyzed using both frequentist and Bayesian techniques in JASP (JASP Team, 2019). For the latter, note that a Bayes Factor (BF_{10}) between 3 and 20 indicates positive evidence for the alternate hypothesis (and therefore evidence against the null hypothesis), a BF_{10} between 20 and 150 indicates strong evidence, and a BF_{10} greater than 150 indicates very strong evidence (Kass & Raftery, 1995). Similarly, BF_{01} indicates evidence for the null hypothesis. In the Bayesian analyses main-effect models were evaluated with respect to the default random-effects error model, and interaction models were evaluated with respect to the default main-effects model. Default priors were used. Note that in all analyses non-integer degrees of freedom indicate that the Greenhouse-Geisser sphericity correction was applied.

Participant responses for Experiments 1A and 1B were run through an automated spellchecker for spelling and obvious typing errors, e.g., entering "WISDON" instead of "WISDOM". Participant type groups had similar error rates. For the in-person participants, the mean percentage of errors were 2.00%, 1.78%, and 1.94% in the quiet, steady-state, and changing-state sound conditions, respectively. Correcting the error changed a response from incorrect to correct for 27 (out of 1800) responses in the quiet condition compared to 17 and 21 in the steady-state and changing-state sound conditions. For the online participants, the percentage of errors were 1.94%, 2.11%, and 1.61% in the quiet, steady-state, and changing-state sound conditions, respectively. Correcting the error changed a response from incorrect to correct for 26 (out of 1800) responses in the quiet condition compared to 24 and 17 in the steady-state and changing-state sound conditions. Because there was no difference in the statistical results between the raw and corrected responses, only the raw responses are reported.

Planned paired samples t-tests indicated that both an ISE and changing-state effect were present in Experiment 1. The proportion of words recalled in the correct order in the quiet condition ($M = .51$; $SD = .16$) was significantly greater than the proportion recalled in the correct order in the steady-state sound condition ($M = .45$; $SD = 0.15$, $t(59) = 4.68$, $p < .001$, $d = 0.60$, $BF_{10} = 1105.28$). Similarly, the proportion of words recalled in the correct order in the steady-state sound condition were significantly greater than in the changing-state sound condition ($M = 0.40$; $SD = 0.15$; $t(59) = 4.30$, $p < 0.001$, $d = 0.56$; $BF_{10} = 327.98$), see Table 1.

These results show a textbook ISE; the sound conditions significantly impaired serial recall performance compared to the quiet condition. Crucially, a changing-state effect was also found, changing-state sound significantly impaired serial recall performance compared to the steady-state sound condition. Having confirmed that the developed irrelevant sounds elicit both the ISE and changing-state effect in serial recall, they can be confidently used in subsequent experiments in this series.

The presence of the changing-state effect confirms that the irrelevant sound is doing more than just diverting attention, as had that been the case steady-state and changing-state sound conditions would have been relatively equal. That there is a significant difference between these two types of sounds suggests that the ISE is disrupting memory directly. Experiments 2-4 will use these same auditory stimuli in surprise recognition tasks, the details of which are outlined in the appropriate methods sections.

The secondary goal of Experiment 1 was to compare Prolific Academic and in-lab participant responses. Experiment 1, which acted as an irrelevant sound manipulation check, was run in-person with MUN students (Experiment 1A) and online with Prolific Academic participants (Experiment 1B), see Table 1.

A repeated measures ANOVA reported a main effect of sound condition, $F(1.79,103.86) = 34.764$, $MSE = 0.006$, $\eta_p^2 = 0.375$, $p < 0.001$, $BF_{10} = 6.73 \times 10^9$. Analysis of a main effect of participant type differs subtly between approaches, $F(1,58) = 2.66$, $p = .11$, $BF_{10} = 1.01$; while frequentist analysis reports a null effect, Bayesian analysis was ambiguous. No evidence of an interaction between sound condition and participant type was found, $F(1.79,103.86) =.11$, $p = .87$, $BF_{01} = 8.73$. Both frequentist and Bayesian accounts agree that sound condition is the main factor that explains the differences in the data. In contrast, participant type has little to no impact on performance in irrelevant sound tasks. Future irrelevant sound experiments can safely recruit participants from Prolific Academic knowing that this factor will have little to no impact on task performance.

Table 1
Proportion Correct Responses in Experiment 1 as a Function of Sound Condition and Participant Type

	Quiet			Steady-state sound			Changing-state sound		
	In-Person	Online	Combined	In-Person	Online	Combined	In-Person	Online	Combined
Mean	0.49	0.54	0.51	0.42	0.48	0.45	0.37	0.44	0.40
Std. Deviation	0.16	0.15	0.16	0.13	0.16	0.15	0.12	0.17	0.15

Note. $N = 60$, In-person $n = 30$, Online $n = 30$.

Experiment 2

Experiment 2 replicated and extended Stokes and Arnell's (2012) surprise yes/no recognition task design. Due to their design, yes/no recognition tasks are only able to provide a single false alarm rate. This means that the same false alarm rate was used when calculating *d'* for each of Stokes and Arnell's sound conditions. This has the potential to be hiding a small changing-state effect as it is possible that false alarm rate did differ across sound conditions. By introducing a 2AFC recognition task *d'* can be calculated using false alarm rates unique to each sound condition. This is because each test trial consists of an old word and a new word. The old word is both the correct choice and a means of tagging the test trial with the sound condition the old word appeared in during the study phase.

This experiment, like that of Stokes and Arnell (2012), used a lexical-decision task as a cover for the 2AFC study phase. The test phase of the 2AFC therefore came as a surprise to participants. This decision was made in attempts to minimise rehearsal and therefore reduce task reliance on order information. The models outlined above offer different predictions for the results of Experiment 2. The O-OER model predicts that given the task does not rely on order information, no ISE will be found. In contrast, the Feature model, Working Memory model, and Attentional Account predict that an ISE will be found as they do not consider reliance on order information essential to the effect.

In addition to recognition memory performance, accuracy and response times (RTs) for the lexical-decision task were also recorded and analyzed. Stokes and Arnell (2012) reported a decrease in RTs under irrelevant sound conditions. This is not the only time such a trend has been reported when using irrelevant sound in decision making tasks (e.g., Smith, 2010). It is predicted that this replication will also follow this trend.

Methods

Participants. Thirty Memorial University of Newfoundland students volunteered to participate in exchange for course credit or financial compensation ($10 for 1 hour). All were native speakers of English. The mean age was 20.97 (SD 3.17, range 18-31) and 22 self-identified as female and 8 as male. Fifty-eight volunteers from Prolific Academic participated and were compensated £8 per hour (prorated). Prolific Academic participant inclusion criteria were identical to Experiment 1B. The mean age was 28.53 (SD = 4.89, range 19-38), and 31 self-identified as female and 27 as male. Location information was not recorded.

Sample size was determined through analysis of Stokes and Arnell's fourth experiment (2012). As they were unable to elicit a changing-state effect with a sample size of 44 participants, the sample size was increased to 88 participants.

Stimuli/Materials. Experimental stimuli consisted of 720 monosyllabic words gathered from the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm; Kučera & Francis, 1967). All words were between four and six letters in length, emotionally neutral, and of moderate to high familiarity and concreteness. A pool of 256 phonotactically correct non-words were also assembled. All non-words were four to six letters in length and followed the phonetic and morphemic rules of the English language. None of the non-words were homonyms of actual words.

The same irrelevant sound stimuli were used as in Experiment 1.

Design. A within-subjects design, participants were individually presented with a series of visual word and non-word stimuli. Each presented stimuli required completion of a lexical-decision task. Trials were randomly and evenly divided between the three sound conditions (quiet, steady-state sound, and changing-state sound).

Procedure. Participants were invited to participate in two experiments investigating the impact of irrelevant sound on task performance. Instructions were given immediately before each task. In the first task, participants completed a series of 144 lexical-decision tasks in three evenly divided and counterbalanced irrelevant sound blocks. Participants were informed through a written message on the screen when one sound block had ended and the next was about to begin. When completing the lexical-decision tasks participants were instructed to categorize the visual stimuli as quickly and accurately as possible all while ignoring any auditory information emitting through their headphones. They were to indicate their decision with a keypress ("Z" for word, "M" for non-word). Visual stimuli remained on the screen until a categorization decision key was pressed. A 1000 ms intertrial interval preceded the subsequent trial. At no point was an impending recognition memory test mentioned or alluded to.

The second task was a 2AFC recognition memory task. Participants were directed to select the word that had been a part of the lexical-decision task by pressing a key. The "Z" key indicated the old word was on the left of the screen and the "M" key indicated the old word was on the right. As in the lexical-decision trials no time limit was given and visual stimuli remained on the screen until a key was pressed. The recognition task consisted of 72 trials, 24 words from each condition block were paired with 24 new foils. Words were randomly selected and ordered.

Results and Discussion

Means and standard deviations of hit and false alarm rate by sound condition can be found in Table 2.1. A repeated measures ANOVA of hit rate found a main effect of condition, $F(2,174) = 3.962$, $MSE = 0.008$, $\eta_p^2 = 0.044$, $p = 0.021$, $BF_{10} = 1.404$; however, Bayesian analysis was ambiguous. Follow up post hoc tests found that quiet and changing-state sound conditions differed significantly ($p = 0.021$, $BF_{10} = 2.488$), yet Bayesian evidence remained equivocal. A repeated measures ANOVA of false alarm rate was more clear, finding no significant differences and Bayesian evidence supporting the null hypothesis, $F(2,174) = 2.360$, $MSE = 0.032$, $\eta_p^2 = 0.026$, $p = 0.970$, $BF_{01} = 3.0152$.

Table 2.1.
Hit and False Alarm Rates as a Function of Sound Condition

Sound Condition	Hits		False Alarms	
	M	*SD*	*M*	*SD*
Quiet	0.89	0.10	0.16	0.13
Steady-state sound	0.87	0.11	0.18	0.15
Changing-state sound	0.86	0.13	0.20	0.15

Note. N = 88

As discussed earlier, the use of a 2AFC recognition task allows for measurement of both sensitivity (how well the observer can detect the correct stimulus, called "d prime") and response bias (called "c"). Following the theory laid out by Macmillan and Creelman (pp. 167-168; 2005) and choosing left side as old, both of these measures can be calculated from the hits and false alarms. The distance between these rates is estimated by the sensitivity measure *d'*. A larger *d'* score indicates greater sensitivity, meaning that participants are able to discriminate the signal (old word) from the noise (new word). Means and standard deviations for *d'* and c scores can be found in Table 2.2.

Table 2.2
d' and *c* in Experiment 2 as a Function of Sound Condition

Sound Condition	*d'*		c	
	M	*SD*	*M*	*SD*
Quiet	1.92	0.73	-0.12	0.44
Steady-state sound	1.69	0.74	-0.10	0.39
Changing-state sound	1.65	0.73	-0.12	0.45

Note. $N = 88$

Initial inspection of mean *d'* scores (Table 2.2) suggests that a small ISE may be present. Recognition memory for items in sound conditions, (steady-state, *d'* = 1.69, *SD* = 0.74; changing-state, *d'* = 1.65, *SD* = 0.73) was worse than it was for items in the quiet condition (*d'* = 1.92, *SD* = 0.73). A small numerical difference was also visible between steady-state sound (*d'* = 1.69, *SD* = 0.74) and changing-state sound conditions (*d'* = 1.65, *SD* = 0.73). A repeated measures ANOVA confirmed a main effect of sound condition, $F(2,174) = 6.35$, $MSE = 1.887$, η_p^2 0.068, $p = 0.002$, $BF_{10} = 10.744$. Follow up post hoc testing using the Holm-Bonferonni correction revealed that this difference was between the quiet and steady-state sound condition ($p = 0.01$, $BF_{10} = 4.672$) and the quiet and changing-state sound condition ($p < 0.01$, $BF_{10} = 18.376$). Steady-state and changing-state sound conditions did not differ ($p = 0.59$, $BF_{01} = 7.409$). A repeated measure ANOVA indicated that the decision criterion *c* was unaffected by sound condition, $F(2,174) = 0.064$, $MSE = 0.010$, $\eta_p^2 = 7.359 \times 10^{-4}$, $p = 0.938$, $BF_{01} = 23.918$.

Experiment 2 set out to determine if the ISE occurs in recognition tasks by replicating and extending the work of Stokes and Arnell (2012). Compared to the quiet condition, recognition memory performance was significantly impaired in both the steady-state and changing-state sound conditions. This is sufficient evidence under most models to report an ISE. These results replicate Stokes and Arnell's (2012) findings of statistical equivalency between steady-state and changing-state sound recognition memory performance. This means that there is no changing-state effect and as such, under the O-OER model Experiment 2 does not provide evidence of an ISE.

In addition to recognition memory performance, lexical decision task accuracy and response times were also analysed. For each participant, the median response time for a correct response was determined for each condition. These response times were then analyzed by a 2 lexicality (word vs. nonword) × 3 sound condition (quiet, steady state, and changing state) repeated measures factorial ANOVA. There was a significant main effect of lexicality, $F(1,87) = 21.711$, $MSE = 21848.5$, $\eta_p^2 = 0.200$, $p < 0.001$, $BF_{10} = 2.01 \times 10^9$, with faster responding to words ($M = 704.386$, $SD = 130.808$) than nonwords ($M = 763.831$, $SD = 142.709$), replicating the usual result (e.g., Forster & Chambers, 1973). The effect of sound condition was not significant, $F(1.90,165.24) = 2.592$, $MSE = 7254.29$, $\eta_p^2 = 0.029$, $p = 0.081$, $BF_{01} = 7.692$. The mean response time (and standard deviation) for the quiet, steady state, and changing state sound conditions was 745.74 (140.71), 728.20 (126.56), and 728.39 (128.63), respectively. The interaction was not significant, $F(1.84,159.98) = 0.191$, $MSE = 3169.00$, $\eta_p^2 = 0.002$, $p = 0.809$, $BF_{01} = 24.23$.

A mean accuracy score was also determined for each participant. Similarly, these scores were then analyzed by a 2 lexicality (word vs. nonword) x 3 sound condition (quiet, steady state, and changing state) repeated measures factorial ANOVA. There was a significant main effect of lexicality, $F(1,87) = 38.004$, $MSE = 0.819$, $\eta_p^2 = 0.304$, $p < 0.001$, $BF_{10} = 4.936 \times 10^{19}$, with higher accuracy for nonwords ($M = 0.937$, $SD = 0.086$), than words ($M = 0.857$, $SD = 0.117$). The effect of sound condition was significant by the frequentist test, $F(2,174) = 5.034$, $MSE =$

0.022, η_p^2 = 0.055, p = 0.007, but BF_{10} = 0.192. The mean accuracy score (and standard deviation) for the quiet, steady state, and changing state sound conditions was 0.888 (0.106), 0.910 (0.086), and 0.894 (0.113), respectively. The interaction was not significant, $F(2,174)$= 0.914, MSE = 0.003, η_p^2 = 0.010, p = 0.403, BF_{01} = 16.393.

The lexical decision results partially replicate those of Stokes and Arnell (2012; and Smith, 2010), in that responses were faster to words than nonwords. However, whereas Stokes and Arnell found significantly faster responding in sound conditions than in quiet, the current experiment only trended in that direction, having not reached significance.

Experiment 3

Experiment 3 sought to replicate the findings of Stokes and Arnell (2012) and Experiment 2 under a different cover task. A pleasantness rating task was substituted for the lexical-decision task initially used by Stokes and Arnell (2012). This particular orienting task has been used quite frequently in studies using incidental encoding. Hyde and Jenkins (1973) showed that the pleasantness rating task resulted in more elaborate encoding than a surface task such as counting "E's". Thus, this change will determine whether the results of Stokes and Arnell (2012) and Experiment 2 are not unique to the use of a cover lexical-decision task

Methods

Participants. Thirty Memorial University of Newfoundland students volunteered to participate in exchange for course credit or financial compensation ($10 for 1 hour). All were native speakers of English. The mean age was 20.45 (SD 3.12, range 18-34) and 19 self-identified as female and 11 as male. Fifty-eight volunteers from Prolific Academic participated and were compensated £8 per hour (prorated). Inclusion criteria were identical to Experiment

1B. The mean age was 30.29 (SD = 5.95, range 19-39), and 35 self-identified as female and 21 as male. Logic for sample size followed that of Experiment 2.

Stimuli/Materials. Stimuli and materials were identical to Experiment 2. Given the change in cover task, the non-word pool was not utilized.

Design. Design was very similar to Experiment 2, differing only in the cover task used. While Experiment 2 used a lexical-decision task, the current experiment had participants complete a pleasantness rating task as a cover for the study phase.

Procedure. Instructions and stimuli were presented as in Experiment 2. Of course, participants were given instructions for the word pleasantness task, rather than a lexical-decision task. When completing the word pleasantness tasks participants were instructed to rate the relative word pleasantness of the visual stimuli as quickly and accurately as possible all while ignoring any auditory information emitting through their headphones. They were to indicate their decision with a keypress ("Z" for more positive or "M" for more negative). The test phase of the recognition task followed the procedure outlined in Experiment 2.

Results and Discussion

As in Experiment 2, hit and false alarm rates were calculated and can be found in Table 3.1. Repeated measures ANOVAs conducted on both hits, $F(1.86, 162.06) = 0.687$, $MSE = 0.004$, $\eta_p^2 = 0.008$, $p = 0.495$, $BF_{01} = 13.347$, and false alarms, $F(2,174) = 0.920$, $MSE = 0.005$, $\eta_p^2 = 0.010$, $p = 0.400$, $BF_{01} = 10.859$, showed no significant differences between conditions and supported the null hypothesis.

Table 3.1
Hit and False Alarm Rates in Experiment 3 as a Function of Sound Condition

Sound Condition	*Hits*		False Alarms	
	M	*SD*	*M*	*SD*
Quiet	0.92	0.10	0.09	0.09
Steady-state sound	0.93	0.08	0.10	0.10
Changing-state sound	0.92	0.09	0. 09	0.10

Note. N = 88

Initial inspection of the descriptive statistics found in Table 3.2 do not suggest an ISE. Recognition memory for items in the sound conditions (*d'* = 2.31, *SD* = 0.74) was numerically comparable to the quiet condition (*d'* = 2.36, *SD* = 0.77).

Table 3.2
d' and *c* in Experiment 3 as a Function of Sound Condition

Sound Condition	*d'*		c	
	M	*SD*	*M*	*SD*
Quiet	2.36	0.77	-0.05	0.39
Steady-state sound	2.31	0.72	-0.10	0.39
Changing-state sound	2.32	0.77	-0.04	0.36

Note. N = 88

A repeated measures ANOVA, $F(2, 174) = 0.154$, $MSE = 0.050$, $\eta_p^2 = 0.002$, $p = 0.857$, $BF_{01} = 21.790$, found no significant differences between sound conditions. Bayesian analysis corroborates these findings reporting strong evidence in support of the null hypothesis. The decision criterion *c* was also unaffected by sound condition, $F(2,174) = 0.651$, $MSE = 0.086$, $\eta_p^2 = 0.007$, $p = 0.523$, $BF_{10} = 13.402$.

The substitution of a cover pleasantness rating task for a cover lexical-decision task has eliminated the trend towards an ISE. No differences were found between any of the three sound conditions and therefore there was no evidence of the ISE or changing-state effect. This will be discussed in detail in the general discussion.

Unlike in the lexical-decision task, the pleasantness judgment task does not have an objective correct answer. Even if a word was rated as pleasant by a large sample, a particular individual might find the word negative. Therefore, for each person, the median response time for all responses, whether they agreed with the rating in Warriner et al. (2013) or not, was determined for each condition. These response times were then analyzed by a 2 pleasantness (positive vs. negative) × 3 noise condition (quiet, steady-state, and changing) repeated measures factorial ANOVA. There was a significant main effect of valence, $F(1,87) = 104.320$, $MSE = 7081.59$, $\eta_p^2 = 0.545$, $p < 0.001$, $BF_{10} = 7.22 \times 10^{13}$, with faster responding to positive ($M = 803.82$, $SD = 142.19$) than negative ($M = 878.63$, $SD = 153.25$) words, an expected congruency effect (Simon, 1990). The effect of noise condition was also significant, $F(1.84,160.46) = 4.649$, $MSE = 18310.79$, $\eta_p^2 = 0.051$, $p = 0.013$, $BF_{10} = 15.95$. Post hoc tests indicated that response times in the quiet condition ($M = 865.23$, $SD = 175.25$) were significantly slower than in both the steady-state sound condition ($M = 825.47$, $SD = 144.97$), $t(87) = 2.870$, $p = 0.014$, $d = 0.306$, $BF_{10} = 30.53$, and the changing-state sound condition ($M = 832.98$, $SD = 164.94$), $t(87) = 2.328$, $p = 0.042$, $d = 0.248$, $BF_{10} = 4.66$. However, there was no difference between the steady-state and changing-state sound conditions, $t(87) = 0.542$, $p = 0.589$, $d = 0.058$, $BF_{01} = 8.70$. The interaction was not significant, $F(1.97,181.09) = 2.980$, $MSE = 3178.96$, $\eta_p^2 = 0.033$, $p = 0.054$, $BF_{01} = 10.39$.

Response times were significantly faster in the noise conditions than in the quiet conditions, the same result reported by Stokes and Arnell (2012) for their lexical-decision task.

Numerically, the pattern is also the same as that observed in Experiment 2, although that result was not significant.

Experiment 4

Like Experiment 3, Experiment 4 attempted to replicate the findings of Stokes and Arnell (2012) and Experiment 2 under a different cover task. A word frequency task was substituted for the lexical-decision task used by Stokes and Arnell (2012).

Methods

Participants. Eighty-eight volunteers from Prolific Academic participated and were compensated £8 per hour (prorated). Inclusion criteria were identical to Experiment 1B. The mean age was 28.43 (SD = 5.52, range 19-39), and 53 self-identified as female and 35 as male. Logic for sample size followed that of Experiment 2.

Stimuli/Materials. Stimuli and materials were identical to Experiment 3.

Design. Design was very similar to Experiment 2, rather than a lexical-decision task, participants completed a word frequency task as a cover for the study phase.

Procedure. Instructions and stimuli were presented as in Experiment 2 with the exception that participants were given instructions for a word frequency task rather than a lexical-decision task. When completing the word frequency tasks participants were instructed to rate the relative word frequency of the visual stimuli as quickly and accurately as possible all while ignoring any auditory information emitting through their headphones. They were to indicate their decision with a keypress (“Z” for less frequent or “M” for more frequent). The test phase of the recognition task followed the procedure outlined in Experiment 2.

Results and Discussion

Means and standard deviations of hit and false alarm rate by sound condition can be found in Table 4.1. Repeated measures ANOVAs of hit rate, $F(2,174) = 0.428$, $MSE = 0.005$, $\eta_p^2 = 0.005$, $p = 0.653$, $BF_{01} = 17.074$, and false alarm rate, $F(2,174)$ 0.343, $MSE = 0.004$, $\eta_p^2 = 0.004$, $p = 0.710$, $BF_{01} = 18.586$, found no significant differences between sound conditions.

Table 4.1
Hits and False Alarms in Experiment 4 as a Function of Sound Condition

Sound Condition	Hits		False Alarms	
	M	*SD*	*M*	*SD*
Quiet	0.82	0.16	0.21	0.16
Steady-state sound	0.82	0.17	0.21	0.16
Changing-state sound	0.81	0.17	0.20	0.15

Note. $N = 88$

Means and standard deviations for *d'* scores and c scores for Experiment 4 can be found in table 4.2. Initial inspection raises doubts about further analysis finding significant differences between the three sound conditions. The quiet condition (*d'* = 1.55, *SD* = 0.93) and steady-state sound condition (*d'* = 1.55, *SD* = 0.91) are almost numerically equal, and the changing-state sound condition is similar as well (*d'* = 1.52, *SD* = 0.94).

Table 4.2
d' and *c* in Experiment 4 as a Function of Sound Condition

Sound Condition	*d'*		*c*	
	M	*SD*	*M*	*SD*
Quiet	1.55	0.93	-0.09	0.41
Steady-state sound	1.55	0.91	-0.09	0.43
Changing-state sound	1.52	0.94	-0.05	0.32

Note. $N = 88$

A repeated measures ANOVA confirmed suspicions of no significant difference between sound conditions, $F(2, 174) = 0.127$, $MSE = 0.041$, $\eta_p^2 = 0.001$, $p = 0.880$, $BF_{01} = 22.516$. A Bayesian repeated measures ANOVA provides further support for this assessment. Lastly, the decision criterion c was also unaffected by sound condition, $F(2,174) = 0.328$, $MSE = 0.042$, $\eta_p^2 = 0.004$, $p = 0.721$, $BF_{01} = 18.072$. Like the pleasantness rating cover task, the word frequency cover task did not elicit an ISE in the subsequent surprise recognition test. This will be discussed along with the null results of Experiment 3 in the general discussion.

As with the pleasantness judgments in Experiment 3, for each person the median response time for any response was determined for each condition. These response times were then analyzed by a 2 frequency (high vs. low) × 3 noise condition (quiet, steady-state, and changing) repeated measures factorial ANOVA. There was a significant main effect of frequency, $F(1,87) = 46.483$, $MSE = 10642$, $\eta_p^2 = 0.348$, $p < 0.001$, $BF_{10} = 6.13 \times 10^8$ with faster responding to low ($M = 709.03$, $SD = 159.88$) than high ($M = 770.25$, $SD = 154.53$) frequency words. The effect of noise condition was significant, $F(1.86,161.96) = 4.623$, $MSE = 15510$, $\eta_p^2 = 0.050$, $p = 0.013$, $BF_{10} = 6.15$. As in Experiment 3, response times were slower in the quiet condition, although the only significant difference was between the quiet ($M = 760.57$, $SD = 164.07$) and changing ($M = 722.05$, $SD = 182.32$) conditions, $t(87) = 3.007$, $p = 0.009$, $d = 0.321$, $BF_{10} = 21.45$. There was no difference between response times in the quiet and steady-state sound ($M = 736.31$, $SD = 152.06$) conditions, $t(87) = 1.894$, $p = 0.120$, $d = 0.202$, $BF_{01} = 0.60$, or between the two noise conditions, $t(87) = 1.113$, $p = 0.267$, $d = 0.119$, $BF_{01} = 4.74$. The interaction was not significant, $F(1.92,167.00) = 0.797$, $MSE = 5266.03$, $\eta_p^2 = 0.009$, $p = 0.448$, $BF_{01} = 18.09$.

The numerical ordering of response times as a function of noise condition once again followed the general pattern observed by Stokes and Arnell (2012) and also seen in Experiments

2 and 3. The differences are whether each specific comparison was or was not significant, but numerically, response times were always slower in the quiet condition than in the noise conditions.

General Discussion

The present study aimed to clarify whether the ISE could be observed in a recognition memory task. This served as a test of the O-OER model as unlike a serial recall task, recognition memory tasks do not require order information. Previous work by LeCompte (1994), which attempted the same, was criticized for allowing subvocal rehearsal to seep into the task. Word lists were short, and trials were many, increasing the likelihood of participants to engage in subvocal rehearsal for the shortly anticipated recognition memory test. Stokes and Arnell (2012) designed a methodology to address this criticism, greatly increasing the word list and reducing the trials to one. However, their choice of a yes/no recognition task meant that they were unable to provide unique false alarm rates for sound conditions, leaving hit rate as the only unique component in *d'*. Stokes and Arnell (2012) were unable to elicit the changing-state effect and so despite a reported ISE, they were unable to challenge the O-OER model. The current study incorporated a forced-choice design allowing for an improved *d'* signal detection analysis. This allows for a more thorough analysis which is more capable of picking up a weak changing-state effect if there is one.

Statistically, the results of all four experiments are clear. The sound condition manipulation check (Experiment 1) confirmed that the irrelevant auditory stimuli used could produce both an ISE and a changing-state effect. This is essential information to have before continuing with less robust tasks. Experiment 1 was also able to confirm that participant type (in-person or online) had little impact on serial recall performance. Experiment 2, which used a lexical-decision cover task, showed an ISE but no changing-state effect. Like the results of Stokes and Arnell (2012), recognition memory performance was significantly better in the quiet condition than the sound conditions, yet performance on steady-state and changing-state sound conditions was statistically equivalent. Experiments 3 and 4, employed pleasantness rating and word frequency cover tasks, respectively, and found neither an ISE nor a changing-state effect. Response times also trended in the direction reported by Stokes and Arnell (2012), with faster response times reported in irrelevant sound conditions in all three experiments.

These results are not due to inadequate power: Stokes and Arnell (2012) did not find a changing-state effect with a sample size of 44 participants, so the sample size used in the current studies was increased to 88 participants. This should be more than adequate to find an effect if one were truly present. One might note that the online participants were older and from a wider geographic catchment area than the usual on-campus participants. These differences might allow for wider generalization than is normally the case. However, since we did not collect location, nor did we sample participants based on demographics, we cannot state this with confidence.

As an aside, of the three cover tasks, lexical decision, pleasantness rating, and word frequency rating, only lexical decision resulted in an ISE in recognition. Overall discrimination was numerically better with the two new cover tasks. All of these tasks have been used in the past as cover tasks for incidental encoding (e.g., Hyde & Jenkins, 1973). One might speculate that tasks requiring "deeper" or more elaborate encoding might insulate the participant from the effects of irrelevant background noise. This speculation could be tested in the future using a design that deliberately manipulates cover tasks.

Of the models of the ISE, Working Memory's phonological loop hypothesis struggles the most with these results. Recall that this theory places the mechanism of action of the ISE in the phonological loop. Regardless of task type, TBR words are theorized to become muddled with irrelevant sound in the phonological store, resulting in a decline in serial recall or recognition memory performance (Baddeley, 1986, 1992). Given this, the phonological loop hypothesis predicted a robust ISE would be found in all four experiments as they all presented irrelevant sound and TBR words together. While an ISE was indeed found in both Experiment 1's serial recall task and Experiment 2's lexical-decision surprise 2AFC recognition task, it was not present in Experiment 3 or Experiment 4. In part due to its vague description of a mechanism of action, the phonological loop hypothesis is unable to explain this non-effect in the later experiments.

In strong contrast, the non-effects reported in Experiments 3 and 4 were consistent with the O-OER model. Information in this model is represented as abstract amodal objects which are strung together in an ordered stream by interference-vulnerable pointers that contain encoded order information (Jones et al., 1996). While TBR items and changing-state sound each create their own streams, steady-state sound is instead perceived as a single object and acts as a simple auditory distractor. As such, changing-state sound impairs serial recall performance to a greater extent (referred to as the changing-state effect). This is because theoretically the impairment is a result of the interference caused by stream intersection which can occur when there are multiple streams present during rehearsal. This interference breaks the links in the TBR stream and impairs recall (Jones et al., 1996). It is important to note that without order information there are no pointers to turn single objects into a stream and therefore there can be no stream intersections and no ISE. The limited need for order information in the cover task phases in Experiments 2-4 means that these TBR items would, like steady-state sound, be coded as single objects. As such, the O-OER model only predicted an ISE to emerge in Experiment 1's serial recall task. This prediction was supported. Experiment 1 showed a robust ISE and changing-state effect, and Experiments 2 through 4 did not. Although recognition memory performance in Experiment 2 was significantly better in the quiet condition than the steady-state and changing-state sound conditions, this is not sufficient evidence under the O-OER model. Without the changing-state effect this impairment is considered a simple distraction effect.

Unconcerned with changing-state effects - parsimoniously interpreting it as an example of how complex distractors recruit more attentional processing resources and so further impair serial recall and recognition memory performance – the Attentional Account predicted an ISE in all four experiments. At its core, this account theorizes that irrelevant sound is not filtered out during the Selective Attention process and instead is processed alongside target information, draining limited attentional processing resources (Cowan, 1995). Less resources means the TBR information receives less processing, leading to worse serial recall and recognition memory performance at test. While Experiments 1 and 2 fall in rank, the lack of an ISE in Experiments 3 and 4 are difficult for the Attentional Account to explain. These experiments used word pleasantness and word frequency cover tasks, both tasks that require thinking beyond the word itself. Participants must think about when and how the word is encountered opposed to merely determining lexicality as in a lexical-decision task. Irrelevant sound recruiting limited attentional processing resources from the target words in these more complex cover tasks should have led to greater recognition memory impairment. Instead, these two tasks were unaffected, and only the simpler lexical-decision task (Experiment 2) bore an ISE.

In terms of the Feature model these results provide mixed support. The Feature model places the ISE's mechanism of action within the concept of feature adoption. In this model items enter both primary and secondary memory (Neath, 2000). While the item in primary memory (referred to as a cue) becomes degraded over time due to retroactive interference, the item in secondary memory remains intact. Successful item retrieval requires matching the degraded primary memory cue to the item that best matches it from secondary memory (Neath, 2000).

Features from irrelevant sound can be adopted by the TBR item cues, distorting them and making it less likely that they will be matched with the appropriate item in secondary memory. This process of feature adoption leads the Feature model to predict that an ISE would be found in all experimental conditions with speech-like sound and therefore all four experiments in this series. Experiments 1 and 2 are easily explained, however the Feature model has a harder time explaining Experiments 3 and 4. It is possible that the emphasis placed on the TBR words and their phonemes in the lexical-decision task (Experiment 2) may have encouraged feature adoption in the sound conditions, impairing recognition memory performance. Pleasantness rating and word frequency tasks require participants to think beyond the individual items and would not have placed the same emphasis on the words themselves.

A finding of note in this series is the high recognition memory performance of participants. In all three surprise recognition tasks participants performed far above chance, with hits averaging above 80%. This is true even in Experiment 2 where an ISE was found. This has highlighted the strength of implicit memory and casts doubts on the necessity of serial rehearsal.

In summary, the ISE did not consistently appear in surprise recognition memory tasks. It was only found when using a cover lexical-decision task. Changing the cover task to one requiring participants to think of where and how study words are encountered, i.e., word pleasantness and frequency tasks, eliminated the effect. This inconsistency is difficult to explain for most models of the ISE. The phonological loop hypothesis and the Attentional Account struggle to reconcile these results, both models having predicted an effect regardless of cover task. The Feature model, having made the same prediction, is also challenged; however, the emphasis placed on words in the cover lexical-decision task may have been more likely to encourage feature adoption and therefore an ISE. It is the O-OER model that best reflects the reported findings. Having predicted that none of the recognition experiments would elicit the

ISE, which in the O-OER model requires a changing-state effect, O-OER is the only model left standing.

References

Baddeley, A. D. (1966). Short term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*,362-365. https://doi.org/10.1080/14640746608400047

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

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36*(2), 233 – 252. https://doi.org/10.1080/14640748408402157

Baddeley, A., & Salame, P. (1986). The unattended speech effect: Perception or memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(4), 525. https://doi.org/10.1037/0278-7393.12.4.525

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*,*14*(6), 575 – 589. https://doi.org/10.1016/S0022-5371(75)80045-4

Beaman, C. P., & 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*(2), 459 – 471. https://doi.org/10.1037/0278-7393.23.2.459

Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(3), 615-636. https://doi.org/10.1080/713755774

Bell, R., Röer, J. P., & Buchner, A. (2013). Irrelevant speech disrupts item-context binding. Experimental Psychology. https://doi.org/10.1027/1618-3169/a000212

Bell, R., Röer, J. P., Dentale, S. and Buchner, A. (2012). Habituation of the irrelevant speech effect: Evidence for an attentional theory of short-term memory disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1542 – 1557. https://doi.org/10.1037/a0028459

Bridges, A. M. (1996). Word dose in the disruption of serial recall by irrelevant speech: Phonological confusions or changing state?. *The Quarterly Journal of Experimental Psychology: Section A*, *49*(4), 919 – 939. https://doi.org/10.1080/713755663

Buchner, A., Bell, R., Rothermund, K., & Wentura, D. (2008). Sound source location modulates the irrelevant-sound effect. *Memory & Cognition, 36*(3), 617-628. https://doi.org/10.3758/MC.36.3.617

Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, *32*(5), 722-731. https://doi.org/10.3758/BF03195862

Campbell, R., & Dodd, B. (1984). Aspects of hearing by eye. In H. Bouma and D. G. Bouwhuis (Eds.), *Attention and Performance X* (pp. 299 – 312). Hillsdale, NJ: Erlbaum.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of verbal learning and verbal behavior, 15*(1), 17 – 31. https://doi.org/10.1016/S0022-5371(76)90003-7

Colle, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behavior*, *19*(6), 722-735. https://doi.org/10.1016/S0022-5371(80)90403-X

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology,* 55, 75-84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Crowder, R. G., & Melton, A. W. (1965). The Ranschburg phenomenon: Failures of immediate recall correlated with repetition of elements within a stimulus. *Psychonomic Science, 2*(1-12)*,* 295 – 296. https://doi.org/10.3758/BF03343462

Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in" irrelevant speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1406 – 1414. https://doi.org/10.1037/0096-1523.24.5.1406

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the “irrelevant speech effect”. *The Journal of the Acoustical Society of America, 102*(4), 2191-2199. https://doi.org/10.1121/1.419596

Ellermeier, W., & Zimmer, K. (2014). The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology, 35*(1), 10-16. https://doi.org/10.1250/ast.35.10

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition, 30*(3), 478-487. https://doi.org/10.3758/BF03194948

Farley, L. A., Neath, I., Allbritton, D. W., & Surprenant, A. M. (2007). Irrelevant speech effects and sequence learning. *Memory & Cognition, 35*(1), 156-165. https://doi.org/10.3758/BF03195951

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12*, 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Hanley, J. R. (1997). Does articulatory suppression remove the irrelevant speech effect?. *Memory*, *5*(3), 423 – 431. https://doi.org/10.1080/741941394

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review, 10*(2), 435 – 444. https://doi.org/10.3758/BF03196503

Hanley, J. R., & Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology, 78*(3), 287 – 297. https://doi.org/10.1111/j.2044-8295.1987.tb02247.x

Hellbrück, J., Kuwano, S., & Namba, S. (1996). Irrelevant background speech and human performance: Is there long-term habituation?. *Journal of the Acoustical Society of Japan (E)*, *17*(5), 239-247. https://doi.org/10.1250/ast.17.239

Henson, R. N. A. (1998). Item repetition in short-term memory: Ranschburg repeated. *Journal of Experimental Psychology: Learning. Memory, & Cognition, 24*(5)*,* 1162 – 1181. https://doi.org/10.1037/0278-7393.24.5.1162

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (39 – 56). Hillsdale, NJ: Erlbaum.

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 736. https://doi.org/10.1037/0278-7393.31.4.736

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1050. https://doi.org/10.1037/0278-7393.33.6.1050

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks, *Journal of Verbal Learning and Verbal Behavior, 12*(5), 471 – 480. https://doi.org/10.1016/S0022-5371(73)80027-1

JASP Team (2019). JASP (Version 0.11.1) [Computer software] https://jasp-stats.org/

Jones, D. (1993). Objects, streams, and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (p. 87–104). Clarendon Press/Oxford University Press.

Jones, D. M. (1994). Disruption of memory for lip-read lists by irrelevant speech: Further support for the changing state hypothesis. *The Quarterly Journal of Experimental Psychology*, *47*(1), 143-160. https://doi.org/10.1080/14640749408401147

Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. In Gathercole, S. E. (Ed.), *Models of short-term memory* (pp. 209 – 238). Psychology Press. https://doi.org/10.4324/9780203775868

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*(2), 369 – 381. https://doi.org/10.1037/0278-7393.13.2369

Jones, D. M., & Macken, W. J. (1995). Phonological similarity in the irrelevant speech effect: Within-or between-stream similarity?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 103 – 115. https://doi.org/10.1037/0278-7393.21.1.103

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*(3), 318-328. https://doi.org/10.3758/BF03208264

Jones, D., 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 Section A*, *44*(4), 645 – 669. https://doi.org/10.1080/14640749208401304

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-795. https://doi.org/10.1080/01621459.1995.10476572

Kjellberg, A., & Sköldström, B. (1991). Noise annoyance during the performance of different nonauditory tasks. *Perceptual and motor skills, 73*(1), 39-49. https://doi.org/10.2466/pms.1991.73.1.39

Klatte, M., Kilcher, H., & Hellbrück, J. (1995). Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen [The effects of the temporal structure of background noise on working memory: Theoretical and practical implications]. *Zeitschrift für Experimentelle Psychologie, 42*(4), 517–544.

Kučera, H., & Francis, W. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1396 – 1408. https://doi.org/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, and Cognition*, *22*(5), 1154 – 1165. https://doi.org/10.1037/0278-7393.22.5.1154

LeCompte, D. C., Neely, C. B., & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 472. https://doi.org/10.1037/0278-7393.23.2.472

LeCompte, D. C., & Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *The Quarterly Journal of Experimental Psychology Section A*, *50*(1), 100 – 118. https://doi.org/10.1080/713755679

Longoni, A. M., Richardson, J. T., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition*, *21*(1), 11 – 22. https://doi.org/10.3758/BF03211160

Mackworth, J. F. (1963). The duration of the visual image. *Canadian Journal of Psychology*,17(1), 62 – 81. https://doi.org/10.1037/h0083263

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, second edition*. Mahwah, NJ: Erlbaum.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of verbal Learning and verbal Behavior*, *2*(1), 1 – 21. https://doi.org/10.1016/S0022-5371(63)80063-8

Miles, C., Madden, C., & Jones, D. M. (1989). Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron (1987). *Perception & Psychophysics, 45*(1), 77-81. https://doi.org/10.3758/BF03208036

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, 579 – 584. https://doi.org/10.1037/0278-7393.17.3.578

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*(4, Pt.1), 679–684. https://doi.org/10.1037/h0026641

Murray, D. J., Rowan, A. J., & Smith, K. H. (1988). The effect of articulatory suppression on short-term recognition. *Canadian Journal of Psychology*, *42*(4), 424 – 436. https://doi.org/10.1037/h0084204

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*(4), 343 – 352. https://doi.org/10.3758/BF03197045

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251 – 269. https://doi.org/10.3758/BF03213879

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*(3), 403 – 423. https://doi.org/10.3758/BF03214356

Neath, I., Surprenant, A. M., & LeCompte, D. C. (1998). Irrelevant speech eliminates the word length effect. *Memory & Cognition*, *26*(2), 343 – 354. https://doi.org/doi.org/10.3758/BF03201145

Nittono, H. (1997). Background instrumental music and serial recall. *Perceptual and Motor Skills*, *84*(3_suppl), 1307-1313. https://doi.org/10.2466/pms.1997.84.3c.1307

Peterson, L. R., & Johnson, S. T. (1971). Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 346 – 354. https://doi.org/10.1016/S0022-5371(71)80033-6

Röer, J. P., Bell, R., & Buchner, A. (2014). Evidence for habituation of the irrelevant-sound effect on serial recall. *Memory & Cognition, 42*(4), 609-621. https://doi.org/10.3758/s13421-013-0381-y

Rouleau, N., & Belleville, S. (1996). Irrelevant speech effect in aging: An assessment of inhibitory processes in working memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *51*(6), 356 – 363. https://doi.org/10.1093/geronb/51B.6.P356

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of verbal learning and verbal behavior*, *21*(2), 150 – 164. https://doi.org/10.1016/S0022-5371(82)90521-7

Salamé, P., & Baddeley, A. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, *30*(8), 1185 – 1194. https://doi.org/10.1080/00140138708966007

Salamé, P., & Baddeley, A. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society*, *28*(6), 540 – 542. https://doi.org/10.3758/BF03334073

Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008). Does irrelevant music cause an irrelevant sound effect for auditory items? European Journal of Cognitive Psychology, 20, 252-271. https://doi.org/10.1080/09541440701427838

Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, & Psychophysics, 74*(1), 194-203. https://doi.org/10.3758/s13414-011-0230-7

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus–response compatibility: An integrated perspective* (pp. 31- 86). Amsterdam: Elsevier, North-Holland.

Smith, P. A. (2010). Ageing, auditory distraction, and grammaticality judgement. *Aphasiology, 24*(11), 1342-1353. doi:10.1080/02687030903490533

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Elmsford, NY: Pergamon.

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica,* 27, 285–292. https://doi.org/10.1016/0001-6918(67)90070-4

Stokes, K. A., & Arnell, K. M. (2012) New considerations for the cognitive locus of impairment in the irrelevant-sound effect. *Memory & Cognition, 40*(6), 918 – 931. https://doi.org/10.3758/s13421-012-0194-4

Surprenant, A. M., Neath, I., & LeCompte, D. C. (1999). Irrelevant speech, phonological similarity, and presentation modality. *Memory, 7*(4), 405 – 420. https://doi.org/10.1080/741944920

Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 659. https://doi.org/10.1037/0278-7393.24.3.659

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1750. https://doi.org/10.1037/0278-7393.26.6.1750

Tremblay, S., Macken, W. J., & Jones, D. M. (2000). Elimination of the word length effect by irrelevant sound revisited. *Memory & Cognition, 28*(5), 841-846. https://doi.org/10.3758/BF03198419

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods, 45*(4), 1191-1207. https://doi.org/10.3758/s13428-012-0314-x