

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/165552/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zhang, Qiyuan, Williams, Craig and Morgan, Phillip L. 2024. Partial habituation to disruption by irrelevant emotive speech – evidence for duplex-mechanism account. *Journal of Cognitive Psychology* 36 (1) , pp. 42-60. 10.1080/20445911.2024.2310881

Publishers page: <https://doi.org/10.1080/20445911.2024.2310881>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



**Partial Habituation to Disruption by Irrelevant Emotive Speech –  
Evidence for Duplex-Mechanism Account**

Qiyuan Zhang<sup>a,b,c</sup>, Craig Williams<sup>a,b</sup> and Phillip L. Morgan<sup>a,b,c,d</sup>

*<sup>a</sup>School of Psychology, Cardiff University, Cardiff, UK, CF10 3AT*

*<sup>b</sup>Human Factors Excellence Research Group (HuFEx)*

*<sup>c</sup>Cardiff University Centre for Artificial Intelligence, Robotics and Human-Machine Systems  
(IROHMS)*

*<sup>d</sup>Visiting Professor - Luleå University of Technology - Psychology, Division of Health, Medicine  
& Rehabilitation, Sweden, Regnbagsallen 5, 977 54, Lulea, Sweden*

*Corresponding Author: Phillip L Morgan, School of Psychology, 70 Park Place, Cardiff  
University, UK (morganphil@cardiff.ac.uk)*

## **Acknowledgement**

*We dedicate this article to great colleagues and dear friends, Professor William 'Bill' J Macken (16th May 1965 – 17th Feb 2020) and Professor Dylan M Jones OBE (30th Mar 1948 – 8th Apr 2022).*

**Disclosure of Conflict of Interest**

*The authors report there are no competing interests to declare.*

## **Abstract**

There is accumulating evidence that the degrading effect of irrelevant speech on short-term memory task performance is modulated by its post-categorical properties, including emotional valence, pointing to an attention-driven process as its underlying mechanism, as opposed to pre-attentive, involuntary serializations as its key component (suggested by the interference-by-process hypothesis). Within the current experiments, these opposing theories are investigated by examining habituation to emotive background speech, the extent of which would be taken as evidence for the involvement of attentional capture. Our findings show that attentional capture cannot fully account for the disruptive effect of emotive speech. First, negative and positive background speech were in general more disruptive than neutral speech. Second, repeated exposure to negative speech resulted in habituation and improvement of serial recall performance, but only to the point where the disruptiveness was akin to that of neutral speech. Third, participants displayed no habituation to neutral speech compared to control (no speech) condition. Together, these findings support the duplex-mechanism account of the irrelevant sound effect and reveal two components of the disruption of emotive speech: a valence component driven by attentional shift, and, an acoustic component explicable by interference-by-process. Practical implications for emotionally charged workplace settings are also discussed.

**Keywords:** Auditory distraction, irrelevant sound, emotive background speech, disruption, serial recall, memory

## **Introduction**

Background noises - including speech - that are irrelevant to the focal task, are prevalent in workplaces. Their degrading effect on performance of various cognitive tasks (e.g. involving short-term memory recall, categorisation, mental arithmetic) - referred to as the irrelevant sound effect (ISE) - has been well documented in laboratory settings (Buchner et al., 2004; Elliott, 2002; Hughes et al., 2013; Hughes & Marsh, 2019; Jones et al., 1993; Jones & Macken, 1993; Macken et al., 2009; Marsh et al., 2009, 2018; Salamé & Baddeley, 1982; Vachon et al., 2020). The nature of this degrading effect, as well as its determinant factors, continues to be of great interest to researchers not only because of the practical implications (given pervasive auditory distractions in workplaces – and therefore subsequent human errors and the like) but also due to its theoretical importance – paradigms, findings, models and theories have for decades helped psychologists better understand the inner working of several cognitive processes including attention and short-term memory.

A point receiving much debate is that of the role and extent of involvement of attentional capture as part of the underlying mechanisms of the ISE on performance of serial recall tasks (which concerns memorising and reproducing a sequence of digits or letters and in their presented order). For example, whether the disruptiveness of irrelevant sound can be explained by attentional resources being temporarily directed away from the focal memory task by irrelevant sounds (i.e., the attentional capture account) (e.g., Bell, Röer, et al., 2019; Cowan, 1999; Röer et al., 2014b), or a pre-attentive process in which the processing of the order information of the to-be-remembered items is interfered by irrelevant sounds that gain automatic access to the same representational scene (i.e., the interference-by-process account) (e.g., Jones & Tremblay, 2000; Marsh et al., 2009).

One line of evidence that favours the attentional capture account came from Buchner and colleagues (2004, 2006) who demonstrated that the disruptiveness of irrelevant speech was influenced by its emotional valence. That is, emotive speech (of positive or negative valence) is more disruptive than non-emotive (or neutral) speech. Such findings are potentially applicable to real-life (including highly-pressurised) collaborative working environments such as healthcare settings (e.g. staff working in hospitals), control centres (e.g. emergency services dispatch staff), military contexts (e.g. onboard air-, land- and sea- craft) and so on. In these and other contexts (safety critical and otherwise), speech (e.g. conversations) in the foreground and/or background can be emotionally-charged, whether in terms of aspects such as tone and/or semantic content of the speech. The attention-grabbing property of the emotive speech has also been applied to developing speech-based auditory warning systems (Edworthy & Hellier, 2006). However, Marsh et al. (2018) demonstrated that only the semantic - *valent* - aspect of disruption caused by emotive speech can be explained by attentional capture while the acoustic - *speech* - aspect is better understood in terms of interference-by-process account. The current paper further examines the role of attentional capture in the disruptive effect of emotive speech (i.e., utterances of words that have emotional connotations) on serial recall tasks by investigating the extent to which people might adapt and habituate to them: that is, to determine if the disruptiveness of irrelevant emotive background speech decreases after repeated exposure. In doing so, the paper aims to provide novel insights into the two possible mechanisms underlying ISE in general, with potential implications for how valent irrelevant background speech could be better managed within a number of workplace settings.

### *Theories of the Irrelevant Sound Effect*

Although irrelevant sound has been shown to be disruptive to a variety of cognitive tasks, we focus here on short-term memory serial recall tasks, which involve memorising and reproducing visually or auditorily presented items (Jones et al., 1993; Jones & Morris, 1992; Macken et al., 2009; Marsh et al., 2009; Salamé & Baddeley, 1982). In a classic ISE paradigm, participants undertake a serial recall task where they are usually presented with six to nine items (often digits or letters and sometimes words) – the to-be-remembered (TBR) stimuli – one at a time in succession (with inter-stimulus-intervals) and are to recall these items in the order in which they were presented. Background noises, or to-be-ignored (TBI) stimuli (speech, non-speech, etc.) are played to the participant either during the presentation phase, retention phase, or both. Recall performance – usually measured by the number or percentage of correct responses in order (and sometimes reaction times) – is recorded under these conditions and compared with a silent (no background sound/speech) condition to establish the magnitude of the disruption caused. This effect was first demonstrated by Colle and Welsh (1976) and replicated in numerous studies over an almost 50-year period (Beaman & Jones, 1998a; Bridges & Jones, 1996; Jones & Macken, 1993, 1995; Marsh et al., 2018; Perham & Banbury, 2008; Salamé & Baddeley, 1982). It has also been extended to tasks involving the processing of semantic information – for example, memory for prose and mental arithmetic (Banbury & Berry, 1998; Perham et al., 2016), proof reading (Jones et al., 1990) and text comprehension (Lehmann & Seufert, 2017; Martin et al., 1988).

Despite being a well-documented phenomenon, the underlying mechanism of the ISE is still a matter of much debate (Elliott, 2002; Elliott & Briganti, 2012; Georgi et al., 2022; Hughes & Marsh, 2019; Marsh et al., 2023; Marsh et al., 2018; Perham et al., 2023; Vachon et al., 2020) and a number of theories have been developed over many decades to account for the processes that possibly underlies the disruption caused. These theories can be roughly subsumed under two

broad categories with respect to the assumed role of attentional capture in the mechanism of disruption. One school of thought regards the ISE as a process that does not involve the recruitment of attention from the TBR items. For example, one account stemmed from the classic model of working memory which regards ISE as a consequence of the irrelevant speech gaining automatic access to the phonological loop, which interferes with the TBR items that occupy the same store through deliberate subvocal articulation (Salamé & Baddeley, 1982). One important prediction of this model is that speech should be more disruptive than non-speech (Salamé & Baddeley, 1989).

This view was later challenged by many researchers and alternative theories have since been established. Most prominently, Jones (1993) and Jones and Morris (1992) offered early evidence that non-speech can be as disruptive as speech as long as it has acoustic variations within the auditory signal(s). This led to the proposition of the *changing-state hypothesis*, and the position that it is the degree of acoustic dissimilarity within a stream of sounds, not its similarity to the TBR items, that determines the degree of disruptiveness (Bridges & Jones, 1996; Jones et al., 1993; Jones & Macken, 1995). Compatible with this hypothesis, an alternative theory – the interference-by-process account of the ISE – was proposed (Jones & Tremblay, 2000; Marsh et al., 2009). According to this account, the disruption caused by irrelevant speech is underpinned by the pre-attentive, involuntary streaming and seriation process of the irrelevant auditory stimuli interfering with the deliberate, articulatory-based, seriation of the TBR items. The interference-by-process account was derived from the Object-Oriented Episodic Record (O-OER) memory model (Jones et al., 1996), according to which, a stream of auditory stimuli generated by the continuous sound waves of the distractor, is – involuntarily – segmented into distinctive *objects* based on how dissimilar one auditory *event* is to the previous one. Then links or *pointers* are

assigned between *objects* to indicate their serial order. It was speculated that these *pointers* will help to signify boundaries which serve as distinctive markers for retrieval and the successful recall of the serial order information depends on the integrity of these linkages. It was assumed that the disruption on serial recall tasks results from this involuntary process interfering with a similar process applied to the TBR items through voluntary, deliberate articulate rehearsal. That is, the two streams of objects intertwine and interfere with each other and the *pointers* assigned to the acoustic sequence get confused with those assigned to the TBR items. Thus, auditory distractors that vary acoustically are disruptive because more “objects” are identified from a stream and hence more *pointers* need to be assigned.

Three key predictions are made by the interference-by-process account. First, the disruptive potency of irrelevant sounds is purely determined by acoustical properties (e.g., pitches, tones, etc.) – namely, the level of acoustic variation in the signal(s). Any post-categorical properties of the speech – including its semantic meaning, valence and self-relevance should not impact the magnitude of the resultant disruption to the primary task. Second, disruption caused by the varying sound properties should only be observed in memory tasks that involve seriation (such as serial recall) and not e.g. free recall. Third, and perhaps most important to the purpose of the current paper, this theory predicts no habituation to irrelevant sound: that is, disruptive effects should hold true no matter how much exposure a person has to it.

The other school of thought, in contrast, posits that irrelevant sound disrupts cognitive tasks through the mechanism of attentional capture (e.g. Bell, Röer, et al., 2019; Cowan, 1999; Röer et al., 2014b). That is, serial recall performance is degraded by attention being momentarily diverted away to irrelevant auditory stimuli from focal cognitive tasks (Bell, Mieth, et al., 2019; Bell, Röer, et al., 2019; Buchner et al., 2004; Elliott, 2002; Elliott & Briganti, 2012; Röer et al.,

2014b, 2018). This view descends from the embedded-processes model of working memory (Cowan, 1998, 1999) as well as the *feature model* (Farley et al., 2007; Nairne, 1990; Neath, 2000), both of which have an orientational component in their account of forgetting. For example, according to the embedded-process model, the functioning of working memory involves retaining sets of features in long term memory that are relevant to the current task in an activated state and a subset of these features may enter the spotlight of attention to participate in the cognitive task at hand. The notion of orienting response (Sokolov, 1963) and habituation play important roles in this model, according to which attention automatically orients to a new stimulus when there are discrepancies detected between the stimulus and an pre-established *neural model* based on previous stimuli. The neural model is regularly being updated in light of new stimuli and serves as the basis for orienting response. According to this model, the disruption caused by the variation of the acoustic features of sound is not the result of acoustic streaming of the TBI sound interfering with the articulatory-based seriation of the TBR items (as envisioned by the interference-by-process account) but instead manifestation of attentional resources being recruited *away* to process novel stimuli. Unchanged, repeated stimuli that conform with the prevailing neural model arguably cause little disruption because they do not demand attention due to a process of habituation. Furthermore, a sequence of changing background stimuli could also lead to habituation if the same sequence is experienced repeatedly (Bell et al., 2012; Röer et al., 2011).

Unlike the interference-by-process account, the attentional capture account predicts that the post-categorical meaning of irrelevant auditory stimuli does matter to the level of disruption.

Irrelevant sound that is meaningful (e.g., speech with self-relevant semantic meaning, such as hearing one's own name, or words with emotional connotations, such as those signifying danger

or distress) is likely to attract attentional resources away from the focal memory task and hence be more disruptive to serial recall performance than irrelevant sound that is meaningless or indeed less meaningful (Buchner et al., 2004, 2006; Neath, 2000; Perham et al., 2023; Röer et al., 2013, 2017a, 2017b). In support of this, a series of studies have demonstrated that the disruptiveness of irrelevant speech is influenced by its semantic content. For example, it has been shown that irrelevant speech that contains taboo words, bears more self-relevance (e.g., one's own name), or has a low-frequency of usage, cause more disruption to serial recall tasks than non-taboo, low-self-relevance, high-frequency words (Buchner & Erdfelder, 2005; Röer et al., 2013, 2017a).

In line with this, Buchner et al. (2004) demonstrated in their first experiment that the valence of speech-based distractors can affect their disruptiveness to serial recall tasks. Participants completed tasks that involved trying to remember three-syllable German words in the presence of background speech played to them concurrently (except for the control condition). Speech was manipulated with respect to self-relevance and emotional valence. It was found that in general utterances of adjectives containing negative (e.g., “aggressive”) or positive emotional connotations (e.g., “honest”) caused more disruption to serial recall performance than neutral adjectives (e.g., “complicated”). (Note, these are English translations.) Negative adjectives were also found to be more disruptive than positive ones. However, their second experiment, which only compared neutral against negative adjectives, failed to fully replicate arguably the main finding of the first experiment in terms of effect sizes. It should be noted that in both experiments, they used the same set of word distractors within every trial for each valent condition. Therefore, participants might have habituated very quickly to the distractors after

some exposure. Although the authors did not report results of trial-by-trial comparisons or e.g. between blocks of trials – such as those experienced earlier versus later.

On the surface, the finding that the degrading effect of background speech is modulated by its valence properties (e.g. Buchner et al. 2004) is in support of the attentional capture account of the ISE. According to the late selection theory (Deutsch & Deutsch, 1963), the semantic meaning of all speech-based stimuli are processed before attention allocation, and words with emotive properties would be more disruptive because they are prioritised for attentional resources. There has already been a large body of evidence for attentional bias towards negative, threat-related stimuli among people with anxiety and other emotional disorders (Eysenck et al., 2007; Williams et al., 1988). Perham and Rosser (2012) as well as Perham and Oaksford (2005) also demonstrated the impacts of mood induction (e.g., through exposure to negatively-valenced words or threatening stimuli) on cognitive reasoning. Similarly, words that vary in valence and arousal have been found to affect cognitive control (e.g., perceptual and conceptual inhibition) (Imbir et al., 2021; Larsen et al., 2006).

Despite this, the level of involvement of attentional capture in the disruptiveness of emotive speech is still a topic of debate. Could attentional capture fully explain the disruption from irrelevant emotive speech to serial recall tasks? Marsh et al. (2018) deconstructed the total effect of emotive speech into the effects occurring due to its emotional *valence* and *acoustical* properties and argued that while the valence element disrupts serial recall via the mechanism of attentional capture, the effect of the acoustic element is governed by the changing-state principle and is explicable by the interference-by-process account of the ISE in which attentional capture does not play a role. They provided evidence for this dissociation by demonstrating that the valence effect of emotive speech can be reduced by increasing participant workload whereas the

effect of the acoustic element was unaffected by this manipulation. Rettie et al. (2023) recently observed similar effects in regard to taboo words (Experiment 2): providing forewarnings to participants about the nature of the upcoming auditory distractor only mitigated the disruptiveness of taboo words on serial recall (to a level akin to neutral, non-taboo words) while the effect of neutral words remained despite this manipulation.

These findings are consistent with what is called the *duplex-mechanism account* of auditory distraction (Hughes, 2014; Hughes et al., 2013) which posits that each of the interference-by-process account and the attentional capture account explains some type of auditory distraction but not all. Regarding the disruption caused by irrelevant sound with varying acoustic property, the duplex-mechanism account distinguishes the classic changing-state sound effect (e.g., ABCABC) from the “deviation” effect, caused by the presence of stimuli that differ significantly from those that precede it (e.g., AAABAA). According to this view, the two phenomena are governed by different processes – while the changing-state sound effect can be explained by the interference-by-process account, the deviation effect can be explained by attentional capture. One important piece of evidence for this dichotomy is that only tasks that require the memory of the serial order of the visually presented items are disrupted by sounds of changing state whereas the tasks that do not (e.g., the missing item task) are only disrupted by deviants (Beaman & Jones, 1998; Hughes et al., 2007). (Also see Hughes et al., 2013; Hughes & Marsh, 2019; Littlefair et al., 2022). Also, Hughes et al. (2005) provided evidence that the deviation effect in serial recall tasks is the result of the deviant stimulus violating the expectation that is formed by the pattern or rules established by the preceding stimuli (i.e., an algorithm-based neural model), instead of mismatching with the stimulus that comes immediately before it (i.e., an aggregation-based neural model). It should be noted that only on an aggregate-based model of the deviation

effect could the changing-state effect plausibly be explained in terms of attentional capture. Therefore the changing-state effect cannot be ascribed to attentional capture on the algorithm-based account because no stimulus in a changing-state sequence violates an established pattern of stimuli..

In line with the duplex-mechanism account and the Marsh et al. (2018) paper, we predict the disruptive effects of irrelevant emotive speech consist of at least two parts. One being the “physical” component – derived from the acoustic aspects of the speech including the phonology of the word(s) being spoken as well as the tone, the pitch and the intensity of the voice. The other being the “valence” component derived from the semantic meaning of the word(s). For example, a word such as laughter is more positively meaningful than a word such as table (neutral in valence), whereas a word such as torture is more negatively meaningful (again in terms of valence) than both previous examples. Distinguishing these two aspects of speech is not uncommon in the field of memory and language studies (e.g., Baddeley, 1966; Leiber, 1977). Unlike Marsh et al. (2018), we examine the two elements of the mechanism that govern the irrelevant emotive speech by observing if and how people habituate to it. According to the duplex-mechanism account of the ISE, partial habituation should be observed – that is, only the semantic, valence component of the emotive speech which is assumed to cause disruption via the attentional capture mechanism, will be habituated away, whereas the physical, acoustic component, which is assumed to cause disruption through the interference-by-process route, will be immune from habituation.

### *Habituation to Irrelevant Emotive Speech*

The theories reviewed above have different predictions regarding habituation to irrelevant sound. While the interference-by-process account of the ISE predicts no habituation, the attentional

capture account predicts complete habituation. The duplex-mechanism account, on the other hand, would predict that habituation depends on the extent to which attention-capture is involved in disruption caused. This divergence creates an opportunity for these theories to be pitted and tested, which is a key aim of the experiments presented in the current paper.

Evidence regarding the habituation to irrelevant sound is mixed. While some studies have demonstrated habituation (Banbury & Berry, 1997; Bell et al., 2012; Elliott & Cowan, 2001; Morris & Jones, 1990; Röer et al., 2014a; Vachon et al., 2012; Waters et al., 1977), others have not (Ellermeier & Zimmer, 1997; Hellbrück et al., 1996; Jones et al., 1997; Röer et al., 2011). Notably, Jones et al. (1997) did not find evidence for habituation even when the trials with the changing-state sounds (i.e., human utterances of a sequence of different words) were presented to participants in blocks instead of interlaced with quiet and steady-state sound trials (i.e., a sequence involving repetition of the same word). However, these modes of presentation could have weakened possible habituation effects. Perham and Banbury (2008) adopted a different paradigm in which participants underwent a phase of exposure to the auditory stimuli before taking on the conventional serial recall task (with different or the same auditory stimuli to the ones used in the exposure phase) and also did not observe signs of habituation. Röer et al. (2014a) suggested several possibilities why some previous studies failed to detect habituation: for example - that the auditory distractors used were too simple (e.g., digits and words) and too repetitive, which may have caused any habituation process to occur too soon to be detected before the recall performance no longer improved. Another possibility is that simple auditory stimuli, compared to more complex auditory stimuli (e.g., natural speech, music), might be less likely to contain deviant acoustic features (e.g., unexpected change in tones or pitches of voice)

or, more relevant to the focus of the current research, less likely to bear strong emotional connotations and relevancy to the participant that can capture attention.

These explanations are in fact in line with the duplex-mechanism account of ISE which would predict a pattern of partial habituation. It should be noted that even in those studies which have demonstrated habituation, total habituation (i.e. elimination of disruptive effects compared to control) was rarely achieved. This pattern indicates that there is an element of the irrelevant sound that can never be fully habituated against. Hence, there is a possibility that there are different components of the disruptive effect of irrelevant speech, which are underpinned by different mechanisms: one component explainable by attentional capture which can be habituated to and another explainable by interference-by-process which is immune to habituation. When exposed to irrelevant speech, people may exclusively habituate to one element of the background sound but not the other.

Within the current experiments, and to formally investigate these predictions regarding partial habituation, valent speech – specifically, series of meaningful (positive, negative, and, control) words were used as TBI auditory stimuli. There are at least two benefits of using valent speech to study partial habituation. First, and as discussed, the disruption of valent speech can be conceptually divided into two components (Marsh et al., 2018) – a semantic, “valence” component which can only be explained by attentional capture and a non-semantic, “acoustic” component which, according to the duplex-mechanism account, can only be attributed to interference-by-process, if there is no deviant element. Second, the disruptive consequence of the two components can be observed by measuring the difference in performance between the valent conditions and the neutral condition (“valence” component) as well as that between the neutral condition and the silent condition (“acoustic” component). This makes deriving specific, testable

predictions possible. Therefore valent speech provides optimal conditions to investigate partial habituation. Rettie et al. (2023) included valent speech in one experiment (Experiment 1) as a comparison point against taboo speech. A cross-trial analysis only revealed a pattern of habituation to the disruption of taboo speech, and not to the valent speech. However, as the authors pointed out, their experiment was primarily designed to investigate the moderating effect of cognitive load on the disruption of taboo and valent speech with habituation examined in a *post-hoc* analysis. Thus the additional variances introduced to the data by their manipulation of cognitive load might have reduced the power of their experiment to detect habituation effect. The two experiments presented here, in comparison, were specifically designed to examine habituation to valent speech.

### *Current Experiments*

The current paper presents two experiments examining the disruptiveness of emotive background speech on serial recall performance and whether and to what extent habituation occurs, and when. In both experiments, we adopted a classic serial recall paradigm with irrelevant speech. Participants viewed eight single digits sequentially and their key task was to try and recall them in the order in which they were presented. The presentation of the digits were accompanied by TBI auditory stimuli which consisted of a series of eight words which varied in terms emotional valence (neutral, positive, negative). These words were selected from the Affective Norms of English Words (ANEW) database (Bradley & Lang, 1999). Experiment 1 used the same set of TBI valent words for the same valence conditions whereas Experiment 2 used different words across all condition trials (e.g. 12 negative valence trials, each containing 12 different negative words) to investigate whether the habituation being observed was to the “valence” of the speech

in general or to the negative or positive meanings of specific words. There were also control trials where no sound was played.

In line with the duplex-mechanism account of auditory distraction, a partial habituation effect was predicted. That is, participants should habituate differently to different aspects of emotive speech. It is predicted that the physical, or acoustic aspects of emotive speech that are likely to be experienced in many environments, are more akin to that of a classic changing-state sounds rather than a monotonic stream of homogeneous sounds that contain deviants. This physical component of emotive background speech will disrupt the performance of a cognitive tasks via an *interference-by-process* pathway, and hence cannot be habituated to. The "valence" component, on the other hand, should be subject to habituation due to the fact that novelty of the emotionally charged semantic meaning should wane and become less attention-demanding after repeated exposure. More specifically, the following pattern of results should be observed: 1) Over repeated trials, recall performance in neutral irrelevant speech trials should not improve in relation to the silent (control, no speech) trials; and 2) recall performance in the negative and the positive trials should improve in relation to the neutral trials but only to the point where the degree of disruption is comparable to the neutral irrelevant speech trials. This is because when the "valence" component of the emotive speech is habituated away, the "changing-state", physical component of the emotive speech should be identical to the neutral speech in respect to its disruptiveness.

## **Experiment 1**

### **Method**

#### *Participants*

Sixty-five Psychology undergraduates at XXX (anonymised) University took part in exchange for course credits but seven were excluded from the final analyses due to not following the instructions for the memory recall task. All of participants were females with the mean age of 19.69 ( $SD = .79$ , Max = 21, Min = 18). Participants needed to have English as their first language or be fluent in English as a second language to take part and needed to have normal (or corrected) vision and hearing.

### *Materials*

The experiment was run in a computer laboratory in the School of Psychology at XXX (anonymised) University, which had 12 cubicles each containing a desktop computer with a 24-inch monitor and a set of sound attenuating headphones. The cubicles were separated by tall sound-attenuating layered panels to minimise possible interference between participants.

### Serial Recall Task

Participants completed multiple trials of a serial recall task (Jones et al., 1993; Jones & Macken, 1993) whilst wearing headphones. The trials were programmed and delivered by *PsychoPy*©. In each trial, participants were presented with eight to-be-remembered (TBR) numerical digits (1-8, in random orders) in succession on the computer screen in random order each time. The digits were in an arial font, white in colour, 20 cm in height and were presented with a dark grey background. Each digit appeared at the centre of the screen for 800 milliseconds with a 450-millisecond inter-stimulus-interval (ISI).

After all eight digits had been shown and following another 450-millisecond interval, the word “recall” would appear on the screen which served as a prompt to start to write down the digits (in response booklets) in their presented serial order. The answer booklet contained multiple rows

each corresponding to a trial. On each row there were eight empty spaces denoted by “\_ \_ \_ \_ \_ \_ \_ \_”. For any digits that could not be remembered, participants were told to either guess or leave a blank space at the corresponding position. They also were instructed to refrain themselves from writing anything on the answer sheet or anywhere else until they saw the word “recall” in each trial. The recall period lasted for 15 seconds, at the end of which an instruction appeared onscreen to cue to ‘press the spacebar to start the next trial’. They were instructed to stop writing and press the space bar as soon as they saw this message. They were also advised that the next trial would start immediately afterwards so it was important that they were looking at the screen when and after pressing the spacebar.

#### To-be-ignored (TBI) Auditory Stimuli

Human speech was played through sound attenuating headphones in all non-control trials, which consisted of eight two-syllable words presented at the same time and for the same duration (800-milliseconds) as the eight TBR digits. These words varied in their valence: in one condition, words were positive, in another negative, and neutral in a third condition. All words were taken from the Affective Norms for English words (ANEW) database (Bradley & Lang, 1999), which contains 1,040 words measured on a 1-9 Likert scale over the dimensions of valence, arousal and dominance. We chose words for our positive speech condition which had the highest valence ratings (i.e. those closet to 9) of all two syllable words within the ANEW ( $M = 7.65$ ) and this criteria was also applied to the negative words albeit with these having the lowest valence ratings (i.e. those closet to 0) ( $M = 2.45$ ). Neutral words were those closest to the midpoint (5) of the ANEW valence scale ( $M = 5.02$ ). An ANOVA was conducted to ensure there were significant differences in valence between the three conditions; neutral and positive, negative and positive and negative and neutral. All were significantly different ( $ps < .05$ ). In addition, mean arousal

ratings were calculated for positive ( $M = 5.54$ ), negative ( $M = 5.83$ ) and neutral ( $M = 4.28$ ) words – noting that by nature, neutral words are likely to be less arousing than positive and negative words. In terms of arousal, an ANOVA determined that positive and negative words did not statistically differ, although both positive and negative words used had, on average higher arousal ratings than neutral words ( $ps < .05$ ). There were in total 12 different versions (chains) of neutral, positive and negative words for each non-silent condition but each participant only experienced one of these versions per valence condition across every trial. That is, for each participant, the same eight neutral words were used in all neutral condition trails and this was also the case for the negative (same eight negative words) and positive (same eight positive words) condition. This was a similar approach to that adopted by Buchner et al. (2004) but in their experiments words did not vary between participants – i.e. the same set of words was used across all participants.

All words were spoken via a pre-recorded UK accent neutral (no obvious regional accent) female voice (800ms p/w) - i.e. the narrator. One word was recorded at a time in a sound-proof chamber. Despite their semantic valence, the narrator was told to utter all words in a calm, neutral tone (and was not told about valence differences) and the final set were agreed (as neutral) and chosen from a number of recording attempts by three raters. The onsets of the utterance of each word and its duration were matched to that of the digit presentation, which means that they all lasted for 800 milliseconds. This was achieved by digitally “stretching” or “squashing” some words slightly using the software *Audacity*© with checks on how the words sounded – again with three raters agreeing that none of the words sounded distorted or unusual. The speech was delivered via 44.1kHz, 64-bit, stereo headphones (75 dB(A)). Sound intensity was determined by using an average base starting level at 60dB. During practice trials, all participants confirmed that the

sound was at a comfortable intensity. It was kept at a consistent level for each participant throughout all the phases. Participants were specifically told to ignore the speech and were guaranteed that they were not going to be tested on them at any point during the experiment.

### *Design*

A 4 (Irrelevant Speech: neutral, positive, negative & silent) X 3 (Block: early, middle & late) repeated-measures design was adopted. There were 48 trials in total, with 12 trials for each speech (and one non-speech, silent) condition. It should be noted that previous studies demonstrating habituation to irrelevant sound typically featured eight trials for each sound condition (Röer et al., 2014a). Buchner et al. (2004) also used eight trials for each sound condition although they did not report any results on comparisons between trials or between blocks of trials (e.g., early trials versus late trials). We increased the number of trials to 12 in our experiments to enhance the likelihood of observing habituation and to potentially determine if it asymptotes. The irrelevant speech conditions were intermixed and the order of trials were counterbalanced. The 48 trials were divided into three blocks – early, middle, late, with 16 trials in each block (i.e. four instances of each condition within each block). Serial recall accuracy – i.e. the average number of TBR digits recalled correctly – was the dependent variable.

### *Procedure*

The experiment was run in groups with nine to twelve participants in each session. On arrival, participants were seated at assigned cubicles where they would find an information sheet and consent form placed in front of the computer. The information sheet stated that the main purpose of the experiment was to “gain a better understanding of factors that influence our ability to remember information over short periods of time and effects that irrelevant background distractor stimuli have on this.” They were not specifically instructed about the emotive nature of

the speech and had no written information about this until reading the debrief form at the end of the experiment. Participants were asked to read the information sheet carefully and sign the consent form if they wished to take part. Then they were handed the written instructions to the serial recall task, which explained the nature of the task and emphasized the following points: 1) During some trials, they would hear background speech (i.e., sequences of words) presented at the same time as the TBR numbers. These words were not important to the experiment and they did not have to try and remember them and should instead focus on the main task of trying to remember the TBR visually presented numbers; 2) they should only start writing their answers down when the word “recall” appeared on the screen, nor before; 3) If there were numbers that they could not recall, they should either make a guess or leave a blank space at the corresponding position on the answer booklet; and 4) The recall period lasts for 15 seconds and they should stop writing as soon as they saw the message “press the spacebar to start the next trial.” These key points were reiterated vocally by the experimenter after they have read the instructions. Participants were also given the chance to ask questions before proceeding to a set of practice trials.

Participants experienced four practice trials (one representing each condition) before moving on to the three blocks of main trials. After each block, participants were given an opportunity to take a break, which was prompted by a message on the computer screen. They could pause for as long as they needed before initiating the next block.

After participants completed all 48 trials, they were handed a post-experiment questionnaire which recorded their basic demographic information including age and gender. They were also asked to report emotional, anxiety or behavioural disorders, or, significant emotional life event(s) experienced over the past 12-months. None were reported. At the end of the experiment,

participants were provided with a debriefing sheet, which contained information about the rationale, methodology, predictions and potential implications of the research and were thanked for their contribution to the experiment.

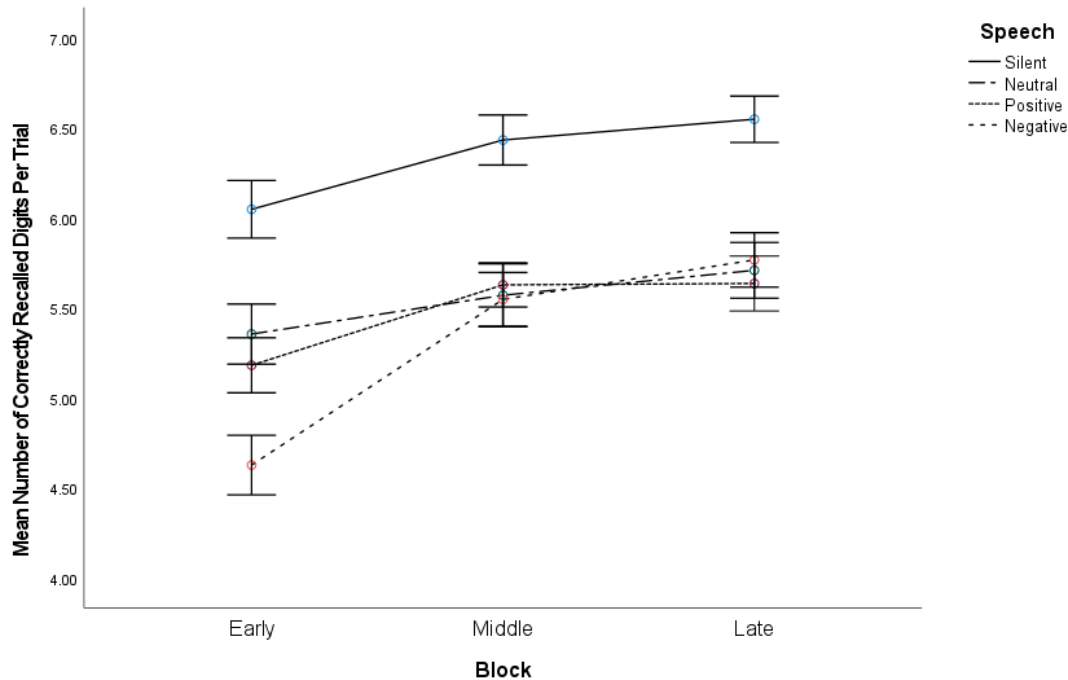
## **Results and Discussion**

Participants' answers in the serial recall task were scored in terms of the number of digits correctly reproduced and the mean scores were first screened for the violation of the assumption of normal distribution. Kolmogorov-Smirnov test was conducted to the mean scores from all four speech conditions and the control condition, none of which had a distribution that was significantly deviant from normality ( $p = >.20$ ). Multivariate repeated measures analyses of variance (ANOVAs) were conducted starting with a 4 (Irrelevant Speech: none, neutral, positive, negative) X 3 (Block: early, middle, late) analysis. As Figure 1 suggests, differences were observed between speech conditions ( $F(3, 192) = 40.397, p < .001, f = .795$ ). Post-hoc comparisons (with Bonferroni adjustments) reveal that in general, recall accuracy in control trials was significantly higher compared to all three non-silent conditions ( $p < .001$ ). However, an overall valence effect was not observed: recall performance in the neutral condition was not significantly different from the positive ( $p > .999$ ) or the negative condition ( $p = .188$ ). The difference between the positive and the negative condition was also not significant ( $p = .278$ ). Nonetheless, on closer inspection, it was found that the differences between the four speech conditions were not consistent across blocks, supported by a significant Irrelevant Speech x Block interaction ( $F(6, 384) = 3.667, p = .001, f = 0.239$ ). In the early block, negative speech was more disruptive than neutral speech ( $p < .001$ ) and positive speech ( $p = .009$ ), providing evidence for the speech valence effect. But it *did not* significantly differ from these conditions in later blocks: with this being a strong indication of a habituation effect.

There was a general trend for the performance to improve across trials (Figure 2), regardless of condition, supported by a significant main effect of Block ( $F(2, 128) = 24.646, p < .001, f = 0.621$ ). This improvement of performance could only be partly attributed to habituation. Note that even performance in silent/control condition had a trend of improving over time, although post-hoc comparisons only showed a significant difference between the early and middle block ( $p < .039$ ), whereas the difference between the middle and late block was not significant ( $p > .999$ ). This might be the result of learning through familiarisation of tasks (e.g., the application of memory strategies). To distil the effect of habituation, we would need to analyse 1) how performance in the neutral condition changes over time in relation to the control/silent conditions; 2) how performance in the negative and positive conditions changes in relation to the neutral and control/silent condition.

### **Figure 1**

*Mean recall performance across four valence conditions over three blocks (Error bars = +/- 1 SE)*



As stated earlier, there was a significant Irrelevant Speech x Block interaction, indicating that the trajectory of performance improvement was not the same among the four Irrelevant Speech conditions. Confirming the hypotheses, the results demonstrate that participants only habituate to the “valence” elements of emotive speech but not the “acoustic” elements. First, after repeated trials, the recall performance in neutral speech trials did not improve in respect to the silent, control trials, confirmed by the lack of Irrelevant Speech-Block interaction when only these two conditions were included in the analysis ( $F(2, 128) = .346, p = .708, f = 0.070$ ). In fact, performance in the neutral condition did not improve at all, as indicated by non-significant differences among the three blocks as determined by post-hoc comparisons ( $p = .773; p > .999; p = .198$ ; respectively). Second, performance in the negative trials improved in respect to silent trails. Although recall performance in both conditions significantly improved from the early to middle block ( $p < .001$  for negative;  $p = .039$  for silent) and did not change from the middle to late block ( $p = .301$  for negative;  $p > .999$  for silent), a significant Irrelevant Speech-Block

interaction involving these two conditions ( $F(2, 128) = 6.963, p = .001, f = 0.330$ ) indicates that the rate of improvement was different in the two conditions, providing evidence for habituation in the negative speech condition. Furthermore, performance in the negative speech condition also improved in relation to the neutral condition, supported by a significant Irrelevant Speech-Block interaction when the analysis only involved these two conditions ( $F(2, 128) = 10.255, p < .001, f = 0.400$ ). However, this improvement seems to have only takes place between early block and middle block, and not after - As Figure 1 illustrates, the performance in negative speech trials was significantly worse than neutral ( $p < .001$ ) in the early block but the difference was reduced and non-significant in the middle block ( $p > .999$ ) and stayed the same in the late block ( $p > .999$ ). Thus, the habituation to negative speech did take place but only to the point when performance was equal to neutral trials, again indicating that only the “valence” element of the emotive speech can be habituated to, not the “acoustic” element.

The results from the positive valence condition displayed a similar pattern - post-hoc comparisons revealed that performance within this condition significantly improved from the early to middle block ( $p = .015$ ) but not from the middle to late block ( $p > .999$ ). However, the Irrelevant Speech x Block interaction involving only the neutral and positive speech conditions was not significant ( $F(2, 128) = .540, p = .584, f = 0.090$ ). Also, there was no significant difference between the positive and neutral speech conditions across each of the three blocks ( $ps > .999$ ). Hence we only found indicative, but not definitive evidence of habituation to positive speech within the current experiment.

In Experiment 1, initial evidence has been found to suggest that participants exclusively habituate to the semantic “valence” element of the emotive speech but not the physical, “acoustic” element. However, evidence for this was only observed in the negative speech

condition It should be noted that although 12 different sets of words were used for every non-silent speech condition, they were distributed across different participants and each participant only experienced one set of words for every non-silent speech condition. That is, each participant experienced the same set of neutral, positive and negative words during each irrelevant speech trial, although subsets of participants experienced different words (but again the same words repeated across their neutral, positive and negative irrelevant speech conditions). Therefore, it is not clear whether it was the specific negative meaning of a particular word that participants habituated to or the negative “valence” of the speech in general. In Experiment 2, we explored this later possibility by using different words for irrelevant speech in every trial.

## **Experiment 2**

### **Method**

#### *Participants*

Seventy-eight Psychology undergraduate students within the School of Psychology at XXX (anonymised) University took part in the Experiment in exchange for course credits. There were 60 females and 17 males and one did not report their gender. The mean age was 19.78 ( $SD = .87$ , Max = 24, Min = 18). Eligibility criteria were the same as Experiment 1.

#### *Materials, Design and Procedure*

Every aspect of Experiment 2 was exactly the same as Experiment 1 except that participants experience different words across every background speech trial. There were 48 trials in total and hence 12 sets of words for each non-silent condition.

### **Results and Discussion**

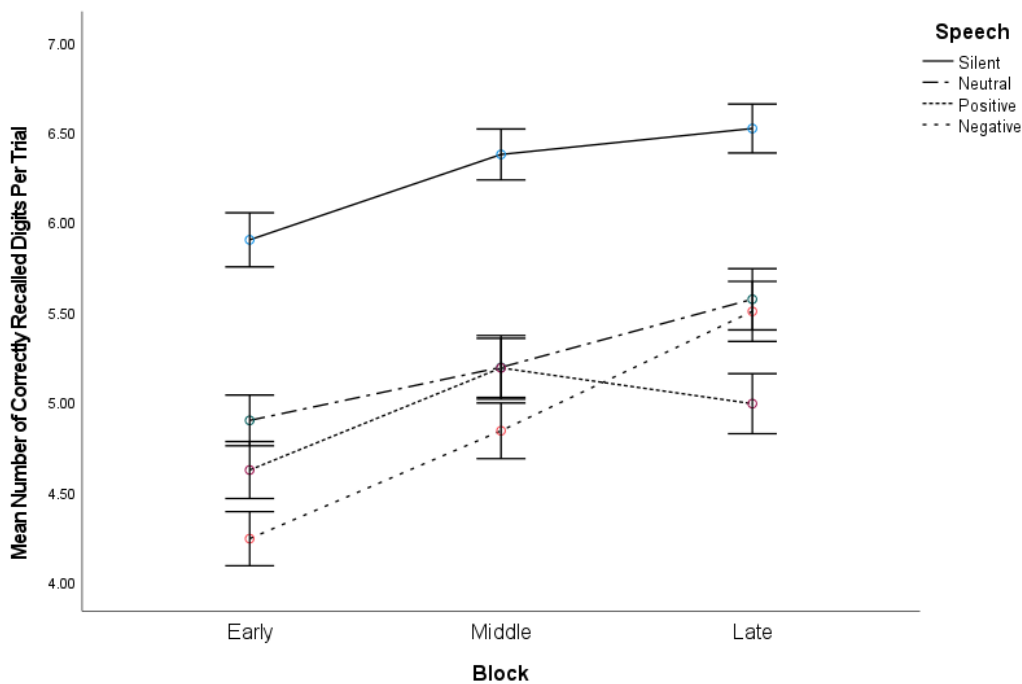
As in Experiment 1, participants' answers in the serial recall task were scored in terms of the number of digits correctly reproduced and the mean scores were first screened for the violation of the assumption of normal distribution. Kolmogorov-Smirnov test was conducted to the mean scores from all four speech conditions and the control condition, none of which had a distribution that was significantly deviant from normality ( $ps > .20$ ). A 4 (Irrelevant Speech: none, neutral, positive, negative) X 3 (Block: early, middle, late) repeated measures ANOVA was conducted. There was a significant main effect of Irrelevant Speech ( $F(3, 231) = 122.181, p < .001, f = 1.259$ ). Post-hoc comparisons (with Bonferroni adjustment) were conducted to explore differences between speech conditions. Recall performance within the no speech control condition was significantly higher than in the three non-silent conditions ( $ps < .001$ ) – see also Figure 2. Among the three non-silent conditions, a valence effect was observed: recall performance in neutral trials was significantly better than the positive ( $p < .001$ ) and the negative condition ( $p < .001$ ). Hence unlike Experiment 1, we found an overall valence effect regardless of blocks. While we did not find an overall significant difference in recall performance between the negative and positive speech condition ( $p > .999$ ) in Experiment 2, there was some evidence that negative speech was more disruptive than positive trials in the early block, although this difference was marginally non-significant ( $p = .054$ ). The said difference was not significant in the middle block ( $p = .109$ ) and reversed in the late block ( $p = .002$ ).

There was also a general trend for recall performance to improve over time (Figure 2), regardless of speech condition, supported by a significant main effect of Block ( $F(2, 154) = 32.958, p < .001, f = 0.655$ ). Moreover, a significant Irrelevant Speech-Block interaction ( $F(6, 462) = 4.567, p < .001, f = 0.244$ ) indicates that the performance differences between speech conditions were not consistent across blocks and performance in all speech conditions did not improve at

the same pace in relation to the silent condition – a sign of habituation. Like in Experiment 1, further analyses confirms that this habituation was partial.

**Figure 2**

*Mean recall performance across four valence conditions over three blocks (Error bars = +/- 1 SE)*



More specifically, as in Experiment 1, recall performance in neutral speech trials did not improve in respect to the silent/control trial, confirmed by the lack of Irrelevant Speech-Block interaction when only these two conditions were included in the analysis ( $F(2, 154) = .727, p = .485, f = .095$ ). Between-block post-hoc comparisons did reveal a significant difference between the early and late block ( $p < .001$ ) in the neutral speech condition, indicating a general trend of performance improvement over time in absolute terms. However, as argued previously, this

cannot be taken as evidence for habituation because the extent of this improvement did not reach beyond that of the silent condition and therefore can only be attributed to general learning or familiarisation effect. There was however, an indication of habituation emerging from the performance increase between the two last blocks – while there was a significant improvement from the middle to late condition in the neutral condition ( $p = .050$ ), the difference was not significant in the silent condition ( $p = .983$ ). This will be discussed in more detail in a later section.

In contrast, clear evidence of habituation has been found with negative speech. Recall performance in the negative trials improved in respect to the neutral trials, supported by a significant Irrelevant Speech-Block interaction when the analysis only involved these two conditions ( $F(2, 154) = 3.746, p = .026, f = .220$ ). In the early block, performance on negative trials was significantly worse than in neutral trials ( $p < .001$ ) but it improved by the middle block ( $p = .019$ ) and became non-significantly different, but not better, than neutral trials in late block ( $p = .650$ ). Recall performance in the negative trials also improved in respect to the silent condition, confirmed by a significant Irrelevant Speech-Block interaction when the analysis only involved these two conditions ( $F(2, 154) = 5.838, p = .004, f = .274$ ). Between-block post-hoc comparisons revealed that the recall performance in the negative speech condition significantly improved from the early to middle block ( $p < .001$ ) and from the middle to late block ( $p < .001$ ), whereas performance in the silent condition only significantly improved from the early to middle block ( $p = .005$ ) but not from the middle to late block ( $p = .983$ ).

Evidence of habituation is less definitive for positive speech. First, recall performance in the positive condition did not significantly vary in respect to silent condition ( $F(2, 154) = 1.353, p = .261, f = .132$ ). Second, recall performance in positive trials did vary in comparison to neutral

trials ( $F(2, 154) = 3.793, p = .025, f = .222$ ). However, this interaction was not characterised by a uniformed trend of improvement – rather, the performance in the positive speech condition seems to have improved relative to the neutral condition from the early ( $p = .224$ ) to the middle block ( $p = .983$ ) (although the interaction was not significant ( $F(1, 77) = 1.677, p = .199, f = .146$ ), but then was worsened in relation to the neutral trials in the late block ( $p < .001$ ) ( $F(1, 77) = 7.375, p = .008, f = .309$ ). However, it should be noted that the absolute performance in the positive speech condition did not significantly change from the middle to late block ( $p = .662$ ), whereas the performance in the neutral condition was found to increase significantly during the same phase ( $p = .050$ ). Hence, the discrepancy between the positive and neutral speech condition in the last condition seems to have been primarily contributed by the performance in the neutral condition getting better rather than the performance in the positive condition getting worse. As mentioned earlier, there was an indication that some degree of habituation has taken place in the neutral speech condition. It is possible that there was an element of the neutral speech disruption that could have been attributed to attentional capture (beside the changing-state effect) and hence could be habituated to. Unlike Experiment 1, participants in Experiment 2 experienced a new set of irrelevant background words for every trial. Therefore a novelty effect could have emerged from the encounter of each speech stimulus (including neutral speech) which captured attention. It can be speculated that this novelty element could be habituated to, but more slowly compared to situations where words were repeated across every trial.

There were at least a few indicative signs to support this speculation: first, performance in all speech conditions was generally worse in Experiment 2 compared to Experiment 1, especially in early trials. Second, the rate of habitation in the negative condition was slower in Experiment 2 compared to Experiment 1, possibly because the novelty element takes more time to habituate to

(Experiment 2). Also, novelty varied from word to word as well. It should be noted that in Experiment 2 we did not control for novelty or frequency of usage between speech conditions, although we did control for arousal between positive and negative condition. It could be that in the last few trials, the positive words used happened to be more novel than those used in the other two speech conditions, which deferred the improvement of performance. Nonetheless, these speculations regarding a novelty effect need to be considered more - within future studies – and in the more general context that the high level ANOVAs did not render any evidence that the performance in either the neutral and positive condition has differed significantly from the silent condition. Therefore from a statistical point of view, the peculiar patterns observed in elements of these two speech conditions could very well be a result of sampling errors. A stronger evidence base is needed for habituation to positive speech and future research should focus on this.

Overall in Experiment 2 we observed additional evidence for partial habituation – participants did not habituate to neutral speech but habituated to negative speech, even when they experienced a different set of words for every trial. This indicated that habituation is transferable to other words of the same valence, although the evidence for this was stronger for negative speech than positive speech.

### **General Discussion**

Two experiments involving serial recall memory tasks were conducted with two main aims. First, to examine whether and to what extent the valence of irrelevant speech is disruptive to recall performance. Second, to determine whether people can habituate to valent irrelevant

speech such that the disruptive effects above and beyond that of a neutral speech condition are alleviated and possibly even mitigated. As predicted, evidence was found for both aspects of the experiments. First, recall performance of visually presented TBR stimuli (digits) was generally worse in the presence of irrelevant speech with marked disruption in a condition with negative speech compared to neutral speech - across both experiments. The difference was particularly pronounced in early trials, before participants had a chance to habituate. The disruptive effect of positive speech was somewhat less clear. Evidence for its valence effect was only observed in Experiment 2. It should be noted that this could not have been caused by the fact that habituation took place more quickly in Experiment 1 because trial-by-trial analyses (not reported here) would reveal that the performance in the positive speech condition was on par with the neutral condition in as early as the first trial.

Second, we found novel and largely consistent evidence for partial habituation to valent irrelevant speech. First, recall performance in neutral speech trials did not improve in relation to silent trials – That is, there were no clear evidence of habituation across experimental blocks whereas clear evidence of habituation was observed in the negative speech condition in both experiments. This is in contrast with the findings of Rettie et al. (2023), from which no evidence of habituation to valent speech was observed. As discussed in the introduction, the discrepancy can be explained by the current studies involving adoption of a more sensitive design to detect habituation as well as the fact that the valent stimuli used in Rettie et al. (2023) were a mixture of positive and negative words. The inclusion of positive words might have diluted the overall valence effect also impacting a possible habituation effect. Second, although improvement was observed in participants' recall performance in negative speech trials compared to neutral trials – after repeated exposure, performance only improved up to the point which it was on par with that

in the neutral trials, and after which it started to at best mimic performance within neutral trials. This pattern of results has two indications. First, the difference in disruptiveness between the neutral and the negative speech, which only differ in valence, can be fully explained by attentional capture and hence can be completely habituated away. Second, once the valence element has been habituated to, negative speech and neutral speech is identical in terms of their disruptiveness because their acoustic, “changing-state” component is the same. The habituation to positive speech was somewhat less clear. In Experiment 1 no evidence of habituation to positive speech was observed – partly due to the fact positive speech was not found to be more disruptive than neutral speech to begin with. In Experiment 2 where different sets of words were used for each trial, there were indications that performance in positive irrelevant speech trials improved compared to neutral trials from the early block to the middle block, but the difference was not proven to reach statistical significance. Also, as noted in the results section of Experiment 2, the recall performance in the positive speech condition seems to have worsened in the late block in respect to the neutral condition. As previously discussed it could simply be the results of sampling errors but it could also have been caused by some confounding factors (e.g., novelty) that we didn’t control for in the word stimuli we used in the emotive speech conditions. Confounding factors will be further discussed in a later section but it’s worth pointing out now that future research is definitely needed to establish a strong evidence base for habituation to positive speech.

Taken together, these findings indicate that only the semantic, “valence” element of the negative speech can be habituated to, while the “acoustical” element cannot, even after repeated exposure. The fact that we observed habituation regardless of whether the same (Experiment 1) or different (Experiment 2) distractor words were used across multiple trials of the same conditions (i.e.

positive, negative and neutral) indicates that the habituation could occur to the valent emotional connotations of the background speech words in general and not to the negative or positive meaning of a specific word.

Our findings of partial habituation, combined with that of Marsh et al. (2018), provides strong evidence that the disruptive effect of emotive speech consists of two components that are governed by two distinctive mechanisms. This is not supportive of either the *attentional capture* account (Bell, Röer, et al., 2019; Elliott & Briganti, 2012; Röer et al., 2014b, 2018), which would predict total habituation, or the *interference-by-process* account (Jones & Tremblay, 2000; Marsh et al., 2009), which would predict the absence of valence effect as well as habituation. Instead, it shows that the disruption of irrelevant emotive speech results from a combination of both, which is in line with the *duplex-mechanism* account of the irrelevant speech effect (Hughes, 2014; Hughes et al., 2013; Hughes & Marsh, 2020), which posits that the both mechanisms underpin the disruptive effects of irrelevant speech. Previously evidence for this account has, to our knowledge, mostly only been gathered from situations concerning auditory distractor stimuli that are free of ‘semantic’ meaning or indeed neutral in terms of valence (Beaman & Jones, 1998; Hughes et al., 2007, 2013; Littlefair et al., 2022), with the exception of Hughes and Marsh (2020), Marsh et al. (2018) and Rettie (2023). Our novel findings add to the mounting evidence that this account can also be extended to explain the effect of emotive irrelevant speech – as operationalized based on valence (mainly), which seems to involve two separate mechanisms – an acoustic, “changing-state” element that is explicable by *the interference-by-process* view, and, a semantic, “valence” element that is explicable by the *attentional-capture* account.

Past research has generated mixed evidence for habituation to irrelevant speech. While some have demonstrated habituation after repeated exposure (Banbury & Berry, 1997; Bell et al., 2012; Elliott & Cowan, 2001; Morris & Jones, 1990; Röer et al., 2014a; Waters et al., 1977), the others didn't (Ellermeier & Zimmer, 1997; Hellbrück et al., 1996; Jones et al., 1997; Röer et al., 2011). Our finding of partial habituation within the current experiments provides a possible reconciliation for this confliction. Röer et al. (2014a) made the observation that studies which demonstrated habituation mostly used complex disrupters like utterances of sentences or music (Banbury & Berry, 1997; Bell et al., 2012; Morris & Jones, 1990). Our finding of partial habituation, provides an explanation for this dichotomy. It could be the case that complex auditory disrupters have a higher likelihood of containing “surprising” elements (e.g., deviant acoustic features) or possess more salient emotional connotations and relevancy to the participant that can result in a change in participants' attentional orientation and after prolonged exposure, participants exclusively habituated to this portion of the effects but not the “changing state” portion. Further, it should be noted that even in those studies that have successfully demonstrated it, complete habituation was rarely observed, indicating that there are some parts of the disruption that cannot be ‘habituated away’. We argue that partial habituation, in line with the duplex-mechanism account of the because it better fits the pattern of the empirical data.

Of course, habituation needs time to occur and another explanation for the partial nature we observed is that perhaps not enough trials (and therefore exposure to valent background speech) were included for ‘complete habituation’ (stabilised across multiple trials or indeed blocks) to be observed. It should be noted that in Röer et al. (2014a), all four experiments featured only 24 trials in total (8 trials for each type of distractor) while in our experiments, 48 trials were used (12 trials for each type of distractor). Even though we employed twice as many trials, there is a

chance that the degree of exposure was not optimal but sufficient nonetheless. However, we argue that the probability is small because our results did not show a clear trend of performance in negative or positive trials to surpass that of neutral trials. Instead, performance in those trials stopped improving in relation to the neutral trials once it reached the same level. Also, habituation to the “valence” element of the emotive speech took place mostly in very early trials but performance in neutral trials did not show any trend of improving in relation to silent trials. Nonetheless, to obtain definitive evidence, future studies should increase the level of exposure – perhaps through inclusion of more trials. However, it must be stressed that the degree of valence associated with the irrelevant speech would be weakened. That is – we employed eight different distractor words within each trial and a total of 96 different words from ANEW per irrelevant speech condition. Despite the inclusion of so many words (288 in total) the valence across each condition was still significantly different (i.e., positive versus negative versus neutral), as noted in the Method. Increasing the number of trials and therefore having to increase the number of words (e.g. doubling to 96 trials would require 576 words) would likely dilute the valence differences between the stimuli used within the experiment, when, normed words are used from sources such as ANEW where there are 1034 words in total. In addition, trying to control for potentially confounding factors such as emotional arousal would become increasingly difficult and potentially impossible.

Our findings have practical as well as theoretical implications. Emotive speech is prevalent in and across many workplace settings (Beaman, 2005; Jones, 1999; Sörqvist & Marsh, 2015) and the finding that people can partly habituate to it is potentially positive news. However, it must be stressed that our findings have been captured using a specific type of primary task – involving serial recall memory and with specific background distractor stimuli, all manipulated and tested

under controlled laboratory conditions. Many real life workplace settings are far more complex with extra degrees of freedom. Further research needs to be conducted within such settings, ideally controlling for the potential effects of extraneous variables but also under more naturalistic conditions. It could well be the case that habituation to emotive background speech is not as straightforward within more naturalistic settings, and this needs to be explored and unpicked to begin to realise the possible positive impact of our novel findings and in order to develop interventions to further alleviate the disruptive effects of (including valent) irrelevant speech. Our findings also seem to suggest that habituation (at least to negative speech) is transferable from one context to another: it was not found to be word-specific – instead, the valence of speech was found to be habituated to in a general manner. But it is unclear how long the protection manifested by exposure to emotive irrelevant speech lasts for. It could be the case that it is ‘session specific’ in that continued exposure on a trial by trial basis in experiments like ours has the powerful attenuation effects found. But, these effects may not hold (or indeed be less powerful) when exposure is more ad hoc (e.g. when exposed to emotive distractor stimuli X times per minute, hour, hours, days etc.) and in line with this – when exposure occurs on an even more infrequent basis (e.g. in bursts but over the space of weeks or months, etc.). Longitudinal studies are also needed to start to determine the ‘transfer effects’ (and indeed if they occur at all) of the habituation observed within the current experiments. Future research should also focus on the factors of habituation speed and its longevity so training can be developed to help mitigate the negative impact of emotive speech in the workplace, and especially in safety-critical contexts.

There are some other limitations that whilst should be considered within future research in this area do not detract from the importance of our findings. The auditory stimuli used in our

experiments were sampled from the Affective Norms for English words (ANEW) database (Bradley & Lang, 1999). While it has been a well cited and tested source for valence manipulation of textual and speech stimuli, the database was created based on data collected ~one quarter of a century ago. Moreover, to avoid exhausting the word pool, we did not control for aspects of the words other than valence (there was however some control over arousal), such as frequency of usage, dominance, concreteness, etc. Although these factors were unlikely to constitute confounding effects for the main findings in our experiments, due to the large quantity of words we were using and random allocation to conditions (to remove possible effects of idiosyncrasies), they might have contributed to the “anomaly” with the positive speech in Experiment 2. Controlling for these dimensions could be important in future research, depending on research questions, aims, samples, and possible application areas.

Participants within both experiments consisted of psychology undergraduate students. While this is common practice in the field of academic-based psychological research, the unbalanced gender distribution as well as the narrow age range (younger adults) featured in the samples need to be noted before research findings can be generalised. Females have been observed to react more strongly to emotive distracting stimuli than males in physiological studies (Garcia-Garcia et al., 2008); therefore the size of valence effect observed in our experiments might be higher than if a more balanced sample was used with an equal ratio of female and male participants.

Research also shows that the capacity of cognitive control, such as functions of inhibition, reduces with age (Lustig et al., 2007) and older adults are more affected by the ISE than younger adults (Getzmann et al., 2013). There is also evidence that older adults displays a slower tendency to habituate to novel auditory stimuli (Richardson et al., 2011). Hence, older adults

might be more susceptible to the disruption of emotive speech and take longer to habituate – and this should also be considered in the future.

## References

- Baddeley, A. D. (1966). The Influence of Acoustic and Semantic Similarity on Long-term Memory for Word Sequences. *Quarterly Journal of Experimental Psychology*, 18(4), 302–309. <https://doi.org/10.1080/14640746608400047>
- Banbury, S. P., & Berry, D. C. (1997). Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied*, 3, 181–195.
- Banbury, S. P., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology*, 89(3), 499–517. <https://doi.org/https://doi.org/10.1111/j.2044-8295.1998.tb02699.x>
- Beaman, C. P. (2005). Auditory distraction from low-intensity noise: a review of the consequences for learning and workplace environments. *Applied Cognitive Psychology*, 19(8), 1041–1064. <https://doi.org/https://doi.org/10.1002/acp.1134>
- 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., Mieth, L., Röer, J., Troche, S., & Buchner, A. (2019). Preregistered Replication of the Auditory Deviant Effect: A Robust Benchmark Finding. *Journal of Cognition*, 2. <https://doi.org/10.5334/joc.64>
- Bell, R., Röer, J., Dentale, S., & Buchner, A. (2012). Habituation of the Irrelevant Sound 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>
- Bell, R., Röer, J., Lang, A.-G., & Buchner, A. (2019). Distraction by Steady-State Sounds: Evidence for a Graded Attentional Model of Auditory Distraction. *Journal of Experimental Psychology: Human Perception and Performance*, 45. <https://doi.org/10.1037/xhp0000623>
- Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*.
- Bridges, A. M., & Jones, D. 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., & Erdfelder, E. (2005). Word frequency of irrelevant speech distractors affects serial recall. *Memory & Cognition*, *33*(1), 86–97. <https://doi.org/10.3758/BF03195299>
- Buchner, A., Mehl, B., Rothermund, K., & Wentura, D. (2006). Artificially induced valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, *34*(5), 1055–1062. <https://doi.org/10.3758/BF03193252>
- 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>
- Colle, H., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*, 17–31. [https://doi.org/10.1016/S0022-5371\(76\)90003-7](https://doi.org/10.1016/S0022-5371(76)90003-7)
- Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.
- Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 62–101). Cambridge University Press. <https://doi.org/DOI:10.1017/CBO9781139174909.006>
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80.
- Edworthy, J., & Hellier, E. (2006). Complex Nonverbal Auditory Signals and Speech Warnings. *Handbook of Warnings*.
- 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>
- Elliott, E. (2002). The irrelevant-speech effect and children: theoretical implications of developmental change. *Memory & Cognition*, *30*(3), 478–487. <https://doi.org/10.3758/bf03194948>
- Elliott, E., & Briganti, A. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, *140*(1), 64–74. <https://doi.org/https://doi.org/10.1016/j.actpsy.2012.02.009>
- Elliott, E., & Cowan, N. (2001). Habituation to Auditory Distractors in a Cross-Modal, Color - Word Interference Task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*, 654–667. <https://doi.org/10.1037/0278-7393.27.3.654>
- Eysenck, M., Derakshan, N., Santos, R., & Calvo, M. (2007). Anxiety and Cognitive Performance: Attentional Control Theory. *Emotion (Washington, D.C.)*, *7*, 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Farley, L. A., Neath, I., Allbritton, D. W., & Surprenant, A. M. (2007). Irrelevant speech effects and sequence learning. *Mem Cognit*, *35*(1), 156–165.

- Garcia-Garcia, M., Domínguez-Borràs, J., SanMiguel, I., & Escera, C. (2008). Electrophysiological and behavioral evidence of gender differences in the modulation of distraction by the emotional context. *Biological Psychology*, *79*(3), 307–316. <https://doi.org/https://doi.org/10.1016/j.biopsycho.2008.07.006>
- Georgi, M., Leist, L., Klatt, M., & Schlittmeier, S. (2022). Investigating the Disturbance Impact of Background Speech on Verbal and Visual-Spatial Short-Term Memory: On the Differential Contributions of Changing-State and Phonology to the Irrelevant Sound Effect. *Auditory Perception & Cognition*, 1–20. <https://doi.org/10.1080/25742442.2022.2127988>
- Getzmann, S., Gajewski, P. D., & Falkenstein, M. (2013). Does age increase auditory distraction? Electrophysiological correlates of high and low performance in seniors. *Neurobiology of Aging*, *34*(8), 1952–1962. <https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2013.02.014>
- Hellbrück, J., Kuwano, S., & Namba, S. (1996). Irrelevant background speech and human performance: Is there long-term habituation? *The Journal of The Acoustical Society of Japan (e)*, *17*, 239–247.
- Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*, 30–41. <https://doi.org/10.1002/pchj.44>
- Hughes, R. W., Hurlstone, M., Marsh, J., Vachon, F., & Jones, D. M. (2013). Cognitive Control of Auditory Distraction: Impact of Task Difficulty, Foreknowledge, and Working Memory Capacity Supports Duplex-Mechanism Account. *Journal of Experimental Psychology Human Perception & Performance*, *39*, 539–553. <https://doi.org/10.1037/a0029064>
- Hughes, R. W., & Marsh, J. (2020). When is forewarned forearmed? Predicting auditory distraction in short-term memory. *Journal of Experimental Psychology Learning Memory and Cognition*. <https://doi.org/10.1037/xlm0000736>
- 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*, 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Imbir, K., Pastwa, M., Duda-Goławska, J., Sobieszek, A., Jankowska, M., Modzelewska, A., Wielgopolan, A., & Żygierewicz, J. (2021). Electrophysiological correlates of interference control in the modified emotional Stroop task with emotional stimuli differing in valence, arousal, and subjective significance. *PLOS ONE*, *16*, e0258177. <https://doi.org/10.1371/journal.pone.0258177>
- Jones, D. M. (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* (pp. 87–104). Clarendon Press/Oxford University Press.

- Jones, D. M. (1999). The cognitive psychology of auditory distraction: The 1997 BPS Broadbent Lecture. *British Journal of Psychology*, *90*(2), 167–187.  
<https://doi.org/https://doi.org/10.1348/000712699161314>
- Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 209–238). Hove: Psychology Press.
- 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.
- 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.
- Jones, D. M., Macken, W. J., & Mosdell, N. A. (1997). The role of habituation in the disruption of recall performance by irrelevant sound. *British Journal of Psychology*, *88*(4), 549–564.  
<https://doi.org/https://doi.org/10.1111/j.2044-8295.1997.tb02657.x>
- 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. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*(2), 89–108.  
<https://doi.org/https://doi.org/10.1002/acp.2350040203>
- Jones, D. M., & Morris, N. (1992). Irrelevant speech and serial recall: implications for theories of attention and working memory. *Scand J Psychol*, *33*(3), 212–229.
- Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, *7*(3), 550–558.  
<https://doi.org/10.3758/bf03214370>
- Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, *6* 1, 62–72.  
<https://api.semanticscholar.org/CorpusID:16114093>
- Lehmann, J. A. M., & Seufert, T. (2017). The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity. *Frontiers in Psychology*, *8*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01902>
- Leiber, L. (1977). Visual, acoustic, and semantic processing of word pairs. *Neuropsychologia*, *15*(2), 217–229. [https://doi.org/https://doi.org/10.1016/0028-3932\(77\)90030-6](https://doi.org/https://doi.org/10.1016/0028-3932(77)90030-6)
- Littlefair, Z., Vachon, F., Ball, L. J., Robinson, N., & Marsh, J. E. (2022). Acoustic, and Categorical, Deviation Effects are Produced by Different Mechanisms: Evidence from

- Additivity and Habituation. *Auditory Perception & Cognition*, 1–24.  
<https://doi.org/10.1080/25742442.2022.2063609>
- Lustig, C., Hasher, L., & Zacks, R. (2007). Inhibitory Deficit Theory: Recent Developments in a "New View. *The Place of Inhibition in Cognition*, 17.
- Macken, W. J., Phelps, F. G., & Jones, D. M. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review*, 16(1), 139–144. <https://doi.org/10.3758/PBR.16.1.139>
- Marsh, J. E., Hughes, R. W., & Jones, D. M. (2009). Interference by process, not content, determines semantic auditory distraction. *Cognition*, 110(1), 23–38.  
<https://doi.org/10.1016/J.COGNITION.2008.08.003>
- Marsh, J. E., Vachon, F., Sörqvist, P., Marsja, E., Röer, J., Richardson, B., & Ljungberg, J. (2023). Irrelevant Changing-state Vibrotactile Stimuli Disrupt Verbal Serial Recall: Implications for Theories of Interference in Short-term Memory. *Journal of Cognitive Psychology*.
- Marsh, J. E., Yang, J., Qualter, P., Richardson, C., Perham, N., Vachon, F., & Hughes, R. W. (2018). Postcategorical auditory distraction in short-term memory: Insights from increased task load and task type. *J Exp Psychol Learn Mem Cogn*, 44(6), 882–897.  
<https://doi.org/10.1037/xlm0000492>
- Martin, R., Wogalter, M., & Forlano, J. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, 27, 382–398.  
[https://doi.org/10.1016/0749-596X\(88\)90063-0](https://doi.org/10.1016/0749-596X(88)90063-0)
- Morris, N., & Jones, D. M. (1990). Habituation to irrelevant speech: Effects on a visual short-term memory task. *Perception & Psychophysics*, 47(3), 291–297.  
<https://doi.org/10.3758/BF03205003>
- 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>
- Perham, N., & Banbury, S. P. (2008). You Cannot Ignore It: Attention to ‘Irrelevant’ Sound during a Pre-exposure Period Does Not Reduce its Disruption of Serial Recall. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19), 1622–1626.  
<https://doi.org/10.1177/154193120805201967>
- Perham, N., Begum, F., & Marsh, J. E. (2023). The Categorical Deviation Effect May Be Underpinned by Attentional Capture: Preliminary Evidence from the Incidental Recognition of Distracters. *Auditory Perception & Cognition*, 1–32.  
<https://doi.org/10.1080/25742442.2023.2167448>

- Perham, N., Marsh, J. E., Clarkson, M., Lawrence, R., & Sörqvist, P. (2016). Distraction of Mental Arithmetic by Background Speech. *Experimental Psychology*, 63(3), 141–149. <https://doi.org/10.1027/1618-3169/a000314>
- Perham, N., & Oaksford, M. (2005). Deontic Reasoning With Emotional Content: Evolutionary Psychology or Decision Theory? *Cognitive Science*, 29, 681–718. [https://doi.org/10.1207/s15516709cog0000\\_35](https://doi.org/10.1207/s15516709cog0000_35)
- Perham, N., & Rosser. (2012). “Not Thinking” Helps Reasoning. *Current Psychology*, 31, 160–167. <https://doi.org/10.1007/s12144-012-9140-7>
- Rettie, L., Potter, R., Brewer, G., Degno, F., Vachon, F., Hughes, R. W., & Marsh, J. (2023). Warning - Taboo Words Ahead! Avoiding Attentional Capture by Spoken Taboo Distractors. *Journal of Cognitive Psychology*. <https://doi.org/10.1080/20445911.2023.2285860>
- Richardson, C., Bucks, R. S., & Hogan, A. M. (2011). Effects of aging on habituation to novelty: An ERP study. *International Journal of Psychophysiology*, 79(2), 97–105. <https://doi.org/https://doi.org/10.1016/j.ijpsycho.2010.09.007>
- Röer, J. P., Bell, R., & Buchner, A. (2013). Self-relevance increases the irrelevant sound effect: Attentional disruption by one’s own name. *Journal of Cognitive Psychology*, 25(8), 925–931. <https://doi.org/10.1080/20445911.2013.828063>
- Röer, J. P., Bell, R., & Buchner, A. (2014a). 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>
- Röer, J. P., Bell, R., & Buchner, A. (2014b). What Determines Auditory Distraction? On the Roles of Local Auditory Changes and Expectation Violations. *PLoS ONE*, 9(1), e84166. <https://doi.org/10.1371/journal.pone.0084166>
- Röer, J. P., Bell, R., Dentale, S., & Buchner, A. (2011). The role of habituation and attentional orienting in the disruption of short-term memory performance. *Memory & Cognition*, 39(5), 839–850. <https://doi.org/10.3758/s13421-010-0070-z>
- Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017a). Attentional Capture by Taboo Words: A Functional View of Auditory Distraction. *Emotion (Washington, D.C.)*, 17. <https://doi.org/10.1037/emo0000274>
- Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017b). Semantic priming by irrelevant speech. *Psychon Bull Rev*, 24(4), 1205–1210. <https://doi.org/10.3758/s13423-016-1186-3>
- Röer, J. P., Rummel, J., Bell, R., & Buchner, A. (2018). Metacognition in Auditory Distraction: How Expectations About Distractibility Influence the Irrelevant Sound Effect. *Journal of Cognition*, 1. <https://doi.org/10.5334/joc.3>

- Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21(2), 150–164. [https://doi.org/https://doi.org/10.1016/S0022-5371\(82\)90521-7](https://doi.org/https://doi.org/10.1016/S0022-5371(82)90521-7)
- Salamé, P., & Baddeley, A. D. (1989). Effects of Background Music on Phonological Short-Term Memory. *The Quarterly Journal of Experimental Psychology Section A*, 41(1), 107–122. <https://doi.org/10.1080/14640748908402355>
- Sokolov, E. N. (1963). *Perception and the Conditioned Reflex*. Macmillan.
- Sörqvist, P., & Marsh, J. (2015). How Concentration Shields Against Distraction. *Current Directions in Psychological Science*, 24. <https://doi.org/10.1177/0963721415577356>
- Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 164.
- Vachon, F., Marsh, J., & Labonté, K. (2020). The Automaticity of Semantic Processing Revisited: Auditory Distraction by a Categorical Deviation. *Journal of Experimental Psychology General*, 149, 1360–1397. <https://doi.org/10.1037/xge0000714>
- Waters, W. F., McDonald, D. G., & Koresko, R. L. (1977). Habituation of the Orienting Response: A Gating Mechanism Subserving Selective Attention. *Psychophysiology*, 14(3), 228–236. <https://doi.org/https://doi.org/10.1111/j.1469-8986.1977.tb01166.x>
- Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1988). Cognitive psychology and emotional disorders. In *Cognitive psychology and emotional disorders*. John Wiley & Sons.

### **Data Availability Statement**

The data that support the findings of this study are openly available in OPENICPSR at <https://www.openicpsr.org/openicpsr/project/190489/version/V1/view>

Reference number: openicpsr-190489