

AUDITORY INTERFERENCE IN HUMAN COGNITION
THEORETICAL FOUNDATIONS AND NEUROCOGNITIVE MECHANISMS
UNDERLYING THE IRRELEVANT SOUND EFFECT
IN CHILDREN AND ADULTS

Vom Fachbereich Sozialwissenschaften
der Rheinland-Pfälzischen Technischen Universität, Campus Kaiserslautern
zur Verleihung des akademischen Grades
Doktor rerum naturalium (Dr. rer.nat.)
genehmigte

D i s s e r t a t i o n

vorgelegt von
Larissa Leist

Tag der Disputation:	Kaiserslautern, 18.12.2024
Dekan:	Prof. Dr. Michael Fröhlich
Vorsitzende/r:	Prof. Dr. Marcus Höreth
Gutachter/in:	1. apl. Prof. Dr. Maria Klatte 2. apl. Prof. Dr. Daniela Czernochowski
	DE 386

März 2025

Remarks

Out of the six studies mentioned in this thesis, four have already been published. The information contained in the publications has been integrated into this thesis. The content matches refer to specific sections of the summary, definition, and theories explaining the ISE, along with the general discussion. The studies are published as follows:

- Study 1 is published as “Irrelevant speech impairs serial recall of verbal but not spatial items in children and adults” by Larissa Leist, Thomas Lachmann, Sabine J. Schlittmeier, Markus Georgi, and Maria Klatte (2023) in *Memory & Cognition*.
- Study 2 is published as “Impact of Irrelevant Speech and Non-speech Sounds on Serial Recall of Verbal and Spatial Items in Children and Adults” by Larissa Leist, Thomas Lachmann, and Maria Klatte (2024) in *Scientific Reports*.
- Study 3 is published as “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” by Markus Georgi, Larissa Leist, Maria Klatte, and Sabine J. Schlittmeier (2022) in *Auditory Perception & Cognition*.
- Study 4 is published as “Differential Effects of Task-Irrelevant Monaural and Binaural Classroom Scenarios on Children’s and Adults’ Speech Perception, Listening Comprehension, and Visual–Verbal Short-Term Memory” by Larissa Leist, Carolin Breuer, Manuj Yadav, Stephan Fremery, Janina Fels, Alexander Raake, Thomas Lachmann, Sabine J. Schlittmeier, and Maria Klatte (2022) in *International Journal of Environmental Research and Public Health*.
- Study 5 and Study 6 are in preparation for publication.

Abstract

The Irrelevant Sound Effect (ISE) describes the phenomenon that serial recall of visually presented verbal items, such as digits, letters, or nouns, is impaired by the presence of specific task-irrelevant background sounds. This effect highlights the auditory system's inherent tendency to automatically process environmental stimuli, regardless of their relevance to the current task. Understanding the ISE is crucial for both theoretical and practical reasons, as most adults and children work or study in noisy environments. Theoretical explanations of the ISE vary. Some theories suggest that background sounds divert attention away from the to-be-remembered items, while others propose that the effect arises from competing seriation processes between the background sounds and the serial order of the memory material. Additionally, the ISE might impact the short-term memory system by disrupting the phonological representations of verbal items in the phonological store.

Research on the ISE has revealed inconsistent findings between children and adults, indicating that the effect may differ across age groups. It is often argued that children might be more vulnerable to task-irrelevant background sounds due to their less developed attention mechanisms. Moreover, while behavioral studies consistently demonstrate the detrimental impact of the ISE on performance, the neurophysiological mechanisms underlying this effect remain less well understood. Investigating event-related potential (ERP) components such as the N1, P3a, and MMN can provide valuable insights into various stages of auditory cognitive processing, from early processing demands, to the brain's pre-attentive auditory discrimination mechanisms and automatic orienting responses. Exploring both age-related variations and the neurophysiological mechanisms of the ISE could offer deeper insights into how auditory stimuli influence attention and cognitive performance.

The overall purpose of this thesis is to disentangle the roles of seriation, phonological processing and attention capture in the ISE, based on how this effect is evoked. It outlines various theoretical explanations for the ISE, highlights inconsistencies in its impact across different age groups, and underscores the importance of investigating both behavioral and neurophysiological aspects to gain deeper insights into how auditory distractions affect serial order retention.

In the present thesis, behavioral experiments (Studies 1–4) combined with EEG methodology (Studies 5–6) were performed to investigate the roles of seriation, phonological processing and attention in the ISE.

Contents

1	The Irrelevant Sound Effect	1
1.1	Definition	1
1.2	Theoretical Concepts Explaining the ISE	3
1.2.1	Phonological Loop Hypothesis	3
1.2.2	The Changing-State Hypothesis	5
1.2.3	The Embedded Processes Model	7
1.2.4	Integrative framework: The Duplex-Mechanism Account	9
1.3	Early Empirical Findings	12
2	Empirical Investigations	25
2.1	Research Questions	25
2.2	Study 1: Effects of Steady-State and Changing-State Speech	29
2.3	Study 2: Comparing Speech and Attention-Capturing Non-Speech Sounds	44
2.4	Study 3: Investigating Changing-State and Phonological Factors	57
2.5	Study 4: Comparing Monaural and Binaural Sound Presentation	78
2.6	Study 5: Neural correlates of the ISE	96
2.7	Study 6: Auditory Discrimination and Background Speech Susceptibility	107
3	General Discussion	123
4	Summary and Conclusion	139

1 The Irrelevant Sound Effect

1.1 Definition

This thesis focuses on the detrimental impact of task-irrelevant auditory material on the serial order retention of unrelated items, a phenomenon widely known as the Irrelevant Sound Effect (ISE). Over time, the terminology used to describe this effect has evolved, reflecting the accumulation of research findings in this area. Initially, the phenomenon was termed the “unattended speech effect” (Salamé & Baddeley, 1982) or the “irrelevant speech effect” (Jones et al., 1990), as early research primarily linked the disruption to background speech. However, subsequent studies revealed that the ISE could also be evoked by non-speech sounds, such as instrumental music (Ellermeier & Hellbrück, 1998; Klatte et al., 1995; Schlittmeier et al., 2008; Schweppe & Knigge, 2020) and tones (Divin et al., 2001; Jones & Macken, 1993). Consequently, the terminology was broadened to “irrelevant sound effect”, which is the term most commonly used today (Beaman & Jones, 1997). However, not all sounds are disruptive; for instance, background sounds that lack acoustic changes between tokens, such as steady-state sounds or pink noise, do not interfere with recall performance (Jones & Macken, 1993; Salamé & Baddeley, 1987).

Research on the ISE spans nearly 50 years, with the earliest evidence presented by Colle and Welsh (1976). In their study, participants were required to remember sequences of eight consonants while a passage from Franz Kafka’s “Ein Hungerkünstler” was played in the background. Despite instructions to ignore the background speech, which was unfamiliar to the participants, their serial recall performance was 12 % worse when task-irrelevant speech was present compared to a silent control condition. Later studies have confirmed the robustness of this phenomenon, with the magnitude of the effect ranging from 5 % to 50 %, depending on the specific conditions and materials used (Neath, 2000).

In typical ISE experiments, the task-irrelevant background sounds are played through speakers or headphones, which the participants are instructed to ignore while performing a serial recall task. A control condition often involves performing the task in silence or with white or pink noise for comparison. The serial recall task usually involves memorizing lists of five to nine unrelated verbal items, such as consonants, digits, words, letters, or pictorial representations of nouns. In contrast, research has shown that background sounds have little to no effect on serial recall of non-verbal items, such as the spatial position of dots. Klatte and Hellbrück (1997) suggested that the ISE is linked to the verbalization of stimuli, as the effect is evident with easily nameable items but absent with spatial items. Supporting this, Bergström et al. (2012) found that as the phonological codability of memory items decreased - from letters to pseudo-letters, patterns, and dot positions - the ISE diminished and eventually disappeared. This suggests that the phonological features

of stimuli play a crucial role in the presence of the ISE. In contrast, Jones et al. (1995) proposed that both verbal and spatial memory tasks are equally susceptible to the ISE, showing that changing-state speech disrupted serial order retention for both verbal and spatial items (visually presented digits and locations of dots on the screen, respectively). However, subsequent attempts to replicate these findings have produced mixed results, with some studies partially confirming the effect (Tremblay et al., 2012), while others have contradicted it (Bergström et al., 2012; Klatte & Hellbrück, 1997; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024).

Interestingly, the magnitude of the ISE in traditional paradigms does not seem to be influenced by the volume of the background sounds, with disruptive effects occurring at sound levels ranging from 40 dB(A) to 76 dB(A) (Colle, 1980; Ellermeier & Hellbrück, 1998; Schlittmeier et al., 2008; Tremblay & Jones, 1999; for conflicting findings see Alikadic & Röer, 2022).

Notably, tasks that do not require serial order processing, such as free recall (Salamé & Baddeley, 1990), item recognition (Beaman & Jones, 1997), paired-associate recall (Beaman & Jones, 1997; Meinhardt-Injac et al., 2022), and the missing item task (Beaman & Jones, 1997; Hughes et al., 2007; Jones & Macken, 1993), generally show little to no interference from irrelevant sounds. However, some studies have found negative effects even in these tasks (LeCompte, 1994; Stokes & Arnell, 2012). Some researchers (Beaman & Jones, 1997, 1998) suggested that these findings could be due to participants inadvertently using rehearsal strategies, which can be mitigated by articulatory suppression (repeating words, letters, or syllables aloud). Nevertheless, recent research shows that even when rehearsal is minimized through fast presentation, the ISE persists (AuBuchon et al., 2020; Samper et al., 2021). These findings indicate that while some progress has been made in understanding the ISE, further research is needed to clarify its impact on tasks that rely heavily on rehearsal.

Various theoretical frameworks have been proposed to explain the mechanisms underlying the ISE. Some theories suggest that the effect arises from specific interference between the obligatory processing of background sounds and the phonological or serial order representations generated during task performance. Other theories propose that irrelevant sounds involuntarily divert attention away from the focal task, thereby impairing performance. These theories will be discussed in greater detail in the following chapter, providing a comprehensive overview of the current theories underlying the ISE.

1.2 Theoretical Concepts Explaining the ISE

The documented study of memory goes back to the 19th century. Back then, Ebbinghaus (1885) learned meaningless syllables to measure how many he could remember a few days later. Just a few years later, the concept of memory was created, based on whether the information was kept for a short or long duration (James, 1890). Several decades later, Atkinson and Shiffrin (1968) developed a model that differentiates human memory based on qualitative features, including the duration of information retention and the capacity of each system: sensory memory, short-term memory, and long-term memory.

The ISE is considered a memory phenomenon that impacts short-term memory processes. In exploring the nature of the ISE, researchers have been debated whether it is a perceptual or a memory-related phenomenon. Broadbent (1983) suggested that the detrimental impact of speech on serial order retention was due to perceptual interference similar to the Stroop effect, rather than an interference with internal memory. To address this question, Baddeley and Salamé (1986) compared the impact of speech on tasks with and without a short-term memory component. In tasks in which participants judged consonant pairs and non-words according to various criteria without the need to memorize items, accuracy and reaction time were not affected by irrelevant speech. In contrast, in the serial-recall task, the detrimental impact was observed. Thus, short-term memorization of items appears to be a necessary condition for the effect, supporting the view that the ISE is a memory-related phenomenon (Baddeley & Salamé, 1986). However, some research suggests that attentional mechanisms may also play a role in contributing to the disruptive effects of irrelevant sounds.

This thesis explores several theoretical concepts that attempt to explain the ISE: the Phonological Loop Hypothesis, the Changing-State Account, and the Attention Capture Account. It also considers the Duplex-Mechanism Account, an integrative framework that combines elements from both the changing-state and attention capture theories. Each theory provides distinct insights into how task-irrelevant sounds disrupt serial order retention.

1.2.1 Phonological Loop Hypothesis

The phonological loop, one subsystem of the Multiple-Component Model of working memory (Baddeley, 2000a, 2021; Baddeley & Hitch, 1974; Baddeley et al., 2021), plays a central role in explaining the ISE. This subsystem is specialized for processing verbal and auditory information and is considered the primary mechanism through which irrelevant speech disrupts the serial retention of verbal items. I will begin by introducing the Multiple-Component Model, followed by a detailed explanation of the phonological loop. An adapted version of the Multiple-Component Model, including the processes of the phonological loop, is shown in Figure 1.1.

Baddeley's Multiple-Component Model of working memory assigns distinct roles to its subsystems (Baddeley, 2000a, 2021; Baddeley & Hitch, 1974; Baddeley et al., 2021).

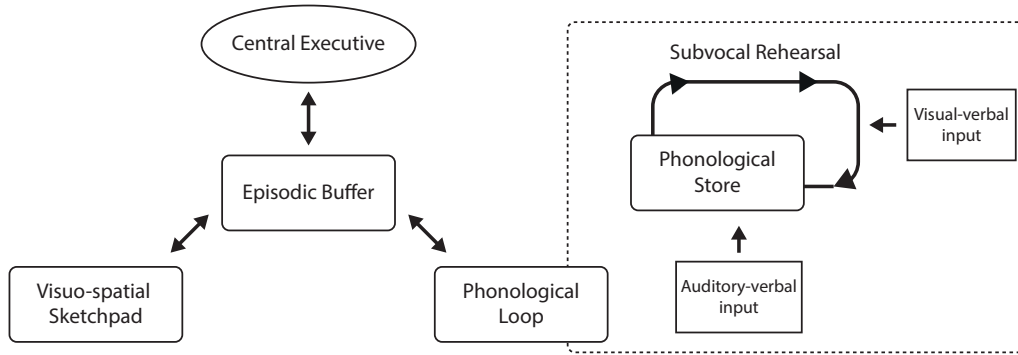


Figure 1.1: A simplified representation of the Multiple Component Model and the sub-processes of the Phonological Loop (Baddeley, 2021). The Phonological Loop consists of a passive store and an active rehearsal process. Articulatory rehearsal refreshes the phonological representations that are held in the storage, which can decay over time. Rehearsal also recodes visual-verbal information, e.g. written language, for storage. In contrast, auditory-verbal information, e.g., spoken language, can directly access the storage.

The central executive is an attentional control system, overseeing and coordinating the activities of the subsystems without acting as a storage unit itself. The episodic buffer separates storage functions from the central executive and is now viewed as a passive storage system, capable of retaining information in a multi-modal format (Baddeley et al., 2021). It integrates information from other components of working memory and long-term memory, forming unified episodic representations accessible to conscious awareness. The visuo-spatial sketchpad temporarily stores spatial and visual information, while the phonological loop is responsible for the storage of verbal information (Baddeley, 2000a, 2021; Baddeley & Hitch, 1974; Baddeley et al., 2021). Today, the model is seen as an interface that connects cognitive processes, such as perception, with long-term memory systems, including episodic, semantic, and procedural memory (Baddeley, 2021; Baddeley et al., 2021). For a detailed report of the model’s development, see Baddeley (2021).

Returning to the phonological loop, identified by Baddeley and colleagues as the primary locus of the ISE (e.g., Salamé & Baddeley, 1982, 1986), I will now explore its mechanisms in greater detail. This subsystem consists of two sub-components: the phonological store, which passively retains verbal information for a few seconds, and the articulatory rehearsal process, which actively refreshes this information through subvocal repetition to prevent decay. According to Salamé and Baddeley’s initial concept (Baddeley, 2000b; Salamé & Baddeley, 1982, 1986, 1989), irrelevant speech gains automatic access to this phonological store due to its speech-like qualities. Once inside, the phonological content of irrelevant speech interacts with the memory representations already held in the store, creating interference and disrupting performance.

Empirical support for the phonological loop’s role in the ISE includes findings from articulatory suppression studies (Salamé & Baddeley, 1982, Exp. 3; Hanley, 1997). For instance, requiring participants to repeat irrelevant sounds while recalling visually presented

lists often reduces the ISE's impact by isolating memory items from the phonological store. Additionally, Salamé and Baddeley (1982) suggested that irrelevant speech disrupts memory more when its phonological features, such as syllables or sounds, resemble those of the target items. This between-stream phonological similarity causes interference in the phonological store, leading to greater impairment in recall performance than when the sounds are less similar.

While the phonological loop account offers a powerful explanation for the ISE, it has limitations. One key issue is its difficulty in explaining why non-speech sounds, like tones (Jones & Macken, 1993) or instrumental music (Salamé & Baddeley, 1989), can also trigger a strong ISE despite lacking phonological content. Additionally, research indicates that the ISE effect of background speech does not necessarily depend on phonological similarity with memory items (Bell et al., 2011; Bridges & Jones, 1996; Larsen et al., 2000), challenging the idea that ISE is driven by between-stream phonological similarity (Salamé & Baddeley, 1982). Moreover, the ISE is most reliably seen in tasks involving serial order retention, suggesting that the effect may be tied more closely to structural properties of sounds, such as their “changing-state” nature (Hughes et al., 2007; Jones et al., 1992). This indicates that auditory disruptions may arise not solely from phonological interference but also from dynamic auditory changes that hinder serial order maintenance.

In sum, the phonological loop account compellingly describes the ISE as arising from content-based interference between phonological codes of task items and irrelevant speech that gains automatic access to the phonological store. However, the model does not fully capture the ISE complexity, especially with non-speech disruptions. According to its predictions, stimuli without speech-like properties, such as tones, would not interfere with verbal short-term memory, nor would irrelevant speech affect tasks requiring short-term memory for non-verbal items, like dot patterns.

1.2.2 The Changing-State Hypothesis

The changing-state hypothesis offers an alternative explanation for the disruptive effects of background sounds on serial order retention. Unlike the phonological loop account, which attributes interference to the content similarity between irrelevant speech and memory items, the changing-state hypothesis emphasizes that disruption arises from the variability or “changing-state” nature of auditory stimuli, which conflicts with serial order processing. This hypothesis was developed in response to findings that not only speech, but also tones disrupt the retention of serial order. Jones et al. (1992) propose that the interference is due to a conflict in the processes required for serial recall, rather than a similarity in content.

A key aspect of the changing-state hypothesis is the distinction between changing-state sounds, which vary over time (e.g., alternating letters like D-H-C-A), and steady-state sounds, which do not vary (e.g., a single letter repeated). The hypothesis suggests that changing-state sounds create a sequence of distinct auditory-perceptual stimuli, which are more disruptive to serial order processing than steady-state sounds. This disruption, known

as the changing-state effect (CSE), is central to the Object-Oriented Episodic Record (O-OER) Model (Jones et al., 1996). For an illustration of the model, see Figure 1.2.

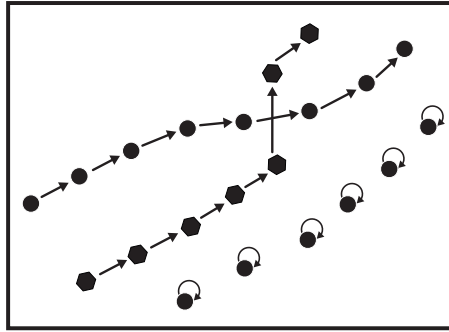


Figure 1.2: Adaption of the O-OER model as proposed by Jones et al. (1996). In this model, events are represented as streams of objects that are serially linked together through pointers on a metaphorical blackboard. One stream could include the changing-state sounds and another the list items that a participants tries to keep in mind. According to this model, the pre-attentive processes of auditory perceptual organization for changing-state sounds include the order of the auditory tokens that comprise the irrelevant stream. The resulting order cues disrupt the deliberate seriation processes involved in rehearsing memory items, resulting in impaired serial recall. The six self-referential objects represent a steady-state stream.

The O-OER Model, introduced by Jones et al. (1996), provides a unitary framework for understanding how the ISE arises from conflicting serial processes. According to the model, all stimuli, whether auditory or visual, are immediately represented as objects on a metaphorical blackboard (Elliott, 2002; Jones et al., 1996). These objects are organized into streams, connected by pointers that decay over time (Jones & Macken, 1993; Jones et al., 1996). Task-irrelevant sounds are pre-attentively organized into a connected stream, while visually presented items enter the blackboard through deliberate rehearsal. Once on the blackboard, these objects lose any distinction related to their original modality and becoming amodal. The competing streams—one formed by irrelevant sounds and the other by the to-be-recalled items—interfere with each other, leading to disrupted recall performance.

The changing-state hypothesis posits that sounds characterized by variability, such as fluctuating pitch, rhythm, or intensity, are more likely to interfere with cognitive processes involved in serial order retention than others (Hughes & Jones, 2001; Jones et al., 2010). This idea is rooted in the notion that changing-state sounds create competing cognitive representations, disrupting the rehearsal and retrieval processes crucial for maintaining information in working memory.

Substantial evidence supports the changing-state hypothesis. For example, Jones et al. (1992) conducted foundational experiments demonstrating that serial recall is significantly impaired by background sounds with a changing-state nature, such as alternating tones, compared to steady-state sounds like continuous white noise. These findings have been

consistently replicated, reinforcing the idea that the structure of auditory stimuli, rather than their content, plays a crucial role in the ISE. Moreover, Jones et al. (Exp. 4 1995) found that serial recall performance declines for both verbal and spatial items, such as black dots shown on various screen positions, suggesting that the ISE is tied to the disruption of serial order processes rather than being solely a phonological phenomenon. Furthermore, tasks that do not require serial order processing, such as free recall (Salamé & Baddeley, 1990), item recognition (Beaman & Jones, 1997), paired-associate recall (Beaman & Jones, 1997), and the missing item task (Beaman & Jones, 1997; Hughes et al., 2007; Jones & Macken, 1993), typically show little to no interference from irrelevant sounds, suggesting that the evocation of the ISE could rely on serial processing.

Despite its strengths, the changing-state hypothesis has faced criticism. First, some studies have observed negative effects on tasks like free recall (LeCompte, 1994), recognition (LeCompte, 1994; Stokes & Arnell, 2012), and the missing item task (LeCompte, 1996), indicating that the ISE may not be solely dependent on serial order processing. Second, efforts to replicate the findings of equivalent effects on verbal and spatial items have been unsuccessful (Bergström et al., 2012; Klatte & Hellbrück, 1997; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024). Klatte and Hellbrück (1997) suggested that the ISE is linked to the verbalization of stimuli, as the effect is evident with easily nameable items but absent with spatial items. Supporting this, Bergström et al. (2012) found that as the phonological codability of memory items decreased — from letters to pseudo-letters, patterns, and dot positions — the ISE diminished and eventually disappeared. This suggests that the phonological features of stimuli play a crucial role in the presence of the ISE. Additionally, the changing-state hypothesis does not fully explain why some steady-state sounds can also cause significant interference in certain tasks. These findings suggest that while the ISE is closely related to the need for serial order processing, it is not strictly limited to it. This clearly challenges the assumption that the ISE is a result of interfering seriation processes, indicating that factors other than the changing-state nature of sounds might also contribute to the ISE.

In sum, based on their research, Jones and colleagues proposed two key ideas: first, the changing-state hypothesis, which posits that auditory distractors that vary over time are more disruptive than repetitive sounds; and second, the concept of functional equivalence, suggesting that both speech and non-speech sounds (Jones & Macken, 1993), as well as different types of representational codes (verbal and spatial), have similar effects on serial order retention (Jones et al., 1995; Tremblay et al., 2012).

1.2.3 The Embedded Processes Model

The Embedded Processes Model (Cowan, 1988, 1999; Cowan et al., 2020) differs from other working memory models by proposing that working memory is not a separate system but rather a subset of activated long-term memory that is currently in focus. A simplified version of the Embedded-Processes Model, as proposed by Cowan (1988, 1999) and Cowan

et al. (2020), is depicted in Figure 1.3. Cowan’s Embedded Processes Model is based on his original model of information processing in the human mind (Cowan, 1988). This Model emphasizes the role of attention and links it to hierarchically organized memory stores.

In Cowan’s framework, working memory consists of mental components that temporarily store a limited amount of information in an active state. It includes several layers: long-term memory, representing the vast store of knowledge accumulated over a lifetime; activated long-term memory, a subset of long-term memory brought into an accessible state to meet current cognitive demands; and the focus of attention, a highly restricted portion of activated memory, usually holding only a few items, that is immediately available for conscious processing. Attention, in Cowan’s model, is crucial for maintaining and manipulating information within activated memory. Unlike models proposing distinct working memory stores, Cowan’s Embedded Processes Model ties working memory directly to how attention is allocated. The focus of attention is not only the seat of conscious awareness but also the gateway to deeper cognitive processing. Disruptions to this attentional focus can impair memory performance, especially in tasks that depend on maintaining serial order.

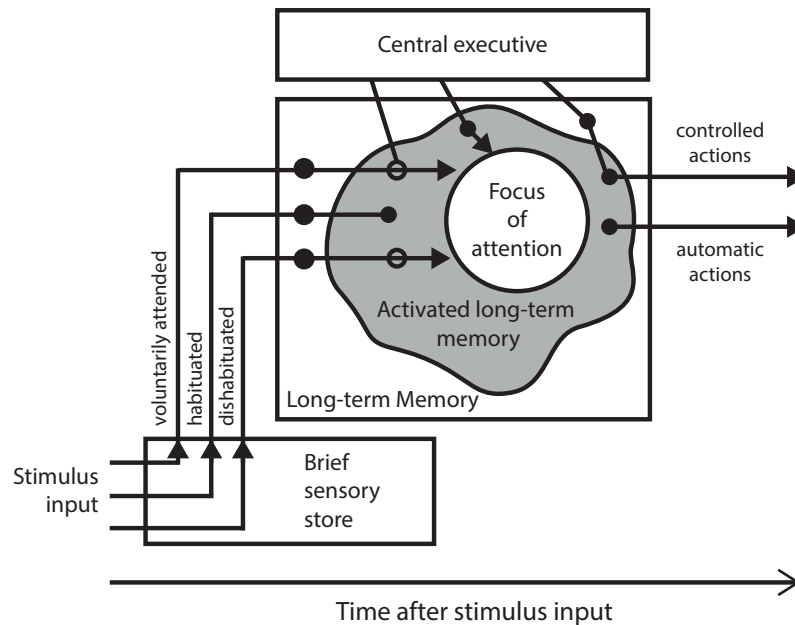


Figure 1.3: Simplified adaption of the Embedded-Processes Model as proposed by Cowan (1988, 1999) and Cowan et al. (2020). The model illustrates the chronological order of storing and processing information after a stimulus is presented. Processing occurs between functionally distinct components, which are represented by enclosed areas, either running in parallel and continually.

Regardless of the mode of presentation, any incoming information is briefly stored in the sensory register for a few hundred milliseconds, during which its physical characteristics are processed. This sensory information automatically activates matching representations in long-term memory, making it readily accessible for further processing. Some of these activated elements may enter the focus of attention. If the activated information is not the current focus, it may still become part of implicit long-term memory but cannot

be randomly or arbitrarily retrieved. For episodic long-term memories to form, active processing within the focus of attention is required. Novel or particularly salient stimuli entering the sensory store automatically capture attention through an orienting response.

In the Embedded Processes Model, the focus of attention is highly vulnerable to interference from irrelevant auditory stimuli. When attention is captured by such distractions, it shifts away from task-relevant items in working memory. This disruption not only hinders the encoding and rehearsal processes required for maintaining serial order but also allows irrelevant information to infiltrate the focus of attention, further complicating retention. The model explains the ISE by emphasizing how auditory distractions undermine the efficient management of attention. To maintain a longer sequence of items, they must be periodically refreshed through the focus of attention—a process known as “attentional refreshing”. This mechanism supports both verbal and nonverbal items. According to the attentional account, task-irrelevant sounds involuntarily intrude on this refreshing process, leading to impaired recall performance.

Proof for the attention capture, and thus the Embedded Processes Model, is provided by a body of research. Numerous studies (e.g., Hughes, Vachon, & Jones, 2005; Rettie et al., 2023; Röer et al., 2017; Vachon et al., 2012) have shown that auditory stimuli, particularly those that are unexpected or novel, can capture attention involuntarily. One key example is the so called deviant effect, which refers to the disruption within an otherwise identical sequence. For example, when a sequence includes a single item that deviates from a regular auditory pattern (e.g., “B, B, B, B, A, B, B”), it significantly impairs serial recall compared to sequences without such a deviant. Examples of auditory deviants include distractors spoken by a different voice (Sörqvist, 2010), or a distractor presented with a brief delay (Hughes, Vachon, & Jones, 2005).

Further evidence for attention capture comes from developmental studies, which explore age-related differences in the ISE, and investigations into habituation effects—that is, the reduction in ISE magnitude over the course of a task. The attentional account predicts that the ISE should decrease as participants become habituated to the irrelevant sounds with repeated exposure. Habituation is built on the idea that a mental model of a repeated stimulus is progressively refined (Sokolov, 1963). As the stimulus is presented repeatedly, it becomes less likely to differ from the mental model, reducing the likelihood of triggering an orienting response, leading to habituation (Hughes & Jones, 2001). However, findings from both research approaches remain inconclusive. For an overview on studies into these topics, see the sections on Habituation and Age effects.

1.2.4 Integrative framework: The Duplex-Mechanism Account

The Duplex-Mechanism Account (Hughes, 2014; Hughes et al., 2007) combines earlier explanatory models into one framework to identify two mechanisms of auditory distraction: interference-by-process and attention capture. This integrative framework explains both the specific interference of changing sounds with serial rehearsal processes, and the general

1 The Irrelevant Sound Effect

diversion of attention away from any focal task to the sounds.

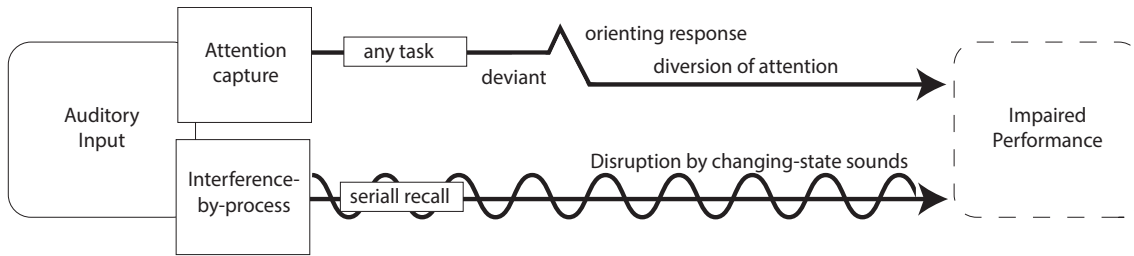


Figure 1.4: The diagram illustrates the Duplex-Mechanism Account of Auditory Distraction, which includes two distinct mechanisms: interference-by-process and attention capture. From the auditory input, two pathways branch out. The upper pathway depicts attention capture, a task-independent process in which a deviant sound, such as one’s own name or a change in voice, diverts attention from the focal task. Both pathways converge at the final outcome: impaired performance. The lower pathway represents interference-by-process, showing how task-specific processes, like serial recall, are disrupted by changing-state sounds (e.g., speech).

There are two distinct types of attention capture: specific attention capture, triggered by meaningful content (e.g., one’s name or a crying baby for its mother), and aspecific attention capture, which depends on the context (e.g., the interruption of a sequence of spoken letters). An example of attention capture is the deviant effect, in which a single deviating sound (e.g., in pitch, intensity, or content) within a regular sequence disrupts serial order retention. As long as a task requires attention and is not automated, it can be disrupted by specific interruptions (Hughes, 2014). For example, a letter spoken by a female can be interrupted by the same letter spoken by a male (Hughes, 2014; Hughes et al., 2007). Similarly, a sequence of the letter *A* can be disrupted by an infrequent letter, such as *B* (“A-A-A-B-A-A”).

In contrast, the other type of auditory distraction may result from direct interference between the automatic, obligatory processing of irrelevant sounds and the deliberate processes involved in the focal task. Specific interference occurs when obligatory sound processing depletes resources relevant to the task. Examples are the changing-state effect (Hughes, 2014), see the chapter on the Changing-State Hypothesis, and the semantic distraction effect (Marsh & Jones, 2010). The semantic distraction effect occurs when semantically similar speech triggers automatic activation within the long-term semantic network, interfering with the retrieval processes required for the focal task (Hughes, 2014). For example, in a semantic fluency task, in which participants generate words from a category like “fruit”, irrelevant spoken words from a related category, such as “vegetables”, are more disruptive to performance than unrelated words like “furniture” (Jones et al., 2012).

Evidence supporting the distinction between different mechanisms of auditory distraction comes from a variety of research approaches. These include studies on differences across differing age groups (for an overview, see the section on Age effects), task demands,

as well as anticipation of distractions (Hughes, 2014). Empirical research has shown that manipulating task difficulty through perceptual changes, such as adding noise to visually degrade stimuli, differentially impacts auditory distraction mechanisms. These manipulations attenuate or eliminate the disruptive effects of attentional capture while leaving the interference from changing-state auditory stimuli unaffected (Hughes, 2014; Hughes et al., 2013). Additionally, studies on the role of foreknowledge have demonstrated that warning participants about an impending auditory distraction can eliminate the deviant effect but not the changing-state effect (Hughes, 2014; Hughes & Marsh, 2020; Hughes et al., 2013). However, contradictory findings have been reported in other studies (Bell et al., 2017; Röer, Bell, & Buchner, 2015). Research on developmental changes indicates that while there are no age differences in the changing-state effect, the deviant effect does vary with age (Hughes, 2014; Joseph et al., 2018). For more information on this topic, see section on Age effects.

In sum, the Duplex-Mechanism Account explains auditory distraction through two mechanisms: interference-by-process, which disrupts task-specific cognitive processes like serial rehearsal, and attention capture, which diverts attention away from the task regardless of the specific task demands. Interference is task-dependent, while attention capture is task-independent.

1.3 Early Empirical Findings

For almost fifty years, the ISE has consistently demonstrated reproducibility across diverse studies, utilizing various experimental materials, designs, and procedures (e.g., Colle, 1980; Ellermeier & Zimmer, 1997; Hanley & Broadbent, 1987; Jones & Macken, 1993; Salamé & Baddeley, 1982, 1989). Empirical studies indicate that several factors influence the detrimental effects of specific background sounds on serial order retention. The following sections provide an overview of research findings accumulated over the past half-century, covering task-specificity, sound characteristics, developmental and subjective aspects, as well as related neurocognitive findings.

Acoustic Features

There is broad consensus that task-irrelevant background sounds impair serial recall performance. However, not all types of noise have the same effect (Marsh, Hanczakowski, et al., 2024). A number of studies have been conducted for investigating various characteristics of background sounds, including the distinction between speech and non-speech sounds, as well as auditory complexity. These studies aim to better understand the specific conditions under which the ISE is most pronounced and how different types of auditory stimuli contribute to this sound-induced disruption.

Speech and Non-Speech Various studies provide compelling evidence that not only speech, but also non-speech sounds can disrupt serial recall. This includes tones (Antonietti et al., 2018; Divin et al., 2001; Hadlington et al., 2004; Jones & Macken, 1993; Jones et al., 1999; Klatte et al., 1995; LeCompte et al., 1997), instrumental music (Ellermeier & Hellbrück, 1998; Klatte et al., 1995; Salamé & Baddeley, 1989; Schlittmeier et al., 2008; Schweppe & Knigge, 2020), sine-wave speech (Tremblay et al., 2000; Viswanathan et al., 2014), backward speech (Jones et al., 1990; LeCompte et al., 1997; Viswanathan et al., 2014), and environmental sounds (Buchner et al., 2008). Although these non-speech sounds are disruptive, the magnitude of interference tends to be highest with speech sounds (Schlittmeier, Weißgerber, et al., 2011; Tremblay et al., 2000).

But why does speech have such a strong impact on serial order retention? Schlittmeier, Weißgerber, et al. (2011) argue that speech signals “best fulfill the criteria to gain automatic and obligatory access to short-term memory” due to the fluctuations in these signals, as “heard speech is just a complex time-varying acoustic signal” (p. 202).

This raises the question of whether the semantic content of speech sounds plays a special role in evoking the ISE. Research has shown that serial recall performance can be impaired even when the distractors lack meaningful content. For instance, foreign language distractors, which listeners do not understand, still cause disruption (Colle, 1980; Colle & Welsh, 1976; Salamé & Baddeley, 1986, 1987, 1989). Similarly, the effect is observed with reversed speech (Jones et al., 1990; LeCompte et al., 1997; Viswanathan et al., 2014) or sine-wave speech (Tremblay et al., 2000; Viswanathan et al., 2014), in which the normal

flow of language is distorted to the point of incomprehensibility.

Studies comparing familiar speech to incomprehensible foreign speech, where participants do not understand the language, show comparable disruption in serial recall (Ueda et al., 2019). Furthermore, meaningless words, which carry no semantic value, have also been found to cause significant disruption (Buchner et al., 1996). Specifically, Buchner et al. (1996) investigated the effect of semantic similarity between to-be-remembered material (e.g., digits) and to-be-ignored material (also digits), showing no difference in the disruptive effect across numbers, pseudo-numbers (e.g., “drünfzig”, a combination of the German digits thirty and fifty), and phonologically related words, indicating that semantic similarity does not influence the ISE.

The influence of semantic content on auditory disruption appears to depend on the type of task and the participant’s foreknowledge (Rettie et al., 2023; Röer et al., 2017). Recent research suggests that taboo words have a greater impact than neutral words, implying that task-irrelevant speech is still processed to some extent. Some studies demonstrate that taboo words exert a stronger disruptive effect (Rettie et al., 2023, Exp.1; Röer et al., 2017). However, this taboo-distractor effect is eliminated when participants are made aware of the distractors beforehand (Rettie et al., 2023, Exp.2). These findings suggest that the effect may be due to attention capture rather than semantic processing (Klatte, Lachmann, et al., 2010; Schlittmeier, Weißgerber, et al., 2011). In line with this, Marsh, Ljung, et al. (2018) examined the influence of “halfalogues”, which occur when overhearing only one side of a phone conversation. Their findings suggest that the semantic unpredictability of these conversations contributes to their disruptive impact on a search task, attributing this to specific attention capture.

Moreover, meaningful but irrelevant speech can disrupt tasks that rely on the use of meaning for retrieval, such as proofreading (Jones et al., 1990) and reading comprehension (Martin et al., 1988). The research on semanticity suggests that its impact may be task-dependent (Marsh & Jones, 2010; Marsh, Ljung, et al., 2018; Marsh, Yang, et al., 2018). For a detailed review of the effects of semanticity, see Marsh and Jones (2010).

Recent studies have shown that the emotional content of speech can also play a role in disruption. Zhang et al. (2024) demonstrated that both negative and positive emotional valence are generally more disruptive than neutral speech in serial recall tasks.

Finally, it is important to note that repeated speech sounds, such as steady-state sounds, whether a repeated letter (Jones et al., 1992) or a tone (Jones & Macken, 1993), do not or only mildly disrupt serial order retention. This suggests that variability in the auditory stream, i.e. the fluctuation strength of a background sound, is a key factor in the degree of disruption caused by background sounds (Schlittmeier, Weißgerber, et al., 2011).

Modes of Sound Presentation From an acoustics perspective, sound presentation can be categorized based on whether the sound is delivered to one ear or both ears and whether the sound differs between the ears (Blauert & Braasch, 2008). Monaural (or diotic) presentation involves delivering the same sound signal to both ears, meaning that one

1 The Irrelevant Sound Effect

ear alone is sufficient for hearing. In contrast, binaural presentation provides sound to both ears in a way that includes spatial information, accurately reproducing the interaural differences that help us locate sounds in space. Dichotic presentation, on the other hand, involves delivering different sounds to each ear.

In most studies examining the ISE, irrelevant sounds have been presented monaurally, with few studies considering the effects of spatially distributed sound sources. Buchner et al. (2008), for example, presented non-speech sounds (such as footsteps and water splashes) and speech (unrelated word sequences) from loudspeakers positioned at various locations. The results showed that disruption was greatest when sounds were played from the front, near the visual target field, while the distance of the sound source had only a minimal effect on the overall level of interference. Jones and Macken (1995) explored the effects of background speech produced by six voices delivered through loudspeakers arranged in a circle around the participant. Serial recall performance was worse when each voice was assigned to a separate loudspeaker (creating six distinct single-talker streams) compared to when a mix of the six voices was played from each loudspeaker (creating identical streams of babble speech).

Despite the complexity of real-world acoustic environments, most ISE research has relied on monaural presentations, which oversimplifies the rich features of binaural hearing and the effects of spatially distributed sounds. Typically, these studies use headphone-based diotic presentations of auditory distractors, such as speech, tones, or environmental sounds. When loudspeakers are used, the same signal is often sent to each speaker, neglecting the complex acoustics of environments like open-plan offices or noisy classrooms. To accurately reproduce these conditions, it is essential to consider spatial, binaural, and room acoustics. Hitherto, within the ISE paradigm, only one study has utilized binaural reproduction methods to more accurately replicate interaural differences (Yadav et al., 2023). Specifically, Yadav et al. (2023) investigated the effects of monaural versus binaural reproduction in a simulated open-plan office environment and found no significant difference between the two methods.

In summary, most ISE studies using headphone-based reproduction have relied on monaural presentations, with only one study investigating binaural presentations that accurately represent spatial and interaural cues. The impact of realistic room acoustics and representative background sounds remains largely unexplored, pointing to a need for future research that more closely mimics real-world auditory environments.

Developmental Factors and Habituation

This chapter explores age-related differences in susceptibility to auditory distraction and the role of habituation, both key factors in understanding attention control within competing models of working memory.

Age effects While specific task-irrelevant background sounds have been shown to impair short-term memory performance for serial order retention in adults, it is unclear if children are more prone or equally susceptible to auditory distractions. Hitherto, twelve studies have investigated developmental changes in the ISE across fifteen experiments. Among these, three studies exclusively examined the detrimental effects of auditory distraction in children with no comparison to different age groups (AuBuchon et al., 2019; Elliott et al., 2007; Klatte et al., 2007). Another study analyzed the impact of auditory distraction separately for children and adults (Elliott & Cowan, 2005). The remaining studies investigated differences in visual-verbal serial order retention, with a limited number also examining non-serial short-term memory tasks across various age groups to identify developmental differences. For a full overview, see Table 1.1. Examining the results of these investigations yields mixed findings: while some studies (Elliott et al., 2016, Exp.1; Elliott, 2002; Elliott & Briganti, 2012, Exp. 3; Joseph et al., 2018) reported a stronger ISE in children compared to adults, others (Elliott et al., 2016, Exp.2; Joseph et al., 2018; Klatte, Lachmann, et al., 2010; Meinhardt-Injac et al., 2022; Röer et al., 2018; Schwarz et al., 2015) report no significant developmental differences in susceptibility to auditory distraction. The extent and nature of the age effect may differ based on age groups, methodological variations (such as varying the list length), and the auditory distractors.

Despite numerous attempts to clarify developmental changes in the ISE, the overall results remain mixed and difficult to interpret due to several factors (Röer et al., 2018): Firstly, one difficulty arises from equating task difficulty across age groups. In many studies, the length of the lists for serial order retention is shortened for children to avoid over-straining. However, using proportion correct as the dependent measure can distort results. Since children and adults are recalling different numbers of items, each additional error impacts children's performance in proportion to total recall more than adults. This may create a false impression that children are more susceptible to distractions, when, in reality, the absolute effect of distraction is identical for both groups. Secondly, the procedures across studies have varied widely, making it difficult to generalize findings or understand how background sounds impact memory across diverse populations. For instance, some studies (Elliott & Briganti, 2012; Elliott et al., 2016) have focused solely on top performers, such as children from magnet elementary schools who passed an admission screening or college students in the upper quartile of their cohort. By restricting samples to high achievers, these studies may overlook key insights into how different performance levels might influence susceptibility to auditory interference. Moreover, some studies used differing number of trials across age groups (Joseph et al., 2018; Klatte, Lachmann,

1 The Irrelevant Sound Effect

et al., 2010). Finally, some experimental conditions were excluded from the analysis in some studies. For example, Elliott (2002) collected sixty trials to cover a wide range of span lengths; however, only one-quarter of these trials were analyzed. Without direct comparisons between different conditions, it is difficult to isolate the specific impacts measured in the Experiment.

These factors collectively contribute to the mixed and sometimes ambiguous findings in research, leaving open the empirical question of whether children are more susceptible to auditory distraction in serial recall than adults. Future studies are needed to directly compare the effects of task-irrelevant sounds across age groups under controlled conditions to determine if children indeed experience greater interference.

Table 1.1: Studies on developmental differences in the ISE. Adapted and Expanded from Röer et al. (2018) to include additional information not present in the original source and to include recent studies.

Publication	Exp.	N	Age groups (Mean age)	Task	Trials (practice)	To-be-remembered material	List length	Auditory conditions	Age effect
Elliott, 2002	1	128	32 2nd-graders (8;3) 32 3rd-4th graders (9;6) 32 5th-graders (11;8) 32 adults (19;0)	SR	60 (5)	Digits	adjusted	Silence St.-state words Ch.-st. words St.-st. tones Ch.-st. tones	yes
Elliott & Cowan, 2005	2	127	64 adults (na) 63 2nd-graders (8;2)	SR	na	Digits	adjusted	Silence Ch.-st. words Ch.-st. sine-wave tones	not analyzed
Elliott & Briganti, 2012	3	57	23 adults (19.5) 34 2nd-3rd graders (7.7)	SR	30 (3)	Words	Ad.: fixed Ch.: adjusted	Silence Low-frequency words (Ch.-st.) High-frequency words (Ch.-st.)	yes
Klatte, Lachmann, et al., 2010	1	100	47 adults (Med.=25) 53 1st-graders (Med.=7;0)	SR	Ad: 24 (na) Ch: 16 (na)	Pictures	fixed	Silence Classroom noise without speech Speech	no
	2	45	24 adults (Med.=22) 21 2nd-3rd graders (Med.=8;5)	SR	30	Pictures	adjusted	Silence Classroom noise without speech Speech	no
Schwarz et al., 2015	-	98	68 adults (20) 30 children (9;1)	SRC	128 (8)	Nouns	fixed	Pink noise Speech	no
Elliott et al., 2016	1	62	37 adults (na) 25 children (7.4)	SR	-	Digits	fixed	Silence words	yes
	2	128	64 adults (19.7) 35 children (7.7) 29 children (7.5)	POR MI	36 (9)	Digits	fixed	Silence St.-st. letters Ch.-st. letters	no (POR) yes (MI)
Joseph et al., 2018	-	178	89 adults (20) 89 children (8;2)	SR POR MI	Ad.: 32 (na) Ch. 24 (na)	Digits	fixed and adjusted	Silence Ch.-st. letters St.-st. letters Deviant-Ch.-st. letters Deviant-St.-st. letters	yes (deviant) no (Ch.-st.)
Röer et al., 2018	1	121	28 Younger adults (24.1) 30 Older adults (68.8) 35 3rd-graders (8.03) 28 5th-graders (10.8)	SR	24 (8)	Digits	fixed	Ch.-st. words St.-st. words Deviant-Ch.-st. words	no
	2	68	38 adults (23.6) 30 4th-graders (9.1)	SR	24 (8)	Digits	adjusted	Ch.-st. words St.-st. words Deviant-Ch.-st. words	no
Meinhardt-Injac et al., 2022	-	182	52 adults (23.6) 57 younger children (9.2) 73 older children (11.3)	SR PAR Stroop	60 (5)	Digits	adjusted	Silence Classroom noise with speech	no
Elliott et al., 2007	-	33	11 (C)APD (10;11) 11 Span-Controls (8;6) 11 Age-Controls (10;10)	SR	30 (3)	Digits	adjusted	Silence Ch.-st. words Ch.-st. sine-wave tones	-
Klatte et al., 2007	3	21	2nd-graders (8;4)	SR	64 (na)	Digits	mixed	Silence Unfiltered train Filtered train Speech	-
AuBuchon et al., 2019	-	60	children (age range: 7;0-8;7)	SR	30 (na)	Digits	adjusted	Silence Ch.-st. words St.-st. words	-

Med.= Median, MI = Missing item task, POR = Probed-order recall task, PAR = Paired-associate recall task, SR = Serial recall task, SRC = Serial recognition task, Ch.-st. = Changing-state, St.-st. = steady-state., fam. = familiar

Habituation Habituation refers to the reduction of the ISE with prolonged exposure to distracting auditory stimuli (Bell et al., 2012). Understanding whether individuals can habituate to task-irrelevant background sounds is crucial for both theoretical and practical reasons. Among the models explaining the ISE, the attentional capture account uniquely predicts full habituation, suggesting that people can adapt and diminish the impact of irrelevant sounds over time. This raises the important question of whether performance habituates to the continuous processing of auditory information (Röer et al., 2013a).

Research on this topic, however, yields mixed results. Some studies report evidence of habituation (Elliott, 2002; Littlefair et al., 2022; Morris & Jones, 1990; Pelletier et al., 2016; Röer et al., 2013a, 2013b, 2014a, 2018; Vachon et al., 2012; Zhang et al., 2024; Röer et al., 2014b, Exp. 1, 3–4), while others do not find any effect of habituation (Beaman & Roeer, 2009; Ellermeier & Zimmer, 1997; Hellbrück et al., 1996; Hughes, Vachon, & Jones, 2005; Jones et al., 1997; Röer et al., 2011, 2017, 2019; Tremblay & Jones, 1998). Only two studies have examined the effect of habituation in children, yielding mixed results. Elliott (2002) found evidence of habituation, which was comparable between children and adults, indicating that the performance of both age groups improved over time. In contrast, Röer et al. (2019) did not observe a habituation effect.

Researchers have attributed the inconsistency in habituation effects to variations in methodologies, such as the varying presentation duration of background sounds. It has been suggested that habituation may depend on whether the sounds are played continuously in a blocked design (i.e. specific habituation) or varied on a trial-by-trial basis (i.e., unspecific habituation; Röer et al. 2019). For a full list of studies on habituation in the ISE and the methods used, see Table A1. However, studies comparing this presentation duration revealed mixed results. Table 1.2 shows the different presentation duration formats used to study habituation in the ISE and how they are analyzed. For example, while some research found no effect of habituation neither in the trial-by-trial nor in a longer blocked presentation of background sounds (Jones et al., 1997; Röer et al., 2019), others reported that participants attenuate to the distraction levels between trial-by-trial and longer blocked presentation of background sounds (Röer et al., 2013a). This suggests that additional factors may mediate the relationship between habituation effects and auditory interference.

Another factor contributing to these inconsistencies is the type of auditory distraction, with Sörqvist (2010) suggesting that habituation might vary depending on whether the distraction arises from changing-state background sounds or a deviant embedded in a continuous sound sequence. However, empirical evidence supporting this distinction is limited. For instance, Röer, Bell, Marsh, and Buchner (2015) compared a changing-state speech condition to an auditory deviant condition and found no habituation effects. Similarly, a study using taboo words in an auditory deviant condition found no evidence of habituation (Röer et al., 2017). In contrast, Röer et al. (2018) conducted a study investigating habituation effects in children and adults, and reported habituation for both changing-state and auditory deviant conditions, but not for a steady-state condition. Likewise, Littlefair et al. (2022) observed habituation to acoustic deviants but not to

categorical deviants, suggesting a potential difference in how these types of distractions are processed.

Most recently, Zhang et al. (2024) explored habituation to emotional speech, including neutral, positive, and negative expressions. They found partial habituation to negative speech, with its initial disruptive effect diminished over time. This suggests that while emotional valence initially captures attention, its impact may lessen with repeated exposure.

In summary, the evidence on habituation to irrelevant sounds remains mixed, with factors such as the presentation method and type of auditory distraction playing key roles in determining the outcome.

Table 1.2: Studies on habituation in the ISE. For a full overview, see Appendix.

Publication	Presentation format	Analysis	Habituation effect
Röer et al., 2013a, 2013b, 2014a	trial-by-trial	trial-by-trial	yes
Röer, Bell, Marsh, & Buchner, 2015; Röer et al., 2014b, 2017, 2019	trial-by-trial	trial-by-trial	no
Ellermeier & Zimmer, 1997; Hughes, Vachon, & Jones, 2005; Jones et al., 1997; Tremblay & Jones, 1998	trial-by-trial	blocked	no
Elliott, 2002; Morris & Jones, 1990; Pelletier et al., 2016	blocked	blocked	yes
Jones et al., 1997; Röer et al., 2011; Tremblay & Jones, 1998; Vachon et al., 2012	blocked	blocked	no
Röer et al., 2013a	blocked	trial-by-trial	yes
Hellbrück et al., 1996	blocked	trial-by-trial	no
Röer et al., 2014b, Exp. 1, 3–4	trial-by-trial	trial-by-trial	*
Zhang et al., 2024	trial-by-trial	blocked	*
Littlefair et al., 2022	trial-by-trial	blocked	*
Röer et al., 2018	blocked	trial-by-trial	*

* depends on auditory condition

Neurocognitive Mechanisms of the ISE

One approach to explore neurocognitive correlates of auditory processing involves utilizing electroencephalography (EEG). EEG is a non-invasive neurophysiological technique that monitors and records the electrical activity of the brain. During EEG measurement, small electrodes are positioned on the scalp to capture the electrical signals produced by neurons, manifesting as brainwaves measured in microvolts due to continuous changes in neuronal arousal (Gazzaniga et al., 2014). This enables researchers to investigate brain activity with exceptional temporal resolution, revealing subtle changes in neural activity associated with the processing of, for example, the onset of a picture on a computer screen, a sound presented over headphones or a the participants' button presses (Kappenman & Luck, 2011). By averaging EEG signals across multiple segments, researchers can isolate and analyze the brain activity associated with a specific event, filtering out irrelevant brain activity and noise. The segmentation of the EEG signal to time-locked events results in so-called event-related potentials (ERPs).

Neuronal basis of the ISE Different ERP components are related to distinct cognitive events. Therefore, researchers have the ability to examine various aspects of brain function. Various neural responses play a crucial role in clarifying how the brain perceives and interprets auditory stimuli. So far, three ERP components have been examined in relation to the ISE: the N1, the Mismatch Negativity (MMN) and the P3a (Bell et al., 2010; Valtonen et al., 2003).

The N1 component is an early negative deflection at fronto-central sites that typically occurs within the first 100 ms after stimulus onset. It is associated with the distribution of attentional resources. Research suggests that an increased amplitude signifies enhanced attention (Little et al., 2010).

The MMN is an ERP component typically observed as a negative deflection occurring when an infrequent deviant stimulus disrupts a series of otherwise identical frequent stimuli. It is assumed to reflect the brain's pre-attentive detection of changes in the auditory environment (Näätänen et al., 2005). It is observed within a time frame of 150 to 200 milliseconds after the presentation of a deviant stimulus. Furthermore, previous research findings indicated that infrequent deviants triggering an MMN can divert attention away from an ongoing task (Campbell et al., 2007; Escera et al., 2000). Campbell et al. (2007) indicated that MMN does not appear to play a role in memory impairment during serial order retention, while the N1 is linked to memory disruption.

The P3a component is a positive deflection at fronto-central sites that occurs around 250–300 ms after stimulus onset. It is associated with an orienting response to novel or salient auditory stimuli, indicating the involuntary capture of attention by unexpected events (Bell et al., 2010; Polich, 2007; Sutton et al., 1965).

Research background on the neuronal basis of the ISE A series of studies have explored the ERP correlates of the ISE. While most of these studies have focused on investigating the

processes time-locked to the auditory distractors (Bell et al., 2010; Campbell et al., 2003, 2007; Valtonen et al., 2003), fewer have explored the detrimental effects of these distractors on the memory items that need to be remembered (Little et al., 2010; Martín-Loeches et al., 1997; Schlittmeier, Weisz, & Bertrand, 2011; Weisz & Schlittmeier, 2006). The following section provides an overview of the studies that have investigated the neural basis of the ISE.

Bell et al. (2010) explored the impact of changing-state sequences compared to steady-state speech tokens on serial recall. They found that the N1 amplitude was greater in the changing-state condition compared to the steady-state condition. This was also shown by Campbell et al. (2003), who examined the involvement of the auditory N1 component while participants heard nonsense syllables. They reported a greater auditory N1 component. Moreover, Campbell et al. (2003) investigated how increasing the number of different sound tokens, comparing one to two to five syllables, affected the N1 response. Behaviorally, the study observed that memory performance was most disrupted when exposed to five tokens, and there was no difference between one and two tokens. The N1, on the other hand, increased along with increasing number of tokens (Campbell et al., 2003). In an additional study, Campbell et al. (2007) replicated these findings by using a token-set size of nine items. Valtonen et al. (2003) investigated how memory load affects the auditory N1m¹ response to irrelevant stimuli. They found that increasing the list length of visually presented items resulted in a higher amplitude of the auditory N1m response compared to a no-task condition (Valtonen et al., 2003).

Bell et al. (2010) and Valtonen et al. (2003) have examined whether the auditory N1 is affected differently when task-irrelevant background sounds are presented either during the encoding phase or the retention interval. A study conducted by Valtonen et al. (2003) showed that the amplitude of the N1m varied based on the phase of the memory task. Specifically, the N1m amplitude was larger during encoding compared to the retention interval. In contrast, Bell et al. (2010) observed no significant difference in the auditory N1 amplitude as a function of phase. Their findings indicate that the N1 amplitude throughout encoding and the retention interval was comparable.

Only a limited number of studies have explored the neural activity associated with the processing of the memorized material in the presence of task-irrelevant sounds. Weisz and Schlittmeier (2006) observed a diminished visual N1 during task-irrelevant fluent speech condition compared to control conditions. This is in line with the findings reported by Little et al. (2010). During a serial recall task of seven visually presented consonants, Little et al. (2010) observed an increased visual N1 amplitude during the presence of changing words compared to noise and silence. This led to the interpretation that task-irrelevant background speech may automatically capture attentional resources of the to-be-remembered material (Bell et al., 2010; Little et al., 2010; Weisz & Schlittmeier, 2006). However, when evaluating

¹The N1m refers to the ERP component measured through magnetoencephalography (MEG). MEG is a non-invasive neuroimaging method, measuring the magnetic fields of the brain. The N1m and N1 component reflect comparable cognitive processes.

1 The Irrelevant Sound Effect

the N1 response of the auditory presentation of learned material, the magnitude of the response appears to be comparable across sound conditions. In a serial recall task involving auditorily presented lists of 9 digits, there was no distinct difference in N1 amplitude of the to-be-remembered items between the changing-state and steady-state conditions (Schlittmeier, Weisz, & Bertrand, 2011).

Taken together, three ERP components, namely the N1, the MMN and the P3a, have been examined in the irrelevant sound paradigm. Overall, research findings consistently observed that the visual and the auditory N1 is more negative in changing-state speech condition compared to a control condition. However, there is no difference in the N1 response across different sound conditions when the memory items are presented auditorily. The auditory N1m response seems to be affected by the memory load, as evidenced by the comparison of long and short list lengths. Divergent findings emerge regarding the N1 response in relation to the presentation phase of background sounds. Research on the MMN suggests that this component may be unrelated to the ISE. An enhancement of the P3a in response to background speech has been linked to an orienting reaction, suggesting that task-irrelevant background speech may automatically capture attentional resources.

Overview of Early Empirical Findings

To summarize the extensive empirical findings on the ISE discussed in previous sections, I will now highlight key observations across various domains of research, including task-specific characteristics, sound properties, subjective factors, and neurocognitive findings.

Task-specific factors Task characteristics play a crucial role in determining the extent of the ISE. Studies have shown that the memory material, and seriation processes within a task might affect the susceptibility to auditory distraction. Visually presented memory lists typically consist of verbal items, such as letters, digits, words, and pictorially represented nouns. Most studies examining the ISE with spatially distributed items—like dot patterns displayed at various locations on a computer screen—generally find that background sounds have little to no impact on performance (Bergström et al., 2012; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024). Additionally, background sounds tend to have minimal effects on tasks that do not require retaining items in a specific order, such as the Missing Item Task and Free Recall Task.

Sound Characteristics and Presentation Sound Characteristics and Presentation appear to play a minor role. In the traditional paradigm, the volume of the background sound does not affect ISE magnitude. Background sounds from 40 to 76 dB(A), even with volume variations, do not affect the level of disruption (Neath, 2000; cf., Alikadic & Röer, 2022). Research suggests that the ISE can be elicited by a variety of sounds, including unfamiliar speech (Colle & Welsh, 1976), backward speech (Jones et al., 1990; LeCompte et al., 1997; Surprenant, 2007; Viswanathan et al., 2014), tones (Divin et al., 2001; Jones & Macken, 1993) and music (Ellermeier & Hellbrück, 1998; Klatte et al., 1995; Schlittmeier et al., 2008; Schweppe & Knigge, 2020). However, the disruptive impact of speech is unsurpassed (Schlittmeier, Weißgerber, et al., 2011). Most studies on the ISE have focused on monaural presentations, with only one (Yadav et al., 2023) examining binaural presentations that reflect spatial and interaural cues, showing no differences between monaural and binaural sound presentation.

Developmental factors and Habituation The investigation of developmental change and habituation provides important insights into inter-individual differences in susceptibility to auditory distraction. The research findings regarding whether or not children are more susceptible to sound-induced performance decline have been inconclusive. Several studies (Elliott et al., 2016, Exp.1; Elliott, 2002; Elliott & Briganti, 2012, Exp. 3; Joseph et al., 2018) have observed an increased ISE in children. However, other studies have not detected any age-related differences in the magnitude of the ISE (Elliott et al., 2016, Exp.2; Joseph et al., 2018; Klatte, Lachmann, et al., 2010; Meinhardt-Injac et al., 2022; Röer et al., 2018; Schwarz et al., 2015). Habituation refers to the reduction of the ISE with repeated exposure to auditory stimuli. The evidence on habituation to irrelevant sounds is inconsistent. While some researchers attribute these mixed findings to methodological variations (e.g., Röer et al., 2019)—such as whether background sounds are presented continuously or vary from trial to trial—others suggest the type of auditory distraction plays a role, such as

1 The Irrelevant Sound Effect

whether it involves a deviant or changing-state effect. However, empirical support for both explanations remains limited. Studies investigating habituation in children within the ISE paradigm have also produced inconsistent results, further adding to the complexity of understanding the habituation effect.

Neurocognitive Findings EEG measures, such as the investigation of the ERP components N1, MMN, and P3a, provide insights into the neural mechanisms underlying the ISE. The amplitude of the visual N1, which corresponds here to the visual-verbal target items, that are need to be remembered in serial order, and the auditory N1, which corresponds to the auditory distractors, is enhanced when irrelevant speech is presented. This enhancement indicates heightened attention to the auditory distractor and may reflect interference with memory. Research on the MMN suggests that this component may be unrelated to sound-induced memory disruption. Additionally, an enhancement of the P3a due to background speech has been linked to an orienting reaction.

2 Empirical Investigations

2.1 Research Questions

The ISE, well-documented for almost 50 years, describes the consistent negative effect of certain task-irrelevant background sounds on the ability to recall visual-verbal items in serial order. Various theories have sought to explain the mechanisms behind this interference. This thesis aims to clarify the roles of phonological processing, attention capture, and serial order retention by exploring how different background sounds, such as speech, non-speech, and classroom noise, affect the serial recall of visual-verbal and visual-spatial items. Additionally, EEG is used to examine the neural mechanisms underlying the ISE. The research questions are outlined and explained in the following sections.

Research Question I: Are verbal and spatial serial recall equally impaired by background sounds? The first research question explores whether serial recall of verbal and spatial items presented visually are equally impaired by task-irrelevant background sounds. The working memory model posits that verbal and spatial materials are processed through distinct storage systems, leading to specific interference effects based on the type of material involved. According to this model, task-irrelevant speech should primarily disrupt verbal tasks (Salamé & Baddeley, 1982) while leaving spatial tasks unaffected. However, findings from Jones et al. (1995, Exp. 4) challenged this assumption by demonstrating that changing-state speech causes comparable impairments in the serial order retention of both verbal items (visually presented digits) and spatial items (locations of dots on the screen). This suggests a functional equivalence between verbal and spatial materials, as the interference from changing-state speech appears to operate similarly across both domains.

Investigating whether changing-state speech produces similar impairments in both types of materials is essential for understanding the underlying processes that contribute to the elicitation of the ISE. If both verbal and spatial items are equally affected, it implies that interference may not be as domain-specific as previously thought. This has significant implications for our understanding of attention capture, serial order retention and phonological processing. By exploring this question, we can gain deeper insights into the interplay between verbal and spatial processing and potentially refine existing models to account for these findings.

Research Question II: Are children more susceptible to the ISE than adults? The second research question explores whether susceptibility to the ISE changes across developmental stages. Specifically, it addresses whether children are more vulnerable to the disruptive effects of background sounds. While extensive research has focused on the ISE in adults,

2 Empirical Investigations

there is limited understanding of how it affects other age groups, particularly children. Furthermore, the existing findings on the ISE in children are inconclusive. Investigating children's serial task performance in noisy environments is essential for understanding the development of attention control and short-term memory. Serial recall tasks engage the phonological loop, a key component of working memory responsible for maintaining the order of verbal information. By examining how children perform these tasks amidst background noise, we can reveal developmental changes in their ability to filter out irrelevant auditory stimuli, crucial for attention regulation and memory retention. The studies were based on the premise that if the ISE arises from sound-induced attention capture, then children should be particularly susceptible to disruption due to their developing attention control. This premise aligns with developmental research on cross-modal selective attention, which indicates that children's ability to concentrate on visual categorization tasks in noisy environments progressively improves throughout early and middle childhood, becoming comparable to that of adults around the age of 10 (Gumenyuk et al., 2004; Wetzel et al., 2016, 2018). Understanding how age influences the magnitude of the ISE is critical for both practical and theoretical reasons. By comparing primary school children with young adults, Studies 1, 2, and 4 investigate how different background sounds impact serial order retention, phonological processing, and attention control across age groups.

Research Question III: Does habituation to task-irrelevant background sounds occur over time? The third research question addresses habituation to background sounds, which is relevant for understanding both theoretical concepts and potential applications in minimizing auditory distractions. Within Studies 1, 2, and 4 the role of habituation is examined. By comparing habituation effects across different background sounds and age groups, these studies aim to provide insights into the dynamics of attention over the course of an experiment and its potential role in mitigating the adverse effects of background sounds on serial recall. Only the attentional account posits that the ISE will diminish over time as participants become habituated to task-irrelevant sounds through repeated exposure. Research in this field has yielded inconsistent findings. Some studies have provided evidence for habituation and attention capture, while others have not. This is particularly significant for children, as their attention control is not fully developed. Their heightened susceptibility to surrounding sounds may be associated with attention capture, necessitating an investigation into potential differences in habituation between children and adults.

Research Question IV: How do acoustic properties, such as changing-state and phonological properties, and presentation modes affect the irrelevant sound effect? The fourth research question explores the impact of acoustic properties, such as changing-state and phonological properties, and presentation modes, on serial order retention. By manipulating the complexity of the auditory stimuli, these studies aim to discern how factors such as phonological content (Study 3) and sound presentation (Study 4) influence serial recall.

In Study 3 the mechanisms underlying the ISE were investigated by manipulating the phonological content and changing-state characteristics of the task-irrelevant speech. Specifically, the pseudowords used in Study 1 were transformed to sine wave speech and compared to its' unaltered counterparts. Prior studies (Tremblay et al., 2000) have provided evidence indicating that sinewave speech are less disruptive compared to natural speech. However, it is still unclear whether this decline in performance is a result of phonological interference or distinct attributes of changing state. The acoustical complexity of sinewave speech may be reduced, which could affect its changing-state content and consequently affect its disruptive effect on performance. From the changing-state account perspective, both unaltered changing-state syllables and sinewave-transformed versions should impair serial order retention for verbal and spatial items. However, if phonology-based interference is involved, the disruptive effects should be evident for both steady-state and changing-state sequences, with natural speech causing greater disruption than sinewave speech. This interference is expected to affect only the verbal task, with no impairment anticipated for spatial items. In Study 4, the focus is on how presentation mode of the sounds, specifically the binaural integration of spatially distributed sound sources within a classroom noise scenario, affects attention capture from irrelevant sounds. Many laboratory studies inadequately reflect real-world auditory environments because they utilize monaural sound presentation, thereby oversimplifying the various aspects of binaural hearing in complex acoustic settings. Hence, Study 4 (Exp.2) employs a realistic classroom-noise scenario presented both monaurally and binaurally in a serial recall task of visual-verbal items in both children and adults. Binaural sound presentation is anticipated to exert a stronger impact than monaural presentation for two key reasons: (1) it allows for a clearer separation of speech streams, resulting in increased speech-based interference, and (2) it enhances attention capture due to spatially distributed, shifting sound sources. Moreover, if attention capture is the primary source of disruption, it is expected that children will experience greater impairment than adults, particularly in binaural compared to monaural conditions.

This helps to further clarify the role of attention, as the spatial information provided by the binaural presentation should further divert the attention away.

Research Question V: Which ERP components can be linked to the ISE? The last research question investigates the neural correlates associated with the ISE. To date, research into the underlying neural mechanisms has been limited. Prior EEG studies have explored ERPs linked to auditory distractors, yielding insights into the processes affected by background sounds. However, few studies (Little et al., 2010; Weisz & Schlittmeier, 2006) have investigated how the encoding of the visual-verbal memory material is affected by unfamiliar, task-irrelevant speech. Study 5 examines the ERPs, specifically the N1, P2 and P3a, related to the influence of background speech compared to a control condition on visual-verbal items in a serial recall task. An enhancement in N1 amplitude would reflect higher processing demands. Moreover, this study disentangles the role of P2 and

2 Empirical Investigations

P3a, as the P3a response is particularly significant as it suggests an attention switch to the auditory modality, highlighting how auditory distractions may affect attention allocation. Previous research has linked an enhanced N1 response to the ISE (Campbell et al., 2007), however, the relationship between pre-attentional shifts due to auditory discrimination, as indicated by the MMN, and the P3a, which reflects an orienting response, has not been thoroughly investigated. Thus, Study 6 investigates the potential relationship between ERP components associated with auditory discrimination (MMN), and attentional mechanisms (P3a) in a traditional oddball-task with speech sounds, and the susceptibility to task-irrelevant speech in a visual-verbal serial recall task. The hypothesis posits that participants with high sensitivity to pre-attentive change detection will exhibit greater impairment in serial recall when exposed to changing-state speech compared to those with lower sensitivity. The distinct relationships between susceptibility to task-irrelevant speech and the MMN and P3a ERP components may enhance our understanding of the theoretical concepts of specific interference and attentional capture.

2.2 Study 1: Effects of Steady-State and Changing-State Speech on Verbal and Spatial Recall in Children and Adults

Several studies have shown that task-irrelevant background speech has a negative effect on the ability to serially recall verbal items presented visually. This effect has been attributed to attention capture and specific interference between automatic sound processing and deliberate memory retention processes. However, the precise roles of attentional capture, serial order retention, and phonological processing are still under debate. While research among adults consistently shows a negative impact of background speech on visual-verbal serial recall, limited studies have explored this effect in visual-spatial serial recall. Furthermore, investigating children yielding mixed results.

Study 1 investigates the roles of attention capture, serial order retention, and phonological processing in the ISE. While earlier research on the changing-state effect suggested that background speech equally impairs serial order retention for both verbal and spatial items (Jones et al., 1995, Exp. 4), later studies have shown that background speech does not negatively affect serial order retention of spatial items (Bergström et al., 2012; Klatte & Hellbrück, 1997; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024). To further explore this, we designed tasks requiring participants to reconstruct the serial order of verbal items (words presented as pictures) and spatial items (sequences of dot locations on a screen). The background sounds used were steady-state speech (repeating the syllable “ba”) and changing-state speech (e.g., “beh-tie-peh-buh-toh-gie-tuh-keh-pah” with pauses between syllables). Given the potential impact of developmental changes and children’s higher susceptibility to task-irrelevant sounds due to immature attention control, both primary school children and adults were included in the study. Based on the theoretical background the following predictions arise: According to the attention-capture account (unitary view), the ISE should occur in both verbal and spatial tasks, be more pronounced in children due to their less developed attention control, and gradually diminish with exposure to changing-state speech as habituation occurs. In contrast, the changing-state account predicts that the ISE will appear in both verbal and spatial tasks for both age groups, with the disruptive effect of changing-state speech remaining constant throughout the tasks. Finally, the phonological loop account suggests that the ISE should only affect the verbal task, leaving the spatial task unaffected in both children and adults.

Study 1, being part of this thesis, is presented as an original manuscript (see Leist et al., 2023) of an article published by Springer Nature in *Memory & Cognition* on 3 October 2022, available at: <https://doi.org/10.3758/s13421-022-01359-2>.



Irrelevant speech impairs serial recall of verbal but not spatial items in children and adults

Larissa Leist¹ · Thomas Lachmann^{1,2} · Sabine J. Schlittmeier³ · Markus Georgi³ · Maria Klatte¹

Accepted: 7 September 2022 / Published online: 3 October 2022
© The Author(s) 2022

Abstract

Immediate serial recall of visually presented items is reliably impaired by task-irrelevant speech that the participants are instructed to ignore (“irrelevant speech effect,” ISE). The ISE is stronger with changing speech tokens (words or syllables) when compared to repetitions of single tokens (“changing-state effect,” CSE). These phenomena have been attributed to sound-induced diversions of attention away from the focal task (attention capture account), or to specific interference of obligatory, involuntary sound processing with either the integrity of phonological traces in a phonological short-term store (phonological loop account), or the efficiency of a domain-general rehearsal process employed for serial order retention (changing-state account). Aiming to further explore the role of attention, phonological coding, and serial order retention in the ISE, we analyzed the effects of steady-state and changing-state speech on serial order reconstruction of visually presented verbal and spatial items in children ($n = 81$) and adults ($n = 80$). In the verbal task, both age groups performed worse with changing-state speech (sequences of different syllables) when compared with steady-state speech (one syllable repeated) and silence. Children were more impaired than adults by both speech sounds. In the spatial task, no disruptive effect of irrelevant speech was found in either group. These results indicate that irrelevant speech evokes similarity-based interference, and thus pose difficulties for the attention-capture and the changing-state account of the ISE.

Keywords Visuo-spatial short-term memory · Verbal short-term memory · Serial recall · Irrelevant sound effect · Auditory distraction · Changing-state effect · Children · Attention

Introduction

In everyday life, cognitive tasks are often performed in the presence of irrelevant background sounds such as speech, traffic noise, or music. It is therefore important to know whether and how cognitive processes are affected by task-irrelevant environmental sounds in order to avoid or attenuate possible strain and performance impairments. Consequently, numerous studies have been conducted over the recent decades to examine the characteristics of sounds, tasks, and the

individuals exposed that determine the effects of sounds on mental performance (for reviews, see Klatte et al., 2013; Schlittmeier & Marsh, 2021; Szalma & Hancock, 2011).

The current study focuses on the cross-modal effects of irrelevant sounds, i.e., effects of irrelevant sound on tasks that require processing of visually presented information. According to the *duplex-mechanism framework* proposed by Hughes et al. (2007), such effects may result from two separable mechanisms. On the one hand, attention may be diverted away from the focal task and towards the sound (attentional capture). A well-studied manifestation of sound-induced attention capture is the auditory deviation effect, i.e., the decrement in visual task performance due to an unexpected change in the auditory environment (e.g., a spoken letter A embedded in a sequence of Bs: BBBB-BABB). Detrimental effects of auditory deviants have been demonstrated across a range of cognitive tasks, indicating that the auditory deviation effect is not dependent on specific task requirements (for reviews, see Hughes, 2014; Vachon et al., 2017). On the other hand, sound-induced performance

✉ Larissa Leist
lleist@rhrk.uni-kl.de

¹ Cognitive and Developmental Psychology, University of Kaiserslautern, Erwin-Schrödinger-Str. 57, 67663 Kaiserslautern, Germany

² Centro de Investigación Nebrija en Cognición, Facultad de Lenguas y Educación, Universidad Nebrija, Madrid, Spain

³ Teaching and Research Area of Work and Engineering Psychology, RWTH Aachen University, Aachen, Germany

decrements may result from direct interference between specific processes involved in the automatic, obligatory processing of the irrelevant sound, and deliberate processes involved in the execution of the focal task. Specific interference is expected to occur whenever obligatory sound processing depletes processing resources that are relevant for the current task.

Cross-modal effects of irrelevant sounds have been intensively studied in the context of verbal short-term memory. The task used in these studies requires immediate serial recall of sequences of five to nine unrelated items (e.g., digits, letters, words) presented one-by-one on a screen, with a rate of one to two items per second. Performance is appreciably impaired by task-irrelevant speech that the participants are instructed to ignore. This so-called “irrelevant speech effect” (ISE) is reliable even with low-intensity speech sounds (Colle, 1980; Ellermeier & Hellbrück, 1998), and when sound presentation is confined to a rehearsal phase after encoding of the list items (Elliott et al., 2016; Macken, Mosdell, & Jones, 1999b).

Since its discovery by Colle & Welsh, 1976, the ISE has been considered primarily as a manifestation of specific interference. Early studies conducted within the framework of Baddeley’s multi-component model of working memory (WM) focused on speech as irrelevant sound and attributed its disruptive effect on verbal short-term memory to interference between phonological representations of visual and auditory origin in the phonological loop (Salamé & Baddeley, 1982, 1986, 1989). Especially, it was assumed that speech – even if it is task-irrelevant – gains automatic, obligatory access to the phonological store component of the loop, where it then interferes with the phonological representations of the visually presented list items (“interference-by-content”; Jones & Tremblay, 2000). According to this account, phonological recoding of the visually presented items is a precondition for ISE evocation. However, the phonological loop account was unable to explain later findings demonstrating that (i) the disruptive effect of background speech is independent from its phonological similarity to the list items (e.g., Bell et al., 2011), (ii) nonspeech sounds such as tones (Jones & Macken, 1993) and instrumental music (Schlittmeier et al., 2008) also cause disruption, (iii) the ISE is reliable only with tasks that require serial order retention (Hughes et al., 2007), and (iv) the ISE is determined by the “changing-state”-nature of the sound (Jones et al., 1992). Concerning the latter, serial recall performance is especially impaired by irrelevant sounds with a changing-state characteristic, i.e., by auditory streams consisting of distinct auditory-perceptive objects that vary consecutively. For example, sounds consisting of different syllables or tones evoke an ISE, whereas steady-state sounds, for example, babble noise or repetitions of single syllables or tones, have little or no effect (for an overview, see Schlittmeier et al., 2012). The

increase in disruption of serial recall performance evoked by changing-state sounds (e.g., sequences of different syllables) when compared to steady-state sounds (e.g., repetitions of a single syllable) is termed the *changing-state effect* (CSE).

Aiming for a comprehensive account of these findings, Jones and co-authors (Jones et al., 1996, 1999; Jones & Tremblay, 2000) suggested that, for changing-state sounds, the pre-attentive processes of auditory perceptual organization involve seriation of the order of the auditory tokens that make up the irrelevant stream. The resulting order cues interfere with the deliberate seriation processes involved in rehearsing the sequence of memory items, leading to impaired serial recall. In contrast to the phonological loop account, the changing-state account states that it is not the similarity of *contents*, but the similarity of *processes* that leads to disruption (“interference-by-process,” Hughes et al., 2007; Macken, Tremblay, et al., 1999a). In contrast to Baddeley’s multi-component WM model, the changing-state account is based on a non-modular model of serial short-term memory in which events of different origin or modality share a common level of representation (object-oriented episodic memory (O-OER) model, Jones et al., 1996; Jones & Macken, 2018). Rather than postulating domain-specific short-term stores and maintenance mechanisms, the O-OER model follows an embodied approach, and argues that serial short-term memory is a by-product of general-purpose perceptual and motor functions. When serial recall of unconnected items is required (e.g., unrelated words, digits, or spatial positions), these functions are adopted to establish inter-item associations that are necessary for serial order retention. This is achieved by mapping the perceptual input (the to-be-remembered sequence) to a motor output plan (e.g. articulatory in case of verbal items, oculomotor in case of spatial positions), a record of which is then cyclically repeated in support of serial recall. This process is impaired by task-irrelevant changing-state sounds, as they are automatically organized into streams of ordered objects (Bregman, 1994), which then act as competing candidates for motor sequence planning (Hughes & Marsh, 2017; Macken et al., 2015).

According to this view, the disruption of serial recall performance by changing-state sounds is not modulated by the participants’ attention control or focal task engagement. Furthermore, because the changing-state account attributes the disruptive effect to a domain-general rehearsal mechanism, it predicts that irrelevant speech should impair serial recall of both verbal and nonverbal, visuo-spatial items. Data consistent with this assumption of “functional equivalence” were reported by Jones et al. (1995, Exp. 4). They showed that changing-state speech (i.e., sequences of different syllables) evoked comparable impairments in serial order retention of verbal items (visually presented digits) and spatial items (locations of dots on the screen), whereas steady-state

speech (i.e., repetitions of a single syllable) had no effect in either task. This finding is at odds with the classical WM model, which assumes separate storage systems and rehearsal mechanisms for visuo-spatial and verbal materials and thus predicts specific interference of irrelevant speech with only the latter (for a thorough discussion, see Meiser & Klauer, 1999).

Taken together, both the phonological loop and the changing-state account attribute the ISE to specific interference between obligatory sound processing and task demands, but they differ considerably with respect to the nature of the assumed sound-induced interference. Neither of these two accounts explicitly specifies a role of attention in the ISE.

The duplex-mechanism framework has been challenged by other authors who prefer a more parsimonious account, arguing that the ISE – and its main characteristic, the CSE – are simply another manifestation of sound-induced attention capture. The auditory deviation effect and the ISE are thus attributed to one and the same mechanism (Bell et al., 2019a; Cowan, 1995, 1999; Elliott, 2002; Körner et al., 2017, 2019). Specifically, in Cowan's (1995) embedded processes model, working memory is the activated part of long-term memory, and items are kept active in working memory through attentional refreshing. Attentional refreshing is a domain-general maintenance mechanism that allows multimodal representations to be refreshed by bringing them cyclically into the focus of attention. This process is impaired by involuntary diversions of attention towards the task-irrelevant sound.

Different experimental strategies have been used in order to test the contrasting accounts. In a series of studies, Hughes et al. (2005, 2007, 2013) provided evidence for the duplex-mechanism account by confirming the predicted dissociation between the auditory deviation effect and the CSE. In these studies, the CSE is considered instead of the ISE, since the latter may result from the combined action of both mechanisms, i.e., a steady-state effect indexing attention capture (Bell et al., 2019b), and a changing-state effect indexing interference-by-process (AuBuchon et al., 2019). The findings revealed independent, additive effects of changing-state speech and unexpected voice deviants on verbal serial recall performance (Hughes et al., 2007), as well as differential effects of moderating variables: The CSE, but not the deviation effect, was confined to a task requiring serial order retention (Hughes et al., 2007), whereas the deviation effect, but not the CSE, was confined to the encoding phase of the task and was moderated by manipulations of the participants' attention control, i.e., by providing warnings on the upcoming irrelevant sound, by focal task engagement, and by individual working memory capacity (WMC) (Hughes et al., 2013; Sörqvist, 2010). Other studies indicated dissociable effects of habituation to the auditory deviation effect and the CSE. While habituation to the

deviation effect has consistently been reported (for review, see Sörqvist, 2010), little evidence was found for habituation to the disruptive effects of changing-state speech on serial recall (Ellermeier & Zimmer, 1997; Hellbrück et al., 1996; Röer et al., 2011; Tremblay & Jones, 1998).

However, more recent attempts to replicate these findings failed. For example, Körner et al. (2019) found that the deviation effect and the CSE were equally affected by the timing of the auditory distractors in the experimental task. Both effects were most pronounced when the sounds were presented in the second half of the encoding phase, presumably because of the necessity to coordinate rehearsal of already presented items with encoding of the subsequent ones. Other studies found evidence for habituation to changing-state sounds (Pelletier et al., 2016; Röer et al., 2014) and equivalent habituation rates for auditory deviants and changing-state speech (Röer, Bell, Marsh, & Buchner, 2015b). Furthermore, Röer, Bell, and Buchner (2015a) found that the impairment evoked by changing-state speech was attenuated when specific foreknowledge on the upcoming sound was provided and, in two studies with considerable statistical power, the auditory deviation effect and the CSE were equally uncorrelated to the participants' attention control, measured by WMC (Körner et al., 2017; Röer, Bell, Marsh, & Buchner, 2015b). These findings are at odds with the duplex-mechanism account and are in line with the unitary view that both the CSE and the deviation effect result from a capture of attention.

Another strategy for exploring the role of attention in the ISE has been pursued in studies of potential differences in the ISE between children and adults (Elliott, 2002; Elliott et al., 2016; Elliott & Briganti, 2012; Joseph et al., 2018; Klatter et al., 2010; Meinhardt-Injac et al., 2022; Röer et al., 2018). The rationale behind these studies was that, if the ISE results from sound-induced attention capture, children should be especially prone to disruption due to their underdeveloped attention control. This argument seems reasonable in view of developmental studies on cross-modal selective attention that show that children's ability to focus on visual categorization tasks in the presence of irrelevant sounds improves continuously during early and middle childhood and is adult-like at about age 10 years (Gumenyuk et al., 2004; Wetzel et al., 2016, 2019). However, the findings concerning developmental change in the effects of irrelevant sounds on serial short-term memory are inconsistent. Some studies showed significantly greater impairments due to changing-state speech in 7- to 10-year-old children when compared to young adults (Elliott, 2002; Elliott & Briganti, 2012, Exp. 3; Elliott et al., 2016, Exp. 1), whereas others found equivalent effects across age groups (Elliott et al., 2016, Exp. 2; Joseph et al., 2018; Klatter et al., 2010; Meinhardt-Injac et al., 2022; Röer et al., 2018; Schwarz et al., 2015). Furthermore, as Elliott et al. (2016) point out, an enhanced ISE in children does not unequivocally confirm the

attention capture account, but may also result from increased susceptibility to interference-by-process due to children's immature, less robust rehearsal skills.

Research interest

In the current study, we aimed to investigate the roles of attention capture, serial order retention, and phonological processing in the ISE. To this end, we tested the effects of task-irrelevant steady-state and changing-state speech on serial order reconstruction of verbal and spatial items in children and adults. We recruited children aged about 9 years, as we may assume that children of this age use verbal rehearsal when memorizing verbal items (Elliott et al., 2021; Lehmann & Hasselhorn, 2012). As memory items in the verbal and spatial task we used words presented pictorially and locations of dots on the screen, respectively.

Whether or not serial recall of spatial items is impaired by irrelevant speech is still unclear. In view of its theoretical importance, the number of studies addressing this question is surprisingly small. As stated above, Jones et al. (1995) reported equivalent effects of changing-state speech on serial recall of verbal and spatial items. However, later attempts to replicate the findings reported in Jones et al. (1995) yielded inconsistent results (for a partial replication, see Tremblay et al., 2012; for contradicting evidence, see Bergström et al., 2012; Klatte & Hellbrück, 1997; Kvetnaya, 2018).

The predictions for the current study are straightforward. If the ISE results solely from attention capture (unitary view), it should occur irrespective of item type (verbal vs. spatial), and it should be more pronounced in children due to their immature attention control. Furthermore, the disruptive effect of changing-state speech should diminish over the course of the experiment as a result of habituation. In contrast, if the ISE results from specific interference between order information derived from obligatory sound segmentation and deliberate rehearsal (changing-state account), the effect should be evident in both the verbal and the spatial task in both age groups (provided that the children make use of serial rehearsal in the same way as the adults), and remain constant during the course of the experiment. Finally, if phonological processing is a precondition for ISE evocation (phonological loop account), the ISE should be reliable for the verbal task, whereas the spatial task should be unimpaired in both children and adults.

Methods

Participants

A total of 83 adults¹ (students and staff of the University of Kaiserslautern, 50 females, aged between 19 and 32 years),

and 81 third-grade children from three primary schools in the Kaiserslautern region (44 females, mean age 8 years, 10 months, *SD* 9 months) took part in the study. Following prior studies (Bergström et al., 2012; Jones et al., 1995; Klatte & Hellbrück, 1997), we treated sound condition as the within-subjects factor and task as the between-subjects factor. The verbal task was performed by 43 adults (24 females) with a mean age of 23.5 years (*SD* = 2.63 years), and 30 children (19 females) with a mean age of 9 years, 4 months (*SD* = 4 months). The spatial task was performed by 40 adults (26 females) with a mean age of 24.0 years (*SD* = 2.49 years) and 51 children (29 females), with a mean age of 8 years, 7 months (*SD* = 6 months). All participants were native German speakers and had both normal or corrected-to-normal vision and normal hearing according to either self-reports (adults) or parental reports (children). The study was approved by the Rhineland-Palatinate school authority and by the ethics committee of the German Psychological Society (Deutsche Gesellschaft für Psychologie; DGPs). Informed written consent was provided by the adult participants and by the parents of the children. Adults received either course credit or payment for participation (10€). Children received small gifts (writing materials, e.g., erasers, pencils).

Tasks and materials

Apparatus The experiment was programmed in Python 3.7/ PsychoPy 3.1.5 (Peirce, 2007), and controlled by 15.6-in. laptop computers (ProBook 650 G1). Display resolution was 1,920 × 1,080, with a refresh rate of 60 Hz. Sounds were presented via headphones (children: Sennheiser 210; adults: Sennheiser HD650) with a Focusrite Scarlett 2i2 2nd Generation audio-interface.

Irrelevant sounds Consonant-vowel syllables were produced by a trained female speaker in a sound-attenuated laboratory and recorded as .wav files with a sampling rate of 44.1 kHz and a 16-bit resolution. The set comprised the syllables /ba:/, /be:/, /bi:/, /bo:/, /bu:/, /de:/, /ga:/, /gi:/, /go:/, /gu:/, /ka:/, /ke:/, /ki:/, /ko:/, /pa:/, /pe:/, /pi:/, /pu:/, /ta:/, /ti:/, /to:/, /tu:/ . The syllables were approximately 500 ms in length. In changing-state trials, a random sequence of the syllables was presented with the restriction that adjacent syllables differed in both consonant and vowel. In steady-state trials, the syllable /ba:/ was repeated. A 200-ms silent interval was introduced between adjacent syllables.

Serial order reconstruction tasks In both the verbal and the spatial task, each trial consisted of a presentation phase, a retention interval, and a recall phase. In the presentation phase, items were presented visually one after another with a presentation duration of 1,500 ms and an interstimulus interval of 500 ms. The final list item was followed by a 5,000-ms

retention interval. The onset of the recall phase was signaled by the simultaneous re-presentation of all stimuli. Participants then had to reconstruct the serial order by clicking with the mouse on the items in the order of presentation. Clicking an item changed its shading, indicating that it had been selected. There was no time limit for responding and no possibility of error correction. After selecting the final item, participants were presented with a visual cue to start the next trial by pressing the space bar.

For the spatial task, stimuli were presented as black dots (10-mm diameter) on a white background in a 304 × 183 mm black frame. Dot locations were drawn from a total of 80 possible locations that were arranged in a 8 × 10 grid (not shown to the participants) in the frame (see Fig. 1a). Prior studies with this task (Parmentier, 2011; Parmentier et al., 2005) have shown that the difficulty of reproducing a sequence depends on the complexity of the spatial path formed by the successive dots, and that path complexity results from three parameters: the number of crossings, the path length, and the angular degree. In order to verify equal task difficulty across sound conditions, we first constructed eight lists of five (children) and 12 lists of seven (adults) dot locations. We then produced two parallel versions of each list by horizontally and vertically mirroring the resulting “path” (for illustration, see Fig. 1b–d). With this strategy, locations differed from trial to trial, but the parameters affecting difficulty were constant for each of the three lists.

For the verbal task, colored drawings representing the monosyllabic German words *Bett*, *Bus*, *Eis*, *Frosch*, *Kamm*, *Mond*, *Pilz*, *Schal*, *Schiff*, and *Zaun* (*bed*, *bus*, *ice*, *frog*, *comb*, *moon*, *mushroom*, *scarf*, *ship*, and *fence*) were used for both children and adults. For the adults, the set included the additional items *Brief*, *Haus*, *Herz*, *Hut*, *Nuss*, and *Schwein* (*letter*, *house*, *heart*, *hat*, *nut*, and *pig*). Four lists of five items (drawn out of 10) and six lists of eight items (drawn out of 16) were constructed for the children and adults, respectively. List construction was quasi-random with the restriction that each item occurred in either four

lists (children) or six lists (adults). Random permutations of the list items were used to construct two more versions of each list. The drawings were presented in a rectangular black frame (102 × 73 mm) in the center of the white screen. In the recall phase, the pictures were simultaneously re-presented and randomly positioned in a fixed array of five (children) and eight (adults) black frames.

Pictures were used instead of written words to rule out potential impacts of children’s reading ability on performance. Prior studies using pictorial presentation proved detrimental effects of irrelevant speech in children and adults (Klatte et al., 2010), and significant effects of phonological similarity and length of the pictures’ verbal labels in 8-year-old children (Poloczek et al., 2019; Steinbrink & Klatte, 2008), confirming the use of phonological coding and rehearsal. In adults, word length and phonological similarity affect serial order reconstruction of written words and pictorially presented words to the same degree (Schiano & Watkins, 1981), confirming that pictorial presentation does not alter participants’ strategies.

Procedure

Adults were tested in groups of two to four in a sound-attenuated booth at the University of Kaiserslautern. Children were tested in groups of two to four in a large lecture room in their school. Four computer workplaces were arranged in the room, with a distance of about 10 m between them. Adults received written instruction. Each child was instructed by a researcher or trained student assistant that sat next to the child. In both age groups, the steady- and changing-state sounds were played for 4 s each at the beginning of the experiment, followed by one practice trial per sound condition. All pictures used in the verbal tasks were presented and named by a female speaker. Participants were informed that they should ignore the sounds and focus solely on the serial order reconstruction task.

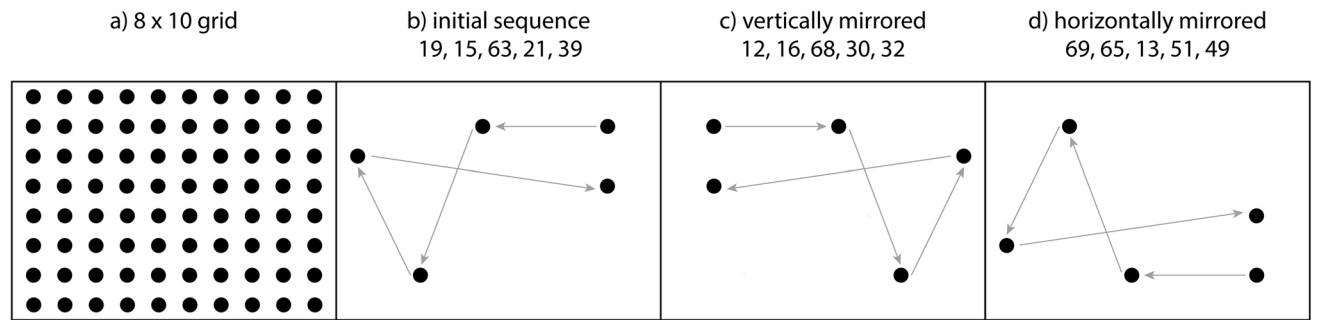


Fig. 1 Illustration of the construction of sequences for the spatial task. **a** Total set of locations in an 8 × 10 grid. **b–d** Exemplary five-item sequence with two parallel versions derived by vertical and horizontal mirroring

For both tasks, the three versions of each list were randomly assigned to the sound conditions. Children performed a total of 24 experimental trials, eight per sound condition. Adults completed 16 trials for each sound condition, for a total of 48 experimental trials.

In both groups, sound conditions varied from trial to trial and were quasi-randomized: All sound conditions were presented in random order before being randomized again. In the steady-state and changing-state conditions, the sound started when the participant initiated the trial with the space bar. A random interval between 1,200 and 1,800 ms was introduced before the presentation of the first list item, to avoid correlations between sound and item onset (e.g., Bell et al., 2010). The sound was played throughout both the item presentation and retention interval and stopped at the beginning of the order reconstruction phase. The average sound level (LAeq) of the syllable sequences was 62 dB, as measured by an artificial ear (Brüel & Kjær, Type 4153). The test session lasted about 25 min for children and 35 min for adults.

Results

All analyses were based on scores representing the proportion of items recalled at the correct serial position. Proportion correct scores with respect to age group and sound condition are depicted in Fig. 2. Performance in the silent control condition did not differ between age groups in the verbal task (adults: $M = 0.70$, $SD = 0.15$, children: $M = 0.64$, $SD = 0.16$), $t(71) = 1.61$, $p = .11$, but was higher in the adults in the spatial task (adults: $M = 0.70$, $SD = 0.14$, children: $M = 0.60$, $SD = 0.19$), $t(89) = 2.76$, $p < .01$. Performance in the silent control conditions did not differ between tasks, for either the adults or for the children ($t < 1$ in both groups). Thus, a potential age difference in the ISE on the verbal task cannot be attributed to differences in task

difficulty between age groups, and a potential difference in the ISE between tasks cannot be attributed to differences in task difficulty within age groups.

The data were analyzed using a $3 \times 2 \times 2$ mixed-design ANOVA with sound condition (silence, steady-state, changing-state) as a within-subject factor and age group (children, adults) and task (verbal, spatial) as between-subjects factors. Greenhouse-Geisser corrections were performed in case of violations of the sphericity assumption according to Mauchly's test. Power calculations were conducted using G*Power (Faul et al., 2007). The analysis revealed significant main effects of sound condition, $F(2, 320) = 19.65$, $p < .001$, partial $\eta^2 = .11$, and age group, $F(1, 160) = 18.46$, $p < .001$, partial $\eta^2 = .10$. The main effect of task and the task $\times$ age interaction were not significant, $F(1, 160) = 1.65$, $p = .20$, and $F(1, 160) < 1$, respectively. Importantly, there was a significant interaction of task $\times$ sound condition, $F(2, 320) = 15.30$, $p < .001$, partial $\eta^2 = .09$, reflecting a stronger effect of sound on the verbal when compared to the spatial task (see Fig. 2). The sound condition $\times$ age group interaction approached significance, $F(2, 320) = 2.50$, $p = .08$, partial $\eta^2 = .02$, and there was a significant three-way interaction, $F(2, 320) = 4.53$, $p < .05$, partial $\eta^2 = .03$. In order to explore the significant interactions, separate analyses were performed for the verbal and spatial task.

Verbal task

The mixed-design ANOVA revealed significant main effects of sound condition, $F(2, 142) = 38.15$, $p < .001$, partial $\eta^2 = .35$, and age group, $F(1, 71) = 13.19$, $p < .001$, partial $\eta^2 = .16$, due to better overall performance in the adults. Furthermore, there was a significant sound $\times$ age interaction, $F(2, 142) = 6.98$, $p < .01$, partial $\eta^2 = .09$, reflecting stronger sound-induced impairments in the children when compared to the adults. Figure 2 indicates that, in children, both changing-state and steady-state speech evoked

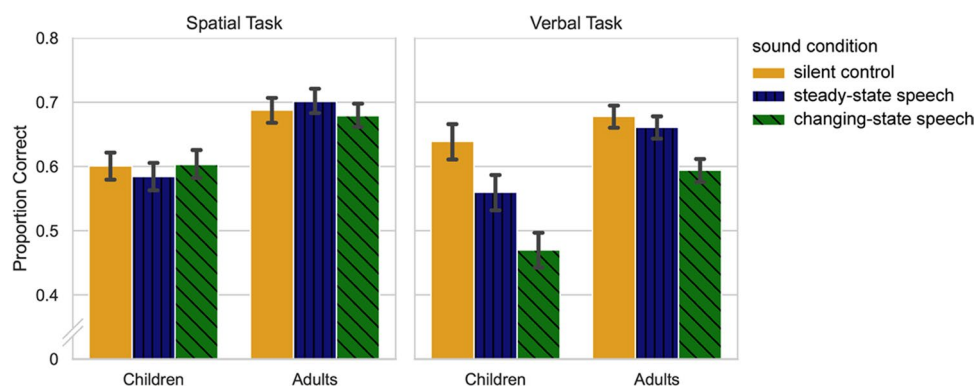


Fig. 2 Mean proportion correct scores with respect to task, age group, and sound condition. Error bars denote the bootstrapped confidence intervals

a stronger impairment relative to quiet, whereas the difference between the steady-state and the changing-state conditions (i.e., the CSE) is comparable between the age groups. This impression was confirmed statistically: When the silent control condition was dropped from the analysis, the age group $\times$ sound interaction proved non-significant, $F(1, 71) = 2.50$, $p = .12$. A power analysis confirmed that, with a total size of $N = 73$, $\alpha = .05$, $1 - \beta = .80$, and a correlation between repeated measurements of $r = .70$ (calculated from the data), an age $\times$ sound interaction of effect size $f = .12$ could be detected.

The sound $\times$ age interaction was further explored by separate analyses in both age groups, which confirmed a significant effect of sound condition in the adults, $F(2, 84) = 13.65$, $p < .001$, partial $\eta^2 = .25$, and in the children, $F(2, 58) = 20.77$, $p < .001$, partial $\eta^2 = .42$. Bonferroni-corrected post hoc tests revealed that, in the adults, performance in the changing-state condition was lower when compared to steady-state and silence (both $ps < .001$), which did not differ ($p = .93$). In the children, performance with changing-state speech was lower when compared to silence and steady-state speech (both $ps < .001$), and performance in the steady-state condition was lower when compared to silence ($p < .05$). Thus, a steady-state effect was evident in the children, but not in the adults.

In a study with 7- to 8-year-old children, AuBuchon et al. (2019) found that the ISE increased with short-term memory span assessed through a digit span test performed in silence. Especially, children's digit span explained 29% of variance in the disruption evoked by changing-state speech. This finding is not in line with the view that underdeveloped rehearsal skills make children especially vulnerable to the ISE. In contrast, the ISE in children around 7 years of age seems to *increase* with their rehearsal abilities. Respective studies

with adults found no evidence for an association between short-term memory capacity and the ISE (Ellermeier & Zimmer, 1997; Elliott & Cowan, 2005, Exp. 1; Elliott et al., 2020), with the exception of Elliott and Cowan (2005, Exp. 2), who reported a significant correlation. Following these studies, we used serial recall performance in quiet as an estimate of individual short-term memory capacity, and difference scores (performance in quiet – performance with changing-state speech) as a measure of ISE (AuBuchon et al., 2019; Ellermeier & Zimmer, 1997). Both measures were normally distributed in children and adults. Replicating AuBuchon et al. (2019), we found that high-performing children exhibited stronger disruption when compared to low performers ($r = .63$, $p < .001$). A similar but weaker association was evident in the adults ($r = .35$, $p = .05$).

Addressing the role of habituation in the ISE, a further analysis was run to find out whether the impairment decreases over the course of the experiment. For this analysis, the data of the adults were trimmed to the first eight trials per sound condition, in order to match the number of trials included between age groups. For each sound condition, proportion correct scores were calculated for two consecutive blocks of four trials each. Means and standard errors are depicted in Fig. 3. A 3 (sound condition) $\times$ 2 (block) $\times$ 2 (age group) ANOVA confirmed significant main effects of sound condition, $F(1, 71) = 29.76$, $p < .001$, partial $\eta^2 = .30$ and age group, $F(1, 71) = 7.50$, $p = .01$, partial $\eta^2 = .10$, and a significant sound condition $\times$ age interaction, $F(1, 71) = 3.54$, $p < .05$, partial $\eta^2 = .05$. The main effect of block approached significance, $F(1, 71) = 3.64$, $p = .06$, partial $\eta^2 = .05$, but the sound condition $\times$ block interaction, the block $\times$ age interaction, and the three-way interaction were not significant (all $F_s < 1$). The lack of a sound $\times$ block-interaction was confirmed in a further analysis on the full

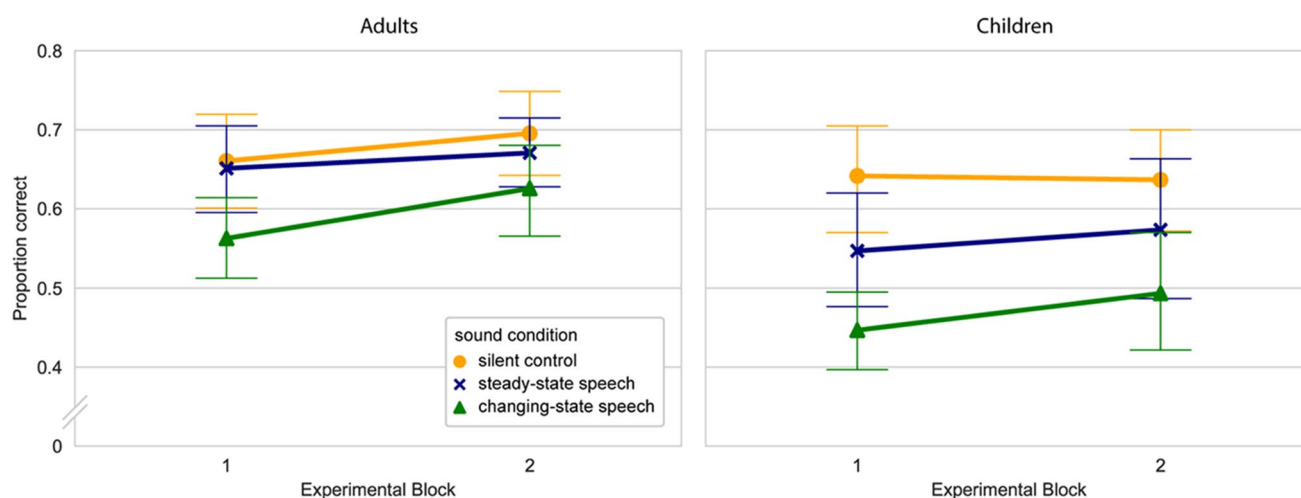


Fig. 3 Mean proportion correct scores in the verbal task in two consecutive blocks of four trials each. Error bars denote the bootstrapped confidence intervals

data from the adults, using four consecutive blocks of four trials each. This analysis yielded significant main effects of sound condition, $F(2, 84) = 13.65$, $p < .001$, partial $\eta^2 = 0.25$, and block, $F(3, 126) = 7.84$, $p < .001$, partial $\eta^2 = .16$ but no interaction, $F(6, 252) < 1$. Thus, there was no sign of habituation to irrelevant speech found, neither in the children nor in the adults.

Spatial task

The mixed-design ANOVA revealed a significant main effect of age group, $F(1, 89) = 7.30$, $p < .01$, partial $\eta^2 = .08$, resulting from better overall performance in the adults. The main effect of sound condition and the sound $\times$ age interaction were not significant ($F(2, 178) < 1$ in both cases). A power analysis confirmed that, with a total size of $n = 91$, $\alpha = .05$, $1 - \beta = .80$ and a correlation between repeated measurements of $r = .75$ (estimated from the data), a main effect of sound condition and an age $\times$ sound - interaction of effect size $f = 0.09$ could be detected.

In summary, in the spatial task, there was no evidence for a disruptive effect of changing-state speech when compared to silence, and no evidence for a changing-state effect (i.e., better performance in steady-state when compared to changing-state speech) for either age group.

Discussion

With the aim of further exploring the role of attention capture, serial order retention, and phonological coding in the ISE, the effects of steady-state and changing-state irrelevant speech (CV-syllables) on serial recall of verbal and spatial items were assessed in third-grade children and adults. The tasks required serial order reconstruction of words presented pictorially (verbal task) and locations of dots on the screen (spatial task).

In the verbal task, both age groups exhibited performance decrements under changing-state speech when compared to silence (ISE), and when compared to steady-state speech (CSE). The ISE was stronger in children, but the magnitude of the CSE did not differ between age groups. Only the children were significantly impaired by steady-state speech. In both age groups, the ISE increased significantly with increasing performance in the silent control condition. The disruption evoked by changing-state speech did not diminish over the course of the experiment, neither in the children nor in the adults. In the spatial task, no evidence for a disruptive effect of irrelevant speech was found, in either the children or the adults.

The immunity of spatial serial recall to disruption through changing-state speech is difficult to reconcile with both the changing-state account and the attentional account of

the ISE. According to the changing-state account, the ISE results from pre-attentive processing of order in changing-state sounds, which interferes with the deliberate serial rehearsal of the to-be-remembered sequence. The changing-state account is based on a unitary model of serial short-term memory, assuming that sequences of unrelated items – irrespective of their sensory modality or representational code – share a common maintenance mechanism, i.e., generating and cyclically repeating a record of a motoric reconstruction of the to-be-remembered list. Accordingly, the ISE and the CSE should occur irrespective of item modality (e.g., visual, auditory) or code (e.g., verbal, spatial). Supporting evidence for the latter was provided in Jones et al. (1995, Exp. 4), who reported equivalent effects of irrelevant speech on serial recall of verbal and spatial items. In both tasks, performance was impaired by changing-state speech, but not by steady-state speech. This finding is crucial for the changing-state account of the ISE. However, in the current study, no detrimental effect of changing-state speech was found in the spatial task, whereas in the verbal task, the ISE and CSE were clearly replicated. Prior studies also reported null effects of changing-state speech on spatial serial recall (Bergström et al., 2012; Klatte & Hellbrück, 1997; Kvetnaya, 2018). These failures to replicate the Jones et al. (1995, Exp. 4) finding cannot be attributed to methodological differences or a lack of statistical power, as the tasks and sounds were rather similar, and the numbers of participants were higher when compared to the original study. We cannot provide a clear explanation for the non-replicability of the original finding, but we may conclude that, if an ISE on spatial serial recall exists at all, it differs vastly from the ISE on verbal tasks with respect to robustness and replicability. This is clearly not in line with the claim of “functional equivalence.”

The attention-capture account attributes the ISE to an impairment of the attentional refreshing of the list items due to involuntary diversions of attention towards the task-irrelevant auditory stimuli. Attentional refreshing means that multimodal representations are kept active in working memory by repeatedly drawing them into the focus of attention (Cowan, 1995). As attentional refreshing is a domain-general maintenance mechanism, the attention-capture account predicts an ISE for both the verbal and the spatial task. Furthermore, children should be more impaired than adults because of their immature attention control. In fact, according to the attentional capture account, the sound-induced impairment should be even greater in the spatial when compared to the verbal task in both age groups. There is considerable evidence that visuo-spatial short-term memory tasks – including those requiring serial order retention – rely more than their verbal equivalents on domain-general attentional resources in adults (Morey & Miron, 2016; for review, see Morey, 2018) and children (Alloway et al., 2006; Campos et al., 2013; Michalczyk et al., 2013). Thus, visuo-spatial

tasks should be reliably disrupted when attentional resources are bound by task-irrelevant sounds. In line with this, it has been shown that serial spatial recall is susceptible to attention capture through auditory deviants (Morey & Miron, 2016). The null effect of irrelevant speech found in the current study adds to the evidence that the ISE and the auditory deviation effect are dissociable. A further problem for the attentional account is the absence of habituation to the ISE in the verbal task.

In regard to developmental effects, the current study replicated the disruptive impact of changing-state speech on serial recall of verbal items in both age groups, the children were more affected than the adults, and only the children were significantly impaired by steady-state speech. Similar results have been reported by Elliott (2002), Elliott and Briganti (2012, Exp. 3), and Elliott et al. (2016). In the latter study, in addition to serial recall, a further task was included that did not require serial order retention (missing item task, Hughes et al., 2007). The effect of changing-state speech on serial recall was more pronounced in children (Exp. 1), and children but not adults were significantly impaired by both steady-state and changing-state speech in the non-serial task (Exp. 2). The authors discussed these age effects in the framework of the duplex mechanism account, arguing that children and adults are equally prone to interference-by-process, but children are more susceptible to attentional diversions by task-irrelevant sounds, regardless of whether or not they change over time.

Thus, in these studies (Elliott, 2002; Elliott & Briganti, 2012, Exp. 3; Elliott et al., 2016), the enhanced effect of irrelevant speech on serial recall in children has been attributed to children's increased susceptibility to sound-induced attention capture. This argument is difficult to reconcile with the current finding that serial recall of spatial items was unimpaired by irrelevant speech in children and adults. If irrelevant speech captures children's attention, it should impair performance for both verbal and spatial items.

Taken together, the finding that irrelevant speech impairs verbal, but not spatial serial recall contradicts the assumption that the ISE results from interference with a domain-general maintenance mechanism, i.e., attentional refreshing (as proposed by the attention-capture account), or serial rehearsal through perceptual-motor sequence planning (as proposed by the changing-state account). Instead, the current results indicate that irrelevant speech interferes with mechanisms or memory traces that are *specific* to the verbal domain, and that children are more susceptible to this kind of interference than adults. Potential candidates as points of attack for irrelevant speech are therefore the efficiency of articulatory rehearsal, and/or the integrity of phonological representations.

Articulatory rehearsal is the repetitive subvocal pronunciations of verbal list items in a sequential manner, using

articulatory motor programs that are also involved in language production. Articulatory rehearsal is regarded as the dominant maintenance strategy in verbal serial recall tasks, and is integrated in modular models of working memory as a mechanism specifically dedicated to the prevention of time-based decay of phonological representations within a phonological store (e.g., Baddeley, 2012; Camos et al., 2009). In view of the evidence that the ISE is most pronounced with verbal tasks that require serial order retention (Beaman & Jones, 1997; Elliott et al., 2016; for review, see Hughes et al., 2007), or at least suggest a serial rehearsal strategy (Hughes & Marsh, 2020), articulatory rehearsal seems to be a promising route of entry for irrelevant speech. However, explaining the ISE through specific interference with articulatory rehearsal is not without problems. First, there is evidence that cumulative articulatory rehearsal does not improve serial recall performance in adults (Souza & Oberauer, 2018), and that children below the age of 10 years employ various strategies in visual-verbal serial recall tasks (Koppenol-Gonzalez et al., 2014), with cumulative rehearsal being not the dominant one (Poloczec et al., 2019). It is difficult to explain how the disruption of a strategy that is inefficient or rarely used should lead to a performance decrement. Second, in line with this argument, preventing articulatory rehearsal through articulatory suppression (i.e., continuously uttering an irrelevant syllable) during the retention interval (Toppino & Pisegna, 2005), or rapid presentation of the visual list items (AuBuchon et al., 2020; Samper et al., 2021) does not diminish the magnitude of the ISE. Third, the rehearsal account of the ISE is mainly based on findings that non-serial tasks are not or less affected by irrelevant speech. However, the non-serial tasks used in the respective studies, i.e., free recall (Beaman & Jones, 1998; Salamé & Baddeley, 1990) and the missing item task (e.g., Beaman & Jones, 1997; Elliott et al., 2016; Hughes et al., 2007), rely heavily on semantic strategies and thus differ from serial recall not only with respect to serial order retention, but also with respect to the role of phonological processing. It has been shown that working memory tasks that do not involve serial rehearsal (or at least strongly discourage its use), but require phonological processing, i.e., the maintenance, analysis, or manipulation of phonological representations, are reliably affected by irrelevant speech in children (Klatte et al., 2007) and adults (Klatte et al., 2019; Samper et al., 2021; Smith et al., 1995). These findings open the possibility that it is not the serial order component, but the reliance on phonological representations that makes verbal serial recall vulnerable to disruption by irrelevant speech.

Attributing the ISE to a corruption of phonological traces might also explain the differences between children and adults with respect to the magnitude of the ISE, and its relation to overall serial recall performance. Children might be more affected by irrelevant speech than adults because their

phonological representations are less well specified (Claessen et al., 2009; Hazan & Barrett, 2000; Metsala & Walley, 1998), and therefore more vulnerable to disruption. The correlation of performance in the silent control condition and the ISE magnitude might be explained by assuming that high performers rely more than low performers on phonological strategies (Poloczek et al., 2019), making them more prone to disruption by irrelevant speech.

The view that irrelevant speech evokes “interference-by-content” in that it impairs phonological codes has originally been taken by the phonological loop account of the ISE (Salamé & Baddeley, 1982), and was later adopted and mathematically specified in the feature model of working memory (Neath, 2000). In the latter, items are represented by vectors of modality-dependent and modality-independent features. Modality-dependent features are generated during encoding and represent raw, physical aspects of the stimuli. Modality-independent features are produced through internal processes such as identification and categorization. The detrimental effect of irrelevant speech is modelled by replacing modality-independent features of the list items by similar modality-independent features of the irrelevant speech (“feature adoption”). These accounts were more or less abandoned, mainly because of studies showing that the overlap of phonemes in the irrelevant speech tokens and the list items (“between-stream-similarity”) does not affect the magnitude of the disruption (Bell et al., 2011; Hughes et al., 2005; Larsen et al., 2000; LeCompte & Shaibe, 1997), and that non-speech sounds with a changing-state characteristic, such as tone sequences or instrumental music, also disrupt verbal serial recall (Jones & Macken, 1993; Schlittmeier et al., 2008).

However, to reject the assumption of content-based interference might have been premature, given the evidence that supports a role of similarity in the disruption of serial recall performance by irrelevant sounds. Specifically, it has been shown that speech or speech-like sounds are more disruptive to verbal serial recall than non-speech sounds (Schlittmeier et al., 2012), even with acoustic complexity controlled (Dorsi et al., 2018; Viswanathan et al., 2014), that the magnitude of the ISE depends on the overlap of phonetic features (though not of the phonemes) in the irrelevant speech tokens and the list items (Eagan & Chein, 2012), and that speech interferes more with the serial order retention of verbal when compared to tonal items, whereas irrelevant tones interfere more with the retention of tonal when compared to verbal sequences (Defilippi et al., 2019; Kattner & Meinhardt, 2020; Williamson et al., 2010). These findings suggest an important role of between-stream-similarity in the disruptive effects of irrelevant sounds, and thus content-based interference.

To summarize, the current study showed that the disruptive effect of irrelevant speech on serial recall performance

is confined to the verbal domain, and is more pronounced in children compared to adults. These results are difficult to explain with the dominant accounts of the ISE, i.e., specific interference with serial order retention, or attention capture. We argue that phonological processing, but not serial order retention is a precondition for ISE evocation, and that irrelevant speech impairs verbal serial recall and other tasks that rely on phonological processing by corrupting phonological traces in working memory.

This study is limited, as it does not include nonspeech sounds nor phonological processing tasks that do not require serial order retention. Furthermore, the current findings do not allow a clear specification of the mechanism that evokes the proposed interference with phonological codes. Feature-based models of working memory may provide a parsimonious account of similarity-based interference effects, as they are able to explain specific interference without postulating specialized working memory modules (Morey, 2018). It has already been shown that Nairne’s feature model (Nairne, 1990; Neath, 2000) successfully simulates specific interference through verbal and spatial secondary tasks on verbal and spatial working memory (Poirier et al., 2019). As pointed out by Eagan and Chein (2012), with some modifications, the feature model could also account for the effects of irrelevant sounds, and their dependencies on participants’ individual strategies.

Acknowledgements We wish to thank the children for participating, and their teachers for support and organizational help. Many thanks to Kirsten Graewe, Maxim Fuchs, Julia Lange, and Halaa Arif for help in data acquisition and Markus Kurtz for technical support.

Funding Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; SCHL 1876/4, KL 1305/4). Sabine Schlittmeier’s contribution was supported by a grant from the HEAD-Genuit-Stiftung (Head-Genuit Foundation; P-16/10-W).

Data availability The data and materials are available at: <https://osf.io/pyqvr/>

Declarations

In the first version of this manuscript, we reported the data from a different sample of adults ($N = 80$). These participants were tested with two further sound conditions, i.e., five sound conditions in total. During the review process, it became clear that the imbalance in sound conditions makes valid comparisons between the two age groups difficult. Therefore, we recruited a new sample of adults ($N = 83$) in early 2022, and used only the same three sound conditions that were presented to the children. The pattern of results was perfectly replicated with the new sample. Here, only the data from the new sample are reported.

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Deutsche Gesellschaft für Psychologie (DGPs, German Psychological Society). For the schoolchildren, study approval was also

obtained from the Rhineland-Palatinate school authority (Aufsichts- und Dienstleistungsdirektion, ADD).

Disclosure statement The authors have no competing interests to declare that are relevant to the content of this article.

Consent to participants Informed written consent was provided by the adult participants and by the parents of the children.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77(6), 1698–1716. <https://doi.org/10.1111/j.1467-8624.2006.00968.x>
- AuBuchon, A. M., McGill, C. I., & Elliott, E. M. (2019). Auditory distraction does more than disrupt rehearsal processes in children's serial recall. *Memory & Cognition*, 47(4), 738–748. <https://doi.org/10.3758/s13421-018-0879-4>
- AuBuchon, A. M., McGill, C. I., & Elliott, E. M. (2020). Decomposing the Role of Rehearsal in Auditory Distraction during Serial Recall. *Auditory Perception & Cognition*, 3(1-2), 18–32. <https://doi.org/10.1080/25742442.2020.1842996>
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Beaman, C., & Jones, D. M. (1997). Role of Serial Order in the Irrelevant Speech Effect: Tests of the Changing-State Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 459–471. <https://doi.org/10.1037/0278-7393.23.2.459>
- Beaman, C. P., & Jones, D. M. (1998). Irrelevant Sound Disrupts Order Information in Free Recall as in Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*, 51(3), 615–636. <https://doi.org/10.1080/17470218.2010.483769>
- Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology*, 47, 1182–1191. <https://doi.org/10.1111/j.1469-8986.2010.01029.x>
- Bell, R., Mund, I., & Buchner, A. (2011). Disruption of short-term memory by distractor speech: Does content matter? *Quarterly Journal of Experimental Psychology*, 64, 146–168. <https://doi.org/10.1080/17470218.2010.483769>
- Bell, R., Röer, J. P., Lang, A.-G., & Buchner, A. (2019a). Reassessing the token set size effect on serial recall: Implications for theories of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(8), 1432–1440. <https://doi.org/10.1037/xlm0000658>
- Bell, R., Röer, J. P., Lang, A. G., & Buchner, A. (2019b). Distraction by steady-state sounds: Evidence for a graded attentional model of auditory distraction. *Journal of Experimental Psychology: Human Perception and Performance*, 45(4), 500–512. <https://doi.org/10.1037/xhp0000623>
- Bergström, K., Lachmann, T., & Klatte, M. (2012). Wann stört Lärm das geistige Arbeiten? Einfluss von Aufgaben- und Geräuschcharakteristiken bei der Wirkung moderaten Lärms auf Arbeitsgedächtnisleistungen. In: *Fortschritte der Akustik. Beiträge zur 38. Jahrestagung für Akustik, DAGA*, Darmstadt. Berlin: DEGA.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.
- Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, 61(3), 457–469. <https://doi.org/10.1016/j.jml.2009.06.002>
- Campos, I. S., Almeida, L. S., Ferreira, A. I., & Martinez, L. F. (2013). Working memory as separable subsystems: A study with Portuguese primary school children. *The Spanish Journal of Psychology*, 16. <https://doi.org/10.1017/sjp.2013.6>
- Claessen, M., Heath, S., Fletcher, J., Hogben, J., & Leitão, S. (2009). Quality of phonological representations: a window into the lexicon? *International Journal of Language & Communication Disorders*, 44(2), 121–144. <https://doi.org/10.1080/13682820801966317>
- Colle, H. A. (1980). Auditory encoding in visual short-term recall: effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behavior*, 19, 722–735. [https://doi.org/10.1016/S0022-5371\(80\)90403-X](https://doi.org/10.1016/S0022-5371(80)90403-X)
- Colle, H. A., & 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. (1995). *Attention and memory: An integrated framework*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195119107.001.0001>
- Cowan, N. (1999). An embedded-process 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/10.1017/CBO9781139174909.006>
- Defilippi, A. C. N., Garcia, R. B., & Galera, C. (2019). Irrelevant sound interference on phonological and tonal working memory in musicians and nonmusicians. *Psicologia: Reflexão e Crítica*, 32. <https://doi.org/10.1186/s41155-018-0114-z>
- Dorsi, J., Viswanathan, N., Rosenblum, L. D., & Dias, J. W. (2018). The role of speech fidelity in the irrelevant sound effect: Insights from noise-vocoded speech backgrounds. *Quarterly Journal of Experimental Psychology*, 71(10), 2152–2161.
- Eagan, D. E., & Chein, J. M. (2012). Overlap of phonetic features as a determinant of the between-stream phonological similarity effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 473–481. <https://doi.org/10.1037/a0025368>
- Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio and binaural masking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(5), 1406–1414. <https://doi.org/10.1037/0096-1523.24.5.1406>
- Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the 'irrelevant speech effect'. *Journal of the Acoustical Society of America*, 102(4), 2191–2199. <https://doi.org/10.1121/1.419596>
- Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory and Cognition*, 30(3), 478–487. <https://doi.org/10.3758/BF03194948>
- Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, 140(1), 64–74. <https://doi.org/10.1016/j.actpsy.2012.02.009>

- Elliott, E. M., & Cowan, N. (2005). Coherence of the irrelevant-sound effect: individual profiles of short-term memory and susceptibility to task-irrelevant materials. *Memory & Cognition*, 33(4), 664–675. <https://doi.org/10.3758/bf03195333>
- Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., & Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental differences. *Journal of Memory and Language*, 88, 39–50. <https://doi.org/10.1016/j.jml.2015.12.008>
- Elliott, E. M., Marsh, J. E., Zeringue, J., & McGill, C. I. (2020). Are individual differences in auditory processing related to auditory distraction by irrelevant sound? A replication study. *Memory & Cognition*, 48(1), 145–157. <https://doi.org/10.3758/s13421-019-00968-8>
- Elliott, E. M., Morey, C. C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E. J., Attwood, M., Bayram, B., Beeler-Duden, S., Blakstvedt, T. Y., Büttner, G., Castelain, T., Cave, S., Crepaldi, D., Fredriksen, E., Glass, B. A., Graves, A. J., Guitard, D., Hoehl, S., ... Voracek, M. (2021). Multilab Direct Replication of Flavell, Beach, and Chinsky (1966): Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Advances in Methods and Practices in Psychological Science*, 4(2). <https://doi.org/10.1177/25152459211018187>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Gumenyuk, V., Korzyukov, O., Alho, K., Escera, C., & Näätänen, R. (2004). Effects of auditory distraction on electrophysiological brain activity and performance in children aged 8–13 years. *Psychophysiology*, 41(1), 30–36. <https://doi.org/10.1111/1469-8986.00123>
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396. <https://doi.org/10.1006/jpho.2000.0121>
- Hellbrück, J., Kuwano, S., & Namba, S. (1996). Irrelevant background speech and human performance: is there long-term habituation? *Journal of the Acoustical Society of Japan (E)*, 17(5), 239–247. <https://doi.org/10.1250/ast.17.239>
- Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, 3(1), 30–41. <https://doi.org/10.1002/pchj.44>
- Hughes, R. W., & Marsh, J. E. (2017). The functional determinants of short-term memory: Evidence from perceptual-motor interference in verbal serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 537. <https://doi.org/10.1037/xlm0000325>
- Hughes, R. W., & Marsh, J. E. (2020). When is forewarned forearmed? Predicting auditory distraction in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(3), 427–442. <https://doi.org/10.1037/xlm0000736>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 736–749. <https://doi.org/10.1037/0278-7393.31.4.736>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of Short-Term Memory by Changing and Deviant Sounds: Support for a Duplex-Mechanism Account of Auditory Distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6), 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Jones, D. M., & Vachon, F. (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 and Performance*, 39(2), 539–553. <https://doi.org/10.1037/a0029064>
- Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 369–381. <https://doi.org/10.1037/0278-7393.19.2.369>
- Jones, D. M., & Macken, B. (2018). In the beginning was the deed: Verbal short-term memory as object-oriented action. *Current Directions in Psychological Science*, 27(5), 351–356. <https://doi.org/10.1177/0963721418765796>
- Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin and Review*, 7(3), 550–558. <https://doi.org/10.3758/BF03214370>
- Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology Section A*, 44(4), 645–669. <https://doi.org/10.1080/14640749208401304>
- Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 1008–1018. <https://doi.org/10.1037/0278-7393.21.4.1008>
- Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. *Models of Short-Term Memory*, 209–238.
- Jones, D., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 464–473. <https://doi.org/10.1037/0278-7393.25.2.464>
- Joseph, T. N., Hughes, R. W., Sörqvist, P., & Marsh, J. E. (2018). Differences in auditory distraction between adults and children: A duplex-mechanism approach. *Journal of Cognition*, 1(1), 1–11. <https://doi.org/10.5334/joc.15>
- Kattner, F., & Meinhardt, H. (2020). Dissociating the Disruptive Effects of Irrelevant Music and Speech on Serial Recall of Tonal and Verbal Sequences. *Frontiers in Psychology*, 11, 346. <https://doi.org/10.3389/fpsyg.2020.00346>
- Klatte, M., & Hellbrück, J. (1997). Effects of irrelevant speech on serial recall of verbal and spatial materials. In A. Schick & M. Klatte (Eds.), *Contributions to Psychological Acoustics. Results of the 7th Oldenburg Symposium on Psychological Acoustics* (pp. 531–538).
- Klatte, M., Meis, M., Sukowski, H., & Schick, A. (2007). Effects of irrelevant speech and traffic noise on speech perception and cognitive performance in elementary school children. *Noise & Health*, 9(36), 64–74. <https://doi.org/10.4103/1463-1741.36982>
- Klatte, M., Lachmann, T., Schlittmeier, S., & Hellbrück, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology*, 22(8), 1168–1191. <https://doi.org/10.1080/09541440903378250>
- Klatte, M., Bergström, K., & Lachmann, T. (2013). Does noise affect learning? A short review of noise effects on cognitive performance in children. *Frontiers in Developmental Psychology*, 4, 578. <https://doi.org/10.3389/fpsyg.2013.00578>
- Klatte, M., Bergström, K., & Lachmann, T. (2019). Effects of task-irrelevant background speech on verbal working memory tasks. Proceedings of the 23rd International Congress on Acoustics, Aachen, Germany. Retrieved September 22, 2022, from <http://pub.dega-akustik.de/ICA2019/data/articles/001223.pdf>
- Koppenol-Gonzalez, G. V., Bouwmeester, S., & Vermunt, J. K. (2014). Short-term memory development: Differences in serial position curves between age groups and latent classes. *Journal of Experimental Child Psychology*, 126, 138–151. <https://doi.org/10.1016/j.jecp.2014.04.002>

- Körner, U., Röer, J. P., Buchner, A., & Bell, R. (2017). Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. *Journal of Memory and Language*, 96, 122–137. <https://doi.org/10.1016/j.jml.2017.05.005>
- Körner, U., Röer, J. P., Buchner, A., & Bell, R. (2019). Time of presentation affects auditory distraction: Changing-state and deviant sounds disrupt similar working memory processes. *Quarterly Journal of Experimental Psychology*, 72(3), 457–471. <https://doi.org/10.1177/1747021818758239>
- Kvetnaya, T. (2018). Registered Replication Report: Testing Disruptive Effects of Irrelevant Speech on Visual-Spatial Working Memory. *Journal of European Psychology Students*, 9(1), 10–15. <https://doi.org/10.5334/jeps.450>
- Larsen, J. D., Baddeley, A., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory*, 8(3), 145–157. <https://doi.org/10.1080/096582100387579>
- LeCompte, D. C., & Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology Section A*, 50(1), 100–118. <https://doi.org/10.1080/1713755679>
- Lehmann, M., & Hasselhorn, M. (2012). Rehearsal dynamics in elementary school children. *Journal of Experimental Child Psychology*, 111(3), 552–560. <https://doi.org/10.1016/j.jecp.2011.10.013>
- Macken, W., Tremblay, S., Alford, D., & Jones, D. (1999a). Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. *International Journal of Psychology*, 34(5-6), 322–327. <https://doi.org/10.1080/002075999399639>
- Macken, W. J., Mosdell, N., & Jones, D. M. (1999b). Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3), 810–814. <https://doi.org/10.1037/0278-7393.25.3.810>
- Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, 6, 293. <https://doi.org/10.3389/fpsyg.2015.00293>
- Meinhardt-Injac, B., Imhof, M., Wetzel, N., Klatte, M., & Schlittmeier, S. J. (2022). The Irrelevant Sound Effect on serial recall is independent of age and inhibitory control. *Auditory Perception & Cognition*, 5(1-2), 25–45. <https://doi.org/10.1080/25742442.2022.2064692>
- Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1272–1299. <https://doi.org/10.1037/0278-7393.25.5.1272>
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Lawrence Erlbaum Associates Publishers.
- Michalczyk, K., Malstädt, N., Worgt, M., Könen, T., & Hasselhorn, M. (2013). Age differences and measurement invariance of working memory in 5-to 12-year-old children. *European Journal of Psychological Assessment*, 29(3), 220–229. <https://doi.org/10.1027/1015-5759/a000149>
- Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin*, 144(8), 849–883. <https://doi.org/10.1037/bul0000155>
- Morey, C. C., & Miron, M. D. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(12), 1907–1918. <https://doi.org/10.1037/xlm0000280>
- 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>
- Parmentier, F. B. R. (2011). Exploring the determinants of memory for spatial sequences. In A. Vandierendonck & A. Szmalec (Eds.), *Spatial working memory* (pp. 67–86). Psychology Press.
- Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional Information in Spatial Serial Memory: Path Characteristics Affect Recall Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 412–427. <https://doi.org/10.1037/0278-7393.31.3.412>
- Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Pelletier, M.-F., Hodgetts, H. M., Lafleur, M. F., Vincent, A., & Tremblay, S. (2016). Vulnerability to the Irrelevant Sound Effect in Adult ADHD. *Journal of Attention Disorders*, 20(4), 306–316. <https://doi.org/10.1177/1087054713492563>
- Poirier, M., Yearsley, J. M., Saint-Aubin, J., Fortin, C., Gallant, G., & Guillard, D. (2019). Dissociating visuo-spatial and verbal working memory: It's all in the features. *Memory & Cognition*, 47(4), 603–618. <https://doi.org/10.3758/s13421-018-0882-9>
- Poloczek, S., Henry, L. A., Messer, D. J., & Büttner, G. (2019). Do children use different forms of verbal rehearsal in serial picture recall tasks? A multi-method study. *Memory*, 27(6), 758–771. <https://doi.org/10.1080/09658211.2018.1563615>
- 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 and Cognition*, 39(5), 839–850. <https://doi.org/10.3758/s13421-010-0070-z>
- Röer, J. P., Bell, R., & Buchner, A. (2014). Evidence for habituation of the irrelevant-sound effect on serial recall. *Memory & Cognition*, 42(4), 609–621. <https://doi.org/10.3758/s13421-013-0381-y>
- Röer, J. P., Bell, R., & Buchner, A. (2015a). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception and Performance*, 41(3), 692–702. <https://doi.org/10.1037/xhp0000028>
- Röer, J. P., Bell, R., Marsh, J. E., & Buchner, A. (2015b). Age equivalence in auditory distraction by changing and deviant speech sounds. *Psychology and Aging*, 30(4), 849–855. <https://doi.org/10.1037/pag0000055>
- Röer, J. P., Bell, R., Körner, U., & Buchner, A. (2018). Equivalent auditory distraction in children and adults. *Journal of Experimental Child Psychology*, 172, 41–58. <https://doi.org/10.1016/j.jecp.2018.02.005>
- Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21(2), 150–164. [https://doi.org/10.1016/S0022-5371\(82\)90521-7](https://doi.org/10.1016/S0022-5371(82)90521-7)
- Salamé, P., & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, 24(4), 263–265. <https://doi.org/10.3758/BF03330135>
- Salamé, P., & Baddeley, A. (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>
- Salamé, P., & Baddeley, A. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society*, 28(6), 540–542. <https://doi.org/10.3758/BF03334073>

- Samper, J. R., Morrison, A., & Chein, J. (2021). Doubts about the role of rehearsal in the Irrelevant Sound Effect. *Experimental Psychology*, 68(5), 229–242. <https://doi.org/10.1027/1618-3169/a000527>
- Schiano, D. J., & Watkins, M. J. (1981). Speech-like coding of pictures in short-term memory. *Memory & Cognition*, 9, 110–114. <https://doi.org/10.3758/BF03196956>
- Schlittmeier, S.J. & Marsh, J.E. (2021). Review of research on the effects of noise on cognitive performance 2017-2021. Proceedings of the 13th ICBEN Congress on Noise as a Public Health Problem. Retrieved September 22, 2022, from http://www.icben.org/2021/ICBEN%202021%20Papers/full_paper_28062.pdf
- Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008). Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology*, 20(2), 252–271. <https://doi.org/10.1080/09541440701427838>
- Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, & Psychophysics*, 74(1), 194–203. <https://doi.org/10.3758/s13414-011-0230-7>
- Schwarz, H., Schlittmeier, S., Otto, A., Persike, M., Klatte, M., Imhof, M., & Meinhardt-Injac, B. (2015). Age differences in the irrelevant sound effect: A serial recognition paradigm. *Psihologija*, 48(1), 35–43. <https://doi.org/10.2298/PSI1501035S>
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33(11), 1433–1454. [https://doi.org/10.1016/0028-3932\(95\)00074-D](https://doi.org/10.1016/0028-3932(95)00074-D)
- Sörqvist, P. (2010). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory and Cognition*, 38(5), 651–658. <https://doi.org/10.3758/MC.38.5.651>
- Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology*, 107, 1–21. <https://doi.org/10.1016/j.cogpsych.2018.09.002>
- Steinbrink, C., & Klatte, M. (2008). Phonological working memory in German children with poor reading and spelling abilities. *Dyslexia*, 14(4), 271–290. <https://doi.org/10.1002/dys.357>
- Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin*, 137(4), 682–707. <https://doi.org/10.1037/a0023987>
- Toppino, T. C., & Pisegna, A. (2005). Articulatory suppression and the irrelevant speech effect in short-term memory: Does the locus of suppression matter? *Psychonomic Bulletin & Review*, 12(2), 374–379. <https://doi.org/10.3758/BF03196387>
- Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 24(3), 659–671. <https://doi.org/10.1037/0278-7393.24.3.659>
- Tremblay, S., Parmentier, F. B. R., Hodgetts, H. M., Hughes, R. W., & Jones, D. M. (2012). Disruption of verbal-spatial serial memory by extraneous air-traffic speech. *Journal of Applied Research in Memory and Cognition*, 1(2), 73–79. <https://doi.org/10.1016/j.jarmac.2012.04.004>
- Vachon, F., Labonté, K., & Marsh, J. E. (2017). Attentional Capture by Deviant Sounds: A Noncontingent Form of Auditory Distraction? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 43(4), 622–634. <https://doi.org/10.1037/xlm0000330>
- Viswanathan, N., Dorsi, J., & George, S. (2014). The role of speech-specific properties of the background in the irrelevant sound effect. *Quarterly Journal of Experimental Psychology*, 67(3), 581–589. <https://doi.org/10.1080/17470218.2013.821708>
- Wetzel, N., Schröger, E., & Widmann, A. (2016). Distraction by Novel and Pitch-Deviant Sounds in Children. *Frontiers in Psychology*, 7, 1949. <https://doi.org/10.3389/fpsyg.2016.01949>
- Wetzel, N., Scharf, F., & Widmann, A. (2019). Can't Ignore—Distraction by Task-Irrelevant Sounds in Early and Middle Childhood. *Child Development*, 90(6), e819–e830. <https://doi.org/10.1111/cdev.13109>
- Williamson, V. J., Mitchell, T., Hitch, G. J., & Baddeley, A. D. (2010). Musicians' memory for verbal and tonal materials under conditions of irrelevant sound. *Psychology of Music*, 38(3), 331–350. <https://doi.org/10.1177/0305735609351918>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.3 Study 2: Impact of Unfamiliar Speech and Environmental Sounds on Verbal and Spatial Recall Across Age Groups

In Study 1, the impact of changing-state and steady-state speech on verbal and spatial serial order retention in children and adults was investigated. The findings from Study 1 that irrelevant speech impairs verbal but not spatial serial recall contradict the notion that the ISE results from interference with a domain-general maintenance mechanism, such as attentional refreshing (attention capture account) or serial rehearsal through perceptual-motor sequence planning (changing-state account). Instead, the results suggest that irrelevant speech specifically interferes with verbal memory mechanisms, with children being more susceptible than adults. Since Study 1 did not include non-speech stimuli, it remains unclear whether the detrimental impact observed is specific to speech stimuli.

The objective of Study 2 is to investigate the disruptive effects of both speech and non-speech sounds on serial order retention of verbal and spatial items, which are identical to the tasks from Study 1, in children and adults. To achieve this, we used narrative unfamiliar speech and a mix of environmental non-speech sounds, such as animal noises (e.g., dog barking, gackling chickens), as well as traffic and construction sounds (e.g., approaching car, bicycle bell, hammering).

Previous research (Klatte, Lachmann, et al., 2010) shows that non-speech classroom noise disrupts serial recall in children but has little effect on adults, while speech impairs performance across all ages. These findings suggest different mechanisms for attentional distraction and specific interference, which Study 2 aims to explore. We hypothesize that the unpredictability of environmental sounds effectively captures attention, as supported by earlier studies (Gumenyuk et al., 2004; Klatte, Hellbrück, et al., 2010; Wetzel et al., 2021).

We predict that if attention capture plays a key role in the evocation of the ISE, both speech and environmental sounds will cause significant disruption in both tasks, with a greater effect in children than adults, and there should be habituation over time. From an interference-based perspective (Phonological Loop Account), speech should impair performance on the verbal task equally in both groups, with no effect on the spatial task. Environmental sounds are expected to cause a minimal detrimental impact in the verbal task due to the lack of phonological content. There should be no habituation. According to the changing-state account, speech should impair both tasks across ages, while environmental sounds, being a mix of unrelated noises, should cause little disruption, as they do not form a coherent auditory stream. Disruption should remain stable throughout the tasks.

Study 2, being part of this thesis, is presented as an original manuscript (see Leist et al., 2025) of an article published by Springer Nature in *Scientific Reports* on 14 January 2025, available at: <https://doi.org/10.1038/s41598-025-85855-w>.



OPEN Impact of irrelevant speech and non-speech sounds on serial recall of verbal and spatial items in children and adults

Larissa Leist^{1✉}, Thomas Lachmann^{1,2,3} & Maria Klatte¹

Short-term memory for sequences of verbal items such as written words is reliably impaired by task-irrelevant background sounds, a phenomenon known as the “Irrelevant Sound Effect” (ISE). Different theoretical accounts have been proposed to explain the mechanisms underlying the ISE. Some of these assume specific interference between obligatory sound processing and phonological or serial order representations generated during task performance, whereas other posit that background sounds involuntarily divert attention away from the focal task. To explore the roles of phonological processing, serial order retention, and attention control, we analyzed the effects of environmental non-speech sounds and unfamiliar speech on serial recall of verbal items (pictures representing German nouns) and spatial items (dot locations) in children ($n=137$) and adults ($n=98$). In the verbal task, both age groups were equally affected by background sounds, with speech impairing recall more than environmental sounds. In the spatial task, no ISE was found in adults and fourth graders, but third graders exhibited significant performance impairment from both sounds. There was no habituation to the sound effects across the experimental trials. The findings indicate that both specific interference and attention capture may contribute to the ISE, with the impact of attention capture potentially decreasing with age.

In our daily lives, we are frequently exposed to environmental noise, such as traffic noise, music, and other peoples’ conversations. Mental work in noisy environments is becoming increasingly prevalent in our digitalized society as people are taking greater advantage of the opportunity—or the obligation—to work virtually anywhere, e.g., at home, in cafés, on trains and airplanes. The issue of environmental noise is also increasingly important for children. For them, learning in classrooms is often accompanied by task-irrelevant sounds from classmates’ activities or noise pollution from outside, which has been shown to have a negative impact on mental performance and well-being^{1–4}. To avoid or reduce such impairments, we need to know more about the mechanisms underlying noise effects in different age groups. Therefore, in the present study we investigated the effects of different types of environmental noise on the short-term memory performance in children and adults.

Research on noise-induced impairments of short-term memory goes back more than 50 years, and has focused on the so-called “Irrelevant Sound Effect” (ISE). The ISE refers to the impairment of short-term memory for verbal items (e.g., digits, words, letters) presented visually due to task-irrelevant background sounds that participants are instructed to ignore. The usual paradigm for studies on the ISE uses a standard short-term memory task which requires immediate serial recall of lists of 5 to 9 items presented in succession for about one second each. The ISE is a robust and reliable phenomenon which has been replicated numerous times using a wide range of materials, designs, and procedures^{5–10}. There is widespread agreement that both speech and non-speech sounds have a detrimental impact on verbal serial recall in adults.

Several theoretical frameworks have been proposed to explain the mechanisms underlying the ISE. Some of these emphasize automatic, content- or process- based interference between sound characteristics and specific task demands, whereas others attribute the ISE to more general diversions of attention away from the focal task (for a comprehensive review, see Leist et al.¹¹).

Early studies performed by Baddeley and his colleagues^{5,7,12,13} focused on speech as irrelevant sound and attributed its detrimental effect to interference between phonological representations in the phonological loop

¹Center for Cognitive Science, Cognitive and Developmental Psychology Unit, University of Kaiserslautern-Landau (RPTU), 67663 Kaiserslautern, Germany. ²Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, 28007 Madrid, Spain. ³Brain and Cognition Research Unit, Faculty of Psychology and Educational Sciences, University of Leuven (KU), 3000 Leuven, Belgium. ✉email: lleist@rptu.de

component of working memory¹⁴. The phonological loop is a storage system for verbal information. It consists of two parts: a phonological store, which holds phonological representations in a passive form, and an articulatory rehearsal process, which prevents trace decay in the store through subvocal repetitions of the to-be-remembered (TBR) list items. According to the phonological loop account, the ISE results from specific (content-based) interference between the phonological codes of the visually presented items and the irrelevant speech that gains automatic access to the phonological store. Consequently, sounds lacking speech-specific properties such as tones or instrumental music should not impair verbal short-term memory, and tasks requiring short-term memory for nonverbal items should be unimpaired by irrelevant speech.

However, a number of later findings did not support these predictions. First, it was shown that the ISE is also evoked by reversed speech^{15–18}, tones^{10,19}, and music^{20–23}, even though the disruption is most pronounced with speech²⁴. Second, the disruptive effect of background speech is independent from its phonological similarity to the list items²⁵, which contradicts the assumption that similarity between the irrelevant stimuli and the TBR items determines the magnitude of the disruption. Third, the phonological loop account was unable to explain findings demonstrating a dominant role of seriation in ISE evocation. Concerning task characteristics, verbal memory tasks without a serial component were shown to be unaffected or only mildly affected by irrelevant sounds²⁶, whereas a nonverbal task requiring the recall of spatial locations in serial order was significantly affected, see Jones et al.²⁷ (Exp. 4). Concerning sound characteristics, it was shown that sounds representing coherent sequences of distinct, different auditory tokens (“changing-state” sounds, e.g., sequences of different syllables produced by a single speaker) produce stronger impairments when compared to continuous, unstructured sounds like broadband noise or sounds consisting of single, repeated tokens (“steady-state” sounds, e.g., repetitions of single syllables or tones). Aiming to integrate these findings, Jones et al.²⁸ provided another explanation of the ISE. According to this “changing-state account”, sounds with a changing-state characteristic are automatically structured and organized into streams of ordered objects via pre-attentive auditory streaming processes. Sequential information generated during this process interferes with the serial rehearsal of the list items that participants rely on during serial recall tasks. Thus, in contrast to “interference-by-content” proposed by the phonological loop account, the changing-state account assumes interference-by-process (seriation) as the mechanism underlying the ISE.

The changing-state account has been integrated into a broader theoretical framework of serial recall and respective impairments, which assumes that serial order retention is not accomplished through specialized memory structures or processes, but through general-purpose motor and perceptual functions (perceptual-gestural view of short-term memory, see^{29,30}). According to this model, ordered recall of unrelated items is achieved through a deliberate rehearsal process that involves mapping the perceived item sequence to a motor output plan, a record of which is then cyclically repeated until recall is completed. This holds for items of any origin or modality, as they all share a common representational space. Therefore, this model predicts equivalent effects of speech and non-speech changing-state sounds, and equivalent effects of changing-state sounds on serial recall of verbal and nonverbal (e.g., spatial) items.

Neither the phonological loop nor the changing-state account incorporates a specified role of attention in ISE evocation. In contrast, Cowan’s Embedded Processes Model^{31,32} emphasizes the role of attention in memory performance and its impairments. Cowan^{31,32} posits that memory representations vary with respect to activation. Working memory is considered as a subset of long-term memory, comprising those representations that are currently in a heightened state of activation. Representations with the strongest activation enter the so-called “focus of attention”, which is limited in capacity to about four items. Maintaining a longer sequence of items is achieved by recirculating them through the focus of attention. This mechanism is termed “attentional refreshing”, and is available for verbal as well as nonverbal items. According to the attentional account, refreshing is impaired through involuntary intrusions of task-irrelevant sounds into the focus of attention, leading to poorer recall.

Aiming to explore the role of attention in the ISE, several studies compared the effects of irrelevant sounds between children and adults. Given their immature attention control, children are likely to be less able than adults to focus attention on the task in the presence of distracting stimuli. Thus, if the ISE were to result from attention capture, children should be more affected than adults. However, the findings on developmental change in ISE magnitude are inconsistent. While some studies^{11,33–35} identified a larger ISE in children when compared to adults, others found no differences between groups^{35–40}.

Another approach to examine the impact of attention capture is to investigate the magnitude of the ISE over the time-course of task completion. According to the attentional account, the ISE should diminish with time, as participants habituate to the irrelevant sound with repeated exposure. Research in this area also revealed mixed results. Some studies provided evidence for habituation (and thereby attention capture)^{41,42}, while others did not^{43,44}. Researchers have attributed the inconsistency in habituation effects to variations in methodologies.

Especially, it has been suggested that habituation effects are more pronounced when the same sound sequence is presented in each trial of the respective condition (specific habituation) when compared to paradigms where a sound condition is represented by different sound sequences, e.g., different sequences of syllables representing a changing-state condition (unspecific habituation)⁴⁵. According to this view, unspecific habituation proceeds more slowly and is thus not always observable over the time course of a typical ISE experiment.

In the present study, we aim to get further insight into the roles of phonological interference, seriation, and attention control in the ISE. For this aim, we investigate the effects of environmental sounds and narrative speech on serial recall of verbal and nonverbal (spatial) items in children and adults. As previously stated, in contrast to the assumption of phonology-based interference, Jones et al.²⁷ reported similar effects of changing-state speech on verbal and spatial serial recall, and attributed the disruption to interference between pre-attentive seriation processes involved in obligatory sound processing, and deliberate serial rehearsal of the list items. However, the finding of an ISE in spatial serial recall could not be reproduced in several subsequent studies^{11,46–50}. One of these studies¹¹ compared performance of children and adults, and found a significant impairment due to

changing-state speech (spoken syllables) in the verbal task, which was more pronounced in the children than in the adults. For the spatial task, no sign of a sound-induced disruption was found in either group. The latter finding is in contrast to both the changing-state and the attention-capture account of the ISE, as both assume domain-general maintenance mechanisms as the locus of the ISE.

In this study, we used narrative unfamiliar speech and a mixture of environmental non-speech sounds such as dogs barking, telephone ringing, and traffic sounds as task-irrelevant background sounds. For both sound conditions, the same sequence was used in each of the respective trials, thereby increasing the probability to detect potential habituation effects⁴⁵. Using nonspeech environmental sounds, we deviate from prior studies, which analyzed the role of attention in the ISE by using auditory deviants (sounds that violate expectations, i.e., a rarely presented “deviant” sound within a sequence of repeated “standard” sounds, with deviation defined by acoustic properties like timbre⁵¹, or a change in semantic category⁵²), taboo words^{53,54}, or self-relevant words (e.g., the name of the participants⁵⁵) as irrelevant sounds, and/or by varying the predictability of the respective sequences^{56,57}. These studies provide evidence that irrelevant sounds may impair visual-verbal serial recall performance in adults through attention capture.

We decided for environmental sounds for several reasons. First, performing mental task in the presence of irrelevant environmental sounds is a daily challenge in working environments such as classrooms and open-plan offices. Second, sequences of different environmental sounds should evoke only a minor or no changing-state effect due to their lack of stream coherence. Especially, following the changing-state account, in order to evoke interference with seriation, an irrelevant sound sequence should represent a coherent stream consisting of different auditory tokens emanating from a single source, e.g., a sequence of different syllables uttered by a single speaker, or a melody played by a single instrument. Detrimental effects evoked by a sequence of unrelated environmental sounds may thus be considered as indicators of attention capture. Third, evidence from field studies^{2,58} and from behavioral and EEG experimental studies^{36,59,60} confirms that environmental sounds capture attention, especially in children. One of the studies³⁶ explored age differences in the ISE and found that non-speech classroom noise impaired visual-verbal serial recall performance in 7-year-old first graders, whereas older children and adults were unaffected.

In light of the existing research, the following predictions arise: if attention capture plays a dominant role in the ISE, the speech and environmental sounds should evoke significant disruption in both tasks, with stronger effects in children when compared to adults. Furthermore, we should observe a decrease of disruption over the time-course of task completion due to specific habituation. Concerning the interference-based accounts, the following predictions can be derived. If the ISE results from phonological interference (Phonological Loop Account), task-irrelevant speech should impair performance in the verbal task. This effect should be equivalent in children and adults, and should remain stable over the course of the task. Environmental sounds should evoke a minor or no ISE in the verbal task as they lack phonological properties. The spatial task should not be impaired by any background sound, neither in the adults nor in the children, because spatial items are not processed in the phonological loop. In contrast, according to the changing-state account, speech should evoke impairments in both tasks, regardless of age. The disruptive effect should remain stable over the course of the tasks. The mixture of environmental sounds should evoke no or minor disruption, as it does not represent a coherent auditory stream, but a sequence of unrelated sounds from different sources.

For the verbal task, items were displayed as colorful drawings of monosyllabic German nouns. Drawings were used instead of written words in order to rule out effects of children's reading abilities on serial recall performance. Pictorial presentation does not alter participants' strategies when compared to printed words^{61,62}, for a review see Leist et al.¹¹. We recruited children aged about 9 years, as we assume that they employ verbal rehearsal strategies when memorizing verbal items^{63,64}. However, it is important to recognize that rehearsal strategies can vary significantly among younger children, influencing their recall performance and susceptibility to distractions^{65,66}.

For the spatial task, items were spatial locations presented as uniform black dots at varying positions on the screen. Sequence length for adults (children) was 8 (5) for the verbal task, and 7 (5) for the spatial task. Items were presented one-by-one on the computer screen with an onset-to-onset rate of 2 seconds. After a 5 seconds rehearsal interval, all items presented in the respective trial were shown in random arrangement on the screen. For serial recall, participants clicked on the items in the order of presentation. The experimental factors Task (verbal, spatial) and Sound condition (silence, environmental sounds, speech) were varied between and within participants, respectively (i.e., each participant performed either the spatial or the verbal task in each of the three sound conditions). A recording of a female speaker reading a newspaper article in a language unfamiliar to all participants was used as irrelevant speech.

Results

An item was scored as correct when it was recalled in the correct serial position^{51,67}. All analyses are based on proportion correct scores. Figure 1 illustrates the proportion of correct answers based on age group and sound condition.

The data were analyzed using a $3 \times 2 \times 2$ mixed-design Analysis of Variance (ANOVA) with Sound condition (silence, environmental sounds, speech) as within-subject factor and Age group (children, adults) and Task (verbal, spatial) as between subject factors. The analysis revealed significant main effects of Sound condition, $F(2,462) = 24.770$, $p < .001$, $\eta_p^2 = .097$, and Age group, $F(1,231) = 27.871$, $p < .001$, $\eta_p^2 = .108$. The main effect of Task was not significant, $F(1,231) < 1$. Importantly, there were significant two-way interactions between Sound condition $\times$ Task, $F(2,462) = 8.762$, $p < .001$, $\eta_p^2 = .037$, and between Sound Condition $\times$ Age group, $F(2,462) = 3.889$, $p = .021$, $\eta_p^2 = .017$, demonstrating a stronger effect of sound in the verbal when compared to the spatial task, and a stronger effect of sound in children when compared to adults, respectively.

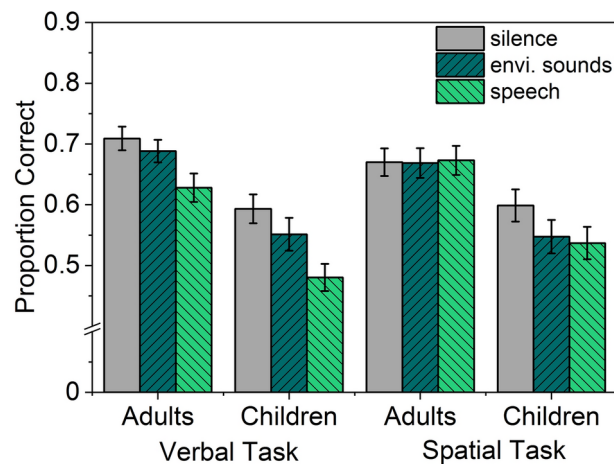


Fig. 1. Mean proportion correct scores of both the verbal and the spatial task with respect to age group and sound condition. Error bars denote standard errors of the mean.

Performance in the silent control condition did not differ between tasks, for either the adults, ($t(96) = 1.308$, $p = .194$, $d = .265$), and the children, ($t(135) = .160$, $p = .873$, $d = .027$). Hence, within age groups, a potential difference in the ISE between tasks cannot be attributed to differences in task difficulty. Across age groups, t -test revealed significantly better performance of the adults in the verbal task ($t(119) = 3.565$, $p < .001$, $d = .656$) and a tendency for better performance in the spatial task ($t(112) = 1.928$, $p = .056$, $d = .367$). The interaction between Task and Age group and the three-way interaction were not significant, $F(1,231) < 1$ and $F(2,462) < 1$, respectively. To evaluate the significant two-way interactions of Sound Condition with Age group and Task, separate analyses were performed for the verbal and spatial tasks.

Verbal task

Data were analyzed using a 3×2 mixed-design ANOVA with Sound condition (silence, environmental sounds, speech) as within-subject factor and Age group (children, adults) as between-subject factor. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 7.11$, $p = .029$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\epsilon = .97$)⁶⁸. There were significant main effects of Sound condition, $F(1.94,230) = 29.274$, $p < .001$,

$\eta_p^2 = .197$, and Age group, $F(1,119) = 19.608$, $p < 0.001$, $\eta_p^2 = .141$. The interaction was not significant, $F(1.94,230) < 1$, indicating comparable sound effects in children and adults. Bonferroni-corrected post hoc tests revealed that performance in the speech condition was lower than the silent and environmental sounds conditions (each $p < .001$). Performance in the environmental sounds condition was lower than in the silent control condition ($p = .019$).

Spatial task

Proportion correct scores were analysed using a 3×2 mixed-design ANOVA with Sound condition (silence, environmental sounds, speech) as within-subject factor and Age group (children, adults) as between-subject factor. There were significant main effects of Sound condition, $F(2,224) = 3.394$, $p = .035$, $\eta_p^2 = .029$, and Age group, $F(1,112) = 9.787$, $p = .002$, $\eta_p^2 = .080$, and a significant interaction, $F(2,224) = 3.394$, $p = .025$, $\eta_p^2 = .032$, indicating that the sound effect differed between age groups.

Separate analyses were performed for each age group in order to analyze the significant interaction. A one-way repeated measures ANOVA of the factor Sound condition (silence, environmental sounds, speech) did not reach significance for the adults, $F(2,92) < 1$, but revealed a significant main effect for the children, $F(2,132) = 6.205$, $p = .003$, $\eta_p^2 = .086$.

Post hoc Bonferroni-corrected tests of the children's spatial task performance revealed lower performance in the speech condition ($p = .006$) and in the environmental sounds condition ($p = .030$) when compared to silence. The sound conditions, environmental sounds and speech, did not differ ($p = 1.00$).

The significant sound effect in children is in contrast to Leist et al.¹¹, who reported no disruptive effect of sounds on children's performance in the spatial task. However, in that study, only third graders were included, whereas in the current study, the spatial task was performed by both third and fourth graders. As attention control develops rapidly during the early years of schooling^{69–73}, we conducted an additional analysis to explore the potential influence of class level on the sound effect. This analysis was performed on 34 third graders (20 female, 14 male, 7;1–9;6 years, $M = 8;4$, $SD = .4$) and 33 fourth graders (20 female, 13 male, 8;1–11;3 years, $M = 9;6$, $SD = .6$). Figure 2 illustrates the proportion of correct answers based on class level and sound condition. A 3×2 mixed-design ANOVA with the within-subject factor Sound condition (silence, environmental sounds, speech) and the between-subjects factor Class level (third grade, fourth grade) revealed significant main effects of Sound condition, $F(2,130) = 6.400$, $p = .002$, $\eta_p^2 = .090$, and Class level, $F(1,65) = 14.141$, $p < .001$, $\eta_p^2 = .179$, and a significant interaction, $F(2,130) = 4.053$, $p = .020$, $\eta_p^2 = .059$. Separate repeated-measures ANOVAs revealed a

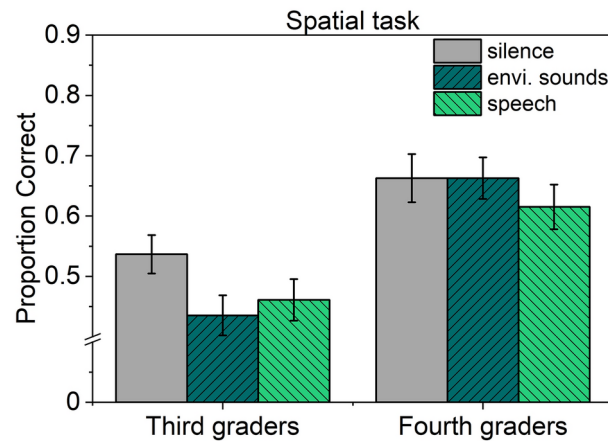


Fig. 2. Mean proportion correct scores of the spatial task with respect to class level and sound condition. Error bars denote standard errors of the mean.

significant effect of Sound condition for the third-graders, $F(2,66) = 8.957$, $p < .001$, $\eta_p^2 = .213$, but not for the fourth-graders, $F(2,64) = 2.061$, $p = .136$, $\eta_p^2 = .060$.

Analyses concerning habituation

To investigate the impact of habituation on the ISE, an additional analysis was conducted to determine whether the sound induced impairments revealed in the preceding analyses diminish over time. If the sound effects are caused by attention capture, one might expect that the first trials with sound would be more disrupted than the last ones, i.e., over time, participants become habituated to the repeated sound sequence (specific habituation, see Röer et al.⁴⁵). Proportion correct scores were calculated for each age group and task for the ordinal trial position, with sixteen trials (adults) and eight trials (children) for each sound condition. Figure 3 depicts the resulting proportion correct scores.

Two-factor repeated measures ANOVAs were performed with Sound condition (silence, environmental sounds, speech) and Trial (1–16) for adults, and Trial (1–8) for children, separately for each task (verbal, spatial).

The analysis for the adults' spatial task performance revealed no significant effects of Sound condition, $F(2,92) < 1$, $p = .939$, Trial, $F(15,690) < 1$, $p = .982$, and no interaction, $F(30,1380) < 1$, $p = .879$. The analysis for the adults' verbal task performance revealed a significant main effect of Sound condition, $F(2,100) = 16.058$, $p < .001$, $\eta_p^2 = .243$, no main effect of Trial, $F(15,750) = 1.24$, $p = .236$, and no interaction, $F(30,1500) = 1.410$, $p = .070$. The same pattern was found for the children's verbal task performance, a significant main effect of Sound condition, $F(2,138) = 17.606$, $p < .001$, $\eta_p^2 = .203$, no main effect of Trial, $F(7,483) < 1$, $p = .434$, and no interaction, $F(14,966) < 1$, $p = .633$. The analysis of the children's spatial task performance revealed significant main effects of Sound condition, $F(2,132) = 6.205$, $p = 0.003$, $\eta_p^2 = 0.086$, and

Trial, $F(7,462) = 2.317$, $p = 0.025$, $\eta_p^2 = 0.034$, but no interaction, $F(14,924) < 1$, $p = .456$, indicating that, regardless of sound conditions, children's overall performance in the spatial task declined over time. Thus, no evidence for habituation was found.

Discussion

This study provides a follow-up extension of earlier research conducted by Leist et al.¹¹, who showed that children's and adults' serial recall performance of verbal but not spatial items is impaired by task-irrelevant changing-state speech. Here, we investigated the impact of task-irrelevant speech and environmental non-speech sounds on verbal and spatial serial recall in primary school children and adults. The tasks were identical to Leist et al.¹¹, requiring the serial reconstruction of pictures displaying monosyllabic German words (verbal task) and of uniform black dots displayed at various positions on the screen (spatial task).

In the verbal task, children and adults were equally affected by both task-irrelevant background sounds. Speech impaired recall performance significantly more than environmental sounds. There was no evidence of habituation across experimental trials, neither in the children nor in the adults. For the spatial task, performance of adults was unaffected by the irrelevant sounds, which is in contrast to the classical finding of Jones et al.²⁷, but in line with more recent findings^{11,50}. However, for children's performance in the spatial task, background speech and environmental sounds evoked significant disruption of equal magnitude. Further analyses revealed that the significant sound effect in the child sample resulted from a reliable impairment in the third-grade children, whereas fourth graders were unaffected. The pattern of results is discussed with respect to the theoretical accounts outlined in the introduction.

In adults, the ISE was evident for visual-verbal serial recall, but was completely absent for serial recall of spatial items. Given the consistent evidence from the current and other studies with considerable statistical power^{11,50}, we may conclude that serial recall of spatial items is unaffected by irrelevant sounds in adults. This poses severe difficulties for theories positing that the ISE results from impairments of domain-general maintenance mechanisms, i.e., perceptual-motor sequencing (changing-state account) and attentional

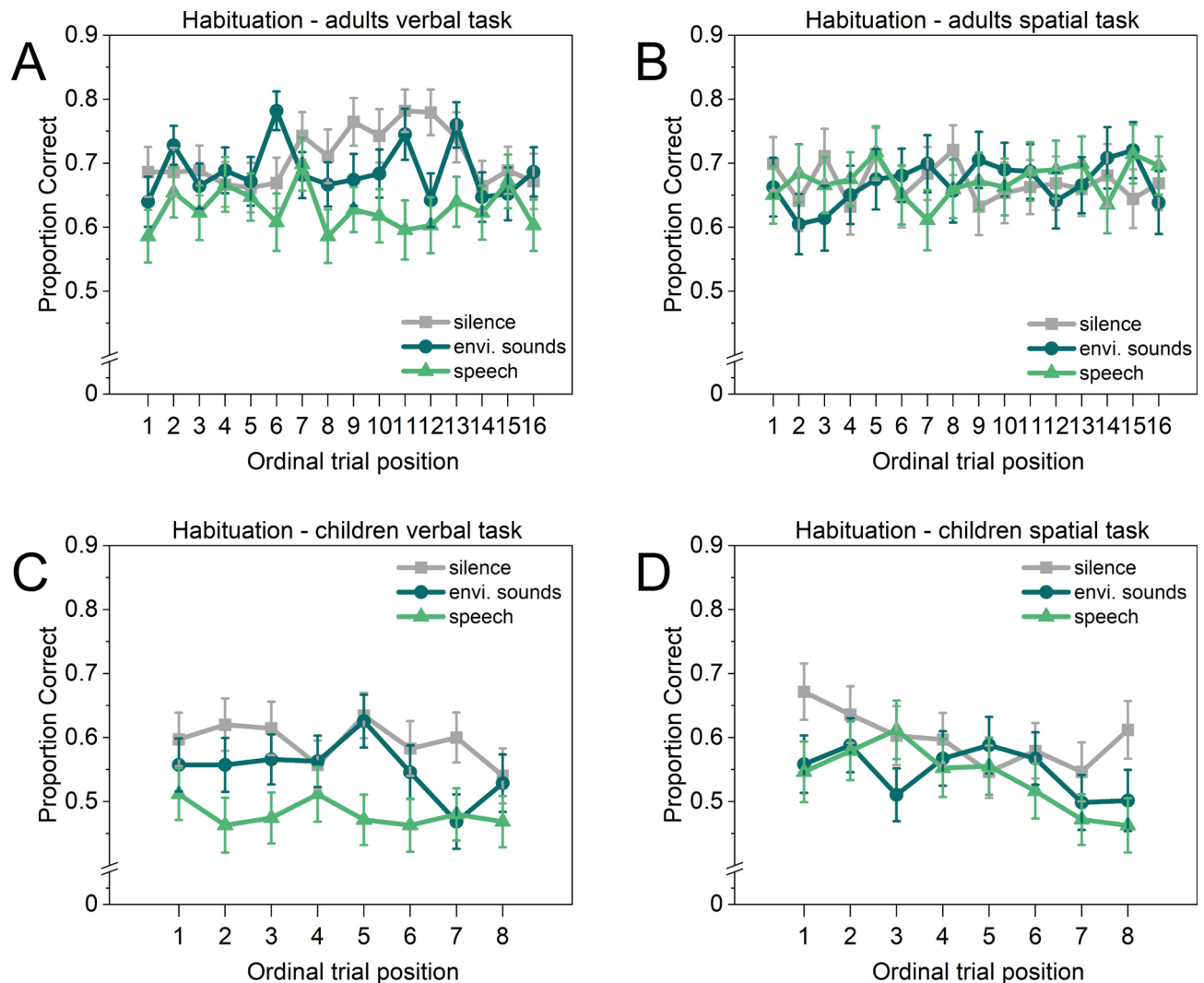


Fig. 3. Mean proportion correct scores as a function of sound condition and ordinal trial position of sixteen (adults) and eight (children) trials for both tasks (verbal, spatial). Error bars denote standard errors of the mean.

refreshing (attention-capture account), as both predict sound-induced performance decrements irrespective of item modality. Importantly, according to the attention capture account, ISE should even be stronger for spatial when compared to verbal serial recall, as it has been shown that visuo-spatial working memory tasks rely much more on domain-general attentional resources^{74,75}. The immunity of spatial serial recall to the ISE is in line with the phonological loop account, which considers the retention of phonological codes in a verbal short-term store as the locus of disruption.

The finding that environmental non-speech sounds also impaired verbal serial recall performance, albeit to a lesser extent than speech, mirrors the results reported in Buchner et al.⁷⁶. These authors explain their findings in the framework of the changing-state account, arguing that speech sounds have a more prominent changing-state characteristic when compared to environmental sounds. In particular, according to the changing-state account, speech is more disruptive than non-speech sounds because it represents a coherent changing-state auditory stream, whereas the sound mix consists of isolated and unrelated auditory tokens. However, the difference in effect magnitude between speech and environmental sounds might also be explained by the phonological loop account, by assuming that complex non-speech sounds are processed in the phonological loop, as was suggested for instrumental music⁷, or that environmental sounds—like spoken words—automatically activate phonological representations which interfere with those of the list items. Recent findings suggest that spoken words and environmental sounds evoke comparable phonological activation⁷⁷, but see Kukona et al.⁷⁸ for contradicting results. For the attention capture account, the reduced effect of environmental sounds is difficult to explain. One would expect that a mixture of heterogeneous, partly alarming sounds has more potential to attract participants' attention when compared to narrative, meaningless speech. In addition, the attention capture account predicts habituation with repeated exposure, which was not found for either sound.

Concerning developmental change, the absence of age-related differences in the ISE on the verbal task is in line with the interference-based accounts, as both predict comparable effects across age groups provided that the same strategies are applied. The lack of an age effect is difficult to handle for the attentional account. Due to the immature attention control in children, they should be more prone to sound-induced attention capture. However, the current results need to be considered in the context of prior studies on developmental change in the ISE on verbal serial recall, which yielded inconsistent results. Some mirrored the current finding of equivalent impairments in children and adults^{35–40,79}, whereas others found stronger impairments in the children^{11,33–35}, thereby confirming a role of attention in the ISE. To account for these discrepancies, following Klatte et al.³⁶ and Elliott et al.³⁵, we suggest that children and adults are equally prone to specific interference through task-irrelevant sounds in visual-verbal serial recall (provided that the children use the same retention strategies), but that the disruptive effect may be superimposed by sound-induced diversions of attention. The contribution of the latter depends on characteristics of the sounds and method, and on the participants' ability to control attention, which increases with age and schooling^{71–73}. Developmental studies on the ISE differ considerably in these respects, for example concerning children's age and class levels, types of background sounds (e.g., meaningless vs. meaningful narrative speech, environmental sounds, tones), sound levels, and methods of sound presentation (blocked vs. trial-by-trial variations of sound conditions, presentation via earphones vs. loudspeakers). All of these aspects have an impact on the sounds' potential to grab attention away from the primary task. In addition, it has to be kept in mind that, even within a restricted age range or class level, there are huge individual differences in children's attention control^{69,70}. This might explain why, in our study, we did not find an age effect on verbal serial recall, whereas a prior study using comparable methods and age groups¹¹ reported greater impairment in children compared to adults.

Concerning the spatial task, the current study revealed disruptive effects of equal magnitude due to speech and environmental sounds in children, but not in adults. We attribute the effect observed in the children to attention capture for several reasons. First, additional analyses revealed that the effect was only evident in the third-grade children, but absent in the fourth graders. In view of the increase in attention control with age and schooling, we may conclude that the fourth-grade children were better able to focus attention on the task in the presence of irrelevant sounds. In line with this argument, Klatte et al.³⁶ found a significant impairment of verbal serial recall due to non-speech classroom noise in 7-year-old first graders, whereas older children were unaffected. Second, the current findings mirror those of a previous study³⁵, which showed that children and adults were equally impaired by irrelevant speech in a visual-verbal serial recall task, but only the children were also affected in a verbal task that has consistently proven immune to the ISE in adults. The authors infer that children and adults are equally prone to sound-induced interference in visual-verbal serial recall, but that, in children, attentional diversions are also at play, making them vulnerable to sound-induced impairments irrespective of task demands and the sounds' changing-state.

To summarize, none of the described theoretical accounts of the ISE is able to integrate the full pattern of results in our study. The findings are best explained by assuming that both specific interference with visual-verbal serial recall and more general effects on attention contribute to the observed effects, as suggested in the "duplex mechanism account" of auditory distraction⁸⁰. Following this account, specific interference and attention capture represent independent mechanisms of sound-induced performance decrements, which may act in isolation or in combination, depending on characteristics of the tasks, the sounds, and the individuals affected. In children, diversions of attention have a stronger impact on the overall disruption, as children are less able to control attention in the presence of task-irrelevant stimuli when compared to adults. However, there is huge interindividual variability in children's attention abilities over the primary school years, which may contribute to the discrepancies in results from studies on developmental change in the ISE.

Several limitations of the present study must be acknowledged. First, the children performed only 8 trials per sound condition, and we thus cannot rule out that habituation effects become evident when the time-course of exposure is extended. Second, with the environmental sounds used here, only limited evidence for performance decrements through attention capture was found in adults. However, as stated in the introduction, a number of studies found clear evidence for task impairments through sound-induced diversions of attention in adults^{51–55}. Obviously, the attention-grabbing potential of the sounds used in those studies (i.e., auditory deviants, taboo-words^{53,54}, self-relevant words⁵⁵) is stronger when compared to the current environmental sounds. Third, as in the majority of respective studies, our child participants aged 8 years or more. As attention control increases strongly during the primary school years and the ability to filter out irrelevant sounds becomes adult-like with about age 10, future studies on age differences in the ISE should include younger children (school starters, aged 6 to 7 years). Fourth, due to the strong correlation between chronological age and class level ($r(67) = .82$), we do not know whether age or schooling is responsible for the stronger sound effect on the spatial task observed in third graders compared to fourth graders.

Experimental research on the ISE in children does not only increase our knowledge on the mechanisms underlying this effect, but may also contribute to practical issues. As verbal short-term memory plays a crucial role in language and reading acquisition⁸¹, sound-induced disruption could result in lasting deficits. Field studies have already demonstrated that noise in classrooms affects reading acquisition and verbal precursors of reading^{1,2,4,82}. Therefore, experimental studies on the ISE in children are critical for creating suitable acoustic environments for learning³⁶. For example, tasks that involve verbal short-term memory, such as reading and spelling in beginning readers, vocabulary learning, or mental arithmetic should not be performed in the presence of background speech or music.

Methods

Participants

An a priori power analysis was conducted using G*Power⁸³ (version 3.1.9.7) to estimate the sample size based on data from Leist et al.¹¹ ($N = 164$). The effect size reported in that study was 0.176. With a significance criterion of $\alpha = .05$ and a power of .80, the minimum sample size required for the mixed ANOVA with this effect size is $N = 108$. A total of 137 children, recruited from two primary schools in the Kaiserslautern region and in North Rhine-Westphalia in Germany (76 female, 61 male, aged between 7 years, 9 months and 11 years, 3 months, mean age of 9;0 years), and 98 students from the University of Kaiserslautern-Landau (45 female, 53 male, aged between 19 and 36 years, mean age of 23.1 years), participated in this study.

The verbal task was performed by 51 adults (22 female, 29 male) aged between 19–26 years ($M = 22.2$ years, $SD = 1.82$ years) and 70 children (36 female, 34 male) aged between 8;0 years to 10 years, 9 months ($M = 8$ years, 6 months, $SD = 7$ months). The spatial task was performed by 47 adults (23 female, 24 male) aged between 19–36 years ($M = 24.0$ years, $SD = 3.48$ years) and 67 children (40 female, 27 male) aged between 7 years, 9 months to 11 years, 3 months ($M = 9$ years, 1 month, $SD = 7$ months).

All participants were native German speakers and had normal or corrected-to-normal vision and normal hearing according to self-reports (adults) or parental reports (children). Adults received either course credit or payment for participation (10 €). Children received writing materials, e.g., pencils, erasers.

Apparatus

A 15.6-inch HP laptop (ProBook 650 G1) running Python 3.7/PsychoPy 3.1.5⁸⁴ on Windows 10 was used to control the experimental trial. Display resolution was 1920×1080, with a refresh rate of 60 Hz. The background sounds were played via open headphones (Sennheiser HD650), connected with a Focusrite Scarlett 2i2 2nd generation audio interface to the laptop.

Irrelevant sounds

Task performance was measured during three different sound conditions: silent control, environmental non-speech sounds, and speech. The environmental sounds encompassed a sequence of diverse common sounds (e.g., dog barking, bicycle bell, church bell, approaching car, chainsaw, cackling hens, siren horn). These environmental sounds were played one after the other, without pause, for two to three seconds per sound. The speech consisted of a female Danish speaker reading an article from a newsletter identical to Klatte et al.³⁶. Danish was unfamiliar to all participants. For both sound conditions, the same sequence was played in each of the respective trials.

To-be-remembered items

For the verbal task, for children and adults, items were drawn from a set of ten colorful drawings representing the German monosyllabic words *Bett* (bed), *Bus* (bus), *Eis* (ice cream), *Frosch* (frog), *Kamm* (comb), *Mond* (moon), *Pilz* (mushroom), *Schal* (scarf), *Schiff* (ship) and *Zaun* (fence). For the adults the set included the additional items *Brief* (letter), *Haus* (house), *Herz* (heart), *Hut* (hat), *Nuss* (nut) and *Schwein* (pig). Four lists of five items (drawn out of 10) and eight lists of eight items (drawn out of 16) were created for the children and the adults, respectively. The lists were created quasi-randomly, with each item appearing in one of four lists (children) or eight lists (adults). To generate parallel versions of each list, random permutations of the list items were used. In addition, one practice list for each sound condition was constructed. All drawings were presented in a rectangular black frame (102×73 mm) in the center of a white screen.

For the spatial task, items were presented as black dots (10 mm diameter) on a white background in a 304×183 mm black frame. Dot positions were drawn from a total of 80 possible locations that were arranged in a 8×10 grid (not shown to the participants) in the frame. As shown by Parmentier and colleagues^{85,86}, the difficulty of reproducing a sequence of dot locations depends on the complexity of the spatial path formed by the successive dots. Path complexity is determined by the number of crossings, the path length, and the angular degree. In order to verify equal task difficulty across sound conditions, we first created eight lists of five (children) and sixteen lists of seven dot positions (adults). The resulting path of each list was mirrored horizontally and vertically, thereby varying dot locations while keeping path complexity constant (for further details, see Leist et al.¹¹). With this strategy, three parallel versions of each list were constructed, which were assigned randomly to the three sound conditions. In addition, three practice lists were constructed.

Serial recall tasks

Both tasks are illustrated in Figure 4. Each trial consisted of a presentation phase, a retention interval, and a reconstruction phase. Items were presented one after the other with a 1500 ms presentation time and a 500 ms interstimulus interval. The final list item was followed by a 5000 ms retention interval. In the reconstruction phase, all items were displayed simultaneously on the screen. In the verbal task, drawings were randomly positioned in a predefined array of 5 (children) or 8 (adults) frames. In the spatial task, the dots were shown in the same locations as in the presentation phase. Participants were required to reproduce the serial order of drawings/dots by clicking on the items in the order in which they were presented. When an object was clicked, its shading changed to indicate that it has been selected. There was no time limit for responding and no possibility for error correction. After selecting the final item, the participants were shown a visual cue instructing them to hit the spacebar to initiate the next trial.

Procedure

Adults were tested in two to five-person groups in a sound-attenuated booth at the University of Kaiserslautern-Landau. Children were tested individually or in groups of up to five at their school in a large classroom, where

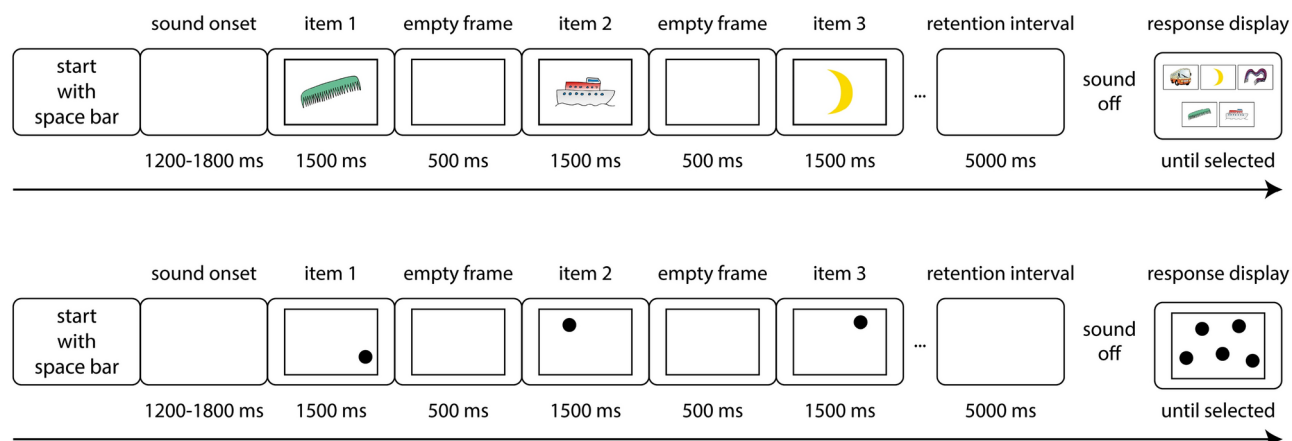


Fig. 4. The experimental procedure for the serial recall tasks. Top: Verbal Task, bottom: Spatial Task. The children were shown five items per trial in both tasks, while the adults were shown eight (verbal task) and seven (spatial task) items per trial. The items in the verbal task were drawings, while the items in the spatial task were uniform black dots. All items shown in the relevant trial were presented again on the response display in the reconstruction phase. The pictures in the verbal task were arranged in an array of 5 (children) and 8 (adults) frames. The spatial task's black dots were displayed in the same location as they had been shown in the trial. Note that the items are enlarged for illustrative purposes.

five computer workstations were set up about ten meters apart. A researcher or trained student assistant instructed each child individually. Adults received written instruction. Participants were instructed to ignore the sounds and focus on the serial recall task. For the verbal task, all pictures used were presented and named by a female speaker. To introduce the background sounds to the participants, both the environmental non-speech sounds and speech were played for 4 seconds each, followed by one practice trial per sound condition. Then, the children completed a total of 24 experimental trials, eight for each sound condition. Adults performed a total of 48 trials, 16 for each sound condition.

For both tasks, the three versions of each list were randomly assigned to the sound conditions. Sound conditions varied from trial to trial and were quasi-randomized: All sound conditions were presented in random order before being randomized again. In the trials with sound presentation, the sound started when the participant initiated the trial with the space bar. A random interval between 1200 and 1800 ms was introduced before the presentation of the first list item, to avoid correlations between sound and item onset⁸⁷. The sound was played throughout both the presentation phase and retention interval and stopped at the beginning of the reconstruction phase. The average sound level (LAeq) of the background sounds was 59 dB, as measured by ITA Artificial Head⁸⁸. The test session lasted about 25 minutes for children and 35 minutes for adults.

Data availability

The datasets generated during and analyzed during the current study as well as the materials are available in the OSF repository, https://osf.io/eayhb/?view_only=441bef7c4f5f487091b7e31f43b684dc.

Received: 26 June 2024; Accepted: 6 January 2025

Published online: 14 January 2025

References

- Klatte, M., Hellbrück, J., Seidel, J. & Leistner, P. Effects of classroom acoustics on performance and well-being in elementary school children: A field study. *Environ. Behav.* **42**, 659–692. <https://doi.org/10.1177/0013916509336813> (2010).
- Klatte, M. et al. Effects of aircraft noise on reading and quality of life in primary school children in Germany: Results from the Norah study. *Environ. Behav.* **49**, 390–424. <https://doi.org/10.1177/0013916516642580> (2017).
- Thompson, R. et al. Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence. *Environ. Int.* **158**, 106905. <https://doi.org/10.1016/j.envint.2021.106905> (2022).
- Gheller, F., Spicciarelli, G., Scimemi, P. & Arfè, B. The effects of noise on children's cognitive performance: A systematic review. *Environ. Behav.* <https://doi.org/10.1177/00139165241245823> (2024).
- Salamé, P. & Baddeley, A. D. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *J. Verbal Learn. Verbal Behav.* **21**, 150–164 (1982).
- Colle, H. A. Auditory encoding in visual short-term recall: effects of noise intensity and spatial location. *J. Verbal Learn. Verbal Behav.* **19**, 722–735. [https://doi.org/10.1016/s0022-5371\(80\)90403-x](https://doi.org/10.1016/s0022-5371(80)90403-x) (1980).
- Salamé, P. & Baddeley, A. D. Effects of background music on phonological short-term memory. *Q. J. Exp. Psychol. Sect. A* **41**, 107–122 (1989).
- Hanley, J. R. & Broadbent, C. The effect of unattended speech on serial recall following auditory presentation. *Br. J. Psychol.* **78**, 287–297. <https://doi.org/10.1111/j.2044-8295.1987.tb02247.x> (1987).
- Ellermeier, W. & Zimmer, K. Individual differences in susceptibility to the “irrelevant speech effect”. *J. Acoust. Soc. Am.* **102**, 2191–2199 (1997).
- Jones, D. M. & Macken, W. J. Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 369. <https://doi.org/10.1037/0278-7393.19.2.369> (1993).

11. Leist, L., Lachmann, T., Schlittmeier, S. J., Georgi, M. & Klatte, M. Irrelevant speech impairs serial recall of verbal, but not spatial items in children and adults. *Mem. Cogn.* **51**, 307–320. <https://doi.org/10.3758/s13421-022-01359-2> (2023).
12. Salamé, P. & Baddeley, A. D. Phonological factors in STM: Similarity and the unattended speech effect. *Bull. Psychon. Soc.* **24**, 263–265. <https://doi.org/10.3758/bf03330135> (1986).
13. Salamé, P. & Baddeley, A. D. Noise, unattended speech and short-term memory. *Ergonomics* **30**, 1185–1194 (1987).
14. Baddeley, A. D. & Logie, R. H. Working memory: The multiple-component model. In *Models of Working Memory* (eds Baddeley, A. D. & Logie, R. H.) (Cambridge University Press, 1999).
15. Jones, D. M., Miles, C. & Page, J. Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory?. *Appl. Cogn. Psychol.* **4**, 89–108. <https://doi.org/10.1002/acp.2350040203> (1990).
16. LeCompte, D. C., Neely, C. B. & Wilson, J. R. Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *J. Exp. Psychol. Learn. Mem. Cogn.* **23**, 472–483. <https://doi.org/10.1037/0278-7393.23.2.472> (1997).
17. Surprenant, A. M. Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging Neuropsychol. Cogn.* **14**, 126–143. <https://doi.org/10.1080/13825580701217710> (2007).
18. Viswanathan, N., Dorsi, J. & George, S. The role of speech-specific properties of the background in the irrelevant sound effect. *Q. J. Exp. Psychol.* **67**, 581–589. <https://doi.org/10.1080/17470218.2013.821708> (2014).
19. Divin, W., Coyle, K. & James, D. T. T. The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *Br. J. Psychol.* **92**, 593–616. <https://doi.org/10.1348/000712601162365> (2001).
20. Klatte, M., Kilcher, H. & Hellbrück, J. Wirkungen der zeitlichen struktur von hintergrundschall auf das arbeitsgedächtnis und ihre theoretischen und praktischen implikationen. *Zeitschrift für Exp. Psychol.* **42**, 517–544 (1995).
21. Ellermeier, W. & Hellbrück, J. Is level irrelevant in “irrelevant speech”? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 1406–1414. <https://doi.org/10.1037/0096-1523.24.5.1406> (1998).
22. Schlittmeier, S. J., Hellbrück, J., Thaden, R. & Vorländer, M. The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance. *Ergonomics* **51**, 719–736. <https://doi.org/10.1080/00140130701745925> (2008).
23. Schweppe, J. & Knigge, J. Irrelevant music: How suprasegmental changes of a melody’s tempo and mode affect the disruptive potential of music on serial recall. *Mem. Cogn.* **48**, 982–993. <https://doi.org/10.3758/s13421-020-01037-1> (2020).
24. Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H. & Hellbrück, J. Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Atten. Percept. Psychophys.* **74**, 194–203 (2011).
25. Bell, R., Mund, I. & Buchner, A. Disruption of short-term memory by distractor speech: Does content matter?. *Q. J. Exp. Psychol.* **64**, 146–168. <https://doi.org/10.1080/17470218.2010.483769> (2011).
26. Marsh, J. E., Hughes, R. W. & Jones, D. M. Interference by process, not content, determines semantic auditory distraction. *Cognition* **110**, 23–38. <https://doi.org/10.1016/j.cognition.2008.08.003> (2009).
27. Jones, D. M., Farrand, P., Stuart, G. & Morris, N. Functional equivalence of verbal and spatial information in serial short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 1008–1018. <https://doi.org/10.1037/0278-7393.21.4.1008> (1995).
28. Jones, D. M., Beaman, C. & Macken, W. J. The object-oriented episodic record model. *Model. short-term memory* 209–238 (1996).
29. Jones, D. M. & Macken, B. In the beginning was the deed: Verbal short-term memory as object-oriented action. *Curr. Dir. Psychol. Sci.* **27**, 351–356. <https://doi.org/10.1177/0963721418765796> (2018).
30. Jones, D., Hughes, R. & Macken, W. Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *J. Mem. Lang.* **54**, 265–281. <https://doi.org/10.1016/j.jml.2005.10.006> (2006).
31. Cowan, N. *Attention and Memory* (Oxford University Press, 1995).
32. Cowan, N., Morey, C. C. & Naveh-Benjamin, M. *An Embedded-Processes Approach to Working Memory: How Is It Distinct From Other Approaches, and to What Ends?* (Oxford University Press, 2020).
33. Elliott, E. M. The irrelevant-speech effect and children: theoretical implications of developmental change. *Mem. Cogn.* **30**, 478–487. <https://doi.org/10.3758/bf03194948> (2002).
34. Elliott, E. M. & Briganti, A. M. Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychol.* **140**, 64–74. <https://doi.org/10.1016/j.actpsy.2012.02.009> (2012).
35. Elliott, E. M. et al. Distraction in verbal short-term memory: Insights from developmental differences. *J. Mem. Lang.* **88**, 39–50. <https://doi.org/10.1016/j.jml.2015.12.008> (2016).
36. Klatte, M., Lachmann, T., Schlittmeier, S. J. & Hellbrück, J. The irrelevant sound effect in short-term memory: Is there developmental change?. *Eur. J. Cogn. Psychol.* **22**, 1168–1191. <https://doi.org/10.1080/09541440903378250> (2010).
37. Schwarz, H. et al. Age differences in the irrelevant sound effect: A serial recognition paradigm. *psihologija* **48**, 35–43. <https://doi.org/10.2298/PSI1501035S> (2015).
38. Röer, J. P., Bell, R., Körner, U. & Buchner, A. Equivalent auditory distraction in children and adults. *J. Exp. Child Psychol.* **172**, 41–58. <https://doi.org/10.1016/j.jecp.2018.02.005> (2018).
39. Meinhardt-Injac, B., Imhof, M., Wetzel, N., Klatte, M. & Schlittmeier, S. J. The irrelevant sound effect on serial recall is independent of age and inhibitory control. *Auditory Percept. Cogn.* <https://doi.org/10.1080/25742442.2022.2064692> (2022).
40. Joseph, T. N., Hughes, R. W., Sörqvist, P. & Marsh, J. E. Differences in auditory distraction between adults and children: A duplex-mechanism approach. *J. Cogn.* <https://doi.org/10.5334/joc.15> (2018).
41. Bell, R., Röer, J. P., Dentale, S. & Buchner, A. Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 1542–1557. <https://doi.org/10.1037/a0028459> (2012).
42. Morris, N. & Jones, D. M. Habituation to irrelevant speech: Effects on a visual short-term memory task. *Percept. Psychophys.* **47**, 291–297. <https://doi.org/10.3758/bf03205003> (1990).
43. Röer, J. P., Bell, R., Dentale, S. & Buchner, A. The role of habituation and attentional orienting in the disruption of short-term memory performance. *Mem. Cogn.* **39**, 839–850. <https://doi.org/10.3758/s13421-010-0070-z> (2011).
44. Jones, D. M., Macken, W. J. & Mosdell, N. A. The role of habituation in the disruption of recall performance by irrelevant sound. *Br. J. Psychol.* **88**, 549–564. <https://doi.org/10.1111/j.2044-8295.1997.tb02657.x> (1997).
45. Röer, J. P., Buchner, A. & Bell, R. Auditory distraction in short-term memory: Stable effects of semantic mismatches on serial recall. *Auditory Percept. Cogn.* **2**, 143–162. <https://doi.org/10.1080/25742442.2020.1722560> (2019).
46. Bergström, K., Lachmann, T. & Klatte, M. Wann stört lärm das geistige arbeiten? einfluss von aufgaben- und geräuschcharakteristiken bei der wirkung moderaten lärms auf arbeitsgedächtnisleistungen. *Fortschritte der Akustik. Beiträge zur 38. Jahrestagung für Akustik DAGA* (2012).
47. Klatte, M. & Hellbrück, J. Effects of irrelevant speech on serial recall of verbal and spatial materials. In Schick, A. & Klatte, M. (eds.) *Contributions to Psychological Acoustics. Results of the 7th Oldenburg Symposium on Psychological Acoustics*, 531–538 (1997).
48. Kvetnaya, T. Registered replication report: Testing disruptive effects of irrelevant speech on visual-spatial working memory. *J. Eur. Psychol. Stud.* **9**, 10–15. <https://doi.org/10.5334/jeps.450> (2018).
49. Tremblay, S., Parmentier, F. B., Hodgetts, H. M., Hughes, R. W. & Jones, D. M. Disruption of verbal-spatial serial memory by extraneous air-traffic speech. *J. Appl. Res. Mem. Cogn.* **1**, 73–79. <https://doi.org/10.1016/j.jarmac.2012.04.004> (2012).
50. Marsh, J. E. et al. Changing-state irrelevant speech disrupts visual-verbal but not visual-spatial serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* <https://doi.org/10.1037/xlm0001360> (2023).
51. Hughes, R. W., Vachon, F. & Jones, D. M. Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *J. Exp. Psychol. Learn. Mem. Cogn.* **33**, 1050. <https://doi.org/10.1037/0278-7393.33.6.1050> (2007).

52. Vachon, F., Marsh, J. E. & Labonté, K. The automaticity of semantic processing revisited: Auditory distraction by a categorical deviation. *J. Exp. Psychol. Gen.* **149**, 1360 (2020).
53. Röer, J. P., Körner, U., Buchner, A. & Bell, R. Attentional capture by taboo words: A functional view of auditory distraction. *Emotion* **17**, 740–750. <https://doi.org/10.1037/emo0000274> (2017).
54. Rettie, L. et al. Warning—taboo words ahead! Avoiding attentional capture by spoken taboo distractors. *J. Cogn. Psychol.* <https://doi.org/10.1080/20445911.2023.2285860> (2023).
55. Röer, J. P., Bell, R. & Buchner, A. Self-relevance increases the irrelevant sound effect: Attentional disruption by one's own name. *J. Cogn. Psychol.* **25**, 925–931 (2013).
56. Röer, J. P., Bell, R. & Buchner, A. Specific foreknowledge reduces auditory distraction by irrelevant speech. *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 692–702. <https://doi.org/10.1037/xhp0000028> (2015).
57. Bell, R., Röer, J. P., Marsh, J. E., Storch, D. & Buchner, A. The effect of cognitive control on different types of auditory distraction: A preregistered study. *Exp. Psychol.* **64**, 359–368. <https://doi.org/10.1027/1618-3169/a000372> (2017).
58. Banbury, S. P. & Berry, D. C. Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics* **48**, 25–37 (2005).
59. Wetzel, N., Widmann, A. & Schröger, E. The cognitive control of distraction by novelty in children aged 7–8 and adults. *Psychophysiology* **46**, 607–616 (2009).
60. Wetzel, N., Schröger, E. & Widmann, A. Distraction by novel and pitch-deviant sounds in children. *Front. Psychol.* **7**, 1949 (2016).
61. Poloczek, S., Henry, L. A., Messer, D. J. & Büttner, G. Do children use different forms of verbal rehearsal in serial picture recall tasks? A multi-method study. *Memory* **27**, 758–771. <https://doi.org/10.1080/09658211.2018.1563615> (2019).
62. Schiano, D. J. & Watkins, M. J. Speech-like coding of pictures in short-term memory. *Mem. Cogn.* **9**, 110–114 (1981).
63. Elliott, E. M. et al. Multilab direct replication of flavell, beach, and chinsky (1966): Spontaneous verbal rehearsal in a memory task as a function of age. *Adv. Methods Pract. Psychol. Sci.* **4**, 251524592110181. <https://doi.org/10.1177/25152459211018187> (2021).
64. Lehmann, M. & Hasselhorn, M. Rehearsal dynamics in elementary school children. *J. Exp. Child Psychol.* **111**, 552–560. <https://doi.org/10.1016/j.jecp.2011.10.013> (2012).
65. AuBuchon, A. M. et al. Lexical access speed and the development of phonological recoding during immediate serial recall. *J. Cogn. Dev.* **23**, 624–643 (2022).
66. Morey, C. C. et al. Is verbal rehearsal strategic? An investigation into overt rehearsal of nameable pictures in 5-to 10-year-old children. *J. Cogn. Dev.* <https://doi.org/10.1080/15248372.2024.2389123> (2024).
67. Vachon, F., Labonté, K. & Marsh, J. E. Attentional capture by deviant sounds: A noncontingent form of auditory distraction?. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 622–634. <https://doi.org/10.1037/xlm0000330> (2017).
68. Girden, E. ANOVA (SAGE Publications Inc., 1992).
69. Romine, C. B. & Reynolds, C. R. A model of the development of frontal lobe functioning: Findings from a meta-analysis. *Appl. Neuropsychol.* **12**, 190–201. https://doi.org/10.1207/s15324826an1204_2 (2005).
70. Burrage, M. S. et al. Age- and schooling-related effects on executive functions in young children: A natural experiment. *Child Neuropsychol.* **14**, 510–524. <https://doi.org/10.1080/09297040701756917> (2008).
71. Brouwers, S. A. et al. Cognitive development through schooling and everyday life: A natural experiment among Kharwar children in India. *Int. J. Behav. Dev.* **41**, 309–319 (2017).
72. Brod, G., Bunge, S. A. & Shing, Y. L. Does one year of schooling improve children's cognitive control and alter associated brain activation?. *Psychol. Sci.* **28**, 967–978 (2017).
73. Morrison, F. J., Kim, M. H., Connor, C. M. & Grammer, J. K. The causal impact of schooling on children's development: Lessons for developmental science. *Curr. Dir. Psychol. Sci.* **28**, 441–449 (2019).
74. Vernucci, S. et al. The structure of working memory during childhood: a systematic review. *J. Cogn. Psychol.* **33**, 103–118 (2021).
75. Morey, C. C. & Miron, M. D. Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **42**, 1907–1918. <https://doi.org/10.1037/xlm0000280> (2016).
76. Buchner, A., Bell, R., Rothermund, K. & Wentura, D. Sound source location modulates the irrelevant-sound effect. *Mem. Cogn.* **36**, 617–628. <https://doi.org/10.3758/mc.36.3.617> (2008).
77. Bartolotti, J. et al. Listening to speech and non-speech sounds activates phonological and semantic knowledge differently. *Q. J. Exp. Psychol.* **73**, 1135–1149. <https://doi.org/10.1177/1747021820923944> (2020).
78. Kukona, A. & Jordan, A. Online mouse cursor trajectories distinguish phonological activation by linguistic and nonlinguistic sounds. *Psychon. Bull. Rev.* **30**, 362–372. <https://doi.org/10.3758/s13423-022-02153-6> (2022).
79. Leist, L. et al. Differential effects of task-irrelevant monaural and binaural classroom scenarios on children's and adults' speech perception, listening comprehension, and visual-verbal short-term memory. *Int. J. Environ. Res. Public Heal.* **19**, 15998. <https://doi.org/10.3390/ijerph192315998> (2022).
80. Hughes, R. W. Auditory distraction: A duplex-mechanism account. *Psych J.* **3**, 30–41 (2014).
81. Baddeley, A. D., Gathercole, S. & Papagno, C. The phonological loop as a language learning device. *Psychol. Rev.* **105**, 158–173. <https://doi.org/10.1037/0033-295x.105.1.158> (1998).
82. Dockrell, J. E. & Shield, B. The impact of sound-field systems on learning and attention in elementary school classrooms. *J. Speech Lang. Hear. Res.* **55**, 1163–1176 (2012).
83. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
84. Peirce, J. W. PsychoPy—psychophysics software in python. *J. Neurosci. Methods* **162**, 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017> (2007).
85. Parmentier, F. B. R., Elford, G. & Maybery, M. Transitional information in spatial serial memory: Path characteristics affect recall performance. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 412–427. <https://doi.org/10.1037/0278-7393.31.3.412> (2005).
86. Parmentier, F. *Exploring the determinants of memory for spatial sequences* (Taylor & Francis, 2014).
87. Bell, R., Dentale, S., Buchner, A. & Mayr, S. Erp correlates of the irrelevant sound effect. *Psychophysiology* **47**, 1182–1191 (2010).
88. Schmitz, A. Ein neues digitales kunstkopfmesssystem [a new digital artificial head measuring system]. *Acustica* **81**, 416–420 (1995).

Acknowledgements

The authors thank all children, parents and participants who took part in the study and all research assistants, especially Aarabis Srijevaghavan, Hannah Plückebaum and Charlotte Mock for their help with data collection and Markus Kurtz for his technical assistance. The illustrations (Figure 4) were created by the first author. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; KL 1305/4).

Author contributions

Conceptualization, M.K. and L.L.; methodology, M.K. and L.L.; software, L.L.; validation, L.L. and M.K.; formal analysis, L.L.; investigation, L.L.; resources, T.L.; data curation, L.L.; writing—original draft preparation, L.L.; writing—review and editing, all; visualization, L.L.; supervision, M.K.; project administration, L.L. and M.K.;

funding acquisition, M.K. and T.L. All authors reviewed the manuscript and agreed to the published version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests

The authors declare no competing interests.

Ethics declarations

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the German Psychological Society (Deutsche Gesellschaft für Psychologie, DGPs). For the children, study approval was also obtained from the Rhineland-Palatinate school authority (Aufsichts- und Dienstleistungsdirektion, ADD). Informed written consent was provided by the adult participants and by the legal guardians of the children.

Additional information

Correspondence and requests for materials should be addressed to L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

2.4 Study 3: Do Changing-State or Phonological Factors Cause Interference?

The results of Studies 1 and 2 revealed distinct processing patterns for steady-state and changing-state speech, as well as for environmental non-speech sounds, and narrative, unfamiliar speech. In Study 1, children were more impaired by both types of speech sounds than adults. In contrast, Study 2 showed a comparable age effect in the verbal task. Additionally, while neither children's nor adults' recall performance for spatial items was impaired by background sounds in Study 1, Study 2 demonstrated a detrimental impact of background sounds on third graders, whereas fourth graders and adults remained unaffected. These findings suggest that both specific interference and more general attentional disruption contribute to the observed patterns.

To further explore the roles of phonological processing, attention capture, and serial order retention, the following study investigates the influence of phonology and changing-state characteristics.

In Study 3, the background sounds were identical to Study 1, but this time the syllables were transformed into sinewave speech. This transformation enabled a direct comparison between the effects of unaltered speech sounds and their sinewave equivalents on serial order retention for both verbal and spatial items. By examining these variations, Study 3 aims to provide a clearer understanding of how both the changing-state effect and the phonological properties of task-irrelevant background speech influence serial recall.

From the changing-state account perspective, both unaltered changing-state syllables and their sinewave-transformed counterparts should impair serial order retention for both verbal and spatial items. However, if phonology-based interference plays a role, the disruptive effects should be observed for both steady-state and changing-state sequences, with natural speech being more disruptive than sinewave speech. Furthermore, this effect should only manifest in the verbal task, with no impairment expected for spatial items.

Study 3, being part of this thesis, is presented as an original manuscript (see Georgi et al., 2022) of an article published by Taylor & Francis in *Auditory Perception & Cognition* on 11 September 2022, available at: <https://doi.org/10.1080/25742442.2022.2127988>.



OPEN ACCESS



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

Markus Georgi^a, Larissa Leist^b, Maria Klatte^b and Sabine J. Schlittmeier^a

^aWork and Engineering Psychology, RWTH Aachen University, Aachen, Germany; ^bCognitive and Developmental Psychology, TU Kaiserslautern, Kaiserslautern, Germany

ABSTRACT

The empirically robust phenomenon of impaired verbal short-term memory due to task-irrelevant background speech is known as the Irrelevant Sound Effect (ISE). Yet, the underlying cognitive mechanisms are still not fully resolved, like whether phonology-based interference contributes to the disruptive effect of speech on verbal serial recall. Thus, we tested the ISE for such a speech-specific aspect by varying the phonological content of the irrelevant speech and the verbal nature of the task. Furthermore, we also varied the changing-state characteristic of irrelevant speech. Therefore, visual-verbal and visual-spatial serial recall of 70 participants each was measured during silence and syllable series varying in phonological (unaltered speech vs. sinewave speech) and in changing-state content (one repeatedly presented vs. successively changing syllables). Both phonology and changing-state characteristic had significant and independent impacts on each of the two serial recall tasks, and both effects were unrelated to participants' working memory capacity. Changing-state sequences were more disturbing than steady-state sounds for both serial recall tasks, but the effect of phonological content was not in the same direction for the two tasks. Whereas unaltered speech sounds impaired verbal serial recall more than their phonologically reduced sinewave versions, this was the opposite for visual-spatial serial recall. In the latter task, however, only changing-state sinewave speech reduced performance significantly compared to silence condition. In summary, our results indicate that there is a speech-specific, disruptive effect of phonological content, which is an attention-independent interference effect and presumably restricted to verbal short-term memory.

ARTICLE HISTORY

Received 22 November 2021
Accepted 11 September 2022

KEYWORDS

Irrelevant sound effect;
changing-state effect;
phonological interference;
verbal serial recall; visual-spatial short-term memory;
working memory capacity

Introduction

Background speech reduces cognitive performance in certain tasks, even if it is irrelevant to the task at hand or the performer of the task knows it should be ignored. This diminishing effect of irrelevant speech and certain non-speech sounds – known as the

CONTACT Markus Georgi  markus.georgi@psych.rwth-aachen.de  Work and Engineering Psychology, RWTH Aachen University, Aachen, Germany

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Irrelevant Sound Effect (ISE) – has been investigated in a multitude of studies (e.g., Jones et al., 2010; Page & Norris, 2003; for an overview see Schlittmeier et al., 2012). However, from a cognitive psychological perspective, it is not yet completely understood which characteristics of background speech are decisive for a specific cognitive function or process to be impaired. Thus, it is still an open question whether the ISE is driven by speech-specific characteristics, such as the phonology of the background speech or the short-term recall of verbal, rather than visual-spatial content, for example. The present study tests these two aspects: the background speech's phonology and the verbal nature of the serial recall task.

Research on the detrimental effect of irrelevant speech on verbal short-term memory dates back to a study by Colle and Welsh (1976). The authors used the visual-verbal serial recall task, which is the standard task to measure verbal short-term memory capacity. In this study, eight consonants were presented visually one after the other and had to be written down in the same order immediately after presentation. When foreign speech was played to the participants during this task, there was a 12% decrease in serial recall performance compared to the quiet control condition. The insight from this study was repeatedly confirmed empirically in later studies: background speech impairs verbal serial recall even if its semantic content is not accessible for the listener (e.g., Martin et al., 1988; Sörqvist et al., 2012). Thus, the decisive factor for the background speech's disturbance impact cannot be its semantic content.

In fact, since the 1980s, a large number of studies have shown that the disturbance impact is not limited to irrelevant speech but can be caused by non-speech sounds too: by music (e.g., Salamé & Baddeley, 1989; Schlittmeier et al., 2008) or by sequences of different tones (e.g., Jones & Macken, 1993; Macken et al., 2003). In order to elicit a disturbance impact, it has been shown to be crucial that the irrelevant sound is composed of distinct auditory-perceptual tokens that vary consecutively. An example would be a sequence of consecutively varying speech syllables (e.g., /de/ /ka/ /ef/ /te/ ...). These are known as changing-state sounds and reduce visual-verbal short-term memory significantly more than steady-state sounds; an example of the latter would be one repeatedly presented speech token (e.g., /de/ /de/ /de/ ...). Relevant studies here include those by Jones et al. (1992) and Jones et al. (1995).

The first explanation of this changing-state effect was the *Object-Oriented Episodic Record* (O-OER) model by Jones and coauthors (Jones, 1993; Jones et al., 1996; W. Macken et al., 1999). Here, the successively varying auditory-perceptive tokens of changing-state sounds are assumed to be encoded in short-term memory in terms of automatically lined-up objects. Performing a serial recall task necessitates the encoding of a further object-sequence for the memory items. Since the presence of order information in short-term memory results in the loss of other order information, serial recall performance is reduced during irrelevant changing-state sound. In contrast to changing-state sounds, steady-state sounds are encoded as a single object with only self-referential, and thus barely concurrent, order information. The detrimental impact of changing-state sounds – whether or not of speech origin – on verbal serial recall is thus attributed to interferences between seriation information. These interference processes, however, are specific neither to background speech, like phonology or semantics, nor to the verbal origin of the retention material: since the O-OER model represents an amodal conception of short-term memory that assumes functional equivalence between, for example,

verbal, visual, spatial codes (Jones et al., 1996), irrelevant speech that exhibits a changing-state characteristic should not only impair the serial recall of verbal material but also affect visual-spatial serial recall. In line with this, Jones et al. (1995) reported one experiment in which changing-state speech disturbed verbal as well as visual-spatial serial recall (Exp. 4).

This finding, and the conception of short-term memory that it supports, stands in sharp contrast to the considerations proposed some years earlier by Salamé and Baddeley (1987, 1989) to explain the ISE. They localized the disturbance effect of irrelevant speech on verbal serial recall in the phonological loop, which is the short-term memory module for verbal material in Baddeley's highly influential *Working Memory* model (Baddeley, 1986, 2000). However, processing and serial recall of visual-spatial items (like a series of positions) is assumed to be achieved by the visual-spatial sketchpad, which is the second short-term storage system within the Working Memory model. Since background speech is not encoded there, it should not reduce visual-spatial serial recall. Indeed, no direct replication has been reported so far for the finding by Jones et al. (1995, Exp. 4) of irrelevant speech impairing visual-spatial serial recall; instead, subsequent studies were not able to verify the detrimental impact of changing-state speech on visual-spatial serial recall (Bergström et al., 2012; Kvetnaya, 2018; Marsh et al., revision under review).

Beyond that, the assumption of an acoustic-phonological processing format of the phonological loop is per se speech-specific. Phonology is the study of the structure of speech sounds; it refers to the sound patterns of a given language and legal combinations of these to form words. In English, for example, /br/ as in /bread/ or /brew/ is a legal phoneme connection at the beginning of a word, while /bt/ is not. Phonological content is inherent in speech sound articulation. Acoustic resonances of the human vocal tract result in broad spectral maxima that are referred to as formants (Peterson & Barney, 1952). The first two formants are critical for vowel identification (Peterson, 1952). These formants are associated with movements of the oral cavity. For example, higher frequency of the first formant is associated with higher opening of the mouth necessary for the /a/ in German, whereas higher frequency of the second formant reflects higher arching of the tongue, e.g., used to build a German /i/ (Slawson, 1968).

The loop's acoustic-phonological processing format suggests that the phonology of the background speech underlies its disturbance impact or at least contributes to it. Although not further specified in the conception of the phonological loop, assuming a phonology-related interference mechanism seems reasonable – and still needs empirical testing. It could be considered that the involvement of phonological interference processes in the disturbance impact of irrelevant speech has already been refuted by studies showing the ISE to be independent of any phonological similarity between the irrelevant sound and the items to be retained (e.g., Jones & Macken, 1995; Larsen et al., 2000; LeCompte & Shaibe, 1997). However, these findings on irrelevance of the so-called the (phonological) between-stream similarity should not be overestimated in terms of their argumentative strength against the role of phonology in ISE evocation. First, the mathematically formulated *Primacy Model*, which is based on the phonological loop's conception, can simulate the ISE independently of the phonological similarity between irrelevant sound and items to be retained (see, Page & Norris, 2003, p. 1297). We will return to this model in the discussion. Second, interference mechanisms do not need to be content-based. Rather, the *Duplex-Mechanism Account of Auditory Distraction* (DMAAD; Hughes et al.,

2013; Hughes, Vachon et al., 2007) assumes interferences to occur when volitional task processing and obligatory, pre-attentive processing of the irrelevant sound “match” in terms of engaging the same cognitive processes (interference-by-process principle; Marsh et al., 2009). Thus, in case of process-based interference relying on phonological processing of both irrelevant speech and verbal items to be remembered, it is expected that irrelevant speech is less disruptive when its phonological content is reduced.

One way to reduce the phonological content of a background speech signal is by sinewave synthesis (Remez & Rubin, 1990; Remez et al., 1981). Here, the main frequency spectra (formants) of natural speech are replaced by sinusoidal tones. Although this kind of signal processing preserves the clear temporal structure and certain spectral regularities of natural speech, it reduces the spectral complexity and the temporal-spectral fine structure our speech production system generates. Since certain information-bearing aspects of the speech signal are still preserved after sinewave synthesis (see, Rosen & Hui, 2015), sinewave speech (SWS) can actually be intelligible for some listeners ad-hoc and for most after training. Nonetheless, it is uncontroversial to state that the phonological content of the speech signal is reduced.

In agreement with the assumption of phonological interference, SWS results in distinctly less disruption than natural speech (Tremblay et al., 2000), whereas foreign language and reversed speech are nearly as disturbing as their corresponding original speech sounds (Jones et al., 1990; Ueda et al., 2019). However, it could also be argued that SWS decreased the background sound’s acoustical complexity, and with this its changing-state content, thus causing less interference by seriation to occur.

Thus, the question arises whether there is a phonology-based disruptive effect that can be distinguished from the detrimental impact of the changing-state. If this is so, there is a further question of whether the performance-reducing impact of background speech is fed by these two interference effects of phonology and seriation changing-state features). There is scarce empirical support for the argument that phonological interference is independent of a sound’s changing-state characteristic. Cautious conclusions can perhaps be drawn from a study by Viswanathan et al. (2014) in which the authors presented participants with background speech conditions that differed in phonological content by comparing unaltered speech with SWS versions, with all being characterized by changing-state features. They found evidence for an independent effect of phonology, but this conclusion should be tested with an empirical study that includes both changing-state and steady-state versions of speech conditions that vary in phonological content.

Research Intent

In our present study, we disentangled disturbance effects due to either phonology or the changing-state characteristic on verbal serial recall by presenting changing-state and steady-state speech as both unaltered speech and SWS in a within-subject design. We hypothesize that both phonological and seriation interferences, the latter driven by a sound’s changing-state characteristic, occur in verbal short-term memory. Thus, we expect SWS with either changing- or steady-state characteristics to reduce performance in verbal serial recall less than the corresponding unaltered speech sounds. Since we assume that an independent changing-state effect is involved, both unaltered speech and

SWS with a changing-state characteristic should be more disruptive than their respective steady-state counterparts.

As a further touchstone for differentiating interference related to phonology from that related to the changing-state characteristic, we also employed a visual-spatial serial recall task in the present study. Since neither the encoding nor the serial recall of visual-spatial material encompasses a phonological processing component in the first place, any phonology-based interference mechanism should not be at work. Moreover, previous evidence on the impact of irrelevant speech in this task is very sparse and at the same time controversial (see, Jones et al., 1995, opposed to Bergström et al., 2012; Kvetnaya, 2018; and Marsh et al., revision under review).

Finally, the individual working memory capacity (WMC) of the participants was also measured to test whether any effects found vary with WMC. This was done to have a checkpoint for discriminating interference effects from attentional capture. These two mechanisms of action of irrelevant sound on cognitive performance can be distinguished according to the DMAAD (Hughes et al., 2007; Hughes et al., 2013). While interferences are assumed to be task-specific (interference-by-process principle; Marsh et al., 2009), attentional capture is assumed to vary with the individual's degree of cognitive control, which can be operationalized via the WMC (Engle, 2002). In fact, there is a group of attentional accounts of the ISE (Bell et al., 2019; Cowan, 1995; Röer et al., 2015) which attribute any disruptive effect of background sound to attentional distraction while interference mechanisms are considered to be dispensable in explaining the ISE.

Materials and Methods

Participants

We took two independent samples, each with 70 native German-speaking participants: one for the verbal and one for the visual-spatial short-term memory task. The sample for the verbal short-term memory task ranged in age between 17 and 40 years (median = 21 years; 44 women). For the visual-spatial short-term memory task, age ranged from 18 to 40 years (median = 21 years; 44 women). However, while completing with 70 participants for each of the two tasks, one participant in the verbal and four participants in the visual-spatial short-term memory task had to be left out as they could recall only 20% or less of the items correctly in at least one of the sound conditions (i.e., made 80% errors). Since we assume limited task understanding here, we excluded these participants.

All participants responded to an e-mail seeking participants inside and outside of RWTH Aachen University and were randomly assigned to one of the two samples. Participants received credit points or 8€/hour for participating in the experiment. Written informed consent was obtained from all participants. All participants reported normal hearing and normal or corrected to normal vision.

Apparatus

All tasks were programmed in PsychoPy v3.1.3 (Peirce, 2007) and presented on a 15.6-inch laptop. Headphones (Sennheiser HD 650) were connected to the notebook

indirectly via a Focusrite Scarlett 2i2 audio interface. The study was conducted in a soundproof booth (Studiobox Premium) at the Institute of Psychology of the RWTH Aachen University.

Stimuli

Sound Conditions

In addition to silence, four irrelevant sound conditions were presented, namely changing-state and steady-state sequences each in unaltered speech and SWS. Irrelevant unaltered changing-state speech consisted of a fixed sequence of 31 varying syllables (/be/, /ti/, /pe/, /bu/, /to/, . . . ; approx. 450–600 ms duration) spoken by a female voice and presented with a 200 ms interstimulus interval. The unaltered irrelevant steady-state speech was spoken by the same female voice and consisted of the syllable “bah” (450 ms duration), which was taken from the changing-state syllables and repeated 33 times with a 200 ms interstimulus interval. All used syllables consisted of phoneme combinations that are legal in German. Auditory material was recorded using an artificial head (HRS II.2, HEAD Acoustics GmbH) via Digital Audio Tape (DAT; 44,100 Hz sampling rate and 16 bit resolution). The sounds were edited with the software SoundEdit 16 (Macromedia, Inc.).

All recorded unaltered speech syllables were then processed in the phonetic software Praat (Boersma & Weenink, 2005) by sinewave synthesis in order to reduce the phonological content of the syllables. Specifically, the first three formants of the original unaltered speech syllables were replaced by time-varying sinusoidal tones tracking the speech signals' frequency and amplitude curves (cf, Remez et al., 1981). The changing-state and steady-state SWS sequences were then composed analogously to the unaltered speech sequences.

The loudness of the unaltered speech sounds was measured binaurally via an artificial head (HRS II.2, HEAD Acoustics GmbH) with a Sound Analyzer Type 110 in L_{Aeq} over 30 minutes. The changing-state speech sound reached $L_{Aeq} = 57$ dB(A), whereas steady-state speech reached $L_{Aeq} = 60$ dB(A). The sound files of the SWS sounds were calibrated with the sound analyzer via ITA-Toolbox (Dietrich et al., 2012) in MATLAB R2019b (MATLAB, 1984) to the same psychoacoustical loudness (DIN 45631, 1991) of their respective unaltered speech sound counterparts, namely $N = 8.8$ sone for both changing-state sounds and $N = 8.0$ sones for both steady-state sounds. However, despite having the same psychoacoustical loudness to the unaltered speech files, changing-state and steady-state SWS were a little louder reaching 58 dB(A) and 62 dB(A), respectively. When no irrelevant sound was played, the sound level was $L_{Aeq} = 28$ dB(A).

Verbal Short-Term Memory

Verbal short-term memory was assessed with a visual-verbal serial recall task. Here, 16 monosyllabic and phonologically dissimilar German nouns (e.g., Pilz (mushroom), Mond (moon), Eis (ice)) were represented by simple pictorial representations in a rectangular black frame (102 × 73 mm) in the center of a white screen. A pretest ensured that these were easily and unambiguously nameable. We used pictorial representation instead of written words so that the setup can also be used for child samples of different ages in the context of developmental studies of the ISE (see, Klatte et al., 2010;

Leist et al., in preaccepted for publication). A trial consisted of eight visual stimuli shown one after the other, for 1500 ms with a 500 ms interstimulus interval, at the center of the screen. The 16 visual stimuli were equally divided on all trials by quasi-randomization: each sound condition received the same number of any visual stimuli. It was not possible for the same visual stimulus to be shown twice within a trial.

Visual-Spatial Short-Term Memory

In a trial of the visual-spatial serial recall task, a total of seven black dots with 10 mm diameter appeared for 1500 ms with 500 ms interstimulus interval successively in different spatial positions on a white screen (304 × 183 mm surrounded by a black frame) in each trial. The screen allowed a total of 80 different dot positions (formed invisibly by 8 rows and 10 columns). The task difficulty can be held constant between all trials by a balanced complexity of the paths formed by the successive dots (Parmentier et al., 2005). Thus, the order of the dot positions for one block (12 trials for one sound condition) was mirrored either horizontally or vertically for each of two other sound condition, respectively, and shuffled within each mirrored position block again for another two sound conditions, to generate further lists with relatively equal task difficulty.

Working Memory Capacity

A version of the Operation Span (OSPAN) Task (Unsworth et al., 2005; first introduced by Turner & Engle, 1989) was used as a measure of WMC (Conway et al., 2005) in line with other studies on the ISE (e.g., Beaman, 2004; Elliott et al., 2020; Sörqvist, 2010b). In the OSPAN task, the participant has to maintain an increasing sequence of consonants (e.g., F, H, J) while judging mathematical equations (e.g., “3/3 + 2 = 1”) as being true or false between each consonant. Here, responses were given by pressing as fast as possible the corresponding key (“M” and “X” were used, on a computer keyboard with a German layout). In one trial, the presentation of one consonant (shown for 800 ms) and a mathematical equation (time limit of 10,000 ms to solve) alternated until the last consonant was shown. In the recall phase, the previously shown consonants needed to be typed in using the computer keyboard, in the same order in which they were presented before. Here, the OSPAN task consisted of three trials for each sequence length of consonants ranging from three up to seven consonants, respectively.

Procedure

Participants were tested individually. First, participants performed the 15 trials of the OSPAN task. Then they put on the headphones for the respective serial recall task that was either the verbal or the visual-spatial version. This was accompanied by five sound conditions altogether: changing-state unaltered speech, steady-state unaltered speech, changing-state SWS, steady-state SWS, and silence. The participants were made familiar with all five sounds by listening to them one after the other for 4000 ms each before starting the trials. Sound conditions varied trial to trial in quasi-randomized order. After all sound conditions were presented in random order, they were randomized again. Each sound condition was tested in 12 out of a total of 60 trials. The participants were told to ignore the sounds. The irrelevant sounds were presented during the item presentation

phase and in a 5000 ms retention interval between list presentation and recall phase. A random interval of sound presentation between 1200 and 1800 ms was introduced before the first list item appeared in the presentation phase. No sound was presented during the recall phase. In the verbal serial recall task, either the 31 changing-state syllables or 33 “bah” syllables occurred per trial for the changing-state or steady-state conditions, respectively. The same happened for the visual-spatial serial recall task but with three syllables less per sound condition due to the shorter trial length.

In the recall phase, all of the stimuli to be remembered were presented at the same time and the participant had to mouse-click all pictures/dots in the correct order as previously shown. Clicking an item changed its shading, indicating that it had been selected. There was no time limit in the recall phase. It was not possible to skip or choose a stimulus once more, or to correct errors by unclicking a stimulus. A visual cue appeared after the selection of the final item indicating the participant should proceed with the next trial by pressing the space bar. Practice trials were given to familiarize participants with the task before the respective testing trials started (OSPAN: two practice trials; verbal serial recall: five practice trials; visual-spatial serial recall task: three practice trials). The whole experiment took about 40 minutes to complete.

Statistical Analysis

For the serial recall tasks, each memory item (picture/dot) correctly recalled on its previously presented serial position was scored as a correct response. Repeated measures ANOVAs were conducted to test the disturbance impact of the different sound conditions in both serial recall tasks. We analyzed the sound effects on serial recall in three steps. First, we included the entire data set and thus all five sound conditions. Here, we checked the comparability of task difficulty of the two serial recall tasks and whether the different sound conditions had any effect at all compared to the silence condition since the latter is the standard baseline condition in ISE research. In the second step, we removed the silence condition from the data set and examined the effects of task, changing-state characteristic and phonological content according to the study’s factorial design. Finally, as a third step to test our ad hoc hypotheses, we examined the effects of phonological content and changing-state characteristics for each task separately.

The performance in WMC was assessed by the so-called all-or-nothing load scoring of the OSPAN values. For any trial, all consonants needed to be recalled in the correct order to get scoring points depending on the number of consonants, i.e., five correctly assigned consonants in a trial with five consonants resulted in 5 points, whereas less than five correctly assigned consonants resulted in 0 points. All points were summed up and divided again by the number of all consonants to get the individual overall mean score value (Conway et al., 2005; Hughes et al., 2013). At least 85% of the equations needed to be judged correctly to justify the ability to remember consonants as being a valid measurement of WMC. This criterion was fulfilled by all participants.

To test whether participants’ WMC contributed to the disturbance effects of phonological content and/or changing-state characteristic, we performed correlations following Sörqvist (2010a). Correlations were tested between OSPAN scores and difference scores of either unaltered speech and SWS or between changing-state and steady-state sound.

The difference scores representing an effect of a sound's phonological content were calculated by subtracting for each participant the mean correct percentage of the two unaltered speech conditions from the corresponding mean of the two SWS conditions. The difference scores representing an effect of changing-state characteristic were obtained in a similar way by subtracting the mean percentage of the correct responses of the two changing-state speech conditions from the corresponding mean of the two steady-state conditions. The correlations were first tested irrespectively of the serial recall task and then for each of the two serial recall tasks separately.

Results

Sound Effects on Verbal and Visual-Spatial Short-Term Memory

The proportions of correctly recalled items during the different sound conditions are depicted in Figure 1 for each of the two tasks. To start with, we performed a 2×5 ANOVA with the between-subjects factor *task* (verbal vs. visual-spatial) and the within-subject factor *sound condition* (changing-state speech, steady-state speech, changing-state SWS, steady-state SWS, silence). Due to non-sphericity of the variance-covariance matrix for this ANOVA, Greenhouse–Geisser corrections were applied to significant effects. The interaction of the two factors *sound condition* and *task* reached statistical significance, $F(3.75, 517.84) = 5.03$, $p = .001$, partial $\eta^2 = .04$, indicating that the sound conditions affected performance differently in verbal and visual-spatial serial recall. The main effect of *task* was not significant, $F(1, 138) = 2.85$, $p = .09$, partial $\eta^2 = .02$, indicating comparable overall task difficulty by appropriate item numbers for each task, i.e., 8 items in verbal serial recall and 7 items in visual-spatial serial recall. However, concerning *baseline* task difficulty, i.e., between both silence conditions, performance was significantly lower in visual-spatial than in verbal serial recall, $t(138) = 2.21$, $p = .03$, Cohen's $d = 0.37$. The main effect of *sound condition* was significant, $F(3.75, 517.84) = 10.33$,

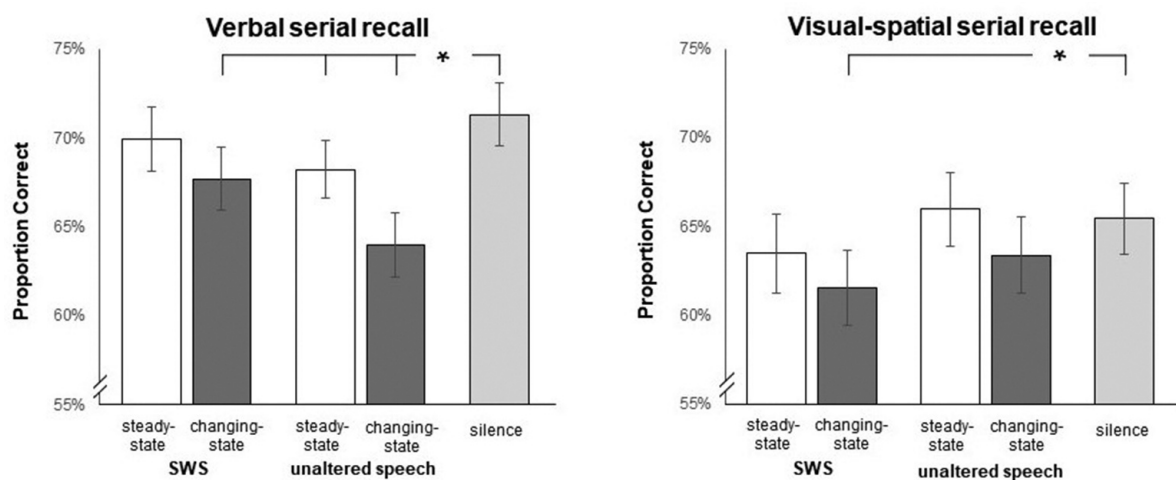


Figure 1. Percentages of correctly recalled items in the verbal (left panel) and visual-spatial (right panel) serial recall task as a function of the type of sound condition. SWS = sinewave speech. Error bars indicate the standard error of the mean. Effect indicators represent the irrelevant sound effect by comparison of all respective sound conditions with silence. $*p < .01$.

$p < .001$, partial $\eta^2 = .07$. It was further analyzed separately for verbal and visual-spatial serial recall because of the significant two-way interaction. Testing was limited to comparing the different irrelevant sounds with performance in silence as a standard baseline condition for the ISE. T-tests revealed that verbal serial recall was significantly reduced by the two changing-state sounds, i.e., unaltered speech, $t(69) = 6.58$, $p < .001$, Cohen's $d = 0.79$, and SWS, $t(69) = 3.59$, $p = .001$, Cohen's $d = 0.43$, and by the steady-state version of unaltered speech, $t(69) = 2.61$, $p = .01$, Cohen's $d = 0.31$ (see, [Figure 1](#), left panel). In contrast, visual-spatial serial recall was exclusively affected by changing-state SWS, $t(69) = 3.03$, $p = .003$, Cohen's $d = 0.36$ (see, [Figure 1](#), right panel). All other comparisons did not reach significance, $t(69) \leq 0.42$, p 's $\geq .68$.

To examine the differential effects of phonological and changing-state content on the two serial recall tasks, a $2 \times 2 \times 2$ ANOVA was conducted considering the between-subjects factor *task* (verbal vs. visual-spatial) and the two within-subject factors *phonology* (unaltered speech, SWS) and *state* (changing-state sequence, steady-state sequence). The three-way interaction of *task*, *phonology*, and *state* was not significant, $F(1, 138) = 0.34$, $p = .56$, partial $\eta^2 = .002$. Neither were the two-way interactions of *state* with either *task*, $F(1, 138) = 0.95$, $p = .33$, partial $\eta^2 = .007$, or *phonology*, $F(1, 138) = 1.41$, $p = .24$, partial $\eta^2 = .01$. These results show that, when considering both tasks simultaneously, the impact of changing-state and steady-state sounds did not vary with either the phonological content of the sound or the task type. However, the effect of irrelevant sound on verbal and visual-spatial serial recall depended on its phonological content: this is shown by the significant interaction of *phonology* and *task*, $F(1, 138) = 13.88$, $p < .001$, partial $\eta^2 = .09$. The main effect for *state*, $F(1, 138) = 29.52$, $p < .001$, partial $\eta^2 = .18$, but not the main effects for *task*, $F(1, 138) = 2.26$, $p = .14$, partial $\eta^2 = .02$, or *phonology*, $F(1, 138) = 0.18$, $p = .67$, partial $\eta^2 = .001$, reached statistical significance.

Because of a significant interaction with *task* as one factor being involved, and to test our hypotheses formulated specifically for verbal serial recall, we now consider each of the two experimental groups on its own. For that purpose, we performed a 2×2 ANOVA separately for verbal and visual-spatial serial recall, each with the two within-subject factors *phonology* (unaltered speech, SWS) and *state* (changing-state sequence, steady-state sequence).

For verbal serial recall, the 2×2 ANOVA indicated no interaction between *phonology* and *state*, $F(1, 69) = 1.59$, $p = .21$, partial $\eta^2 = .02$, but significant main effects of *phonology*, $F(1, 69) = 9.52$, $p = .003$, partial $\eta^2 = .12$, and *state*, $F(1, 69) = 29.22$, $p < .001$, partial $\eta^2 = .30$. Thus, *phonology* and *state* affected verbal serial recall independently of each other: regardless of phonological content, changing-state sounds ($M = 65.85\%$, $SD = 14.27\%$) were more disruptive than steady-state sounds ($M = 69.10\%$, $SD = 13.58\%$), and regardless of the changing-state characteristic, unaltered speech ($M = 66.12\%$, $SD = 13.99\%$) was more disruptive than SWS ($M = 68.83\%$, $SD = 14.38\%$) (see, [Figure 1](#), left panel). Thus, as specifically hypothesized in our research intent, we see that changing- and steady-state SWS reduced performance in verbal serial recall less than their corresponding unaltered speech sounds.

For the visual-spatial serial recall task, there was also no interaction between *phonology* and *state*, $F(1, 69) = 0.18$, $p = .67$, partial $\eta^2 = .003$, as indicated by the corresponding 2×2 ANOVA. And the main effect of *phonology*, $F(1, 69) = 4.97$, $p = .03$, partial $\eta^2 = .07$, as well as the main effect of *state* were also significant, $F(1, 69) = 7.66$, $p = .01$, partial

$\eta^2 = .10$. Looking further at the two main effects that lack interaction, we find that the main effect of *state* was significant because changing-state sounds ($M = 62.46\%$, $SD = 16.87\%$) were more disruptive than steady-state sounds ($M = 64.72\%$, $SD = 17.28\%$), regardless of phonological content. However, and most surprisingly, the main effect of *phonology* was due to the fact that SWS speech ($M = 62.51\%$, $SD = 17.43\%$) was more disruptive than unaltered speech ($M = 64.67\%$, $SD = 16.99\%$) (see, Figure 1, right panel), regardless of the changing-state characteristic. However, as revealed at the very beginning of our data analyses, changing-state SWS was the only irrelevant sound in the visual-spatial serial recall task that differed significantly from performance during silence.

Sound Effects and Working Memory Capacity

Finally, we tested whether participants' WMC, as indicated by performance in the OSPAN task, was related to the disturbance effects of phonological content and/or changing-state characteristic. The overall mean OSPAN score of both experimental groups was $M = .58$ ($SD = .21$), ranging from .00 to 1.00. The overall mean difference score between performance during unaltered speech and SWS sounds irrespective of serial recall task was $M = 0.28\%$ ($SD = 8.09\%$); the mean difference between steady-state and changing-state sound conditions reached $M = 2.76\%$ ($SD = 6.00\%$). There were no significant correlations with the OSPAN scores, neither with the phonological difference scores, $r(138) = .04$, $p = .61$, nor with the state difference scores, $r(138) = .01$, $p = .89$.

Considering the two experimental groups separately, the participants specifically working on verbal serial recall reached for the OSPAN scores a mean performance value of $M = .60$ ($SD = .20$), ranging from .12 to 1.00 (see, Figure 2, left panels). The mean difference score of performance between unaltered speech and SWS was 2.72% ($SD = 7.36\%$), as it was in turn 3.25% ($SD = 5.03\%$) for the difference between steady- and changing-state conditions. Neither the phonology difference scores, $r(68) = -.02$, $p = .88$, nor the state difference scores, $r(68) = .09$, $p = .48$, correlated with the OSPAN scores.

For the participants working on the visual-spatial serial recall task, the mean value for the OSPAN score was $M = .56$ ($SD = .22$), ranging from .00 to .92 (see, Figure 2, right panels). The mean difference score of performance between unaltered speech and SWS, and between steady- and changing-state conditions reached $M = -2.16\%$ ($SD = 8.11\%$) and $M = 2.26\%$ ($SD = 6.84\%$), respectively. In this experimental group, too, there were no significant correlations of OSPAN scores either with phonology difference scores, $r(68) = .04$, $p = .73$, or with the state difference scores, $r(68) = -.05$, $p = .67$.

In total, these results indicate that WMC had no differential contribution to the effects of an irrelevant sound's phonological content or changing-state characteristic on either verbal or visual-spatial serial recall.

Discussion

In our study, we examined visual-verbal and visual-spatial short-term memory during background speech by varying the speech's phonological content and its changing-state characteristic. Foremost, we wanted to test whether phonology-based interference contributes to the disruptive effect of speech on verbal serial recall. The phonological effect

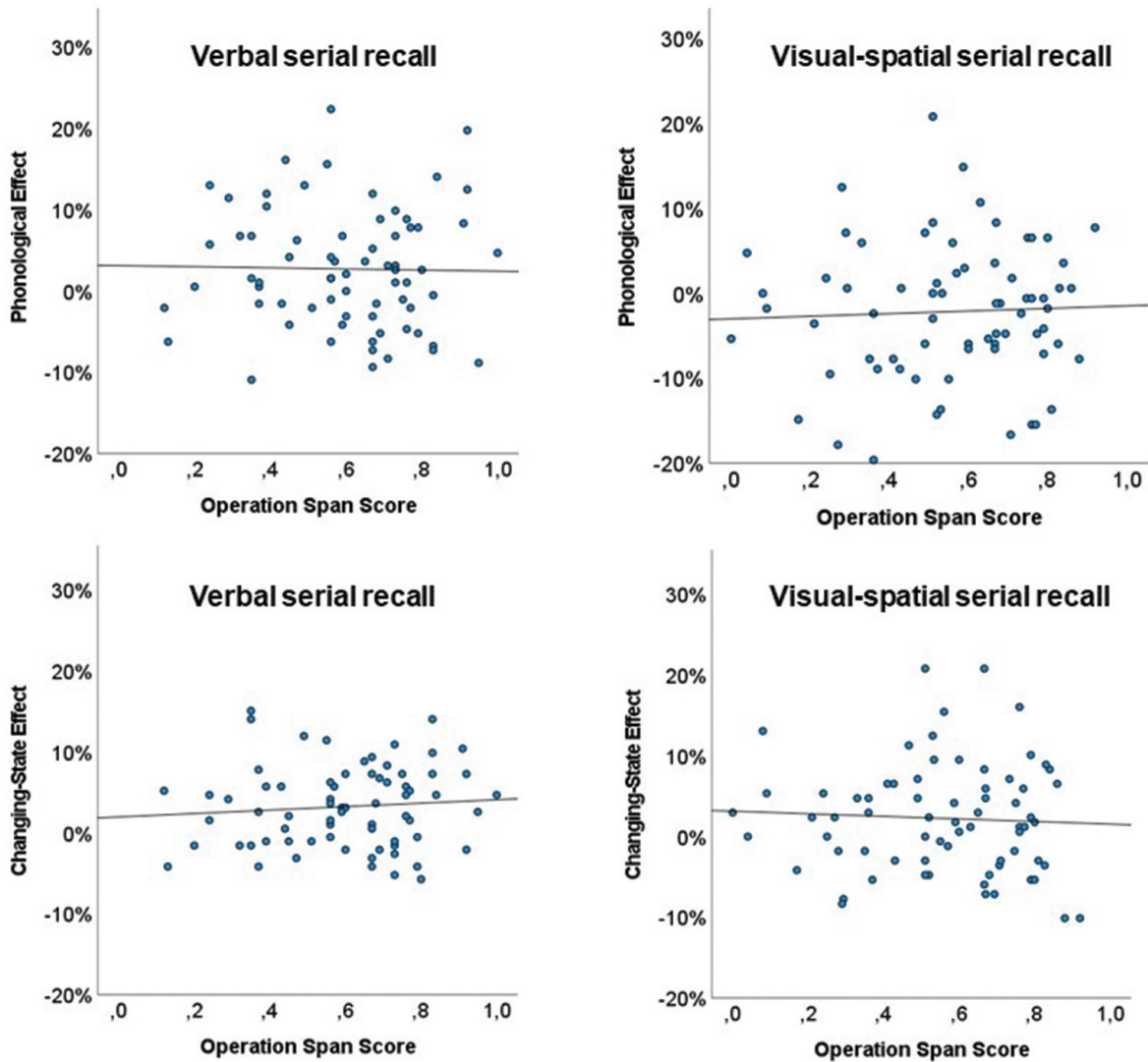


Figure 2. Scatter plot for OSPAN score and the amount of disruption by phonological/changing-state content measured by difference scores with SWS minus unaltered speech (phonological effect; upper panels) and with steady-state minus changing-state sound conditions (changing-state effect; lower panels) for both verbal and visual-spatial serial recall.

was examined by comparing the unaltered versions of the syllable sequences with their phonologically reduced SWS versions, while in turn for the changing-state effect, the steady-state versions of the syllable sequences were compared with their changing-state versions in a within-subject design. As we consider an effect of phonological content to be independent of the changing-state effect, it should occur for changing-state as well as for steady-state sequences. To further elucidate the mechanisms of action on which a potentially given detrimental sound effect is based, we pursued two further experimental strategies. On the one hand, we explored the impact of the same irrelevant sounds also in a visual-spatial serial recall task lacking any phonological processing component to test in particular for phonology-based interference processes. On the other hand, the individual WMC of the participants as a measure for cognitive control was tested for its potential involvement in any given sound effect. Such a correspondence would point to attentional capture being involved instead of interference mechanisms.

According to our main research interest, we start by considering the empirical results in the verbal serial recall task. Here, the present study verified detrimental effects of both phonological and changing-state content. In fact, unaltered speech impaired verbal serial recall significantly more than SWS (effect of phonological content). Furthermore, the changing-state sequences diminished verbal serial recall to a greater extent than the steady-state sequence (effect of changing-state content). Moreover, these two effects were independent of each other. Thus, phonological content and changing-state characteristic contributed individually – if inherent in the background sound’s characteristics – to a given performance reduction in verbal serial recall. According to the rationale put forward by Sternberg (1969), such an additivity of two effects indicates that they are underpinned by different mechanisms.

Please note, that this finding is not blurred by the fact that performance under silence was significantly higher in the verbal than in visual-spatial serial recall task. The absolute difference between the two baseline conditions (71% vs. 65%) can be considered to be practically irrelevant since for such performance baselines, the performance impairments caused by irrelevant sound are hardly masked by floor or ceiling effects. Furthermore, it is difficult to match performance baselines even better in different serial recall tasks because the task can only be varied in integer steps with respect to the number of items.

Our finding of a special role for phonology as a speech-specific sound feature is in line with a previous study by Dorsi et al. (2018). The authors tested whether irrelevant speech’s fidelity (i.e., speech-likeness) and/or signal complexity (i.e., changing-state characteristic) contribute to its disturbance impact on visual-verbal serial recall. Signal complexity was varied by using noise-vocoded speech and varying the number of vocoding channels. Additionally, speech fidelity of the resulting speech signals was varied by time-reversing a selected subset of channels. In this way, the typical signal-based regularities of irrelevant speech, due to the co-vocalization of speech sounds and biological restrictions on voice production, were reduced. Dorsi et al. (2018) found speech fidelity – and thus speech alikeness in a signal-oriented view – to be relevant for the speech signal’s disturbance impact besides signal complexity and thus changing-state content. However, by maintaining changing words as irrelevant speech in all tested sound conditions, Dorsi et al. (2018) actually did not use a definite steady-state sound version to ultimately compare with separated changing-state quality (see, also Viswanathan et al., 2014). In the present study, however, a steady-state sound was realized through a repetitively presented syllable next to successively changing syllables as a changing-state sound. This is a common approach in studies on the ISE to generate irrelevant speech sounds systematically varied by changing-state and steady-state characteristics (e.g., by using *words*: Campbell et al., 2002; Neely & LeCompte, 1999; Salamé & Baddeley, 1982; Tremblay et al., 2000; by using *syllables*: Hughes, Tremblay et al., 2005; Jones et al., 1992; Jones & Macken, 1993; LeCompte, 1995).

To validate the found disturbance impacts of changing-state characteristic and phonological content as being assignable to interference mechanisms only, and that attentional capture by the irrelevant sound is not involved here, we tested the participants’ individual WMC for correlation with the difference scores between the different sound conditions. Such a correlation would suggest that individual attentional control underlies a given detrimental sound effect – or is at least involved in its evocation (cf. the DMAAD; Hughes et al., 2007, Hughes et al., 2013). Nevertheless, we found no association of WMC

either with the disturbance impact of phonological content or with that of a sound's changing-state characteristic. Regarding the changing-state effect, these results are in line with prior studies that found WMC to be unrelated to an irrelevant changing-state sound's disturbance impact on visual-verbal serial recall (Beaman, 2004; Körner et al., 2017; Sörqvist, 2010a, 2010b; for a meta-analysis cf, Sörqvist et al., 2013). Our study expands this finding to the detrimental impact induced by irrelevant speech's phonological content. Thus, our findings argue that the detrimental impact of both a sound's changing-state characteristic and its phonological content are specific and independent interference processes instead of the more general mechanism of attentional capture.

Further clarification of the two interference mechanisms was made by also testing the given sound effects in visual-spatial serial recall as a task with no phonological processing components. Here, there were significant effects of both phonological and changing-state content, too, but for phonology, it was the opposite direction, with SWS being more disruptive than phonologically unaltered speech. One could hypothesize that these SWS sounds were very unusual, even irritating, as reported by participants from both serial recall tasks, and, in contrast to the other sounds, led to a distraction of attention in the visual-spatial serial recall. In verbal serial recall, on the other hand, this type of distraction of attention by these “alien” sounds could have been “masked” by the interference effects of phonological and changing-state content that actually occur there. However, this consideration of irritation or attentional distraction is contradicted by the fact that even in visual-spatial serial recall, individual WMC showed no correlation with individual performance differences between SWS and unaltered speech on the one hand, or changing-state and steady-state sounds on the other. Finally, we would like to address a speculative idea related to the visual-spatial aspects of this serial memory task. It might be that natural speech is more associated with a specific spatial position, namely the position of the talker, while SWS, which is usually not identified as speech, might not appear comparably locally bound. Therefore, SWS might be more challenging for visual-spatial short-term memory than unaltered speech. Whether spatial uncertainty or even changing spatial positions of the sound source (e.g., the talker's location changing after each syllable) affects visual-spatial serial recall more than serial recall without a spatial component could be the subject of future investigation.

As far as we know, this is the first published study examining the possible disturbance of phonological content of background speech on visual-spatial short-term memory. As we have found no such impairment here, we suppose that phonological content may be exclusively disturbing for verbal short-term memory. Our findings support the Working Memory model (Baddeley, 1986, 2000) which assumes irrelevant speech to be exclusively processed in the phonological loop as one short-term store for verbal material, while it gains no access to the visual-spatial sketchpad as the other separated short-term store in this model. Consequently, the model correctly predicts that irrelevant speech is disruptive to short-term recall of visual-verbal but not visual-spatial items. Furthermore, the phonological loop is assumed to operate on an acoustic-phonological processing format which makes it reasonable to assume that phonological content of the background speech is also relevant for its disturbance impact. The latter, however, does not apply to the O-OER model (Jones, 1993; Jones et al., 1996; Macken et al., 1999), which considers exclusively the changing-state characteristic of an irrelevant sound as decisive for its disturbance impact on serial recall. Our findings contradict this assumption of the O-OER model.

However, our study seems to be one of the rare published studies replicating the changing-state effect of Jones et al. (1995, Exp. 4) in visual-spatial serial recall, namely reduced performance during irrelevant changing-state sound than during corresponding steady-state sound. Thus, it contrasts with extant studies of other authors who found no changing-state effect in visual-spatial serial recall (Bergström et al., 2012; Kvetnaya, 2018; Marsh et al., revision under review). Yet our study results also present themselves as less straightforward when considering that only one of the two changing-state conditions disrupted serial recall performance compared with the silence condition, i.e., elicited an ISE at all. Although our results seem to be more in line with the O-OER model that expects that irrelevant changing-state sounds equally affect serial recall independent of modality (e.g., verbal information, spatial information) due to functional equivalence of verbal and non-verbal material in an amodal conception of short-term memory (Jones, 1993; Jones et al., 1996; W. Macken et al., 1999), further studies are needed in this perspective.

Although the interference mechanisms imposed by irrelevant speech are not further detailed within the Working Memory model (Baddeley, 1986, 2000), they are specified in the *Primacy Model* proposed by Page and Norris (1998). This model is based on the phonological loop's architecture and thus takes the “speech is special” view, but it is able to account for the changing-state effect at the same time. The primacy model assumes that performing a serial recall task during irrelevant sound results in the encoding of order information for the to-be-remembered item sequence as well as for the background sound, analogous to the O-OER model. During changing-state sounds, however, higher order information is encoded than during steady-state sounds. The changing-state effect is due to the boosted encoding of order information during changing-state sound that consumes more cognitive resources and additionally leads to more noise among the order information encoded for the item sequence (cf. Page & Norris, 2003, p. 1297, as well; Norris et al., 2004). Because the phonological content of the background sound does not play a role here, the primacy model can correctly model the irrelevance of the (phonological) between-stream similarity (see Introduction), while at the same time it must be assumed that this model cannot cover the phonological interference effect found for verbal serial recall in the present study.

It should be noted that we varied the two sound characteristics of changing-state and phonological content in two rather extreme levels, possibly suggesting a dichotomy of these sound characteristics. However, it can be assumed that these two sound characteristics – and thus possibly also their disturbance impact on verbal serial recall – actually have a dimensional character. Several empirical studies manipulated the irrelevant speech signal in several steps, with corresponding gradual variations of its disturbance impact on verbal serial recall being observed. A reduction in the impact of irrelevant speech has been found when the extent of low-pass filtering (Jones et al., 2000) or the level of continuous noise used for superimposition was increased (Ellermeier & Hellbrück, 1998). Both manipulations, however, not only reduced the acoustic mismatch between successive auditory-perceptive tokens of the irrelevant speech signal, and with this, its changing-state characteristic, but also affected its phonological content. We like to believe that in our study, we managed to vary changing-state and phonological content largely independently of each other. This is supported by the fact that the fluctuation strengths of the two changing-state sounds were very similar to each other (speech: 1.22 vacil; SWS: 1.11 vacil; measured by the sound and vibration analysis software ArtemiS SUITE, HEAD Acoustics GmbH) as to be considered negligibly different for the

magnitude of the interference effect in an ISE modeling study (cf, Schlittmeier et al., 2012). The hearing sensation fluctuation strength describes the percept of fluctuations when listening to slowly modulated sounds (f_{mod} [modulation frequency] <20 Hz) like, for example, narration. It has been found to be an instrumental measure to estimate the disturbance impact of most irrelevant speech and non-speech sounds on visual-verbal serial recall (Schlittmeier et al., 2012). Therefore, fluctuation strength has been said to capture the changing-state content of an irrelevant sound, as it indicates the temporal structure and the prominence of variations in the frequency spectrum, rather than the properties that are exclusive to speech like, e.g., phonology. However, in the aforementioned modeling study with 70 behavioral measurements of visual-verbal serial recall on over 40 irrelevant sounds, it was noticed that the sounds with the highest disruptive effect – and thus the strongest ISE – were almost exclusively speech sounds (Schlittmeier et al., 2012). Our findings contribute to the perspective that this may be the case because speech signals best fulfill the criteria for irrelevant sounds to gain automatic and obligatory access to short-term memory or, in the perspective of the DMAAD (Hughes et al., 2005, Hughes et al., 2013), automatic processing of speech interfere to the highest extent with cognitive processes involved in verbal serial recall. Such criteria might include clear temporal structures and significant changes in spectrum (i.e., changing-state characteristic), but also the presence of phonology, as our results suggest.

In summary, our findings argue for the irrelevant speech effect being driven by certain speech-specific characteristics, such as the phonology and changing-state character of the speech sound. They are also informative for weighing the various proposed conceptions of short-term memory. Even though performance under silence was significantly higher in verbal than in visual-spatial short-term memory, we consider the absolute difference between the two baseline conditions (71% vs. 65%) to be practically irrelevant. First, because for such performance baselines, the performance impairments caused by irrelevant sound are hardly masked by floor or ceiling effects. Second, it should be noted that it is difficult to match performance baselines even better in different serial recall tasks because the task can only be varied in integer steps with respect to the number of items. Nonetheless, the independent role of phonology-based interference processes in the presence of irrelevant speech certainly requires further empirical clarification, for example, through the use of phonology-based short-term memory tasks that do not require serial processing.

Acknowledgments

The authors thank Plamenna Koleva, Jian Pan, Jonathan Plaß, Ann-Sophie Schenk, and Wiebke Stöver for data collection and for preparing the datasets.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This research was financially supported by the Deutsche Forschungsgemeinschaft [DFG, German Research Foundation; SCHL 1876/4, KL 1305/4]. Sabine Schlittmeier's contribution was supported

by a grant from the HEAD-Genuit-Stiftung (Head-Genuit Foundation; P-16/10-W). The software platform ArtemiS SUITE for sound and vibration analysis was kindly provided by the HEAD Acoustics GmbH.

References

- Baddeley, A. (1986). *Working memory*. Oxford University Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Beaman, C. P. (2004). The irrelevant sound effect revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1106–1118. <https://doi.org/10.1037/0278-7393.30.5.1106>
- Bell, R., Röer, J. P., Lang, A.-G., & Buchner, A. (2019). Reassessing the token set size effect on serial recall: Implications for theories of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 45(8), 1432–1440 doi:10.1037/xlm0000658.
- Bergström, K., Lachmann, T., & Klatte, M. (2012). Wann stört Lärm das geistige Arbeiten? Einfluss von Aufgaben- und Geräuschcharakteristiken bei der Wirkung moderaten Lärms auf Arbeitsgedächtnisleistungen [proceeding]. *Jahrestagung für Akustik (DAGA)*, Vol. 38, Darmstadt, Germany: Deutsche Gesellschaft für Akustik. https://pub.dega-akustik.de/DAGA_2012/data/articles/000316.pdf
- Boersma, P., & Weenink, K. (2005). *Praat: Doing phonetics by computer*. (<https://www.fon.hum.uva.nl/praat>)
- Campbell, T., Beaman, C. P., & Berry, D. C. (2002). Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory*, 10(3), 199–214. <https://doi.org/10.1080/09658210143000335>
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, 15(1), 17–32. [https://doi.org/10.1016/S0022-5371\(76\)90003-7](https://doi.org/10.1016/S0022-5371(76)90003-7)
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Negle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. <https://doi.org/10.3758/BF03196772>
- Cowan, N. (1995). *Oxford psychology series*, No. 26. *Attention and memory: An integrated framework*. Oxford University Press.
- Dietrich, P., Guski, M., Pollow, M., Müller-Trapet, M., Masiero, B., Scharrer, R., & Vorländer, M. (2012). ITA-Toolbox – An open source MATLAB toolbox for acousticians [proceeding]. *Jahrestagung für Akustik (DAGA)*, Vol. 38, Darmstadt, Germany: Deutsche Gesellschaft für Akustik. https://pub.dega-akustik.de/DAGA_2012/data/index.html
- DIN 45631. (1991). *Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum – Verfahren nach E. Zwicker* (Berlin: Beuth).
- Dorsi, J., Viswanathan, N., Rosenblum, L. D., & Dias, J. W. (2018). The role of speech fidelity in the irrelevant sound effect: Insights from noise-vocoded speech backgrounds. *Quarterly Journal of Experimental Psychology*, 71(10), 2152–2161. <https://doi.org/10.1177/1747021817739257>
- Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in 'Irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception & Performance*, 24(5), 1406–1414 doi:10.1037//0096-1523.24.5.1406.
- Elliott, E. M., Marsh, J. E., Zeringue, J., & McGill, C. I. (2020). Are individual differences in auditory processing related to auditory distraction by irrelevant sound? A replication study. *Memory & Cognition*, 48(2), 145–157. <https://doi.org/10.3758/s13421-019-00968-8>
- Engle, R. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control auf auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 539–553. <https://doi.org/10.1037/a0029064>

- Hughes, R. W., Tremblay, S., & Jones, D. M. (2005). Disruption by speech of serial short-term memory: The role of changing-state vowels. *Psychonomic Bulletin and Review*, 12(5), 886–890. <https://doi.org/10.3758/BF03196781>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 736–749. <https://doi.org/10.1037/0278-7393.31.4.736>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6), 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention, awareness, and control*. Oxford University Press, 87–104.
- Jones, D. M., Alford, D., Macken, W. J., Banbury, S., & Tremblay, S. (2000). Interference from degraded auditory stimuli: Linear effects of changing-state in the irrelevant sequence. *The Journal of the Acoustical Society of America*, 108(3), 1082–1088. <https://doi.org/10.1121/1.1288412>
- Jones, D. M., Beaman, P., & Macken, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 209–238). Psychology Press.
- Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 1008–1018. <https://doi.org/10.1037/0278-7393.21.4.1008>
- Jones, D. M., Hughes, R. W., & Macken, W. J. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise & Health*, 12(49), 201–209. <https://doi.org/10.4103/1463-1741.70497>
- Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 369–381 doi:10.1037/0278-7393.19.2.369.
- 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), 103–115 doi:10.1037/0278-7393.21.1.103.
- Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology*, 44(4), 645–669. <https://doi.org/10.1080/14640749208401304>
- Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proof-reading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, 4(2), 89–108. <https://doi.org/10.1002/acp.2350040203>
- Klatte, M., Lachmann, T., Schlittmeier, S. J., & Hellbrück, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology*, 22(8), 1168–1191. <https://doi.org/10.1080/09541440903378250>
- Körner, U., Röer, J. P., Buchner, A., & Bell, R. (2017). Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. *Journal of Memory and Language*, 96(1), 122–137. <https://doi.org/10.1016/j.jml.2017.05.005>
- Kvetnaya, T. (2018). Registered replication report: Testing disruptive effects of irrelevant speech on visual-spatial working memory. *Journal of European Psychology Students*, 9(1), 10–15. <https://doi.org/10.5334/jeps.450>
- Larsen, J. D., Baddeley, A., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory*, 8(3), 145–157. <https://doi.org/10.1080/096582100387579>
- LeCompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin & Review*, 2(3), 391–397. <https://doi.org/10.3758/BF03210978>
- LeCompte, D. C., & Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 50(1), 100–118. <https://doi.org/10.1080/027249897392242>

- Leist, L., Lachmann, T., Schlittmeier, S. J., Georgi, M., & Klatte, M. (accepted for publication). *Effects of irrelevant speech on serial recall of verbal and spatial materials in children and adults*, *Memory & Cognition*.
- Macken, W., Tremblay, S., Alford, D., & Jones, D. M. (1999). Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. *International Journal of Psychology*, 34(5/6), 322–327. <https://doi.org/10.1080/002075999399639>
- Macken, W. J., Tremblay, S., Houghton, R. H., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, 29(1), 43–51. doi:10.1037/0096-1523.29.1.43.
- 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., Hurlstone, M. J., Marois, A., Ball, L. J., Moore, S. B., Vachon, F., Schlittmeier, S. J., Röer, J. P., Buchner, A., & Bell, R. (revision under review). *Changing-state irrelevant speech disrupts visual-verbal but not visual-spatial serial recall*, *Journal of Experimental Psychology: Learning, Memory and Cognition*.
- Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, 27(4), 382–398. [https://doi.org/10.1016/0749-596X\(88\)90063-0](https://doi.org/10.1016/0749-596X(88)90063-0)
- MATLAB. (1984). MATLAB. The MathWorks Inc.
- Neely, C. B., & LeCompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & Cognition*, 27(1), 37–44. <https://doi.org/10.3758/BF03201211>
- Norris, D., Baddeley, A. D., & Page, M. P. A. (2004). Retroactive effects of irrelevant speech on serial recall from short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1093–1105. <https://doi.org/10.1037/0278-7393.30.5.1093>
- Page, M. P. A., & Norris, D. G. (1998). Modeling immediate serial recall with a localist implementation of the primacy model. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 227–255). Erlbaum.
- Page, M. P. A., & Norris, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *The Quarterly Journal of Experimental Psychology*, 56(8), 1289–1300. <https://doi.org/10.1080/02724980343000233>
- Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 412–427. <https://doi.org/10.1037/0278-7393.31.3.412>
- Peirce, J. W. (2007). PsychoPy – psychophysics software in python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Peterson, G. E. (1952). The information-bearing elements of speech. *The Journal of the Acoustical Society of America*, 24(6), 629–637. <https://doi.org/10.1121/1.1906945>
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Remez, R. E., & Rubin, P. E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics*, 48(4), 313–325. <https://doi.org/10.3758/BF03206682>
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212(4497), 947–949. <https://doi.org/10.1126/science.7233191>
- Röer, J. P., Bell, R., & Buchner, A. (2015). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception & Performance*, 41(3), 692–702 doi:10.1037/xhp0000028.
- Rosen, S., & Hui, S. N. (2015). Sine-wave and noise-vocoded sine-wave speech in a tone language: Acoustic details matter. *The Journal of the Acoustical Society of America*, 138(6), 3698–3702. <https://doi.org/10.1121/1.4937605>

- 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/10.1016/S0022-5371\(82\)90521-7](https://doi.org/10.1016/S0022-5371(82)90521-7)
- Salamé, P., & Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, 30(8), 1185–1193. <https://doi.org/10.1080/00140138708966007>
- Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology*, 41(1), 107–122. <https://doi.org/10.1080/14640748908402355>
- Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008). Can the irrelevant speech effect turn into a stimulus suffix effect? *Quarterly Journal of Experimental Psychology*, 61(5), 665–673. <https://doi.org/10.1080/17470210701774168>
- Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, and Psychophysics*, 74(1), 194–203. <https://doi.org/10.3758/s13414-011-0230-7>
- Slawson, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *The Journal of the Acoustical Society of America*, 43(1), 87–101. <https://doi.org/10.1121/1.1910769>
- Sörqvist, P. (2010a). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition*, 38(5), 651–658. <https://doi.org/10.3758/MC.38.5.651>
- Sörqvist, P. (2010b). The role of working memory capacity in auditory distraction: A review. *Noise & Health*, 12(49), 217–224. <https://doi.org/10.4103/1463-1741.70500>
- Sörqvist, P., Marsh, J. E., & Nöstl, A. (2013). High working memory capacity does not always attenuate distraction: Bayesian evidence in support of the null hypothesis. *Psychonomic Bulletin & Review*, 20(5), 897–904. <https://doi.org/10.3758/s13423-013-0419-y>
- Sörqvist, P., Nöstl, A., & Halin, N. (2012). Disruption of writing processes by the semanticity of background speech. *Scandinavian Journal of Psychology*, 53(2), 97–102. <https://doi.org/10.1111/j.1467-9450.2011.00936.x>
- Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57(4), 421–457. <https://www.jstor.org/stable/27828738>
- Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1750–1754. <https://doi.org/10.1037//0278-7393.26.6.1750>
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, 28(2), 127–154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Ueda, K., Nakajima, Y., Kattner, F., & Ellermeier, W. (2019). Irrelevant speech effects with locally time-reversed speech: Native vs non-native language. *The Journal of the Acoustical Society of America*, 145(6), 3686–3694. <https://doi.org/10.1121/1.5112774>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. <https://doi.org/10.3758/BF03192720>
- Viswanathan, N., Dorsi, J., & George, S. (2014). The role of speech-specific properties of the background in the irrelevant sound effect. *The Quarterly Journal of Experimental Psychology*, 67(3), 581–589. <https://doi.org/10.1080/17470218.2013.821708>

2.5 Study 4: Comparing Monaural and Binaural Sound Presentation

Study 4 (Exp. 2) aims to explore a realistic auditory environment in which children and adults perform cognitive tasks. This approach demonstrates how auditory distractions in familiar contexts can influence performance, thereby making the results more applicable to everyday situations, such as learning in classrooms.

Furthermore, Study 4 (Exp. 2) investigates whether varying the spatial information of sound sources affects serial recall performance. Specifically, monaural and binaural presentation of a realistic classroom-noise scenario is compared in a visual-verbal serial recall task with children and adults.

Most studies on the irrelevant sound paradigm use monaural presentation, which oversimplifies binaural hearing's spatial processing and ignores the complexity of real-world environments where sounds are spatially distributed. Study 4 addresses this by using binaural presentation to better reflect real-world listening conditions. To achieve this, we designed a classroom-like auditory scenario that consists everyday classroom activities, such as turning the pages of a book, footsteps, furniture use, writing, using zippers, as well as four children speaking in an unfamiliar language.

When comparing monaural and binaural sound presentation, we expect binaural conditions to have a more pronounced impact due to two key factors: (1) increased speech-based interference from a more distinct separation of speech streams, and (2) heightened attention capture resulting from the spatial distribution and shifting locations of sound sources. Additionally, if attention capture is the primary driver of disruption, we anticipate greater impairment in children compared to adults, especially under binaural presentation versus monaural conditions.

Furthermore, if attention capture is the primary factor behind disruption, we expect that habituation will occur over time. In contrast to previous studies (Study 1, 2, and 3), where sound conditions varied from trial to trial, allowing at most two consecutive trials with the same sound, the classroom-like noise scenario in this study was presented continuously within each sound condition across multiple trials. This continuous exposure would allow for habituation, reducing the disruptive effects as participants become accustomed to the sounds, suggesting that the initial disruption stems from the novelty and variability of the auditory stimuli.

Study 4, being part of this thesis, is presented as an original manuscript (see Leist et al., 2022) of an article published by MDPI in *International Journal of Environmental Research and Public Health* on 30 November 2022, available at: <https://doi.org/10.3390/ijerph192315998>.



Article

Differential Effects of Task-Irrelevant Monaural and Binaural Classroom Scenarios on Children's and Adults' Speech Perception, Listening Comprehension, and Visual–Verbal Short-Term Memory

Larissa Leist ^{1,*}, Carolin Breuer ², Manuj Yadav ², Stephan Fremerey ³, Janina Fels ², Alexander Raake ³, Thomas Lachmann ^{1,4}, Sabine J. Schlittmeier ⁵ and Maria Klatte ¹

¹ Cognitive and Developmental Psychology Unit, Center for Cognitive Science, Department of Psychology, University of Kaiserslautern-Landau, 67663 Kaiserslautern, Germany

² Institute for Hearing Technology and Acoustics, RWTH Aachen University, 52074 Aachen, Germany

³ Audiovisual Technology Group, Technische Universität Ilmenau, 98693 Ilmenau, Germany

⁴ Centro de Investigación Nebrija en Cognición, Facultad de Lenguas y Educación, Universidad Nebrija, 28015 Madrid, Spain

⁵ Teaching and Research Area of Work and Engineering Psychology, RWTH Aachen University, 52066 Aachen, Germany

* Correspondence: lleist@rhrk.uni-kl.de



Citation: Leist, L.; Breuer, C.; Yadav, M.; Fremerey, S.; Fels, J.; Raake, A.; Lachmann, T.; Schlittmeier, S.J.; Klatte, M. Differential Effects of Task-Irrelevant Monaural and Binaural Classroom Scenarios on Children's and Adults' Speech Perception, Listening Comprehension, and Visual–Verbal Short-Term Memory. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15998. <https://doi.org/10.3390/ijerph192315998>

Academic Editor: Paul B. Tchounwou

Received: 31 October 2022

Accepted: 26 November 2022

Published: 30 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Most studies investigating the effects of environmental noise on children's cognitive performance examine the impact of monaural noise (i.e., same signal to both ears), oversimplifying multiple aspects of binaural hearing (i.e., adequately reproducing interaural differences and spatial information). In the current study, the effects of a realistic classroom-noise scenario presented either monaurally or binaurally on tasks requiring processing of auditory and visually presented information were analyzed in children and adults. In Experiment 1, across age groups, word identification was more impaired by monaural than by binaural classroom noise, whereas listening comprehension (acting out oral instructions) was equally impaired in both noise conditions. In both tasks, children were more affected than adults. Disturbance ratings were unrelated to the actual performance decrements. Experiment 2 revealed detrimental effects of classroom noise on short-term memory (serial recall of words presented pictorially), which did not differ with age or presentation mode (monaural vs. binaural). The present results add to the evidence for detrimental effects of noise on speech perception and cognitive performance, and their interactions with age, using a realistic classroom-noise scenario. Binaural simulations of real-world auditory environments can improve the external validity of studies on the impact of noise on children's and adults' learning.

Keywords: auditory distraction; children; speech perception; listening comprehension; verbal short-term memory; irrelevant sound effect; binaural; monaural; classroom; learning

1. Introduction

Learning in classrooms is often impeded by unfavorable acoustic conditions, such as noise and reverberation [1]. A recent study in German preschool and school classrooms reported an average sound pressure level (SPL) of 66 dB $L_{A,eq}$ (A-weighted equivalent continuous SPL), and a range of 62–69 dB $L_{A,eq}$ during typical activities [2]. Other studies conducted across Europe and the US reported values ranging from 42 to 100 dB $L_{A,eq}$ [3,4] (see also Table 1 in [2]). Field studies revealed that children instructed in classrooms with high levels of indoor or external (aircraft) noise score lower in achievement tests and in ratings of well-being at school, and exhibit higher levels of annoyance due to noise [5–9]. Numerous experimental studies have analyzed effects of acute noise on children's performance in a range of auditory and non-auditory tasks. Concerning the former, it has consistently been shown that children's

language comprehension is more impaired than adults' by noise and reverberation [10–13]. Concerning non-auditory tasks, findings are less consistent for complex academic tasks such as reading and numeracy [14–16] (for review, see [17], but reliable noise-induced performance decrements have been reported for children's visual–verbal short-term memory [18–21].

However, most of these studies do not represent real-world auditory environments with respect to the type of noise and its perception by the person affected. For example, the noise maskers used in many psychoacoustic studies on speech-in-noise perception have nothing in common with a noisy classroom. In addition, in the vast majority of studies on noise effects on cognitive performance, the noise is presented monaurally (the same signal presented to both ears) over headphones, and in cases of loudspeaker presentation the same signal is sent to each speaker. These presentation formats oversimplify the multiple features of binaural hearing in complex acoustic environments, where sounds are spatially spread across the room, and sound sources change often and unpredictably. In order to represent such complex scenes in laboratory settings, one approach is binaural reproduction, wherein the interaural differences in sound reaching the ears are authentically represented, including spatial cues [22,23]. In the current study, we analyzed the effects of a realistic classroom-noise scenario on tasks requiring processing of auditory (Exp. 1) and visual information (Exp. 2) in children and adults. We aimed to assess the detrimental noise effects across age groups, and to explore whether and to what extent these effects, and the developmental change associated with them, are moderated when a realistic, binaural presentation mode is used instead of the usual, monaural presentation.

2. Experiment 1: Effects of Classroom Noise on Speech Perception and Listening Comprehension

Learning in classrooms relies heavily on oral instruction and listening in the presence of irrelevant sounds. Thus, school children are regularly faced with the requirement of focusing on a specific sound source while ignoring others [24]. Experimental studies on the effects of environmental noise on children's ability to understand speech have focused on simple speech perception tasks requiring identification of isolated speech targets in noise and/or reverberation. However, listening requirements faced by children in classrooms go far beyond pure identification. Effective listening in these situations requires storage and processing of complex oral information in working memory, while constructing a coherent mental model of the information presented [25]. There is evidence that noise may affect the storage and processing of spoken items even when the signal-to-noise ratio (SNR; signal SPL minus noise SPL) is high enough to allow perfect or near-perfect identification [26–28]. Thus, effects of noise on word identification tasks do not allow predictions of decrements in complex listening tasks.

Studying the impacts of noise and reverberation on children's speech perception in a classroom-like setting, Klatte and colleagues [29] found differential effects of foreign, single-talker speech and classroom noise without speech on word identification (word-to-picture matching) and listening comprehension (acting-out complex oral instructions). In both tasks, children were affected more than adults. In the comprehension task, both speech and classroom noise significantly reduced children's performance, with 6- to 7-year-old first-graders suffering the most, whereas adults were unaffected. Speech was more disruptive than classroom noise. In contrast, word identification was much more impaired by classroom noise when compared to speech. The authors proposed that, with the SNRs of -3 dB to 3 dB used in their study, the effects of background speech and classroom noise resulted from different mechanisms. Classroom noise masked the speech targets. Background speech was a less potential masker, but interfered with short-term memory processes that children (but not adults) rely on when listening to complex sentences. The study further revealed that the children's ratings of the sound-induced disruption were unrelated to their objective performance decline. This finding underlines that, while it is undeniable that noise has a negative impact on performance, this does not mean that a person affected is subjectively aware of these effects or feels annoyed; cf. [30].

In Klatte et al. [29], the background sounds were presented via loudspeakers located at the sides of the laboratory room. The same recording was sent to each of the eight loudspeakers, and the target signals were presented through a separate speaker located in front of the room. In the current study, we further increased the realism of the design by Klatte and colleagues [29], by including a classroom-noise scenario that is reproduced binaurally (i.e., authentically representing interaural and spatial cues). Prior studies confirmed that participants' performance in listening tasks is affected when the realism of the auditory scene is increased [31–34].

2.1. Materials and Methods

Participants: The sample consisted of 36 student volunteers (19 female) from the University of Kaiserslautern, aged between 19 and 31 years ($M = 24.9$, $SD = 3.9$ years), and 56 children recruited from a primary school in Kaiserslautern. The child sample comprised 37 second-graders (9 female), aged between 6 years, 3 months and 8 years, 2 months ($M = 7$ years, 5 months, $SD = 3$ months); and 19 third-graders (12 female) aged between 8 years, 3 months and 9 years, 7 months ($M = 8$ years, 9 months, $SD = 4$ months).

All participants were native German speakers and had normal or corrected-to normal vision and normal hearing according to either self-reports (adults) or parental reports (children). The study was approved by the Rhineland-Palatinate school authority and by the Ethics Committee of the University of Kaiserslautern. Informed written consent was provided by the adult participants and by the children's parents. Adults received either course credit or payment for participation (10 €).

Apparatus: The tasks were created in Python 3.7/PsychPy 3.1.5 [35] and executed using a 15.6-inch laptop (HP ProBook 450 G6). The screen's resolution was 1920×1080 pixels, and its refresh rate was 60 Hz. The sounds were presented via Sennheiser HD650 headphones and a Focusrite Scarlett 2i2 2nd generation audio interface.

Speech signals: The words and the instructions were read by a professional female speaker in a sound-attenuated booth and recorded with a Sennheiser MD 421-II-4 Dynamic Studio Microphone at a sampling rate of 44.1 kHz and 16-bit resolution. The recordings were loudness-normalized according to EBU R-128 [36] using Version 3.0.0 of Audacity® recording and editing software [37].

Classroom-noise scenario: The background noise represented a classroom-like auditory environment with sounds from everyday classroom activities, e.g., furniture use; desk noise, including writing and use of other stationary items; footsteps; door opening and closing; use of zippers on bags; undoing a plastic wrapper; and turning the pages of a book. To prevent any learning effects, the different noises were presented at irregular intervals as in real classroom scenarios. Some noises (e.g., writing) were played more often than others (e.g., door) to mimic their typical frequency of occurrence in reality. The selection of nonspeech sounds and the frequency with which they occurred were based on listening to recordings of lessons in medium-sized classrooms. Pink noise (-5 dB/octave decay slope) presented at $L_{Aeq,1m}$ of 41.5 dB provided a steady-state noise that simulated air-conditioning noise throughout the auditory scene.

As children frequently speak in class, multi-talker speech, consisting of four child voices talking in Hindi, which was foreign to all participants, was added to the scene. We recorded (32-bit, 44.1 kHz sampling rate) a child (8 years old) having a natural, unscripted conversation with an adult (the parent) on a range of topics for around two hours in a hemi-anechoic room. The talkers sat on chairs facing each other at a 2 m apart. Each talker wore a DPA 4066 omnidirectional headset microphone positioned 7 cm from the center of lips, as was done in previous research [38]. This recording was post-processed to remove all the adult speech parts, silences, and other artefacts. The remaining speech was segmented into smaller sentences, and the fundamental frequencies of 3/4 sentences were changed. Hence, speech from 4 child voices was created for use in the auditory scene. Individual components of the speech segments were only presented once. In the final auditory scene,

two child voices were active at any time. Following a previous study [38], the order of active talkers changed randomly.

To auralize the binaural background sounds, a classroom model was generated in SketchUp [39]. The classroom has a rectangular floor plate and flat ceiling ($9\text{ m} \times 11\text{ m} \times 2.7\text{ m}$, $W \times L \times H$), and all the room surfaces were assigned an absorption coefficient of 1 (i.e., anechoic conditions). Within the room, 12 desks and chairs were modeled in four rows. Each chair seated a child. Typical sound absorption values were assigned to each child. The listener's position referred to a child sitting in the middle of the last row with an ear height of 1 m. A total of 16 different sound sources were then placed in the room. Twelve sound sources for non-speech sounds were placed at varying distances around the listener; there were equal numbers of sound sources on the left and right sides. The directionality of the last four sound sources represented two child talkers at 3 m away at ± 30 degrees, and two talkers at 5 m away at ± 20 degrees, with an ear height of 1 m. The classroom model was auralized using RAVEN [40]. A generic head-related transfer function (HRTF) from the FABIAN dummy head [41] with a resolution of $1^\circ \times 1^\circ$ was used. Although it is well known that HRTFs differ significantly between adults and children [42], and that this difference can influence cognitive tasks [31], a generic solution of HRTFs was used in the current study, as this was the first attempt to spatially separate the sounds. Thus, in the binaural condition, the sounds were spatially spread across the room, and the order of active talkers and source locations of non-speech sounds changed randomly, as is typical in real classrooms. In the monaural condition, the sounds were presented without any spatial separation, and appeared to come from straight ahead (or inside the head) to the listener. For both sound conditions, the auralized files were mixed down to a 2-channel audio file. The headphone (Sennheiser HD 650, Wedemark, Germany) output per channel was calibrated using Brüel and Kjær Artificial Ear Type 4153 with a Brüel and Kjær Type 4190 omnidirectional microphone capsule.

For both sound conditions, the $L_{A,eq}$ of the target speech signals and the classroom noise were 60 and 63 dB, yielding a SNR of -3 dB [29].

Tasks: We used modified versions of the tasks from Klatte and colleagues [29]. As we included three sound conditions (silence, monaural noise, binaural noise), three equivalent, parallel versions of each task were constructed.

Speech Perception: A word-to-picture matching task requiring discrimination between phonologically similar words was used to measure speech perception. A total of 84 lists of four phonologically similar German nouns (for example, Kopf (head), Topf (pot), Knopf (button), and Zopf (braid)) were constructed. Each word was represented by a simple and easy-to-name colored drawing. Each trial began with a 1.5 s visual cue, followed by a spoken word. Then, the screen displayed four images in a fixed array, one representing the target word and three representing similar-sounding distractor words (see Figure 1). The position of the picture representing the target words was counterbalanced. The participant's task was to mouse-click on the picture that corresponded to the target word. There were 28 trials in each sound condition. The task was the same for both adults and children.

Listening comprehension: A paper and pencil test requiring the execution of complex oral instructions was used to assess listening comprehension. In each of the sound conditions, participants heard 8 oral instructions, such as "Male ein Kreuz unter das Buch, das neben einem Stuhl liegt" ("Draw a cross under the book that is next to the chair"). The task was to carry out the instructions on pre-prepared response sheets. On the response sheets, each instruction was represented by a row of small black-and-white drawings of the target objects (e.g., a book next to a chair) and distractor stimuli (e.g., a book next to a ball) (see Figure 2). The response sheet was also visible on the computer screen in front of the participant. A red arrow indicated the row reflecting the current instruction on the response sheet. Each instruction began with an auditory cue (bell ringing). Participants had 18 s after the end of an instruction to complete the entries on the response sheet. They were instructed to begin carrying out the instructions as soon as feasible. Scoring was based on the number of elements correctly executed according to the respective instructions. In order

to equalize task difficulty across age groups, the adults received longer and syntactically more complex instructions.

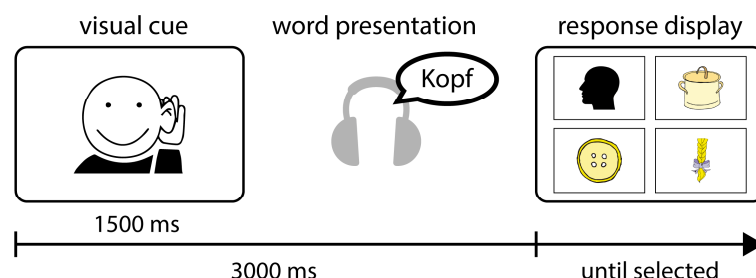


Figure 1. The experimental procedure for measuring speech perception with the word-to-picture matching task. The visual cue was displayed for 3000 ms, indicating the onset of the spoken target word presented over headphones. The target word was played over headphones 1500 ms after the onset of the visual cue. Thereafter, the response display was shown, which comprised four pictures, one representing the target word (here: Kopf (head)) and three representing phonologically similar distractor words (Topf (pot), Knopf (button), and Zopf (braid)).

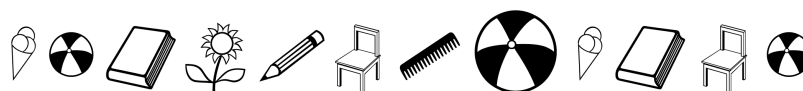


Figure 2. Exemplary trial from the listening comprehension task for the children. “Male ein Kreuz unter das Buch, das neben einem Stuhl liegt” (“Draw a cross under the book that is next to the chair”).

Subjective disturbance was assessed using a smiley scale with 4 points representing the ratings “not at all disturbed” (0), “a little disturbed” (1), “strongly disturbed” (2), or “extremely disturbed” (3).

Procedure: Both children and adults were tested in separate groups of 2 to 4 in a sound-attenuated booth at the University of Kaiserslautern-Landau. Four computer workplaces were arranged in the room, with about 4 m and partition walls between them. The walls around each workstation were equipped with posters of a primary school classroom to create a more classroom-like atmosphere. Adults received written instructions. Children were instructed orally by a researcher. Participants were informed that they should ignore the sounds and focus solely on the execution of the respective task. The experimenter stayed in the back of the laboratory room during the whole session.

Each participant performed both tasks in each of the three sound conditions (silent control, monaural, and binaural auditory classroom scene). Sound conditions were varied in blocks of trials. There were 28 and 8 trials per block in the word identification and listening comprehension tasks, respectively. The order of sound conditions and the allocation of test versions to sound conditions were counterbalanced between participants.

Each session started with a general instruction provided by the experimenter, followed by the monaural presentation of the classroom scene for 4 s to familiarize the participants with the background sound. Then, all the pictures used in the speech perception task were presented and named. Subsequently, participants performed the speech perception task. Thereafter, the listening comprehension task was instructed and performed. Both tasks started with four practice trials. In the sound conditions, the classroom-noise scenario was continuously played during the respective block of trials. Finally, the monaural and binaural auditory scenes were played for 10 s in order to complete the disturbance ratings. The session took about 40 min in total.

2.2. Results

Mean proportion correct scores and standard deviation as a function of task, sound condition, and age group are depicted in Table 1. Difference scores were calculated for each participant by subtracting proportion correct scores in noise from performance in silence.

These scores were used as dependent variables. Figure 3 depicts the mean difference scores with respect to task, sound condition, and age group.

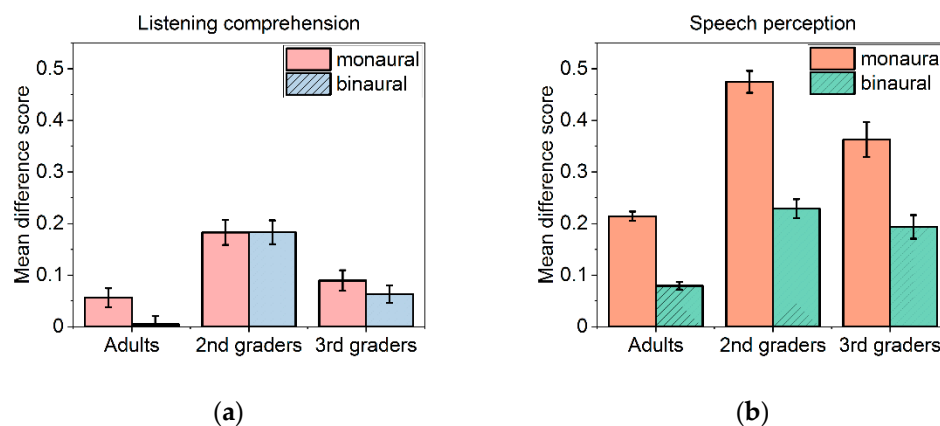


Figure 3. Mean difference scores for speech perception (a) and listening comprehension (b) with respect to age group and sound condition. Error bars denote standard errors of the mean.

Table 1. Mean proportion correct scores for speech perception and listening comprehension, and mean disturbance ratings for Experiment 1 as a function of sound condition and age group (standard deviation in parenthesis).

Task/Ratings	Sound Condition	Adults	2nd Graders	3rd Graders
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Speech Perception	Silence	0.99 (0.00)	0.97 (0.05)	0.96 (0.06)
	Monaural	0.79 (0.05)	0.50 (0.13)	0.61 (0.13)
	Binaural	0.92 (0.04)	0.74 (0.11)	0.77 (0.11)
Listening comprehension	Silence	0.85 (0.10)	0.94 (0.01)	0.94 (0.08)
	Monaural	0.79 (0.12)	0.76 (0.15)	0.85 (0.12)
	Binaural	0.84 (0.11)	0.76 (0.14)	0.87 (0.07)
Disturbance rating	Monaural	1.25 (0.55)	0.97 (0.93)	1.11 (0.66)
	Binaural	2.00 (0.72)	1.35 (0.98)	1.53 (0.84)

For speech perception, a 3×2 mixed ANOVA of the difference scores with age group (adults, second-graders, third-graders) as a between-subjects factor and sound condition (monaural, binaural) as a within-subject factor confirmed a significant main effect of the sound condition— $F(1, 89) = 222.31, p < 0.001, \text{partial } \eta^2 = 0.714$ —reflecting stronger impairment in the monaural when compared to the binaural condition; a significant main effect of age group— $F(2, 89) = 56.29, p < 0.001, \text{partial } \eta^2 = 0.558$ —reflecting stronger impairments in the children when compared to adults; and a significant interaction— $F(2, 89) = 9.19, p < 0.001, \text{partial } \eta^2 = 0.171$. Post hoc *t*-tests revealed that, in each of the age groups, the ability to recognize isolated words was less impaired in the binaural condition when compared to the monaural condition (all *p*'s < 0.001). The interaction reflects a more pronounced difference between age groups in the monaural when compared to the binaural noise condition (see Figure 1). Separate analyses per sound condition revealed significant differences between all age groups for monaural noise, and significant differences between adults and both groups of children for binaural noise (all *p*'s < 0.01), whereas second- and third-graders did not differ ($p = 0.190$).

For listening comprehension, the 3×2 mixed ANOVA revealed a significant main effect of age group— $F(2, 89) = 24.87, p < 0.001, \text{partial } \eta^2 = 0.359$. The effect of sound condition and the interaction were not significant ($F(1, 89) = 2.67, p = 0.106, \text{partial } \eta^2 = 0.029$ and $F(2, 89) = 1.17, p = 0.314, \text{partial } \eta^2 = 0.026$). Concerning the main effect of age, post hoc tests revealed that second-graders were more impaired than adults ($p < 0.001$) and third-graders ($p < 0.001$). No significant differences were found between third-graders

and adults ($p = 0.27$). One-sample t -tests revealed that the performance decrement due to binaural noise in adults did not differ significantly from 0; $t(35) < 1$.

In a further step, we analyzed whether speech perception in noise predicts listening comprehension in noise. In view of the small sample size for the third-graders and the fact that the adults' listening comprehension was largely unaffected by noise, the respective correlation analyses was confined to the second-graders. Correlations between proportion correct scores for speech perception and listening comprehension in noise were calculated. In the binaural condition, speech perception was significantly related to listening comprehension, $r(35) = 0.375$, $p < 0.05$, whereas in the monaural condition, speech perception and listening comprehension were unrelated ($p = 0.19$).

For the disturbance ratings, the 3×2 mixed ANOVA yielded a significant main effect of noise condition— $F(2, 89) = 33.10$, $p < 0.001$, $\text{partial } \eta^2 = 0.271$ —reflecting higher disturbance in the binaural when compared to the monaural condition; a significant main effect of age group— $F(2, 89) = 4.20$, $p < 0.05$, $\text{partial } \eta^2 = 0.09$ —but no interaction: $F(2, 89) = 2.09$, $p = 0.13$. Post hoc tests confirmed that the disturbance ratings were higher in adults when compared to second-graders ($p < 0.05$). Ratings of the third-graders did not differ from those of the second-graders and adults ($p = 1$ and $p = 0.35$, respectively). Further analyses in the second-graders and adults confirmed that, for both noise conditions, the disturbance ratings were unrelated to word identification and listening comprehension in noise (proportion correct scores) and unrelated to the actual performance decrements (difference scores) in adults (all p 's > 0.10) and children (all p 's > 0.40).

2.3. Discussion

Experiment 1 replicated the often-reported finding that children are less able than adults to understand speech in the presence of background noise [10–13,29]. In the monaural condition, word identification was more impaired in second-graders when compared to third-graders, and more impaired in third-graders than in adults. However, in each of the age groups, word identification performance was substantially less affected with binaural noise when compared to monaural presentation of the classroom-noise scenario. This indicates that children and adults may use the spatial cues inherent in the binaural scenario to support separation of the target words from the background noise. The age effect observed for monaural noise was significantly reduced (although still significant) in the binaural condition. These results suggest that the effects of classroom noise on speech perception and the developmental change associated with these effects are strongly moderated by the method of sound presentation. Especially, with a simple, monaural presentation that lacks cues to spatially separate the speech signal from the noise, impairments of speech perception due to real-life environmental noise and their increase with decreasing age might be overestimated. The dominant role of spatial cues is further confirmed by the fact that, in the study of Klatte et al. [29], the impairment of speech perception due to classroom noise was considerably lower than the effects in the monaural condition, but comparable to the binaural condition of the current study (in [29], difference scores were 0.23 in children and 0.12 in adults). This was presumably because in [29], the target words were presented through a separate loudspeaker, thereby allowing spatial separation of signal and background noise.

Concerning listening comprehension, adults showed only minor disruption in the monaural condition (5%) and no significant disruption in the binaural condition. This result replicates the findings of Klatte et al. [29] and can be attributed to the adult listeners' ability to reconstruct noise-masked elements of the speech signals using contextual cues. The age effect observed in the speech perception task was partially replicated; i.e., the second-graders were more affected by background noise when compared to third-graders and adults. However, by contrasting the results in the speech perception task, we can see that the impairment of listening comprehension performance did not differ between sound conditions (monaural vs. binaural). Furthermore, for binaural noise, children's speech perception significantly predicted listening comprehension, whereas for monaural

noise, speech perception and listening comprehension were unrelated. We may therefore conclude that, with binaural presentation of the noise, the effects on word identification allow a more valid prediction of the effects on complex listening tasks when compared to a simple monaural presentation.

Concerning the disturbance ratings, the current results replicate the findings of Klatte et al. [29] that, in view of the noise-induced performance decrements, average ratings of the children were surprisingly low. This was especially true for the monaural noise condition. Despite severe decrements in word identification, children judged this sound condition on average as “a bit disturbing.” The lowest ratings were provided by the second-graders, who showed the strongest performance impairments. Still more surprisingly, across age groups, the binaural sound scenario was judged as more disturbing, even though the speech perception decrements were much stronger in the monaural condition, and the ratings were unrelated to the actual performance decrements. Even though, with age-adequate measurement scales, children are able to provide valid ratings of general annoyance due to noise in their home and school environment [6,9,43,44], the current findings indicate that both children and adults are not aware of the detrimental effects of background noise on their listening performance. In view of this result, it is evident that teachers and researchers cannot rely on students’ ratings when evaluating the acoustic environments of classrooms.

Taken together, Exp. 1 adds to the evidence that children are more impaired than adults by noise in speech perception tasks [10–13,29], and extends this finding to listening comprehension in a realistic classroom-noise scenario. Furthermore, the results revealed that the noise effects and their interaction with age differ considerably between conditions of binaural vs. monaural presentation of the noise. In Exp. 2, we explored whether these differences hold also for cross-modal noise effects, i.e., the effects of noise on the processing of visually presented information.

3. Experiment 2: Effects of Classroom Noise on Visual–Verbal Short-Term Memory

In Experiment 2, the effects of a classroom-noise scenario on short-term memory for visual–verbal items were analyzed in children and adults. Verbal short-term memory is the ability to hold verbal information in an active state for ongoing cognitive processes. A standard task to assess verbal short-term memory requires immediate serial recall of sequences of 5 to 9 verbal items, such as words, digits, or easy-to-name pictures. In such tasks, participants usually employ a strategy called articulatory rehearsal, i.e., repetitive subvocal pronunciations of the list items in a sequential manner. This strategy is evident in children from about age 7, but there is considerable interindividual variation around this age [45,46].

In the current study, we used verbal short-term memory as a cognitive task for two reasons. First, the capacity of short-term memory is a predictor of children’s oral and written language acquisition, and short-term memory processes play a role in many school-based tasks, such as reading and spelling in the early grades, learning new vocabulary, mental arithmetic [47], and following the teachers’ instructions [48]. Second, short-term memory is especially susceptible to the adverse effects of noise. Many studies confirmed that performance in the serial recall task is reliably impaired by task-irrelevant background sounds (for recent reviews, see [21,49]). This “irrelevant sound effect” (ISE) is most pronounced with speech noise, but is also evoked by nonspeech sounds, such as tone sequences or instrumental music (for an overview cp. [50]). The ISE is reliable for coherent sound streams that consist of changing auditory elements emanating from a single source, e.g., fluent speech, sequences of different syllables or tones, and instrumental music. Sounds lacking these “changing-state” characteristics, such as continuous broadband noise, babble speech, or spectrally degraded speech, evoke minor or no disruption. Different theoretical explanations have been provided for the ISE. Some authors [18,51,52] attribute the effect to a diversion of attention away from the task and towards the sound (attention capture-account). Others oppose a role of attention, assuming that the effect results

from specific interference between processes involved in automatic, obligatory processing of the background sound and deliberate processes involved in memorizing the verbal items [53]. Within the interference account, some authors propose serial-order retention as the mechanism of disruption [54], whereas others assume that noise—especially noise with speech—interferes with storage and processing of phonological representations [55,56].

Aiming to disentangle attention-capture from interference-by-process, a number of studies on the ISE included children. In view of children's underdeveloped attention control [57], noise effects resulting from attention capturing should be more pronounced in children when compared to adults. However, this argument is not without problems, since, at least with noise-containing speech, stronger impairments in children may also indicate stronger speech-based interference due to less robust, immature phonological representations and/or maintenance strategies, i.e., articulatory rehearsal of the item sequence [19,21].

While the detrimental effects of irrelevant sounds on children's verbal short-term memory were consistently reported, the findings concerning developmental change are inconsistent. Some studies reported equivalent impairments due to background speech or mixtures of nonspeech sounds with speech in 7- to 10-year-olds and adults [19,20,58–61], whereas others found stronger impairments in the children [18,19,21,62]. Two of these studies included classroom noise [20,59]. In Klatte et al. [20], children's and adults' serial recall performance was equally affected by background speech, but only the youngest children (first-graders) were also impaired by a mixture of classroom sounds without speech. The authors attributed the age-independent effect of background speech to specific interference with the maintenance of phonological representations and the age-dependent classroom noise effect on attentional capture. Meinhardt-Injac et al. [59] used a classroom-noise scenario with speech (bits of conversation between children and adults) and found significant and equivalent impairments of serial recall performance in 8–10-year-olds, 11–12-year-olds, and adults. Across age groups, the noise effects were unrelated to participants' attention control, i.e., their ability to inhibit task-irrelevant information. This finding indicates that specific interference rather than attention capture is the mechanism underlying the noise-induced disruptions.

The vast majority of ISE studies have used simple, monaural presentation of irrelevant sounds. Only a few studies have included spatially spread sound sources. These studies provided evidence that the variation of the source location moderated the sounds' disruptive effects. Buchner et al. [63] showed that the ISE evoked by nonspeech sounds (e.g., footsteps, cries of pain, and squeaking sounds) and speech (sequences of unrelated words) played through loudspeakers from different locations was most pronounced when the sound was played from the front, i.e., from a location near the visual target display to which the participants' attention was directed. However, the source location had only a small impact on the sound-induced disruption, and the disruption evoked by speech was substantially stronger than that evoked by nonspeech sounds. These findings indicate a significant but minor role of attention capturing in the ISE in adults. Jones and Macken [64] analyzed the effects of background speech produced by six voices simultaneously. The speech was presented through loudspeakers located in a circle around the participant. The impairment of short-term memory performance was more pronounced when each voice was assigned to a separate loudspeaker (yielding six single-talker streams), when compared to assigning a mix of the six voices to each of the six loudspeakers (yielding identical streams of babble speech). Comparable results were reported in studies using dichotic vs. monaural headphone presentation of irrelevant syllables [65,66] and interpreted as evidence for specific interference through changing state speech.

In view of these findings, we might expect a stronger effect of the binaural when compared to the monaural classroom-noise scenario, through increased speech-based interference due to clearer separation of the speech streams, and/or increased attention capture due to spatially spread sound sources and changing source locations. If attention

capture is the dominant source of disruption, children should be more impaired than adults, and more impaired by the binaural than the monaural noise scenario.

3.1. Materials and Methods

Participants: The sample consisted of 40 student volunteers (24 female), aged between 19 and 32 years ($M = 24.0$, $SD = 2.5$ years), from the University of Kaiserslautern-Landau; and 69 third- and fourth-grade children recruited from a primary school in Kaiserslautern. Due to technical issues, the data of two children had to be excluded from the analysis. The final child sample consisted of 67 children (36 female), aged between 8 years, 2 months and 10 years, 3 months ($M = 9$ years, 4 months, $SD = 6$ months). Of the children, 19 had taken part in Experiment 1. All participants were native German speakers and had normal or corrected-to-normal vision and normal hearing according to either self-reports (adults) or parental reports (children). The study was approved by the Rhineland-Palatinate school authority and by the Ethics Committee of the University of Kaiserslautern. Informed written consent was provided by the adult participants and by the children's parents. Adults received either course credit or payment for participation (10 €).

Apparatus: Identical to Experiment 1.

Background noise: Identical to Experiment 1.

Task: The task required serial recall of sequences of monosyllabic German nouns presented pictorially. Pictures were used instead of written words in order to avoid confounding by the children's reading abilities. Prior studies confirmed that children and adults use verbal strategies when memorizing words presented pictorially [46,67], and that participants' strategies do not differ between pictorial and written presentation [68]. Each trial consisted of a presentation phase, a retention interval, and a recall phase. Pictures were presented one after another in a 102×73 mm rectangular black frame in the center of a white screen, with a presentation duration of 1500 ms and an interstimulus interval of 500 ms. A random interval between 1200 to 1800 ms passed by before the visual presentation of the first list item. The final list item was followed by a 5000 ms retention interval. The onset of the recall phase was signaled by the simultaneous re-presentation of all stimuli. The pictures were arranged at random in a fixed array of five (children) and eight (adults) black frames (see Figure 4). Participants had to reconstruct the serial order by using the mouse to click on the items in the presentation order. Clicking an item changed its shading, indicating that it had been selected. There was no time limit for responding and no possibility of error correction. After selection of the final item, participants were presented with a visual cue to start the next trial by pressing the space bar.

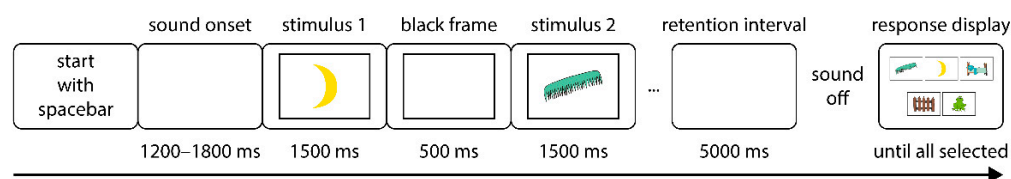


Figure 4. The serial recall task's experimental procedure. Five pictures per trial were shown to the children, whereas eight pictures per trial were presented to the adults. All pictures seen in the respective trial were randomly arranged in an array of 5 (children) and 8 (adults) frames in the response display.

Both children and adults saw colored drawings representing the monosyllabic German words *Bett*, *Bus*, *Eis*, *Frosch*, *Kamm*, *Mond*, *Pilz*, *Schal*, *Schiff*, and *Zaun* (*bed*, *bus*, *ice*, *frog*, *comb*, *moon*, *mushroom*, *scarf*, *ship*, and *fence*). The set for the adults additionally included the items *Brief*, *Haus*, *Herz*, *Hut*, *Nuss*, and *Schwein* (*letter*, *house*, *heart*, *hat*, *nut*, and *pig*). Four lists of five items (drawn out of 10) were created for the children, and six lists of eight items (drawn out of 16) for the adults. Two additional versions of each list were created using random permutations of the list items.

Procedure: Both children and adults were tested in groups of 2 to 4 in a sound-attenuated booth at the University of Kaiserslautern-Landau (see Exp. 1). Adults received written instruction. Children were instructed orally by a researcher. Participants were informed that they should ignore the sounds and focus solely on the serial recall task. At the beginning of each session, the classroom scenarios were played for 4 s, followed by the presentation of all pictures used in the task. Each picture was named by a female speaker. Following the instruction, three practice trials (one per sound condition) were performed. Thereafter, children and adults completed 24 and 48 experimental trials, respectively (8 and 16 trials per sound condition). Sound conditions (silence, classroom noise—monaural, classroom noise—binaural) were varied in blocks of trials. The order of sound conditions was balanced across participants. In the sound blocks, the sound started when the participant initiated the first trial and terminated after finishing the recall phase of the final trial of the respective block. Sounds were presented via headphones at an average level of 60 dB(A). The testing session lasted about 20 min for children and 35 min for adults.

3.2. Results

The dependent variable was the proportion of correct scores based on the number of items recalled at the correct serial position. Proportions of correct scores with respect to age group and sound condition are depicted in Figure 5a. A two-way mixed ANOVA with sound condition (silent control, monaural, binaural) as the within-subject factor and age group (adults, children) as the between-subjects factor revealed significant main effects of sound condition ($F(2, 210) = 8.69, p < 0.001, \text{partial } \eta^2 = 0.08$) and age group ($F(1, 105) = 6.73, p < 0.05, \text{partial } \eta^2 = 0.06$). Bonferroni-corrected post-hoc tests revealed that performance in both noise conditions was significantly lower when compared to the silent control condition, $p < 0.01$, whereas the noise conditions did not differ ($p = 0.99$). The main effect of age group reflects better overall performance of the adults. The sound condition $\times$ age group interaction was not significant ($F < 1$), confirming comparable noise-induced disruption in adults and children. The analyses thus confirmed significant impairments of serial recall performance due to classroom noise. The effects did not differ between age groups, nor between monaural vs. binaural noise.

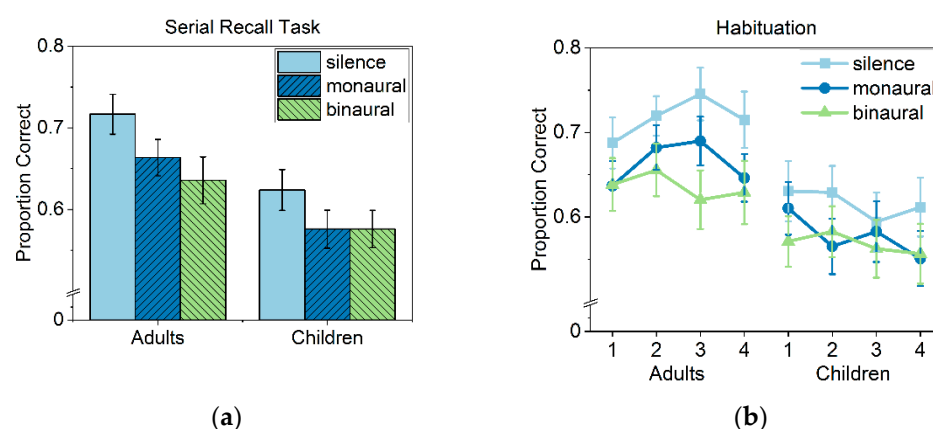


Figure 5. (a) Mean proportion correct scores as a function of age group and sound condition. (b) Mean proportion of correct scores as a function of age group and as a function of sound condition in four consecutive blocks of four (adults) and two (children) trials. Error bars denote standard errors of the mean.

Aiming to assess the role of attention in the noise-induced disruption, further analyses were performed to explore whether or not participants habituated to the noise across trials. If the noise effects result from attention capture, one might expect habituation and thereby a stronger disruption in the first when compared to the final trials performed with noise [69]. For this aim, the proportion of correct scores was calculated for four consecutive blocks of two trials (children) and four trials (adults) for each sound condition. The resulting

proportion correct scores are depicted in Figure 5b. A $3 \times 4 \times 2$ mixed ANOVA with sound condition (silent control, monaural, binaural) and trial block (block 1–block 4) as within-subject factors and age group (adults, children) as the between-subjects factor was conducted. Except for sound condition and age group (reported above), neither trial block nor any interaction reached significance (all $F < 1$). The analysis thus yielded no evidence for habituation to classroom noise in children or adults.

3.3. Discussion

In Experiment 2, the impacts of monaural and binaural classroom-noise scenarios on verbal short-term memory were examined in children and adults. The task required serial recall of words presented pictorially. Children's and adults' performances were significantly and equally impaired in both noise conditions. The magnitude of the noise effects remained stable over the course of experimental trials.

The non-significant effect of age replicates the findings of Meinhardt-Injac et al. [59], who reported significant and equivalent impairments due to classroom noise in children and adults. However, in Klatte et al. [20], only the youngest children were affected by classroom noise, whereas older children and adults were unaffected. The apparent contradiction might be attributed to the kinds of classroom noise and age groups included. In both the current and the Meinhardt-Injac et al. [59] study, the classroom noise contained speech, and the youngest children were age 9 on average, whereas Klatte et al. [20] used nonspeech classroom noise and included 6- to 7-year-old first-graders. In line with the arguments provided in these studies, we propose that the detrimental effects of the classroom-noise scenario result from separate mechanisms. The spoken parts in the noise scenarios evoke specific interference with the maintenance of the list items. This mechanism may evoke stronger disruption in children whose phonological maintenance strategies are not yet fully developed and thus more prone to speech-based interference. In addition, both nonspeech sounds and speech may impair performance through attention capture. The impact of attention capture depends on the sound's potential to grab attention (i.e., personal relevance, emotional valence, and predictability) and on the individuals' attentional abilities. Young children are more vulnerable than older children and adults due to less developed attention control. As schooling contributes considerably to children's development of attention control [70], preschool children and first-graders are especially vulnerable to noise-induced attention capture. Taken together, the findings of the current study add to the evidence provided by Klatte et al. [20] and Meinhardt-Injac et al. [59] that children aged around 9 years or older show adult-like impairments of short-term memory in the presence of classroom noise.

As outlined above, from both the attention capture and the interference account, one might expect stronger disruptive effects with binaural when compared to monaural presentation of the classroom-noise scenario. Following the attention capture account, especially in children, the binaural scenario should evoke stronger disruption because of its attention-grabbing quality (spatially distributed and changing sound source locations). However, contrary to expectation, the disruptive effects of the classroom-noise scenario did not differ with presentation mode, neither in children nor in adults. Furthermore, in both age groups, the detrimental effects of the noise scenarios remained stable across experimental trials—i.e., there was no evidence for habituation. These findings add to the evidence provided by Meinhardt-Injac et al. [59] and Klatte et al. [20] that diversions of attention play a minor role in the ISE, at least in children older than 8 years. This view was further confirmed in a recent study [21], demonstrating that, while 9-year-old third-graders were more impaired than adults by background speech in a verbal serial recall task, serial recall of nonverbal, visuo-spatial items was unaffected in both groups. As storage and processing of visuo-spatial items rely heavily on domain-general attentional resources [71,72], these findings strongly suggest that the impairment in the verbal task, and its interaction with age, reflect speech-based interference rather than a capture of attention.

Following the interference account, binaural sound scenarios comprising speech should evoke stronger disruption because the spatial separation of the sound sources fosters the segregation of the spoken parts into separate, single-talker streams [64]. Evidently, this mechanism was not at play in the current study. This might be due to the fact that, in the classroom-noise scenario used here, all spoken parts consisted of only two talkers simultaneously. Jones and Macken [64] showed that monaural presentations of single-talker speech and a mixture of two speakers evoked similar amounts of disruption in serial recall performance, whereas with six simultaneous speakers, the disruption was significantly reduced. Thus, concerning the current classroom-noise scenario, a clearer separation of the two speech streams through spatial cues might not further increase the interference, because the latter is already at its maximum (resembling that evoked by a single talker). On this assumption, stronger interference effects with binaural presentation should occur with noise scenarios containing more than two simultaneous talkers.

4. Conclusions

In the current study, the effects of a realistic classroom-noise scenario presented either monaurally or binaurally on speech perception, listening comprehension, and verbal short-term memory were investigated in primary school children and adults. In Exp. 1, across age groups, speech perception (identification of spoken words) was more impaired by monaural than by binaural classroom noise, whereas listening comprehension (acting-out complex oral instructions) was equally impaired in both noise conditions. In both tasks, children were more affected than adults. The age effect found here is in line with a number of psychoacoustic studies documenting increasing noise-induced speech perception impairments with decreasing age [10–13] and extends these findings to a realistic noise scenario and a complex listening task that more closely reflects the requirements of children faced during classroom instruction. Disturbance ratings were unrelated to the actual performance decrements, indicating that listeners' subjective reports do not allow a valid evaluation of adverse listening conditions. In Exp. 2, using a paradigm from the domain of the ISE, we found significant detrimental effects of the classroom-noise scenario on verbal short-term memory (serial order reconstruction of words presented pictorially), which did not differ with age or presentation format (monaural vs. binaural). Concerning the age-equivalence of the noise effect, the current finding replicates a recent study demonstrating comparable effects of classroom noise-containing speech in children aged 8 to 10 years and adults [59]. The lack of an effect of presentation format was contrary to our expectation. Especially for the children, we anticipated stronger impairments with the binaural condition due to increased attention capture through spatially separated and changing sound sources. The equivalence of the noise effects across age groups and presentation formats adds to the evidence that the ISE evoked by noise-containing speech results from specific interference between obligatory processing of the speech parts and deliberate processes involved in serial recall performance (i.e., maintenance of phonological representations), whereas attention capture plays a minor role.

Concerning age effects, children were more impaired by noise than adults in a complex listening task requiring processing of oral instructions, but equally impaired as adults in a task requiring processing of visually presented information. This indicates that, in verbal working memory tasks, children are more prone to distraction than adults when the targets and the irrelevant stimuli stem from the same sensory modality (unimodal interference) but are equally affected (or equally unaffected) as adults when the irrelevant stimuli and the targets originate from different sensory modalities (crossmodal interference, i.e., interference between visual and auditory information). It has been shown that selective attention (i.e., the ability to focus on the target stimuli and inhibit irrelevant distractors) is much more easily achieved in crossmodal paradigms when compared to unimodal paradigms. This holds especially for paradigms with visual targets and auditory distractors [73,74]. These findings have been attributed to top-down suppression of auditory distractors at a very early processing stage (at the level of the cochlea). As a consequence, age differences

in attention control should be largely unrelated to auditory distraction when the relevant information is visual. This has been shown in studies including older adults [73–75], but as proposed by Röer et al. [60], may also hold for children. Following this view, age differences in the ISE between children and adults should emerge when the memory items are presented auditorily instead of visually. This prediction might be tested in future studies.

Concerning the impact of the noise presentation format (monaural vs. binaural), differential effects were found only in the speech perception task. The detrimental noise effect was much stronger with monaural when compared to binaural noise, and the difference between noise conditions was especially strong in the youngest (second-grade) children. These findings indicate that, with standard, monaural presentation of the maskers, the effects of noise on speech perception in real-life listening situations and the developmental change associated with speech-in-noise perception might be overestimated. Furthermore, we found that speech perception in binaural noise significantly predicted listening comprehension in binaural noise, whereas with monaural noise, speech perception and listening comprehension were unrelated. These results suggest that studying speech-in-noise perception using binaural scenarios may have the potential to allow valid predictions of detrimental noise effects on language comprehension in everyday situations. Even though the current study yielded no evidence for an effect of the noise presentation format (monaural vs. binaural) on listening comprehension and visual–verbal short-term memory, we cannot rule out that such effects exist. In the classroom-noise scenario used here, no more than two voices are presented simultaneously, and the nonspeech sounds are confined to indoor noise, disregarding noise sources from outside such as road traffic or aircraft noise. With classroom-noise scenarios containing more unexpected and novel sounds and/or more simultaneous voices, or in samples of younger children, a significantly stronger impairment with binaural when compared to monaural noise presentation might emerge.

Despite these limitations, the current study adds to the evidence concerning developmental change in the effects of noise on speech perception, listening comprehension, and short-term memory, and extends it to a realistic classroom-noise scenario. Further research is needed to find out whether and how much the use of binaural sound scenarios in experimental studies allows more valid predictions of the effects of noise in everyday, real-life learning situations.

Author Contributions: Conceptualization, M.K. and S.J.S.; methodology, M.K.; software, L.L.; validation, L.L. and M.K.; formal analysis, L.L.; investigation, L.L.; resources, C.B., M.Y. and T.L.; data curation, L.L.; writing—original draft preparation, L.L. and M.K.; writing—review and editing, all; visualization, L.L.; supervision, M.K.; project administration, L.L., C.B. and S.F.; funding acquisition, S.F., J.F., A.R., M.K., S.J.S. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project ID 444697733 and ID 401278266.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Deutsche Gesellschaft für Psychologie (DGPs, German Psychological Society) and by the Rhineland-Palatine school authority (Aufsichts- und Dienstleistungsbehörde, ADD).

Informed Consent Statement: Informed written consent was provided by the parents of the children and by the adult participants.

Data Availability Statement: All data can be found at: <https://osf.io/b9rua> (accessed on 31 October 2022).

Acknowledgments: The authors would like to thank all of the students, children, teachers, and parents for cooperating to accomplish this study. In addition, we would like to thank Michelle Turner, Fabio Sobotzki, and Markus Kurtz for assisting in data collection and technical support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Mogas-Recalde, J.; Palau, R.; Márquez, M. How Classroom Acoustics Influence Students and Teachers: A Systematic Literature Review. *J. Technol. Sci. Educ.* **2021**, *11*, 245–259. [\[CrossRef\]](#)
2. Loh, K.; Yadav, M.; Persson Waye, K.; Klatte, M.; Fels, J. Toward Child-Appropriate Acoustic Measurement Methods in Primary Schools and Daycare Centers. *Front. Built Environ.* **2022**, *8*, 688847. [\[CrossRef\]](#)
3. Shield, B.; Dockrell, J.E. External and Internal Noise Surveys of London Primary Schools. *J. Acoust. Soc. Am.* **2004**, *115*, 730–738. [\[CrossRef\]](#) [\[PubMed\]](#)
4. McAllister, A.M.; Granqvist, S.; Sjölander, P.; Sundberg, J. Child Voice and Noise: A Pilot Study of Noise in Day Cares and the Effects on 10 Children’s Voice Quality According to Perceptual Evaluation. *J. Voice* **2009**, *23*, 587–593. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Clark, C.; Head, J.; Stansfeld, S.A. Longitudinal Effects of Aircraft Noise Exposure on Children’s Health and Cognition: A Six-Year Follow-up of the UK RANCH Cohort. *J. Environ. Psychol.* **2013**, *35*, 1–9. [\[CrossRef\]](#)
6. Klatte, M.; Spilski, J.; Mayerl, J.; Möhler, U.; Lachmann, T.; Bergström, K. Effects of Aircraft Noise on Reading and Quality of Life in Primary School Children in Germany: Results from the NORAH Study. *Environ. Behav.* **2017**, *49*, 390–424. [\[CrossRef\]](#)
7. Shield, B.M.; Dockrell, J.E. The Effects of Environmental and Classroom Noise on the Academic Attainments of Primary School Children. *J. Acoust. Soc. Am.* **2008**, *123*, 133–144. [\[CrossRef\]](#)
8. Stansfeld, S.A.; Berglund, B.; Clark, C.; Lopez-Barrio, I.; Fischer, P.; Öhrström, E.; Haines, M.; Head, J.; Hygge, S.; van Kamp, I.; et al. Aircraft and Road Traffic Noise and Children’s Cognition and Health: A Cross-National Study. *Lancet* **2005**, *365*, 1942–1949. [\[CrossRef\]](#)
9. Klatte, M.; Hellbrück, J.; Seidel, J.; Leistner, P. Effects of Classroom Acoustics on Performance and Well-Being in Elementary School Children: A Field Study. *Environ. Behav.* **2010**, *42*, 659–692. [\[CrossRef\]](#)
10. Buss, E.; Leibold, L.J.; Porter, H.L.; Grose, J.H. Speech Recognition in One- and Two-Talker Maskers in School-Age Children and Adults: Development of Perceptual Masking and Glimpsing. *J. Acoust. Soc. Am.* **2017**, *141*, 2650–2660. [\[CrossRef\]](#)
11. Talarico, M.; Abdilla, G.; Aliferis, M.; Balazic, I.; Giaprakis, I.; Stefanakis, T.; Foenander, K.; Grayden, D.B.; Paolini, A.G. Effect of Age and Cognition on Childhood Speech in Noise Perception Abilities. *Audiol. Neurotol.* **2006**, *12*, 13–19. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Stuart, A. Development of Auditory Temporal Resolution in School-Age Children Revealed by Word Recognition in Continuous and Interrupted Noise. *Ear Hear.* **2005**, *26*, 78–88. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Wróblewski, M.; Lewis, D.E.; Valente, D.L.; Stelmachowicz, P.G. Effects of Reverberation on Speech Recognition in Stationary and Modulated Noise by School-Aged Children and Young Adults. *Ear Hear.* **2012**, *33*, 731–744. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Boyle, R.; Coltheart, V. Effects of Irrelevant Sounds on Phonological Coding in Reading Comprehension and Short-Term Memory. *Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol.* **1996**, *49*, 398–416. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Halin, N.; Marsh, J.E.; Haga, A.; Holmgren, M.; Sörqvist, P. Effects of Speech on Proofreading: Can Task-Engagement Manipulations Shield against Distraction? *J. Exp. Psychol. Appl.* **2014**, *20*, 69–80. [\[CrossRef\]](#)
16. Kattner, F.; Hanl, S.; Paul, L.; Ellermeier, W. Task-Specific Auditory Distraction in Serial Recall and Mental Arithmetic. *Mem. Cognit.* **2022**. [\[CrossRef\]](#)
17. Klatte, M.; Bergström, K.; Lachmann, T. Does Noise Affect Learning? A Short Review on Noise Effects on Cognitive Performance in Children. *Front. Psychol.* **2013**, *4*, 578. [\[CrossRef\]](#)
18. Elliott, E.M. The Irrelevant-Speech Effect and Children: Theoretical Implications of Developmental Change. *Mem. Cognit.* **2002**, *30*, 478–487. [\[CrossRef\]](#)
19. Elliott, E.M.; Hughes, R.W.; Briganti, A.; Joseph, T.N.; Marsh, J.E.; Macken, B. Distraction in Verbal Short-Term Memory: Insights from Developmental Differences. *J. Mem. Lang.* **2016**, *88*, 39–50. [\[CrossRef\]](#)
20. Klatte, M.; Lachmann, T.; Schlittmeier, S.; Hellbrück, J. The Irrelevant Sound Effect in Short-Term Memory: Is There Developmental Change? *Eur. J. Cogn. Psychol.* **2010**, *22*, 1168–1191. [\[CrossRef\]](#)
21. Leist, L.; Lachmann, T.; Schlittmeier, S.J.; Georgi, M.; Klatte, M. Irrelevant Speech Impairs Serial Recall of Verbal but Not Spatial Items in Children and Adults. *Mem. Cognit.* **2022**. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hammershøi, D.; Møller, H. Binaural Technique—Basic Methods for Recording, Synthesis, and Reproduction. In *Communication Acoustics*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 223–254.
23. Vorländer, M. *Auralization*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020.
24. Gheller, F.; Lovo, E.; Arsie, A.; Bovo, R. Classroom Acoustics: Listening Problems in Children. *Build. Acoust.* **2020**, *27*, 47–59. [\[CrossRef\]](#)
25. Kintsch, W. The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychol. Rev.* **1988**, *95*, 163–182. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Kjellberg, A.; Ljung, R.; Hallman, D. Recall of Words Heard in Noise. *Appl. Cogn. Psychol.* **2008**, *22*, 1088–1098. [\[CrossRef\]](#)
27. Hurtig, A.; van de Poll, M.K.; Pekola, E.P.; Hygge, S.; Ljung, R.; Sörqvist, P. Children’s Recall of Words Spoken in Their First and Second Language: Effects of Signal-to-Noise Ratio and Reverberation Time. *Front. Psychol.* **2016**, *6*, 2029. [\[CrossRef\]](#) [\[PubMed\]](#)

28. Ljung, R.; Kjellberg, A. Recall of Spoken Words Presented with a Prolonged Reverberation Time. In Proceedings of the Performance: 9th International Congress on Noise as a Public Health Problem (ICBEN), Mashantucket, CT, USA, 21–25 July 2008.
29. Klatte, M.; Lachmann, T.; Meis, M. Effects of Noise and Reverberation on Speech Perception and Listening Comprehension of Children and Adults in a Classroom-like Setting. *Noise Health* **2010**, *12*, 270–282. [\[CrossRef\]](#)
30. Shield, B.M.; Dockrell, J.E. The Effects of Noise on Children at School: A Review. *J. Build. Acoust.* **2003**, *10*, 97–106. [\[CrossRef\]](#)
31. Oberem, J.; Lawo, V.; Koch, I.; Fels, J. Intentional Switching in Auditory Selective Attention: Exploring Different Binaural Reproduction Methods in an Anechoic Chamber. *Acta Acust. United Acust.* **2014**, *100*, 1139–1148. [\[CrossRef\]](#)
32. Drullman, R.; Bronkhorst, A.W. Multichannel Speech Intelligibility and Talker Recognition Using Monaural, Binaural, and Three-Dimensional Auditory Presentation. *J. Acoust. Soc. Am.* **2000**, *107*, 2224–2235. [\[CrossRef\]](#)
33. Yost, W.A.; Sheft, S.; Dye, R. (Toby) Divided Auditory Attention with up to Three Sound Sources: A Cocktail Party. *J. Acoust. Soc. Am.* **1994**, *95*, 2916. [\[CrossRef\]](#)
34. Fintor, E.; Aspöck, L.; Fels, J.; Schlittmeier, S.J. The Role of Spatial Separation of Two Talkers' Auditory Stimuli in the Listener's Memory of Running Speech: Listening Effort in a Non-Noisy Conversational Setting. *Int. J. Audiol.* **2022**, *61*, 371–379. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Peirce, J.W. PsychoPy-Psychophysics Software in Python. *J. Neurosci. Methods* **2007**, *162*, 8–13. [\[CrossRef\]](#) [\[PubMed\]](#)
36. European Broadcasting Union. R128-2020: Loudness Normalisation and Permitted Maximum Level of Audio Signals. Technical 734 Report. 2020. Available online: <https://tech.ebu.ch/docs/r/r128.pdf> (accessed on 12 July 2022).
37. Audacity Team Audacity(R): Free Audio Editor and Recorder [Computer Application]. Version 3.0.0. Available online: <https://audacityteam.org> (accessed on 2 December 2021).
38. Yadav, M.; Cabrera, D. Two Simultaneous Talkers Distract More than One in Simulated Multi-Talker Environments, Regardless of Overall Sound Levels Typical of Open-Plan Offices. *Appl. Acoust.* **2019**, *148*, 46–54. [\[CrossRef\]](#)
39. Trimble Inc. SketchUp 3D Design Software. Available online: <https://www.sketchup.com> (accessed on 18 July 2022).
40. Schröder, D.; Vorländer, M. RAVEN: A Real-Time Framework for the Auralization of Interactive Virtual Environments; Forum Acusticum: Aalborg, Denmark, 2011.
41. Brinkmann, F. The FABIAN Head-Related Transfer Function Data Base. 2017. Available online: <https://depositonce.tu-berlin.de/items/bff6568a-5735-4ebc-b3fa-ac10707b7beb> (accessed on 1 November 2022).
42. Fels, J.; Vorländer, M. Anthropometric Parameters Influencing Head-Related Transfer Functions. *Acta Acust. United Acust.* **2009**, *95*, 331–342. [\[CrossRef\]](#)
43. Waye, K.P.; van Kamp, I.; Dellve, L. Validation of a Questionnaire Measuring Preschool Children's Reactions to and Coping with Noise in a Repeated Measurement Design. *BMJ Open* **2013**, *3*, e002408. [\[CrossRef\]](#)
44. Quehl, J.; Bartels, S.; Fimmers, R.; Aeschbach, D. Effects of Nocturnal Aircraft Noise and Non-Acoustical Factors on Short-Term Annoyance in Primary School Children. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6959. [\[CrossRef\]](#)
45. Elliott, E.M.; Morey, C.C.; AuBuchon, A.M.; Cowan, N.; Jarrold, C.; Adams, E.J.; Attwood, M.; Bayram, B.; Beeler-Duden, S.; Blakstvedt, T.Y.; et al. Multilab Direct Replication of Flavell, Beach, and Chinsky (1966): Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Adv. Methods Pract. Psychol. Sci.* **2021**, *4*, 25152459211018187. [\[CrossRef\]](#)
46. Poloczek, S.; Henry, L.A.; Messer, D.J.; Büttner, G. Do Children Use Different Forms of Verbal Rehearsal in Serial Picture Recall Tasks? A Multi-Method Study. *Memory* **2019**, *27*, 758–771. [\[CrossRef\]](#)
47. Alloway, T.P. How Does Working Memory Work in the Classroom? *Educational Research and reviews* **2006**, *1*, 134–139.
48. Gathercole, S.E.; Alloway, T.P. *Working Memory and Learning: A Practical Guide for Teachers*; SAGE Publications: Thousand Oaks, CA, USA, 2008; ISBN 9781412936132.
49. Schlittmeier, S.J.; Marsh, J.E. Review of Research on the Effects of Noise on Cognitive Performance 2017–2021. In Proceedings of the 13th ICBEN Congress on Noise as a Public Health Problem, Stockholm, Sweden, 14–17 June 2021.
50. Schlittmeier, S.J.; Weißgerber, T.; Kerber, S.; Fastl, H.; Hellbrück, J. Algorithmic Modeling of the Irrelevant Sound Effect (ISE) by the Hearing Sensation Fluctuation Strength. *Atten. Percept. Psychophys.* **2012**, *74*, 194–203. [\[CrossRef\]](#)
51. Bell, R.; Röer, J.P.; Lang, A.-G.; Buchner, A. Reassessing the Token Set Size Effect on Serial Recall: Implications for Theories of Auditory Distraction. *J. Exp. Psychol. Learn. Mem. Cogn.* **2019**, *45*, 1432–1440. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Cowan, N. *Attention and Memory: An Integrated Framework*; Oxford University Press: New York, NY, USA, 1995.
53. Hughes, R.W. Auditory Distraction: A Duplex-Mechanism Account. *Psych J.* **2014**, *3*, 30–41. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Jones, D.; Madden, C.; Miles, C. Privileged Access by Irrelevant Speech to Short-Term Memory: The Role of Changing State. *Q. J. Exp. Psychol. Sect. A* **1992**, *44*, 645–669. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Salamé, P.; Baddeley, A. Disruption of Short-Term Memory by Unattended Speech: Implications for the Structure of Working Memory. *J. Verbal Learn. Verbal Behav.* **1982**, *21*, 150–164. [\[CrossRef\]](#)
56. Neath, I.; Nairne, J.S.; Stevenson, A.K.; Surprenant, A.M.; Baddeley, A.; Jones, D.M. *Modeling the Effects of Irrelevant Speech on Memory*; Psychonomic Society: Chicago, IL, USA, 2000; Volume 7.
57. Wetzel, N.; Scharf, F.; Widmann, A. Can't Ignore—Distraction by Task-Irrelevant Sounds in Early and Middle Childhood. *Child Dev.* **2019**, *90*, e819–e830. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Joseph, T.N.; Hughes, R.W.; Sörqvist, P.; Marsh, J.E. Differences in Auditory Distraction between Adults and Children: A Duplex-Mechanism Approach. *J. Cogn.* **2018**, *1*, 13. [\[CrossRef\]](#)

59. Meinhardt-Injac, B.; Imhof, M.; Wetzel, N.; Klatte, M.; Schlittmeier, S.J. The Irrelevant Sound Effect on Serial Recall Is Independent of Age and Inhibitory Control. *Audit. Percept. Cogn.* **2022**, *5*, 25–45. [\[CrossRef\]](#)
60. Röer, J.P.; Bell, R.; Körner, U.; Buchner, A. Equivalent Auditory Distraction in Children and Adults. *J. Exp. Child Psychol.* **2018**, *172*, 41–58. [\[CrossRef\]](#)
61. Schwarz, H.; Schlittmeier, S.; Otto, A.; Persike, M.; Klatte, M.; Imhof, M.; Meinhardt-Injac, B. Age Differences in the Irrelevant Sound Effect: A Serial Recognition Paradigm. *Psihologija* **2015**, *48*, 35–43. [\[CrossRef\]](#)
62. Elliott, E.M.; Briganti, A.M. Investigating the Role of Attentional Resources in the Irrelevant Speech Effect. *Acta Psychol.* **2012**, *140*, 64–74. [\[CrossRef\]](#)
63. Buchner, A.; Bell, R.; Rothermund, K.; Wentura, D. Sound Source Location Modulates the Irrelevant-Sound Effect. *Mem. Cognit.* **2008**, *36*, 617–628. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Jones, D.M.; Macken, W.J. Auditory Babble and Cognitive Efficiency: Role of Number of Voices and Their Location. *J. Exp. Psychol. Appl.* **1995**, *1*, 216. [\[CrossRef\]](#)
65. Jones, D.M.; Macken, W.J. Organizational Factors in the Effect of Irrelevant Speech: The Role of Spatial Location and Timing. *Mem. Cogn.* **1995**, *23*, 192–200. [\[CrossRef\]](#)
66. Jones, D.M.; Saint-Aubin, J.; Tremblay, S. Modulation of the Irrelevant Sound Effect by Organizational Factors: Further Evidence from Streaming by Location. *Q. J. Exp. Psychol.* **1999**, *52*, 545–554. [\[CrossRef\]](#)
67. Steinbrink, C.; Klatte, M. Phonological Working Memory in German Children with Poor Reading and Spelling Abilities. *Dyslexia* **2008**, *14*, 271–290. [\[CrossRef\]](#)
68. Schiano, D.J.; Watkins, M.J. Speech-like Coding of Pictures in Short-Term Memory. *Mem. Cognit.* **1981**, *9*, 110–114. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Wetzel, N.; Widmann, A.; Scharf, F. Distraction of Attention by Novel Sounds in Children Declines Fast. *Sci. Rep.* **2021**, *11*, 5308. [\[CrossRef\]](#)
70. McCrea, S.M.; Mueller, J.H.; Parrila, R.K. Quantitative Analyses of Schooling Effects on Executive Function in Young Children. *Child Neuropsychol.* **1999**, *5*, 242–250. [\[CrossRef\]](#)
71. Alloway, T.P.; Gathercole, S.E.; Pickering, S.J. Verbal and Visuospatial Short-Term and Working Memory in Children: Are They Separable? *Child Dev.* **2006**, *77*, 1698–1716. [\[CrossRef\]](#)
72. Morey, C.C.; Miron, M.D. Spatial Sequences, but Not Verbal Sequences, Are Vulnerable to General Interference during Retention in Working Memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **2016**, *42*, 1907–1918. [\[CrossRef\]](#)
73. Guerreiro, M.J.S.; Murphy, D.R.; van Gerven, P.W.M. The Role of Sensory Modality in Age-Related Distraction: A Critical Review and a Renewed View. *Psychol. Bull.* **2010**, *136*, 975–1022. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Guerreiro, M.J.S.; Anguera, J.A.; Mishra, J.; van Gerven, P.W.M.; Gazzaley, A. Age-Equivalent Top-down Modulation during Cross-Modal Selective Attention. *J. Cogn. Neurosci.* **2014**, *26*, 2827–2839. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Rienäcker, F.; van Gerven, P.W.M.; Jacobs, H.I.L.; Eck, J.; van Heugten, C.M.; Guerreiro, M.J.S. The Neural Correlates of Visual and Auditory Cross-Modal Selective Attention in Aging. *Front. Aging Neurosci.* **2020**, *12*, 498978. [\[CrossRef\]](#) [\[PubMed\]](#)

2.6 Study 5: Background speech affects neural processing associated with the retention of serial order in visual-verbal material

Introduction

There is widespread agreement in behavioral research that background sounds have a detrimental impact on the performance of cognitive tasks, particularly in the context of serial order retention (Colle & Welsh, 1976; Jones & Macken, 1993; Klatte et al., 1995; Salamé & Baddeley, 1982; Schlittmeier, Weißgerber, et al., 2011). Only a limited number of studies have investigated the neuronal processing in relation to the ISE paradigm (Bell et al., 2010; Schlittmeier, Weisz, & Bertrand, 2011). To that end, researchers have used EEG that offers high temporal resolution (Kappenman & Luck, 2011). That makes it possible to precisely observe the timing of neural activity, such as the onset of picture on the computer screen or a sound played over headphones. To investigate these time-locked events resulting in ERPs, stimuli are typically presented multiple times and then averaged.

The ISE paradigm typically involves the visual presentation of verbal items for memorization in serial order (Meinhardt-Injac et al., 2022). These presentations are accompanied by task-irrelevant background sounds in certain trials, while trials without background sounds serve as control conditions. Previous studies have used EEG to investigate ERPs elicited by background sounds, i.e. auditory distractors (Bell et al., 2010; Campbell et al., 2003, 2007), as well as the ERPs elicited by the memory material (Little et al., 2010; Schlittmeier, Weisz, & Bertrand, 2011; Weisz & Schlittmeier, 2006). Compared to control conditions, the N1 ERP component is more negative due to task-irrelevant background sounds (Bell et al., 2010; Campbell et al., 2003, 2007; Little et al., 2010; Weisz & Schlittmeier, 2006).

The N1 and MMN have strong temporal proximity and similar scalp topographies, making it challenging to distinguish these two components. Within the context of the ISE paradigm, only one study has explored the differentiation between the MMN and the N1 components: Campbell et al. (2007) compared the ERPs elicited during a changing-state condition to those during a deviant condition, which an infrequent deviant interrupted a steady-state sequence. While they confirmed the enhanced N1 response to changing-state tones compared to steady-state distractors, the MMN was observed only in the deviant condition (Bell et al., 2010; Campbell et al., 2007). Consequently, they concluded that the N1 and MMN are distinct and functionally separate ERP components within the ISE context (Campbell et al., 2007). The P3a ERP component frequently follows the N1 component (Bell et al., 2010). It is typically linked to an orienting response (Polich, 2007), which aligns with a conscious evaluation of the auditory stimulus (Bell et al., 2010). The impact of task-irrelevant sounds on the P3a component shows mixed results: while Campbell et al. (2003, 2007) did not observe a P3a elicited by auditory distractors, Bell et al. (2010) reported a P3a in response to changing-state sounds.

The results of the previous studies highlight the detrimental impact of background speech and non-speech sounds on visual-verbal serial recall, influenced by phonology and changing-state characteristics, regardless of sound presentation mode and age, with no

habituation observed. These findings suggest that the negative effects on short-term memory performance are primarily due to speech-specific factors rather than attentional factors, necessitating further investigation to fully understand the ISE. The aim of the following studies is to investigate the neuronal mechanisms of the ISE. Study 5 investigates the ERP components associated with the ISE, specifically the N1 and P3a component in response to the memory material.

The current body of neurophysiological research on the ISE paradigm has predominantly focused on the processing of auditory distractors. Only a limited number of studies have examined the neural mechanisms that underlie the negative effects of task-irrelevant background sounds on to-be-remembered items. Furthermore, the findings from these studies are inconclusive. While Little et al. (2010) and Weisz and Schlittmeier (2006) demonstrated a modulation in the N1 amplitude of visually presented items in the presence of background speech, Schlittmeier, Weisz, and Bertrand (2011) found no significant N1 amplitude changes in audibly presented memory material.

The objective of this study was to replicate the findings of Little et al. (2010) using EEG measurements within a traditional ISE paradigm. I sought to broaden the scope of the investigation by analyzing additional ERP components as suggested by Bell et al. (2010). In their study, Bell et al. (2010) linked the P3a to attention capture induced by changing-state distractors. However, the authors noted a temporal overlap between their examination of the P3a and another ERP component, namely the P2. To distinguish between these components, Bell et al. (2010) recommend examining the respective scalp topography.

Behaviorally, I expect to replicate the classical ISE, with lower recall performance in the unfamiliar speech condition compared to the pink noise condition (e.g., Schlittmeier, Weisz, & Bertrand, 2011). Regarding ERP components, I predict a heightened N1 amplitude in the unfamiliar speech condition compared to the pink noise condition. Additionally, I aim to disentangle the P2 and P3a components, considering the potential overlap identified by Bell et al. (2010). If attention capture plays a central role in the negative impact of task-irrelevant background sounds, I expect a discernible difference in the P3a, reflected in a more positive amplitude in the speech condition (e.g., Bell et al., 2010). Studying the P2 independently presents challenges due to its overlap not only with the P3a but also with other components such as the N1 (Crowley & Colrain, 2004). Prior studies have often examined the P2 component in relation to the “P1-N1-P2” complex (Kappenman & Luck, 2011), leaving its functional role not fully understood (Crowley & Colrain, 2004; Sokhadze et al., 2017). While some researchers suggest a connection between the P2 and attentional processes (Novak et al., 1992), alternative viewpoints propose associations with stimulus discrimination (García-Larrea et al., 1992) or task difficulty (Conley, 1999; Sokhadze et al., 2017). However, none of these studies have specifically investigated the P2 component within the ISE context. I assume a link between the N1 and the P2: while the N1 shows increased amplitude, i.e., more negative amplitudes, in the speech condition, the amplitude of the P2 decreases, resulting in a less positive amplitude (Crowley & Colrain, 2004). Thus,

if attention, as indicated by P2, is a contributing factor, its amplitude should decrease when task-irrelevant speech is presented.

Methods

Participants The task was performed by 43 adults (21 female, 22 male) aged between 19 and 29 years ($M = 23.1$ years, $SD = 2.46$ years). All participants were native German speakers. They had both normal or corrected-to-normal vision and normal hearing according to self-reports. They received either course credit or payment for their participation (10€). Informed written consent was provided by all participants. The study was approved by the ethical review board of the Faculty of Social Science of the University of Kaiserslautern-Landau.

Apparatus A 15.6-inch laptop (Thinkpad 590) running Python 3.7/PschoPy 3.1.5 (Peirce, 2007) on Windows 10 was used to control the experimental trial. Display resolution was 1920×1080 , with a refresh rate of 60 Hz. The background sounds were played via open headphones (Sennheiser HD210), connected with a Focusrite Scarlett 2i2 2nd generation audio interface to the laptop.

Irrelevant Sounds Task performance was measured during two different sound conditions: pink noise and fluent speech. Pink noise was selected as a control condition since it provides auditory stimulation without impairing serial recall performance (Schlittmeier, Weisz, & Bertrand, 2011; Weisz & Schlittmeier, 2006). The speech condition comprised a male Korean speaker continuously talking. The language was unfamiliar to the participants.

To-be-remembered items The items are identical to Leist et al. (2023). A total of 14 colored drawings representing monosyllabic German words were used: *Bett* (bed), *Bus* (bus), *Eis* (ice), *Frosch* (frog), *Haus* (house), *Herz* (heart), *Kamm* (comb), *Mond* (moon), *Nuss* (nut), *Pilz* (mushroom), *Schal* (scarf), *Schiff* (ship), *Schwein* (pig), and *Zaun* (fence). Ten lists, each containing seven items (drawn out of 14), were created. The construction of these lists followed a quasi-random approach, ensuring that each item appeared in exactly five out of the ten lists. To obtain a total of 40 lists, additional three sets of random permutations were applied. The drawings were displayed in a rectangular black frame (measuring 102×73 mm) centered on a white screen. During the recall phase, the pictures were simultaneously presented again and randomly positioned within a fixed array of seven black frames.

Serial Recall Task The task is identical to Leist et al. (2023). Each trial comprised a presentation phase, a retention interval, and a recall phase. During the presentation phase, items were visually presented sequentially, each with a duration of 1500 ms and an interstimulus interval of 500 ms. Following the last item, a retention interval of 5000 ms ensued. The recall phase commenced with the reappearance of all items from

the corresponding trial. Participants were then required to reconstruct the serial order by sequentially clicking on the items in the order of their presentation. Clicking an item changed its opacity, indicating its selection. There was no time limit for responding and no possibility of error correction. After selecting the final item, participants were presented with a visual cue to start the next trial by pressing the space bar.

Procedure Participants were tested individually at the University of Kaiserslautern-Landau, engaging in two tasks: a source memory task (not reported here) and the serial recall task. After signing the consent form, participants were prepared for EEG measurement and seated in front of the laptop (approximately 80 cm). Participants received written instructions, explicitly directing them to ignore the auditory distractors and to focus only on the serial recall of the visually presented pictures. Before the experiment started, all pictures used in the task were presented and named by a female speaker. To introduce the task-irrelevant background sound to the participants, it was played for 4 s, followed by one practice trial per sound condition (two in total). Following that, the participants performed a total of 40 trials using the constructed lists, with 30 trials allocated to each sound condition. The sound conditions varied from trial to trial with the constraint that consecutive trials included no more than two instances of the same sound condition. In the speech condition, the sound started when the participant initiated the trial with the space bar. A random interval between 1200 and 1800 ms was introduced before the presentation of the first list item, to avoid correlations between sound and item onset (Bell et al., 2010). The sound was played throughout both the item presentation phase and retention interval and stopped at the beginning of the recall phase. The average sound level (LAeq) of the syllable sequences was 60 dB, as measured by an artificial ear (Brüel & Kjær, Type 4153). The duration of the serial recall task was approximately 30 minutes, while the entire EEG measurement, including preparation time and the source memory task, took approximately 90 minutes.

EEG recording Continuous EEG was recorded using Brain Vision Recorder software (Version 2.25), amplified by a BrainAmp Standard amplifier (BrainProducts GmbH, Gilching, Germany) with 16 bit A/D conversion. The EEG signal was recorded with a sampling rate of 500 Hz. A 32-channel system EEG-actiCap (BrainProducts GmbH, Gilching, Germany) was used to monitor brain signals from active Ag/AgCl electrodes according to the extended 10-20 system (Jasper, 1958) at the positions Fp1, Fp2, F3, F4, F7, F8, Fz, FC1, FC2, FC5, FC6, C3, C4, Cz, CP1, CP2, CP5, CP6, P3, P4, P7, P8, Pz, T7, T8, Oz, M1, M2. Additionally, two electrodes were placed lateral to the outer canthi recording horizontal EOG and two electrodes above and below the right eye recorded vertical EOG. The ground electrode was placed at position AFz and the reference electrode at position FCz. Impedances of the electrodes were kept below 10 k Ω . The study was conducted in a room with low lighting.

EEG data pre-processing To perform the processing and analysis of the ERPs Brain Vision Analyzer-software (Version 2.2.2, BrainProducts, Gilching, Germany) was used. EEG was re-referenced offline to averaged mastoids (M1, M2). A 0.5–40 Hz (24 dB/oct) bandpass-filter (Butterworth) was applied. In the raw data inspection epochs containing large muscle artifacts were excluded. Ocular movements were corrected by an Independent component analysis (ICA)-based correction (as implemented in the Brain Vision Analyzer Software). Artifacts were rejected using a semi-automatic procedure, followed by manual selection based on visual inspection. Artifacts were identified for rejection if they met one of three criteria: (1) amplitudes exceeding 20 $\mu\text{V}/\text{ms}$ at any electrode site, (2) voltage increasing or decreasing more than 75 μV within 200 ms, or (3) activity falling below 0.5 μV for 100 ms.

Stimulus-locked ERPs of the presentation phase were segmented for 1200 ms. These segments started 200 ms before the onset of the visually presented item and continued until 1000 ms after its onset. On average, 114 (range: 20–179) speech segments and 127 (range 26–196) pink noise segments entered statistical analyses. Data was baseline corrected 200 ms before stimulus onset.

Following a visual examination of the grand averages (Wetzel et al., 2011), the ERP amplitudes were individually identified at the regions of interest, specifically central electrode positions (C'; mean of C3, Cz, C4; (Debruille et al., 2019; Wetzel et al., 2011)) and frontal electrode positions (F'; mean of F3, Fz, F4; (Wetzel et al., 2006, 2011)).

The N1 peak amplitude (μV) was locally identified within a time window spanning from 80 to 180 ms, and averaged within 10 ms around the peak for each participant. The P2 peak amplitude (μV) was locally identified within a time window of 170 to 210 ms, and averaged within 10 ms around the peak for each participant. The P3a peak amplitude (μV) was identified within a time window spanning from 210 to 350 ms, and averaged within 15 ms around the peak for each participant. These windows were used to analyze individual peaks of the respective ERP components.

Behavioral analysis Analyses were conducted using scores reflecting the proportion of correctly recalled items at their respective serial positions, based on a strict serial-recall criterion (Bell et al., 2010). Data were statistically analyzed using one-way repeated measures Analysis of Variance (rmANOVA) with the factor Sound Condition (pink noise, speech).

ERP analyses In accordance with the behavioral data, the analysis only considered items from the presentation phase that were correctly recalled later. The statistical significance of each ERP component (N1, P2, and P3a) was separately evaluated using rmANOVA, with factors Sound Condition (pink noise, speech) and Sagittal Electrode Sites (F', C'). Bonferroni-corrected *t*-tests were used to clarify significant interactions.

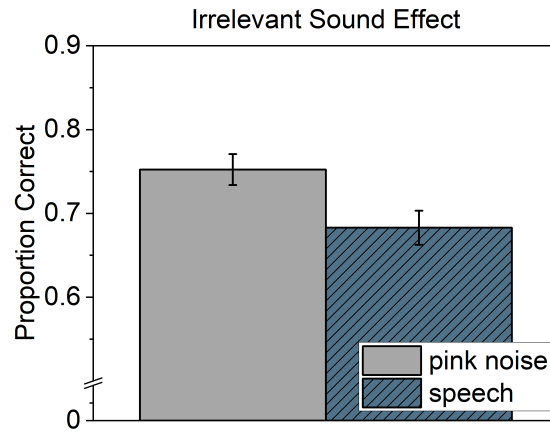


Figure 2.1: Mean proportion correct scores as a function of sound condition (pink noise, speech) in a serial recall task of seven words presented pictorially. Error bars denote standard errors of the mean.

Table 2.1: Mean amplitudes for ERP components (N1, P2, P3a) at frontal (mean of F3, Fz, F4) and central (mean of C3, Cz, C4) electrode positions with respect to the sound conditions (speech, pink noise).

Measure	N1 (μV)		P2 (μV)		P3a (μV)	
	frontal	central	frontal	central	frontal	central
pink noise	-2.08	-1.83	.41	.98	-.30	.89
speech	-2.34	-2.13	.34	.64	-.47	.61

Results

1. Behavioral results of the ISE Proportion correct scores with respect to sound conditions are depicted in Figure 2.1. The one-factorial rmANOVA with the factor Sound Condition (pink noise, speech) revealed a significant main effect, $F(1, 42) = 33.913$, $p < .001$, $\eta_p^2 = .447$. Recall performance in the speech condition ($M = 68.3$, $SD = 13.4$) was more disrupted compared to performance in pink noise ($M = 75.2$, $SD = 12.1$).

2. ERP amplitude measures ERPs are shown in Figure 2.2. Mean amplitudes for averaged N1, P2 and P3a are displayed at Table 2.1.

N1 For N1 amplitudes (μV), the 2×2 rmANOVA revealed a significant main effect of Sound Condition, $F(1, 42) = 7.058$, $p = .011$, $\eta_p^2 = .144$, reflecting that N1 amplitude is more negative in the speech condition than in the pink noise condition. There was no significant effect of Sagittal Electrode Sites ($p = .105$), and no significant interaction of these factors ($p = .599$), suggesting comparable scalp distributions for N1 amplitude.

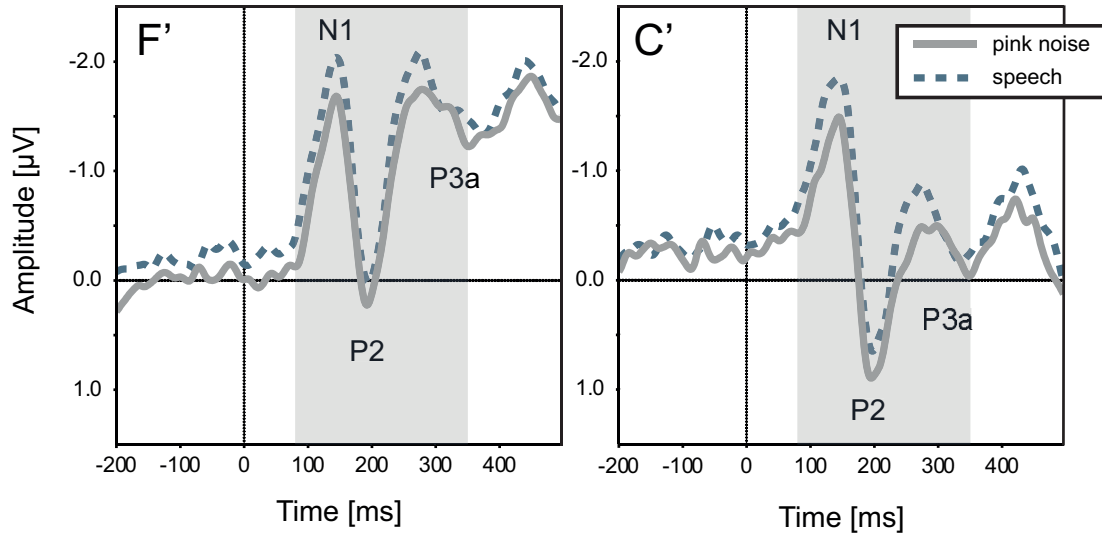


Figure 2.2: Grand averages with respect to the sound conditions (speech, pink noise) of the serial recall task at F' (left panel) and C' (right panel). The grey area shows the time windows of the three separate ERP components that partially overlap. These windows were used to analyze individual peaks. Negative is plotted upward.

P2 For P2 amplitudes (μV), the 2×2 rmANOVA revealed no significant effect of Sound Condition ($p = .139$), a significant main effect of Sagittal Electrode Sites, $F(1, 42) = 11.199$, $p = .002$, $\eta_p^2 = .211$, and a significant interaction of these factors, $F(1, 42) = 4.790$, $p = .034$, $\eta_p^2 = .102$, indicating a more pronounced increase in positivity at central sites. This was further analyzed separately for each electrode site because of the significant two-way interaction. Post-hoc t -tests confirmed that P2 amplitude was more positive in pink noise than speech at central sites, $t(42) = -2.122$, $p = .040$, $d = -.655$. At frontal sites, the sound conditions did not differ significantly ($p = .211$).

P3a For P3a amplitudes (μV), the 2×2 rmANOVA revealed only a trend level of Sound Condition ($p = .066$), and a significant main effect of Sagittal Electrode Sites, $F(1, 42) = 64.452$, $p < .001$, $\eta_p^2 = .605$. There was no significant interaction of these factors ($p = .437$), indicating comparable amplitudes across sound conditions in the P3a.

Discussion

The present study explored the effects of task-irrelevant unfamiliar background speech compared to a pink noise control condition on serial recall of seven words presented pictorially, while simultaneously measuring EEG activity. To summarize, these are the four main findings: 1) In terms of behavioral performance, serial recall performance decreased in the speech condition compared to the pink noise condition. 2) The amplitude of N1 was more negative in the speech condition compared to the pink noise condition. This effect is comparable across both frontal and central electrode positions. 3) The amplitude of P2 was more positive in the pink noise condition compared to the speech condition at central sites, but there was no significant difference in sound conditions at the frontal side. 4) A trend-level difference in P3a amplitude was observed between speech and pink noise at both frontal and central electrode positions, but this finding remains somewhat inconclusive.

At the behavioral level, the disruptive effect on visual-verbal serial recall due to task-irrelevant speech that is unfamiliar to the participants is confirmed, replicating earlier findings (Salamé & Baddeley, 1986, 1987). The disruptive effect, which represents the difference in performance between the control and speech condition, of around 7 % falls within the typical range, which can range anywhere from 5 % to 50 % (Gisselgård et al., 2004; Neath, 2000).

The increase of the N1 amplitude in the speech condition, in contrast to the pink noise condition, aligns with findings from previous studies (Little et al., 2010; Weisz & Schlittmeier, 2006). The N1 is associated with attentional processing (e.g., Bell et al., 2010; Little et al., 2010; Weisz & Schlittmeier, 2006) and memory disruption (Campbell et al., 2003, 2007). The finding that N1 amplitude for visually presented items is modulated by task-irrelevant speech indicates that the brain allocates resources to process and potentially filter out task-irrelevant auditory information, even when it is not the focus of attention. This result emphasizes how sensitive our brain reacts to sounds from our environment, and suggests that the N1 component can serve as a valuable neural marker for investigating the impact of audible environmental distractions.

The increase of the P2 amplitude in the pink noise condition, in contrast to the speech condition, aligns with findings from previous studies (Little et al., 2010). However, this finding challenge Bell et al. (2010)'s results, who temporally attributed their results to the P3a. As previously noted, assessing the P2 and P3a components in isolation poses a considerable challenge. Here, the P2 was more positive in the pink noise control condition, in contrast to Bell et al. (2010)'s finding that the P2/P3a component was more positive in the changing-state speech condition. However, some researchers suggest that with rising attentiveness, the amplitude of the N1 increases while the amplitude of the P2 decreases (Crowley & Colrain, 2004). Thus, the P2's lower positive amplitude in the speech condition supports the notion that attentional mechanisms are at play. Nevertheless, it is important to note some methodological differences between the two studies. Our research focuses

2 Empirical Investigations

on examining the impact of auditory distractors on visually presented material, thereby investigating the ERP components related to the memory material presented visually. In contrast, the study by Bell et al. (2010) explored the ERP components of the auditory distractors. The specific role of the P2 component in isolation remains uncertain (Crowley & Colrain, 2004; Sokhadze et al., 2017), as it is commonly studied within the broader context of the "P1-N1-P2" complex (Kappenman & Luck, 2011) and its' overlay with other ERP components. However, some researchers suggest associations between the P2 and attentional processes (Novak et al., 1992), stimulus discrimination (García-Larrea et al., 1992), or task difficulty (Conley, 1999; Sokhadze et al., 2017).

The results regarding the P3a component were unexpected compared to existing literature. In this study, a trend-level modulation in P3a amplitude was observed between speech and pink noise conditions, though the result remains somewhat inconclusive. This contrasts with the findings of Bell et al. (2010), who reported a higher positive deflection in the changing-state condition compared to the steady-state condition. However, Bell et al. (2010) suggest a potential overlap between the P2 and the P3a components, as the positive wave they investigated began relatively early (200 ms in contrast to 300 ms). In consideration of the scalp topography, Bell et al. (2010) argue that the component they investigated is the P3a, as opposed to the P2. However, it cannot be definitively excluded that Bell et al. (2010) might have examined the P2, given its temporal alignment within the specified timeframe. Through the utilization of peak detection throughout distinct time intervals, we distinguished the P2 from the P3a. Considering the respective time windows, the results from this study challenge Bell et al. (2010)'s findings.

To conclude, the observed increase in N1 amplitude in response to speech compared to pink noise aligns with prior studies, indicating heightened attentional processing and memory disruption in the presence of speech. This suggests that the brain allocates additional attentional processes to filter out irrelevant auditory stimuli, potentially at the expense of encoding and retaining visual-verbal information. The unexpected finding of a more positive P2 amplitude in the pink noise condition challenges previous observations and suggests nuanced attentional mechanisms. While the exact interpretation remains uncertain, it may indicate differences in cognitive processing strategies between speech and pink noise conditions, possibly influenced by task demands or individual differences in auditory processing. Unexpectedly, a trend-level modulation in P3a amplitude were found between speech and pink noise conditions, contrary to previous studies. This may indicate that the neuronal mechanisms involved in re-orientation might be similar across sound conditions, implying that attentional mechanisms, like re-orientation, might have been overestimated in explaining the ISE's elicitation. The functional role of the P2 and P3a remains ambiguous and underscores the complexity of ERP components and the need for further investigation into the specific cognitive processes underlying auditory distraction.

As highlighted by Weisz and Schlittmeier (2006), the limited convergence between cognitive ISE models and neuroscientific methodologies presents a significant challenge in their integration. Defining abstract concepts like 'modality-dependent information' in neuro-

physiological terms is complex, and interpretations may diverge between behavioral and neurophysiological studies. However, our study underscores the importance of integrating these approaches. We found that irrelevant background speech adversely affects the brain's processing and retention of crucial information, likely attributable to diminished attention during both presentation and rehearsal phases.

This study highlights the intricate interplay between auditory distractions, attentional processes, and memory functions. The findings contribute to our understanding of how background speech affects cognitive performance and underscore the importance of considering neural indices for theoretical considerations of the ISE. Further investigation is needed to distinguish the functional, topographical, and temporal characteristics of both P2 and P3, both in the context of visually presented material and auditorily presented distractors. Such research could contribute to a better understanding of the involvement of attentional processes in performance degradation within the serial recall task.

2.7 Study 6: Relation between Auditory Discrimination and Susceptibility to Disruptive Effects of Background Speech

Introduction

Attentional processes play a crucial role in cognitive functions such as working memory performance, which involves the temporary storage and manipulation of information necessary for ongoing tasks (Oberauer, 2019). The impact of irrelevant environmental sounds that interfere with working memory performance has garnered significant interest in cognitive neuroscience (Gisselgård et al., 2004). The processes in the frontal cortex are believed to be responsible for the shift in attention and the conscious perception of changes in stimuli (Näätänen et al., 2005). Prior research has shown a correlation between memory disruption and early sensory processing, i.e., increased N1 amplitudes, as demonstrated by Study 5 and others (Bell et al., 2010; Campbell et al., 2003). The goal of the present study is to investigate the interplay between the detrimental impact of task-irrelevant speech on serial recall performance and the underlying neural mechanisms of auditory discrimination.

Performance in serial recall of verbal items, which is a classic working memory task, is known to be disrupted by task-irrelevant background sounds (Colle & Welsh, 1976). The detrimental impact of background sounds on working memory performance, known as the ISE, occurs in response to both speech (Colle & Welsh, 1976) and non-speech sounds, including music (Klatte et al., 1995; Schweppe & Knigge, 2020) and tones (Jones & Macken, 1993). Traditionally, the ISE has been studied in serial recall tasks in which participants are asked to remember the order of a list of verbal items presented visually on a computer screen while ignoring background sounds played through headphones or loudspeakers. Crucially, the ISE is elicited by background sounds with temporal-spectral variations, such as changing-state sounds. In contrast, continuous noise and repeating speech sounds, like steady-state sounds, do not disrupt serial recall performance (Jones et al., 1996). The ISE has been attributed to specific interference between pre-attentive, automatic sound processing and deliberate processes involved in retention of the memory lists (the changing-state effect) as well as to attentional capture (the deviation effect; (Hughes, 2014)).

Theoretical models propose that the brain automatically processes auditory stimuli, even when they are task-irrelevant, leading to interference with ongoing cognitive processes (Marois et al., 2019). One prominent neural marker of this interference is the MMN, an ERP component associated with pre-attentive detection of deviant auditory stimuli (Näätänen, 1990). Although the MMN was initially observed and extensively studied in the auditory modality, there are reports indicating its elicitation in the visual domain as well (Beck et al., 2021; Kujala et al., 2007). A commonly accepted interpretation of the MMN suggests that it arises from an automatic change-detection process (Näätänen et al., 2005). Traditionally, the MMN has been assessed through the oddball paradigm, which involves presenting a frequent “standard” sound alongside infrequent “deviant” sound. The repeating presentation of the “standard” sound establishes a memory trace in the auditory

system. When a different sound fails to align with this imprint, the MMN is evoked, typically occurring around 100 to 250 milliseconds after stimulus onset at fronto-central regions of the scalp (Näätänen, 2001). This indicates the auditory system's ability to discern between the two sounds (Kujala et al., 2007). It is additionally suggested that this process of change detecting occurs pre-attentively in the auditory cortices. As a result, this process stimulates activity in the frontal cortex, which produces the MMN and contributes to the following P3a.

The P3a, identified as the largest positive peak within the ERP waveform, emerges within a time window spanning approximately 250-500 ms (Wetzel et al., 2006). As part of the broader ERP component known as the P300, both the P3a and P3b subcomponents are delineated by distinct characteristics in localization, timing, and function (Polich, 2007). Specifically, the P3a exhibits its peak activity at central sites, whereas the P3b peaks at parietal locations. Additionally, the P3a typically precedes the P3b in temporal occurrence and is evoked by novel or unexpected stimuli, contrasting with the P3b's responsiveness to task-relevant target stimuli (Polich, 2007). Moreover, while the P3a signifies involuntary attention capture in response to novel or salient stimuli, the P3b is associated with the allocation of attention and subsequent memory processing (Polich, 2007). The current study investigates the involuntary attention capture, i.e., the P3a, and the pre-attentive change-detection, i.e. the MMN.

Although prior studies have thoroughly explored the MMN and P3a in relation to auditory processing (see, for example Koerner et al., 2016, 2017; Wetzel et al., 2006, 2011), fewer studies have investigated their association with working memory performance, especially within the framework of the ISE paradigm (Bell et al., 2010; Campbell et al., 2003, 2007; Little et al., 2010; Weisz & Schlittmeier, 2006).

Auditory signal discrimination, especially speech signals, is impaired by noise exposure, due to alterations in brain function (for a review, see Kujala & Brattico, 2009). Chronic noise exposure induces stress (Evans et al., 1995), which in turn can negatively impact cognitive function, including auditory processing abilities. According to the Umweltbundesamt (Wilke, 2024), the main source of noise in Germany stems from road traffic. Thus, investigating traffic noise and its impact on speech discrimination holds significant ecological validity, as it reflects the auditory challenges people frequently encounter in daily life. By examining how traffic sounds affect individuals' ability to understand speech, researchers can gain valuable insights into the difficulties they face in real-life communication situations. Kujala et al. (2007) has demonstrated that the MMN component is negatively impacted by traffic noise. Until now, there is no study investigating the detrimental impact of traffic noise on the P3a component for speech discrimination. Therefore, this study includes a condition in which traffic noise is presented in order to examine how the MMN and P3a components are affected by noise.

The primary objective of this study is to bridge the gap between auditory processing and serial order retention. This will be achieved by exploring the potential involvement of neural indicators, specifically those associated with pre-attentive auditory change detection

2.7 Study 6: Auditory Discrimination and Background Speech Susceptibility

(MMN) and involuntary attention capture (P3a). The focus is on how these indicators modulate the adverse effects of task-irrelevant background speech on visual-verbal serial order recall.

This investigation follows a two-sided approach. First, the study seeks to replicate the negative effects of task-irrelevant speech on serial order retention. In line with the findings from the verbal task in Study 1, it is assumed that changing-state syllables will have a more significant impact on performance than steady-state syllables or silence.

Second, the study investigates the neural correlates of auditory discrimination through a passive oddball paradigm, measuring the MMN and P3a by comparing consonant and vowel change detection. This further includes an analysis of these ERP components in varying background sound conditions (silent control vs. traffic noise) to assess whether consonants and vowels are processed differently in traffic noise compared to a silent environment.

Based on prior research findings (Koerner et al., 2016; Kozou et al., 2005), it is hypothesized that traffic noise will lead to delayed MMN latency and reduced MMN amplitude, with a greater reduction in MMN amplitude for vowels compared to consonants. Furthermore, aligning with previous studies (Koerner et al., 2017), it is assumed that traffic noise will result in reduced P3a amplitude and delayed peak latencies. Additionally, it is proposed that traffic noise differentially affects P3a responses for consonant and vowel changes, with a greater impact on consonants than on vowels (Koerner et al., 2017).

Finally, the main question addresses the potential relationship between the MMN and the P3a ERP components and susceptibility to background sounds in serial order retention. The hypothesis is, that participants with high sensitivity to pre-attentive change detection will show greater impairment in serial recall when exposed to changing-state speech compared to those with lower sensitivity. Based on the theoretical models, the following implications can be derived for the investigation. Both the changing-state hypothesis and the phonological loop theory suggest that the ISE arises from specific interference. According to the changing-state hypothesis, the ISE arises from impaired rehearsal efficiency due to pre-attentive processing of changing-state sounds (Jones & Tremblay, 2000; Jones et al., 1996). Additionally, the phonological loop account posits that the ISE occurs because task-irrelevant speech overwrites the phonological representations of verbal items in the phonological store. If the MMN is related to sound-induced performance decrements in serial order retention, this could support theories attributing the ISE to specific interference. In contrast, the P3a is often associated with involuntary attention capture and the processing of unexpected or salient stimuli, which aligns well with the concept of attention capture as proposed as a theoretical reason for the elicitation of the ISE (Cowan, 1995). A significant correlation would suggest that changes in the P3a component reflect the extent to which irrelevant sounds impact serial order retention, thereby providing evidence for the role of attentional capture in the ISE.

Methods

Participants A total of 23 student volunteers from the University of Kaiserslautern-Landau participated for either course credits or 10 €. Two participants were excluded from the analysis due to technical issues as no triggers were sent during the recording of the EEG. Data from 21 participants (12 female, 9 male) aged between 19 and 31 years ($M = 24.8$ years, $SD = 4.00$ years) were included in the analysis. All participants were right-handed native German speakers and had both normal or corrected-to-normal vision and normal hearing according to self-reports. Every participant provided written informed consent. The study was approved the ethical review board of the Faculty of Social Science of the University of Kaiserslautern.

Apparatus A 15.6-inch laptop (Thinkpad 590) and a 15.6-inch laptop (HP Pro Book 450) running Python 3.7/PschoPy 3.1.5 (Peirce, 2007) on Windows 10 were used to control the experimental trials of the EEG and the behavioral parts, respectively. Display resolution on both laptops was 1920×1080 , with a refresh rate of 60 Hz. In the behavioral part, sounds were played via open headphones (Sennheiser HD650). In the EEG part, the traffic noise was presented via loudspeakers. The two loudspeakers were placed behind the participant's head, with each loudspeaker positioned approximately 0.7 meters away from the participant's head.

Tasks *Oddball paradigm* A simplified illustration of the task is depicted in Figure 2.3. To measure the MMN ERP component a classical auditory oddball paradigm with a block composition of 75 % standards (/ka:/) and 25 % deviants (vowel change: /ke:/, or consonant change: /ta:/) was employed. Deviants were presented in a pseudo-random order within each block. Each block commenced with the presentation of 5 standard stimuli. There were 4 blocks, each consisting of 220 trials. The sequence of stimuli were pseudo-randomized. Each deviant was preceded by at least 3 standard stimuli. The stimuli were presented with a stimulus-onset-asynchrony (SOA) of 800 ms.

Serial recall task The serial recall task was identical to Study 1 with one exception: instead of words presented pictorially, numbers from 1 to 9 served as to-be-remembered material. The numbers were written in the font Arial and presented in a rectangular black frame (102×73 mm) in the center of a white computer screen. In each trial, lists of eight numbers were presented. In line with Weisz and Schlittmeier (2006), a random generator was used to generate the lists, with the implementation that no consecutive numbers and common number-combinations are included (e.g., 1, 2, 3 and 7, 6, 5 and 2, 4, 6).

Sounds *Oddball paradigm* Three types of consonant-vowel syllables were produced by a trained female speaker in a sound-attenuated laboratory and recorded as .wav files with a sampling rate of 44.1 Hz and a 16-bit resolution. The standard syllable was /ka:/ and the deviant was either /ke:/ (vowel change) or /ta:/ (consonant change). Both deviants were never presented in the same block. The epochs were 175 ms long. The background sound



Figure 2.3: A simplified illustration of the procedure of the passive auditory oddball paradigm. Each black square represents the presentation of a standard (here: /ka:/). The blue squares represent the presentation of a deviant (/ke:/ in vowel change condition, /ta:/ in consonant change condition). Each block contained either the deviants for the consonant change or the deviants for the vowel change.

was a recording of traffic noise (heavy traffic at a distance of 280 m, obtained from DEGA, 2023). The average sound level (LAeq) of the consonant-vowel syllables and the traffic noise was 63 dB, as measured by an artificial ear (Brüel & Kjær, Type 4153), yielding a SNR of 0 dB.

Serial recall task A trained female speaker produced consonant-vowel syllables in a sound-attenuated laboratory, recorded as .wav files with a 44.1 kHz sampling rate and 16-bit resolution. The set included 21 syllables, each approximately 450–600 ms in length (/ba:/, /be:/, /bi:/, /bo:/, /bu:/, /ga:/, /gi:/, /go:/, /gu:/, /ka:/, /ke:/, /ki:/, /ko:/, /pa:/, /pe:/, /pi:/, /pu:/, /ta:/, /ti:/, /to:/, /tu:/). Changing-state trials presented a random sequence with consecutive syllables differing in both consonant and vowel (e.g., ba-ko-tu). In steady-state trials, a single syllable was chosen from the set without replacement and subsequently repeated (e.g., ga-ga-ga). A 200 ms silent gap was inserted between consecutive syllables.

Procedure Participants were tested individually at the University of Kaiserslautern-Landau, engaging in two parts of the study: an EEG part recording the oddball paradigm to measure the MMN and the P3a, as well as a behavioral part to measure serial recall performance. Half of the participants completed the behavioral part one year earlier and were re-invited for the EEG part. The other half of the participants completed the EEG part first, following by the behavioral part within 1–2 weeks. The following section will first describe the procedure of the EEG part, then the procedure of the behavioral part. In the EEG part, after signing the consent form, participants were prepared for EEG measurement and seated in front of the laptop (approximately 0.8 m). Then, a practice block consisting of 12 stimulus presentations (out of 10 standards and 2 deviants) were presented to familiarize participants and verify the functionality of the sound presentation. Participants were asked to confirm their ability to hear the syllables played via headphones. Participants received written instructions. During the task, participants watched a silent video and were instructed to ignore all sounds. Two background sound conditions, silence and traffic noise, were implemented in a consistent order for all participants to avoid carryover effects of the noise. Each condition lasted approximately 30 minutes.

In the behavioral part, participants were tested in groups of up to four individuals within

a sound-attenuated booth at the University of Kaiserslautern-Landau. The procedure was identical to the adult study detailed in section 2.2 (Leist et al., 2023).

EEG recording Continuous EEG was recorded using Brain Vision Recorder-software (Version 2.25), amplified by LiveAmp (BrainProducts GmbH, Gilching, Germany). The EEG signal was recorded with a sampling rate of 1000 Hz. A Stimtrack-device connected the computer to the headphones controlling for correct onset of the event triggers. A 16-channel system EEG-actiCap (BrainProducts GmbH, Gilching, Germany) was used to monitor brain signals from active Ag/AgCl electrodes according to the extended 10-20 system (Jasper, 1958) at the positions F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, M1, M2, and nose. Additionally, two electrodes were placed lateral to the outer canthi recording horizontal EOG and two electrodes above and below the right eye recorded vertical EOG. The ground electrode was placed at position AFz and the reference electrode at position FCz. Impedances of the electrodes were kept below 15 k Ω . The study was conducted in a room with low lighting.

EEG pre-processing To perform the processing and analysis Brain Vision Analyzer-software (Version 2.2.2, BrainProducts, Gilching, Germany) was used. EEG was re-referenced offline to the nose. A 0.1–40 Hz (24 dB/oct) bandpass filter (Butterworth) was applied. In the raw data inspection epochs containing large muscle artifacts were excluded. Artifacts were rejected using a semi-automatic procedure, followed by manual selection based on visual inspection. Artifacts were identified for rejection if they met one of three criteria: (1) amplitudes exceeding 20 μ V/ms at any electrode site, (2) voltage increasing or decreasing more than 75 μ V within 200 ms, or (3) activity falling below 0.5 μ V for 100 ms.

Standards following deviants were excluded from the analysis. Stimulus-locked ERPs were segmented for 700 ms, starting 100 ms before a stimulus until 600 ms after the onset of the stimulus. The average remaining trial numbers that entered statistical analyses in the silent condition were 1812 (range: 1402–1840) standard segments, 428 (range 224–440) deviants with consonant change, and 438 (range 430–440) deviants with vowel change. In the traffic noise background condition 1841 (range: 1762–1840) standard segments, 443 (range 416–440) deviants with consonant change, and 438 (range 429–440) deviants with vowel change entered statistical analyses. Data was baseline corrected 100 ms before a given response.

Following a visual examination of the grand averages (Wetzel et al., 2011), the ERP amplitudes were individually identified at central electrode positions (C'; mean of C3, Cz, C4; (Debrulle et al., 2019; Wetzel et al., 2011)).

Difference waveforms (deviant-standard) were analyzed for both consonant (/ka:/ minus /ta:/) and vowel changes (/ka:/ minus /ke:/) within averaged time windows. The MMN peak amplitude (μ V) was identified locally within a time window spanning from 70 to 140 ms (Endrass et al., 2004), and averaged within 5 ms around the peak. The P3a peak amplitude (μ V) was identified within a time window spanning from 220 to 400 ms, and

averaged within 15 ms around the peak.

Behavioral analysis of the Serial recall task Analyses were conducted using scores that represented the proportion of correctly recalled items at their corresponding serial positions, based on a strict serial-recall criterion. Data were statistically analyzed using a one-way rmANOVA with the factor Sound Condition (silence, steady-state, changing-state pseudowords). Greenhouse-Geisser corrections were applied in instances in which violations of the sphericity assumption were identified, as determined by Mauchly's test. Bonferroni-corrected post hoc tests were used to reveal differences across Sound Conditions.

ERP analysis The statistical significance of MMN and P3a amplitudes and latencies were assessed through rmANOVA, with factors Background Condition (silence, traffic noise) and Phoneme Category (vowel change, consonant change). *T*-tests were used to clarify significant interactions.

Relationship between auditory interference and ERPs Simple linear regression was used to test if both amplitude and latency of each of the two ERP components (MMN and P3a) in silence and in traffic noise significantly predict behavioral performance during changing-state speech in the serial recall task.

Results

1. Behavioral Results of the Serial Recall Task Proportion correct scores are depicted in Figure 2.4. A one-way rmANOVA revealed a significant main effect of Sound Condition, $F(1.33, 26.54) = 8.930$, $p = .003$, $\eta_p^2 = .309$. Bonferroni-corrected post hoc tests revealed that performance in the changing-state condition was lower when compared to silence ($p < .001$). There was a trend suggesting better performance in silence than in the steady-state condition ($p = .060$). Additionally, the steady-state condition did not differ significantly from the changing-state condition ($p = .217$).

2. ERP amplitude and latency measures of the Oddball task ERPs are shown in Figure 2.5. Mean amplitudes, mean latencies and standard error for averaged MMN and P3a components are displayed at Table 2.2.

MMN The 2×2 rmANOVA revealed a significant main effect of Phoneme Category on both MMN latency, $F(1, 20) = 5.582$, $p = .028$, $\eta_p^2 = .218$, and MMN amplitude, $F(1, 20) = 5.037$, $p = .036$, $\eta_p^2 = .201$. No significant effects of Background Condition or interaction between factors were observed for MMN latency or amplitude, with $F(1, 20) < 1$ in each case. The main effect of Phoneme Category was attributed to the finding that vowel change ($M = -1.335 \mu V$, $SE = .138 \mu V$) exhibited a more negative amplitude compared to consonant change ($M = -.989 \mu V$, $SE = .124 \mu V$), and that vowel change peaked later ($M = 122 ms$, $SE = 2 ms$) than consonant change ($M = 114 ms$, $SE = 3 ms$). The MMN

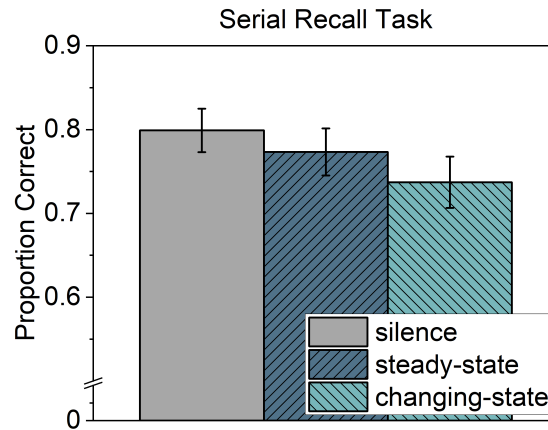


Figure 2.4: Mean proportion correct scores as a function of sound condition (silent control, steady-state and changing-state pseudowords) in a serial recall task of eight visually presented numbers. Error bars denote standard errors of the mean.

Table 2.2: ERP components (MMN, P3a) for mean (standard error) amplitude and latency for vowel change (/ka:/ minus /ke:/) and consonant change (/ka:/ minus /ta:/) in silence and in traffic noise at central electrode positions (mean of C3, Cz, C4).

Measure	Amplitude (μV)		Latency (ms)	
	Background condition			
	silence	traffic noise	silence	traffic noise
MMN vowel change	-1.38 (.15)	-1.29 (.19)	122 (3)	122 (4)
MMN consonant change	-.95 (.16)	-1.03 (.16)	115 (4)	113 (4)
P3a vowel change	.39 (.12)	.36 (.11)	330 (14)	314 (15)
P3a consonant change	.90 (.17)	.39 (.12)	297 (14)	315 (17)

amplitude and latency for both vowel and consonant change were unaffected by traffic noise.

P3a The rmANOVA revealed a significant main effect of Phoneme Category on P3a amplitude, $F(1, 20) = 6.541$, $p = .019$, $\eta_p^2 = .246$. There was also a trend for an effect of Background Condition, $F(1, 20) = 3.205$, $p = .089$, and a significant interaction between factors, $F(1, 20) = 5.940$, $p = .024$, $\eta_p^2 = .229$. Additional t -tests investigating the interaction of Phoneme Category and Background Condition revealed a significant decrease in P3a amplitude for consonant change in noise ($M = .389 \mu V$, $SE = .120 \mu V$) when compared to the silent condition ($M = .899 \mu V$, $SE = .169 \mu V$), $t(20) = 2.253$, $p = .018$, $d = .759$, but no difference for vowel change ($p = .399$). This finding implies that traffic noise had an impact only on the discrimination of consonants, while vowel change remained unaffected. The analysis of P3a latency using a 2×2 rmANOVA did not show any

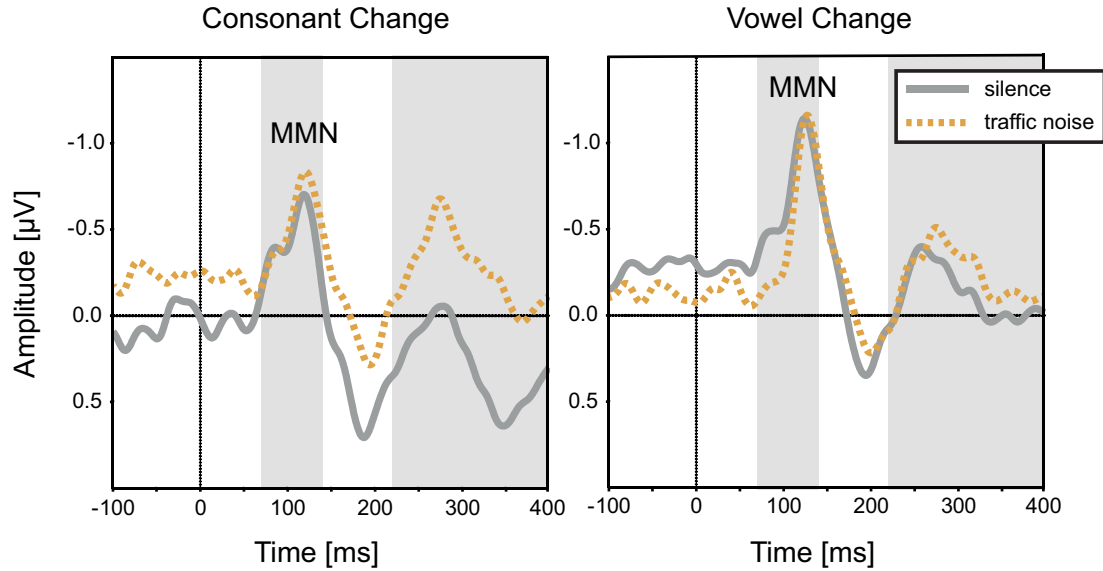


Figure 2.5: Grand averaged difference waveforms (deviant-standard stimuli) for consonant change (left panel) and vowel change (right panel) in the silent and traffic noise condition at C'. Grey boxes represent the time windows chosen for the individual peak detection. Negative is plotted upward.

significant effects, with $F(1, 20) < 1$ in each factor. Additionally, there was no significant interaction between the factors ($p = .132$).

3. Relationship between auditory interference in working memory processes and ERPs

Here, the neural indices underlying auditory discrimination (MMN) and involuntary attention capture (P3a) were investigated to determine the relationship with auditory interference, specifically the disruption caused by task-irrelevant background speech during a serial recall task. The effects are illustrated in Figure 2.6.

For MMN amplitude, a significant regression was found for vowel change in the silent condition, $F(1, 19) = 4.393$, $p = .050$. The R^2 was .19, indicating that MMN amplitude for vowel change in silence accounts for approximately 19 % of the variance in serial recall performance during task-irrelevant changing-state speech. The regression equation was: Proportion correct score during changing state speech = $.615 - .089$ (MMN amplitude vowel change in silence). That is, for each additional decrease in μV , there is an average increase of 0.089 proportion correct accuracy in serial recall during changing-state speech.

For MMN latency, the linear regression analysis revealed a statistically significant model for vowel change in the silent condition, $F(1, 19) = 4.935$, $p = .039$. The R^2 was .21, indicating that MMN latency for vowel change in silence accounts for approximately 21 % of the variance in serial recall performance during task-irrelevant changing-state speech. The regression equation was: Proportion correct score during changing state speech = $.081 + .313$ (MMN latency vowel change in silence). That is, for each additional decrease in μV , there is an average increase of 0.313 proportion correct accuracy in serial recall during changing-state speech.

2 Empirical Investigations

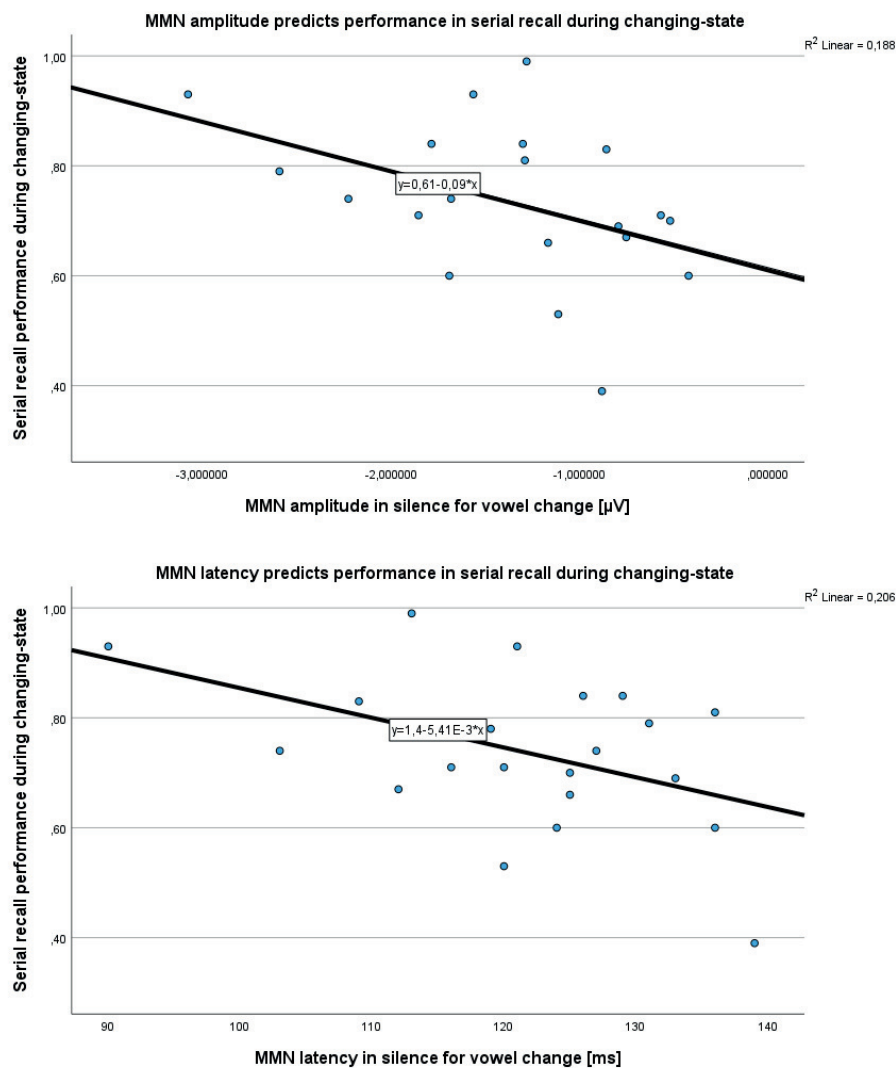


Figure 2.6: Scatter plots of the detrimental impact of changing-state pseudowords and MMN amplitude (upper panel) and MMN latency (lower panel) in silence for vowel change.

These results emphasize the importance of the MMN as a determinant of the ability to serially recall items in the presence of task-irrelevant changing-state speech. The apparent linear trend observed in the scatterplot (see Figure 2.6), combined with the significant regression coefficient, highlights the significance of incorporating MMN amplitude and latency when evaluating the impact of auditory distraction on serial recall during changing-state speech. There were no significant statistical relationships found between any other ERP measures and the serial recall performance during changing-state.

Discussion

The aim of this investigation was to examine the neural mechanisms associated with auditory discrimination (MMN) and attentional processes (P3a), as well as how these

2.7 Study 6: Auditory Discrimination and Background Speech Susceptibility

ERP components might moderate the impact of task-irrelevant speech on serial order retention. The ISE was assessed with a traditional visual-verbal serial recall task. During some trials, task-irrelevant steady-state or changing-state syllables were presented through headphones. For evaluating auditory discrimination, a standard auditory oddball paradigm was employed. This experimental setup involved probing the neural correlates of phonemic discrimination by disrupting a frequent syllable (/ka:/) through an infrequent change in consonant (/ta:/) or vowel sounds (/ke:/). This passive task was conducted both in silence and in the presence of traffic noise to enhance ecological validity.

Before presenting the results of the correlation analysis, the findings on the ISE and the ERP correlates, as these served as a prerequisite for the correlation analysis, are presented. 1) In terms of serial order retention, performance decreased in the changing-state condition compared to silent control condition, indicating the ISE. However, performance in the steady-state condition did not significantly differ from either the silent or changing-state conditions. 2) Regarding the ERP components, the results indicate that the neural mechanisms underlying auditory discrimination are influenced differently by the background condition, depending on the specific ERP component and the measured response, such as amplitude or latency. Notably, the MMN amplitude was more negative for vowel changes than for consonant changes, and traffic noise had no effect on MMN amplitude for either type of change. Regarding the P3a amplitude, the background condition differentially influenced phoneme categories. Traffic noise did not affect the P3a amplitude for vowel changes but did impact the amplitude for consonant changes. In terms of ERP latency, the peak of the MMN for vowel changes occurred later than for consonant changes, with traffic noise having no impact on MMN latency for either change type. For P3a latency, there was no distinction between vowel and consonant change and it was unaffected by traffic noise.

3) Next, the findings of the correlation analysis will be presented. The primary focus is on the potential moderating role of neural correlates associated with pre-attentive auditory change detection (MMN) and involuntary attention capture (P3a) of speech stimuli in moderating the adverse impacts on serial order retention. In the silent condition, both the amplitude and latency of the MMN for vowel change are associated with heightened interference from changing-state sounds during the serial recall task. Conversely, P3a amplitude and latency show no correlation with susceptibility to background sounds during serial recall performance.

The next section will focus on the behavioral findings related to the ISE and their corresponding ERP correlates. A discussion of the correlation analysis will follow subsequently.

The disruptive effect of changing-state speech on the ISE is confirmed, replicating earlier findings (Jelelati et al., 2023; Jones et al., 1992). Performance dropped approximately 6 % in the changing-state condition compared to silence, which is within the typical range of 5 % to 50 % (Gisselgård et al., 2004; Neath, 2000). While Study 1 showed lower performance in the changing-state condition compared to steady-state, the present study found no significant difference between these two conditions, contradicting the changing-state hypothesis, which suggests greater disturbance from changing-state sequences (Beaman

2 Empirical Investigations

& Jones, 1997; Jones et al., 1992). This discrepancy may arise from methodological differences. Study 1 used a repeating syllable “ba” for steady-state sequences, while this study varied the syllable across trials, potentially causing reorientation. Both studies, however, indicated that the steady-state condition did not differ from the silent control condition, suggesting the changing-state nature of background speech is crucial for eliciting the ISE. Further research is necessary to understand why different steady-state syllables are not distinguishable from changing-state sounds.

The MMN paradigm highlights a notable gap in the literature, which has predominantly examined a single deviant, limiting comparisons between vowel and consonant changes. This study addresses that gap by investigating two deviants, enhancing the potential for comparative analysis. Previous research has primarily focused on discrimination of vowel sound changes (Kozou et al., 2005) or consonant changes alongside unfamiliar sounds (Kujala et al., 2007).

Regarding the findings on the ERP component: The results align partially with prior studies, showing that MMN amplitude is generally more negative for vowel than for consonant changes (Koerner et al., 2016). However, traffic noise had no effect on MMN amplitude in this study. Several aspects of the MMN response findings presented here diverge from those reported by Koerner et al. (2016). First, Koerner et al. (2016) examined the neural indices of syllables /bu/ (vowel change) and /da/ (consonant change) compared to a standard /ba/, both in the presence of speech babble and in silence. They found that speech babble resulted in longer latencies and reduced amplitudes for the iMMN. Notably, the peak of the iMMN for /da/ (consonant change) occurred later than for /bu/ (vowel change), and the amplitude of the iMMN was more significantly reduced for /bu/ in noisy conditions compared to /da/. Second, they focused on the “identity MMN” (iMMN). This methodological approach involves subtracting a deviant stimulus (e.g., /da/) from the same stimulus presented as a standard (/da/-/da/-/da/-/da/) in a separate control block of stimulus presentation (e.g., Koerner et al., 2016). Finally, Koerner et al. (2016) utilized speech babble as the background sound, while this study employed traffic noise. Research indicates that the type of background sound can affect the MMN response of speech processing. For instance, Kozou et al. (2005) found that both babble and traffic noise reduced MMN amplitude compared to silence. Notably, babble noise resulted in an earlier peak, whereas the latency of the MMN in traffic noise did not significantly differ from silence (Kozou et al., 2005). This suggests that the traffic noise in the current study likely did not influence MMN latency, which may explain the differing outcomes from those reported by Koerner et al. (2016).

Aside from methodological variations, it is worth considering the structure of the consonant-vowel syllables that were used (e.g., ka, ke, ta), in which the vowel follows the consonant, and the sequential nature of their processing. Based on this, it is plausible to infer that the peak of vowel change occurs later than the peak of consonant change.

The finding that the MMN is unaffected by traffic noise indicates that the brain’s automatic change detection systems in adults remain resilient to this type of auditory

2.7 Study 6: Auditory Discrimination and Background Speech Susceptibility

distraction. This suggests that, even in the presence of traffic noise, the brain can effectively process and differentiate between speech stimuli. Differential amplitudes for vowel and consonant changes indicate different neural processing of these two types of phonemic categories. The greater negativity for vowel change may indicate that the brain allocates more resources or exhibits enhanced sensitivity towards detecting deviations in vowel sounds compared to consonant sounds. Furthermore, this could reflect the linguistic importance of vowels in speech perception and communication, leading to enhanced neural processing for detecting variations in vowel sounds.

Our finding of a reduced P3a amplitude for consonant change is consistent with the results reported by Koerner et al. (2017). However, they found no significant interaction between the background condition and phoneme category for both P3 latency and amplitude. In contrast, our study provides evidence that traffic noise specifically impact the capacity to distinguish consonant sounds, while vowel change remained unaffected by noise, as indicated by the P3a amplitude. The impact of traffic noise on the P3a response for consonant change indicates that the orienting response to novel or unexpected stimuli is selectively affected by the acoustic environment. This finding suggests that the attentional mechanisms involved in detecting changes in consonant sounds are more susceptible to interference from noise than those associated with vowel sounds. Additionally, the varying effects of traffic noise on the P3a response for consonant versus vowel changes imply distinct neural processing mechanisms for these two phonemic categories. It may be that consonant discrimination relies more heavily on attentional resources, rendering it more vulnerable to distraction from background noise compared to vowel discrimination.

Returning to the central question of whether the neural correlates associated with detecting auditory changes (MMN) and attentional mechanisms (P3a) have moderating effects on the susceptibility to task-irrelevant sounds in relation to serial order retention of visual-verbal items. A brief summary of the results is as follows: Correlational analyses revealed no significant relationship between P3a measures and task performance. In contrast, both the amplitude and latency of the MMN for vowel changes measured in silence correlated with increased susceptibility to task-irrelevant sounds in the serial recall task. Specifically, lower performance during changing-state speech in the irrelevant sound paradigm was associated with reduced MMN amplitude and delayed peak latency. These findings suggest that certain neural responses related to phoneme discrimination, particularly those involving vowel changes, significantly influence susceptibility to background speech disturbances during visual-verbal serial recall, potentially manifesting as delays in neural latency as well as decreased activation. The automatic processing of background speech, i.e. performance decrements triggered by changing-state syllables, may result in 1) reduced MMN activation, reflected by a more negative deflection for vowel change and 2) delayed processing, as reflected by prolonged latencies, ultimately impairing working memory performance.

The question arises as to why changes in vowels, but not consonants, correlate with susceptibility in serial recall performance during task-irrelevant speech. Research indicates distinct functional roles for consonants and vowels in language processing (Fogerty &

Humes, 2012). For instance, vowels significantly influence sentence intelligibility (Fogerty & Kewley-Port, 2009) and word recognition (Cole et al., n.d.), more so than consonants (Cole et al., n.d.; Fogerty & Kewley-Port, 2009). This suggests that the acoustic properties of consonants and vowels may serve different roles in auditory processing (Fogerty & Humes, 2012).

In the context of the ISE paradigm, Hughes, Tremblay, and Jones (2005) propose that vowels, rather than consonants, are the dominant cause of the detrimental sound-induced impact (see Jelelati et al. (2023) for conflicting results). The significant correlation of sound-induced susceptibility and the MMN elicited by vowel changes underscores a potential discrepancy in the impact of vowels versus consonants on serial recall tasks during irrelevant speech, emphasizing the necessity of acknowledging their distinct acoustic properties.

The reason for the correlation between sound-induced susceptibility in serial recall and MMN for vowel changes in the silent condition, but not in the traffic noise condition, remains unclear. Notably, descriptive data show a decrease in MMN amplitude during vowel changes in the presence of traffic noise compared to the quiet condition (see Table 2.2). Thus, it can be inferred that the non-significant finding in the correlation analysis may result from the lower MMN amplitude observed in the traffic noise condition.

Hitherto, there has only been one study (Campbell et al., 2007) that examined the relationship between the ISE and ERP components, specifically the MMN and the N1. However, it is important to note that our findings can only be partially compared to this study due to methodological variations. Campbell et al. (2007) modified the way the task-irrelevant background noise was presented in order to measure either the N1 or the MMN during a serial recall task. In their investigation of the MMN, oddball trials were used, featuring a frequent tone that was occasionally interrupted by an infrequent stimulus during the retention interval. This approach is similar to the mechanisms used to examine the deviant effect and attentional capture (Hughes, 2014). In their investigation of the N1, they presented letters that varied continuously. This method, which introduces variation within a single auditory stream, resembles the examination of the changing-state effect. However, because the MMN relies on a memory trace, this approach may not adequately capture the traditional MMN response typically observed in an oddball paradigm. Thus, Campbell's investigation (Campbell et al., 2007) presumably focused more on the deviant effect rather than the standard MMN response, limiting the comparability to the conventional MMN as investigated in our study.

This research has major implications for the theoretical explanations. The findings reported here suggests that serial recall performance in the presence of changing-state sounds is significantly related to the underlying mechanisms of pre-attentive auditory change-detection of speech discrimination, as reflected by the MMN component. However, the processes of involuntary attention capture seem to be less relevant in this specific context.

The findings support a dissociation between specific interference and involuntary attention capture, indicating that theories attributing the ISE solely to attentional mechanisms

2.7 Study 6: Auditory Discrimination and Background Speech Susceptibility

are insufficient. The MMN signifies the brain's automatic, pre-attentive detection and processing of auditory deviations, elicited when a rare or unexpected stimulus deviates from regular auditory patterns. The significant correlation between MMN amplitude and sound-induced disruption in serial recall underscores the role of specific interference, highlighting that such interference occurs independently of conscious attentional shifts. This finding aligns with both the speech-specific interference proposed by the phonological loop account and the changing-state hypothesis, which suggests that cognitive performance suffers from auditory stimuli with frequent changes or variability. Together, these results offer empirical support for both accounts, highlighting that the brain detects and processes deviations in auditory patterns even without conscious attention.

However, it is essential to note that the MMN, as assessed in the passive oddball paradigm, shares similarities with the measurement of the deviant effect within the context of the ISE. According to the Duplex Mechanism (Hughes, 2014), the deviant effect and the changing-state effect belong to distinct forms of auditory distraction. Further research, such as employing a multi-feature paradigm (Pakarinen et al., 2009), could help distinguish between the deviant effect and the changing-state effect and allow for the investigation of other stimuli, like tones, to clarify the nature of speech-specific interference.

It is crucial to acknowledge the limitations of our study, particularly the comparatively small sample size consisting of only 21 participants. The limited sample size diminishes the statistical power of the correlation analyses, potentially compromising our ability to identify genuine effects or associations within the data. Hence, it is advisable to be cautious when interpreting the link between MMN and the susceptibility to changing-state sounds.

To summarize, our study provides insight into the influence of traffic noise on phoneme discrimination. Furthermore, this is the first study to date presenting evidence for a link between the neural indices associated with phoneme discrimination and the vulnerability to task-irrelevant sounds in relation to working memory performance. Further research is needed to improve our understanding of how various types of sounds, including speech babble and environmental noises such as traffic noise, specifically impact speech discrimination. In addition, it would be beneficial to conduct a more thorough examination of the correlation between the MMN and the vulnerability to task-irrelevant auditory stimuli. Considering that the MMN is elicited by various features of a stimulus, it is important to investigate whether the relationship between the MMN and the ISE is speech-specific or if it also applies to other stimuli, such as tones. Moreover, our findings indicate a correlation between the magnitude of the MMN and the susceptibility to task-irrelevant sounds on serial recall performance. It remains open whether the magnitude of the MMN explains this vulnerability.

3 General Discussion

Despite its significant practical implications and nearly half a century of research, the theoretical concepts underlying the ISE remain contentious. While certain background sounds are known to disrupt serial order retention, the extent of this disruption varies based on the characteristics of the noise and the demands of the task, ultimately influencing how effectively individuals process information and retain sequences in short-term memory.

This thesis investigates the ISE, which describes the disruptive impact of specific task-irrelevant background sounds on the serial recall of visually presented verbal material. Since its initial report almost 50 years ago, various theories have attempted to explain the mechanisms underlying the ISE, yet several key questions remain unresolved. Currently, performance impairments can only be predicted for select background sounds and task characteristics, and the cognitive mechanisms involved are only partially understood.

Research identifies distinct cognitive mechanisms contributing to the ISE. The first involves sound-induced attentional distraction, in which background sounds involuntarily shift focus away from the primary task. The second involves specific interference, either by process or content, in which automatic sound processing disrupts intentional, task-related cognitive functions. These mechanisms interact differently depending on the nature of the background sound and the cognitive demands of the task, contributing to the variability in observed performance impairments.

To explore the roles of attention capture, process- and content-specific interference in the ISE, as well as the underlying neural correlates, I employed various research approaches. These approaches across six studies are summarized as follows.

Study 1 In Study 1, the type of interference effect using task-irrelevant steady-state and changing-state speech in comparison to a silent control condition were investigated on the serial order reconstruction of verbal and spatial items in children and adults. The following predictions stem from the theoretical background: According to the attention-capture account, the ISE should occur in both the verbal and spatial task, be more pronounced in children due to their less developed attention control, and gradually diminish with exposure to changing-state speech as habituation occurs. In contrast, the changing-state account predicts that the ISE will appear in both verbal and spatial tasks for both age groups, with the disruptive effect of changing-state speech remaining constant throughout the tasks. Finally, the phonological loop account suggests that the ISE should only affect the verbal task, leaving the spatial task unaffected in both children and adults and there should be no habituation. The results of Study 1 showed that task-irrelevant changing-state speech had a negative impact on the retention of verbal serial order, but not on spatial serial order. The ISE was more pronounced in children. The steady-state condition had a negative impact

3 General Discussion

on children, but it did not affect adults' serial recall performance. The changing-state effect, specifically the comparison of performance between steady-state and changing-state conditions, was comparable across age groups. There was no evidence of habituation. The findings from Study 1 support the phonological loop account while challenging both the changing-state and the attention-capture account.

Study 2 Study 2 is methodologically identical to Study 1, with the exception of the background sounds. In Study 2, the dissociation of interference from distraction effects were investigated by using task-irrelevant narrative speech, unfamiliar speech and environmental non-speech sounds in comparison to a silent control condition on the serial order reconstruction of verbal and spatial items in children and adults. From the theoretical background, these predictions arise: We predict that if attention capture plays a key role in the evocation of the ISE, both speech and environmental sounds will cause significant disruption in both tasks, with a greater effect in children than adults, and there should be habituation over time. From an interference-based perspective (Phonological Loop Account), speech should impair performance on the verbal task equally in both groups, with no effect on the spatial task. Environmental sounds are expected to cause a minimal detrimental impact in the verbal task due to the lack of phonological content. There should be no habituation. According to the changing-state account, speech should impair both tasks across ages, while environmental sounds, being a mix of unrelated noises, should cause little disruption, as they do not form a coherent auditory stream. Disruption should remain stable throughout the tasks. The results of Study 2 demonstrated distinct effects of environmental non-speech sounds and unfamiliar narrative speech on verbal and spatial serial recall. In the verbal task, children and adults were equally impacted by the sound conditions, with speech being more disruptive than environmental non-speech sounds. In the spatial task, children were significantly impaired by both sound conditions, whereas adults were not affected. A further analysis, distinguishing between age groups, revealed that only the third-grade children were affected, while the ISE was absent in the fourth-graders. Both sound conditions had a similar disruptive effect on the younger children. There was no evidence of habituation during the tasks. The findings from Study 2 suggest that both specific interference and attention capture may contribute to the ISE, with the influence of attention capture potentially diminishing with age.

Study 3 Study 3 is identical to the steady-state and changing-state syllables from Study 1, with additional sinewave transformation applied to the speech syllables. Thus, Study 3 examined the effects of phonology-based and seriation-based interference compared to a silent control condition on the serial order reconstruction of verbal and spatial items in adults. The theory suggests the following predictions: From the perspective of the changing-state account, both unaltered changing-state syllables and their sinewave-transformed counterparts should impair serial order retention for both verbal and spatial items. However, if phonology-based interference plays a role, the disruptive effects should be observed for

both steady-state and changing-state sequences, with natural speech being more disruptive than sinewave speech. Furthermore, this effect should only manifest in the verbal task, with no impairment expected for spatial items. For the verbal task, the results of Study 3 showed that unaltered speech was more disruptive to serial recall than the transformed sinewave counterparts. Additionally, the changing-state syllables were more disruptive than the steady-state syllables. The effects of phonology and changing-state were independent of each other. Thus, both phonological factors and the changing-state effect individually contributed to decreased performance in visual-verbal serial recall in adults. Unexpectedly, changing-state sinewave speech impaired serial recall for spatial items. The findings from Study 3 support phonology-based interference processes, which appear to exclusively disrupt visual-verbal serial recall performance.

Study 4 In Study 4, the influence of including spatial information within the background sound was examined. Thus, Study 4 focused on the impact of a realistic classroom-like noise scenario, presented both monaurally and binaurally, on verbal serial recall in children and adults. The following predictions are derived: We expect binaural sound presentation to have a stronger impact than monaural presentation for two main reasons: (1) the clearer separation of speech streams, which leads to increased speech-based interference, and (2) the enhanced attention capture from spatially distributed, shifting sound sources. Furthermore, if attention capture is the main source of disruption, we predict that children will experience greater impairment than adults, particularly under binaural compared to monaural conditions. Furthermore, if attention capture is the primary factor behind disruption, we expect habituation to occur over time. The results of Study 4 showed that the performance was significantly poorer under both sound conditions. The detrimental impact was comparable across both sound conditions. As in Study 2, children and adults were equally impacted by both sound conditions. There was no indication of habituation. The findings from Study 4 support speech-based interference processes over attention capture.

Study 5 In addition to behavioral performance, in Study 5 investigates the neural mechanisms underlying the ISE, specifically aiming to disentangle the roles of ERP components related to early sensory processing, and attentional mechanisms. From related EEG studies on the ISE, the following predictions arise: Behaviorally, the classical ISE should be found, with lower recall performance in the unfamiliar speech condition compared to the pink noise condition. Regarding ERP components, a heightened N1 amplitude in the unfamiliar speech condition compared to the pink noise condition is assumed. Additionally, if attention capture plays a central role in the negative impact of task-irrelevant background sounds, discernible difference in the P3a is to be expected, reflected in a more positive amplitude in the speech condition. Behaviorally, Study 5 successfully replicated the typical ISE, reaffirming previous research demonstrating the detrimental impact of irrelevant speech on serial order retention of verbal items presented visually

(Colle & Welsh, 1976). Regarding the neural mechanisms, the impact of task-irrelevant unfamiliar background speech on serial recall performance was examined. The findings showed an increase in N1 amplitude elicited by the visually presented items in response to speech compared to pink noise. This increase is potentially linked to heightened attention in early sensory processing when speech is present. This implies that additional brain resources are allocated to filter out task-irrelevant auditory stimuli, potentially hindering visual-verbal information processing. While the P2 is possibly influenced by task demands and stimulus discrimination, the functional role of the noise-induced impact remains open. The P3a showed only a trend-level impact from task-irrelevant speech, suggesting that the neural re-orientation processing of the to-be-remembered items may be comparable across background conditions, though this result remains somewhat inconclusive. The finding from Study 5 that task-irrelevant speech modulates visual N1 and P2 but not P3a indicates that task-irrelevant speech influences early sensory processing and attentional modulation but not later cognitive processing such as attentional re-orienting.

Study 6 Study 6 employed a classic ISE paradigm, methodologically comparable to Study 1, along with EEG measurement during a passive oddball paradigm. This approach aimed to relate the impact of task-irrelevant speech during serial recall of visual-verbal items to ERP components associated with auditory discrimination (MMN) of speech items and attentional processes (P3a). The following predictions arise: Behaviorally, the classical ISE should be found, with lower recall performance in the changing-state speech condition compared to the silent control condition, along with a CSE, reflected by lower recall performance under changing-state when compared to steady-state. Regarding ERP components, a heightened N1 amplitude in the unfamiliar speech condition compared to the pink noise condition is assumed. Furthermore, if attention capture is central to the negative impact of task-irrelevant background sounds, we expect to see a noticeable difference in the P3a, indicated by a greater positive amplitude in the speech condition. Behaviorally, Study 6 successfully replicated the typical ISE, reaffirming Study 1 and previous research demonstrating the detrimental impact of changing-state syllables on serial order retention of verbal items presented visually (Marsh, Hurlstone, et al., 2024). However, the changing-state effect, namely the difference between steady-state and changing-state syllables, was not found, possibly due to methodological differences across steady-state trials. Regarding the ERP components elicited during the passive oddball task, we observed specific patterns linked to auditory discrimination and attentional processes. The MMN was elicited reliably, with a higher amplitude and later peak of vowel change compared to consonant change. Traffic noise did not interact with phoneme discrimination. In contrast, traffic noise interacted with P3a, but only affected consonant change. The correlation analysis between susceptibility to changing-state syllables and MMN and P3a revealed that, specifically for vowel changes in silence, better performance in the changing-state speech condition was associated with a higher MMN amplitude and a delayed peak. However, this is one of the first attempts to establish a direct association between neural and behavioral correlates.

More research is needed to refine and replicate the present findings. Based on the present findings from Study 6, the susceptibility to background speech may vary based on neural markers of auditory speech discrimination (MMN), but not of attentional mechanisms (P3a), aligning with specific interference accounts underlying the ISE.

All studies in this thesis consistently confirm the disruptive effect of task-irrelevant unfamiliar speech and non-speech sounds, specifically sinewave-transformed syllables and environmental sounds, on the serial order retention of verbal items, replicating earlier findings (Salamé & Baddeley, 1986, 1987). The disruptive effect on verbal task performance, depending on age group and type of task-irrelevant sound, ranges from approximately 7 % to 17 %, which falls within the typical range (Gisselgård et al., 2004; Neath, 2000). The results of the studies have been outlined above. Below, I will present the main findings related to the central questions of the thesis.

A key question in this thesis is to differentiate interference from distraction effects within the ISE paradigm. None of the existing theoretical accounts of the ISE fully explain the results presented here. Evidence for specific interference effects comes from comparing verbal and spatial material in a serial recall task in adults.

Memory material Study 1, Study 2, and Study 3 showed differential effects of background speech on serial order retention of visually presented verbal and spatial items. While the ISE (and CSE in Studies 1 and 3) was replicated as expected in the verbal task, recall performance of spatial items remained largely unaffected by background speech. The exceptions to this null effect in the spatial tasks were the negative impact of sinewave changing-state speech on spatial serial recall in Study 3, and the detrimental effects of unfamiliar speech and environmental sounds on third-graders, but not fourth-graders, in Study 2.

The immunity of spatial serial recall to disruption by sounds poses a challenge to both the changing-state and attentional accounts. The changing-state account posits that such disruption arises from pre-attentive processing of order in changing sounds, impacting all serial tasks regardless of modality. Jones et al. (1995) (Exp. 4) demonstrated that speech disrupted not only the recall of verbal material but also that of spatial material when items were remembered in their serial order, suggesting a “functional equivalence” between verbal and non-verbal material and attributing the ISE to a domain-general rehearsal mechanism. However, our studies found that background speech impaired verbal recall but not spatial recall, contradicting this theory. In line with our findings, prior studies (Bergström et al., 2012; Klatte & Hellbrück, 1997; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024) have also failed to replicate the original finding from Jones et al. (1995). This indicates that if an ISE for spatial serial order retention exists, it is far less robust than for verbal items.

Moreover, the results are particularly challenging to explain with the attention-capture account, which predicts a stronger ISE for spatial than for verbal items, as visuo-spatial

3 General Discussion

working memory tasks rely more than verbal tasks on domain-general attentional resources in both children (Alloway et al., 2006; Campos et al., 2013) and adults (Morey & Miron, 2016). This prediction is based on the idea that attentional refreshing, through which items are kept active in working memory through focused attention, would be more easily disrupted in spatial tasks, leading to greater impairment from irrelevant sounds. However, the current findings do not support this perspective, as background speech significantly impaired verbal serial recall while having no effect on spatial recall.

Yet, our results, along with prior studies, consistently show that spatial serial recall is largely immune to interference from background sounds, suggesting that the ISE either operates differently in spatial tasks or is absent altogether. This stands in stark contrast to the clear and robust effects observed in verbal tasks, in which irrelevant sounds consistently disrupt performance. This discrepancy highlights a key difference in how attentional capture and specific interference may explain the ISE. While verbal tasks appear highly sensitive to auditory distractions, spatial tasks may rely less on the same attentional mechanisms, or may be less susceptible to sound-induced disruption.

Instead, the absence of sound-induced disruption of spatial serial recall aligns with the phonological loop account, which attributes interference to the verbal short-term store that processes phonological codes. According to this account, irrelevant sounds disrupt the rehearsal of verbal information within the phonological loop, leading to impaired serial recall for verbal tasks. Since spatial tasks rely more on visuo-spatial working memory rather than phonological processing, they are less susceptible to interference from auditory stimuli. This suggests that the disruption caused by irrelevant sounds specifically targets phonological representations, rather than affecting general attentional mechanisms across all types of working memory tasks.

Developmental Change Investigating children’s serial task performance in noise is important for understanding the development of attention control and short-term memory. Serial recall tasks engage the phonological loop, a core component of working memory responsible for maintaining the order of verbal information. Studying how children perform these tasks in noisy environments helps to reveal developmental changes in their ability to filter out irrelevant auditory information, which is critical for both attention regulation and memory retention.

Studying the ISE in children is also essential for understanding how noise affects their cognitive development, particularly in the context of language acquisition (Klatte, Lachmann, et al., 2010). Since verbal short-term memory plays a pivotal role in language and reading development, noise-induced disruptions could result in lasting deficits. Field studies suggest that classroom noise can impact language and reading acquisition, underscoring the need for experimental investigations into the ISE in children to help create optimal acoustic learning environments. Additionally, research on noise disruption in serial tasks contributes to theories of the ISE, offering insights into how noise may interfere with cognitive processes differently in children compared to adults.

Currently, there are conflicting findings in the literature regarding age-related changes in the ISE. The debate about whether children are more susceptible to performance decline due to auditory interference remains ongoing. Several studies indicate age-related differences in the ISE resulting from developmental change (Elliott, 2002; Elliott & Briganti, 2012; Elliott et al., 2016, Exp. 1). In contrast, other studies did not reveal any disparities between children and adults (Klatte, Lachmann, et al., 2010; Meinhardt-Injac et al., 2022; Röer et al., 2018; Schwarz et al., 2015; Elliott et al., 2016, Exp. 2).

This thesis examined the susceptibility to background sounds in children and adults in Studies 1, 2, and 4, providing further evidence that most children are equally affected by background sounds as adults. However, Study 1 and Study 2 showed different effects on children and adults under the influence of background noise retaining verbal and spatial material in serial order.

The findings from Study 1 showed a stronger ISE in children compared to adults, with only children being impaired by steady-state speech. When the silent control condition was excluded, the effects of the changing-state and steady-state conditions were comparable across age groups. In Study 2, children's serial recall performance of spatial items was affected by irrelevant speech and environmental non-speech sounds. At first glance, this seems contradictory to Study 1, which showed no sound-induced impairments in children's spatial task performance. However, considering the age range of the children provides clarity. In Study 2, additional analyses revealed that only the younger children's serial recall performance was impeded by the background sounds, while the older children were unaffected. The findings of Study 4 revealed that both adults and children are equally affected by classroom noise when performing serial recall of visual-verbal items. This aligns with evidence from Meinhardt-Injac et al. (2022), which shows that children around 9 years old and older experience similar short-term memory impairments as adults when exposed to classroom sounds.

Given that children's attention control is less developed and therefore children should generally be more vulnerable to attentional diversions, the absence of an ISE for spatial performance in Study 1 and in the older children in Study 2 further complicates the explanation. If attentional mechanisms were responsible for the ISE, we would expect a more pronounced disruption in children performing spatial tasks. The stronger ISE in children compared to adults in the verbal task in Study 1 cannot be attributed to attentional distraction, given the null effect in the spatial task. Instead, this suggests stronger interference due to less robust phonological processing in children. However, Study 2 showed a detrimental impact on children's spatial task performance, which is best explained by attention mechanisms.

Consistent with this, Klatte, Hellbrück, et al. (2010) found that non-speech classroom noise significantly impaired verbal serial recall in 7-year-old children, whereas older children were unaffected. The findings here also align with those of (Elliott et al., 2016), who reported that both children and adults were equally impaired by irrelevant speech in a visual-verbal serial recall task. However, in a verbal task typically resistant to sound-induced impairments

3 General Discussion

in adults (i.e., the missing item task), only children showed disruption. The authors suggest that while children and adults are similarly affected by sound-induced interference in visual-verbal tasks, children's heightened vulnerability to attentional distractions increases their susceptibility to sound-induced impairments, regardless of task or sound type. These findings, consistent with ours, further support the evidence for specific interference effects in adults while highlighting children's greater sensitivity to attentional disruptions.

Thus, the findings here are best explained by considering both specific interference in visual-verbal serial recall and more general effects on attention, as outlined in the "duplex mechanism account" of auditory distraction (Hughes, 2014; Hughes et al., 2007). According to this model, specific interference and attention capture are independent mechanisms of sound-induced performance deficits that may operate alone or together, depending on task demands, the nature of the sounds, and the characteristics of the individuals involved. In children, attentional diversions play a larger role in overall disruption since they are less adept at controlling attention in the presence of task-irrelevant sounds compared to adults. However, the wide variability in attention abilities across children during the primary school years may explain the inconsistencies found in studies on developmental changes in the ISE.

Habituation Habituation refers to the process by which repeated exposure to background sound leads to a decrease in the ISE (Bell et al., 2012). The question of whether individuals can habituate to background sounds is of practical relevance for developing interventions to mitigate auditory distractions and is fundamental for evaluating theoretical explanations of the ISE.

Research in this area has produced mixed results. Some researchers suggest that it is possible to habituate to the deviation effect but not to the changing-state effect (Sörqvist, 2010). Others challenge the notion that any difference in habituation rates when directly comparing these effects (Röer, Bell, Marsh, & Buchner, 2015). Furthermore, some argue that investigating habituation is challenging due to the methodological use of alternating sounds on a trial-by-trial basis. Instead, they propose studying habituation under conditions, in which background sounds are continuously presented for multiple trials to ensure consistent exposure. Within the three studies in this thesis that examined the effects of habituation, both methodological approaches were used: in Studies 1 and 2 background sounds were alternated on a trial-by-trial basis, while Study 4 investigated habituation with continuously presented background sounds. There was no habituation observed in any of these studies, neither in adults nor in children. The absence of habituation effects across all studies indicates a fundamental independence from sound presentation (trial-by-trial basis vs. blockwise), emphasizing the robustness of the ISE.

Furthermore, the finding that neither Study 1, 2 nor 4 showed any signs of habituation challenges the assumption of the attention capture account. According to this account, repeated exposure to the task-irrelevant background sounds should lead to habituation, resulting in a decrease in the detrimental impact of these sounds over time. This theory

suggests that as the background sounds become more familiar, they should progressively lose their ability to draw attention. Instead, the persistence of disruptive effects suggests that attentional mechanisms alone may not fully account for the observed performance decrements. These findings strongly suggest that interference, rather than solely attentional processes, is a crucial factor contributing to the decline in cognitive performance caused by irrelevant background sounds.

Recent research suggests that training in auditory selective attention can reduce the distraction caused by changing-state sounds. For example, Kattner and Ellermeier (2020) demonstrated that participants who underwent dichotic-listening training showed a substantial improvement (about one third) in serial recall performance for both visually and aurally presented verbal items compared to their pretest levels. This improvement was linked to the participants' ability to suppress the typical orienting response triggered by auditory distractors, which normally diverts attention from the main task. By minimizing this orienting response, the focus of attention remained more resistant to interference from irrelevant speech. However, further research is needed to explore the long-term effects of such training across multiple sessions.

The inability to habituate to background sounds may lead to ongoing cognitive interference and diminished task performance, especially in noisy settings such as classrooms or open offices. This underscores the necessity for an optimized acoustic environment.

Acoustic Properties The investigation into the sound characteristics and presentation of task-irrelevant background sounds enhances our understanding of the roles of attention capture, serial order retention, and phonological processing in the ISE. Here, I focused on two specific aspects: a) The effects of sine wave modification of natural speech syllables on serial order retention, allowing to distinguish between phonological and changing-state characteristics (see Study 3), and b) The distinction between monaural and binaural presentation of the background sounds, comparing the classical laboratory monaural presentation with a more realistic binaural presentation by considering spatial information (see Study 4).

For the verbal task, the results of Study 3 replicated the detrimental impact of changing-state speech on serial order retention. Furthermore, the findings indicate that phonological factors play a significant role. Unaltered speech was found to be more disruptive to serial recall than the transformed sinewave counterparts. Additionally, there was a clear effect of changing-state, meaning that the changing-state syllables were more disruptive than the steady-state syllables. These two effects were independent of each other. Thus, both phonological factors and the changing-state effect individually contributed to decreased performance in visual-verbal serial recall in adults.

Unexpectedly, the sinewave-transformed changing-state sounds significantly reduced serial recall performance for spatial items compared to the silent control condition in adults. This finding contradicts the results reported in Studies 1 and 2, as well as other research (Bergström et al., 2012; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024), which found no

3 General Discussion

detrimental impact of changing-state sounds on the serial order retention of spatial items in adults. However, a closer examination of the main effects of phonology and state reveals distinct impacts from phonological content and the changing-state effect. Specifically, changing-state sequences proved to be more disruptive than steady-state sequences, and natural speech with changing-state characteristics was more detrimental than unaltered speech. Consequently, the findings regarding disruption in the spatial task become less conclusive, as only the sinewave-transformed, but not the natural changing-state speech condition visuo-spatial serial recall performance compared to silence. While these results align more closely with the O-OER model, which suggests that irrelevant changing-state sounds equally impact serial recall regardless of modality (e.g., verbal or spatial material) due to the functional equivalence of verbal and non-verbal material (Jones et al., 1995), further research is needed to fully explore this finding.

This concept of increased disruption from spatial integration within background sounds is addressed in Study 4 (Exp. 2), which investigates whether varying spatial information adds to the significant impact on serial recall. In Study 4 Exp. 2, the impact of presentation mode (monaural vs. binaural) using a realistic classroom-noise scenario on serial recall of verbal items was investigated in children and adults. Contrary to our expectations, the presentation format (monaural vs. binaural) of the classroom noise did not affect disruption of visual-verbal serial recall. We expected that the binaural condition would cause greater impairments in children due to increased attention capture by spatially separated and changing sound sources. The evidence suggests that the interference between obligatory processing of speech parts and deliberate processes involved in serial recall performance, specifically the maintenance of phonological representations, is the cause of the ISE evoked by sounds containing speech. This conclusion is supported by the fact that the noise effects are equivalent across different age groups and presentation formats, indicating that attention capture has a minimal impact on the ISE.

The findings from Study 4 challenge both the attention capture and interference accounts, as stronger disruptive effects were expected with binaural compared to monaural presentation of the classroom-noise scenario. Based on the attention capture theory, one would predict a more pronounced disruptive effect with binaural presentation compared to monaural presentation of classroom noise, especially in children. This is because the spatial distribution and changing locations of the sound sources in a binaural scenario are thought to be more attention-grabbing. Contrary to expectations, the disruptive effects of the classroom-noise scenario were comparable in both adults and children, regardless of whether the presentation was binaural or monaural. According to the interference account, binaural sound scenarios with speech should evoke stronger disruption because the spatial separation of sound sources encourages the segregation of spoken parts into distinct, single-talker streams. However, this mechanism did not seem to operate in the current study. This might be because the classroom-noise scenario used involved only two simultaneous talkers. Jones and Macken (1995) demonstrated that monaural presentations of single-talker speech and a mixture of two speakers caused similar disruption in serial recall performance, while

disruption was significantly reduced with six simultaneous speakers. Therefore, in the current classroom-noise scenario, further separation of the two speech streams through spatial cues might not increase interference, as it is already at its maximum (similar to that evoked by a single talker). Based on this, stronger interference effects with binaural presentation would likely occur in noise scenarios with more than two simultaneous talkers. Instead, the observed interference effects are assumed to be caused by phonologically-based interference between the speech-containing sound scenarios and the availability of the phonological item representations.

Neural Correlates of the ISE Numerous studies consistently indicate that the presence of background noise significantly impairs cognitive task performance, particularly in tasks involving serial order recall of verbal items (Colle & Welsh, 1976). Nevertheless, there has been a lack of research investigating the neural mechanisms that underlie this phenomenon. EEG, a method that provides non-invasive technique that offers precise evaluation of brain activity with a high temporal resolution, has been utilized in Study 5 and Study 6 to investigate this phenomenon, with the main focus on ERPs. Prior research in this field has predominantly focused on investigating the N1, which has been linked to memory impairment resulting from auditory distractions, and the MMN, which was not linked to memory disruption (Campbell et al., 2007). Furthermore, investigating the role of the P2 and P3 ERP components in the ISE does not only fill a gap in the existing research but also contributes to elucidating the theoretical underpinnings, particularly regarding attentional processes.

In Study 5, participants completed a standard serial recall task with verbal items while exposed to either pink noise or unfamiliar speech, with EEG data collected during the task. Behaviorally, this study successfully replicated the ISE, confirming that task-irrelevant background speech negatively impacts the serial order retention of visual-verbal items. Additionally, the findings of Study 5 revealed a cross-modal effect at the neural level: exposure to irrelevant unfamiliar speech resulted in a greater visual N1 amplitude compared to the continuous pink noise control condition. This heightened N1 response suggests that speech presence may enhance processing demands towards the speech, potentially disrupting memory and interfering with the focal task. The visual P2 amplitude was reduced in the presence of speech compared to pink noise. Within the context of selective attention research, the P2 is linked to the inhibition of information processing, protecting against interference (Bell et al., 2010). Thus, the reduction in P2 amplitude indicates the brain's attempt to minimize distraction by suppressing the processing of irrelevant auditory input.

The visual P3a showed a trend-level modulation between the speech and pink noise conditions, though the result remains somewhat inconclusive. This indicates that task-irrelevant speech did not elicit an orienting response. This indicates that, at least in this context, the speech was not perceived as sufficiently salient or disruptive to activate the attentional shift typically reflected by the P3a component. Thus, it did not trigger an

3 General Discussion

involuntary shift of attention, which is typically associated with memory impairment. This implies that the memory impairment observed in the task may not be due to attentional capture or distraction from the speech itself, as indicated by the absence of P3a activation. Instead, the impairment might arise from more specific interference mechanisms, such as disruption of phonological rehearsal or encoding processes, rather than from attentional diversion.

Building upon the findings from Study 5, which demonstrated enhanced N1 amplitude for the visual-verbal serial recall when task-irrelevant unfamiliar speech was presented, future research could explore the N1 component within the context of the visual-spatial serial recall task, as employed in Study 1, Study 2, and Study 3. While it is apparent from the behavioral findings from Studies 1, 2, and 3, as well as other studies (Bergström et al., 2012; Kvetnaya, 2018; Marsh, Hurlstone, et al., 2024), that task-irrelevant speech did not have any detrimental impact on the serial recall performance of spatial items, it may nonetheless be advantageous to explore the visual N1 response as investigated in Studies 5 in the spatial serial recall task. This approach could help elucidate the processing demands underlying spatial task performance at the neural level. If a study demonstrates an enhancement in the visual N1 amplitude for spatial items despite no observable interference effects on behavioral task performance, it would indicate a dissociation between behavioral and neural responses to auditory distraction. Additionally, this finding would suggest that the neural encoding of visual information is sensitive to auditory interference, even when behavioral performance is intact. Such insights could enhance our understanding of the results from Study 5 and clarify the role of the visual N1 in the ISE.

Further investigations may benefit from exploring additional ERP components such as the N400. This ERP component is known to be larger for words that mismatch the current semantic context compared to words that match the context (Kappenman & Luck, 2011; Kutas & Federmeier, 2011). Examining the N400 response could provide insights into how semantic processing is affected by auditory distraction. For instance, examining the N400 response to “taboo” words” (i.e., deviant stimuli) could elucidate the words’ potential for attentional capture. If the N400 response to these deviant stimuli is larger than that observed for changing-state speech items, it could indicate heightened attentional capture by taboo words. This investigation would contribute to our understanding of how auditory distractors, including taboo words, influence semantic processing and attentional mechanisms. This bears the potential to disentangle the roles of specific interference and attention capture at the neural level.

Given the divergence between our findings and the existing body of research, the functional significance of P2 and P3a remains unclear. This underscores the complex nature of ERP components and the need for additional research into the neural mechanisms that underlie the ISE.

Study 6 employed a classic ISE paradigm, methodologically comparable to Study 1, combined with EEG measurement using a passive oddball paradigm. This exploratory study aimed to investigate the relationship between susceptibility to background speech during

visual-verbal serial recall and EEG components associated with auditory discrimination (MMN) and attentional processes (P3a). Participants completed a serial recall task of visual-verbal items under three conditions: silence, steady-state syllables, and changing-state syllables. In a second session, their EEG was recorded during a passive oddball task (consisting of speech syllables), both in silence and while traffic noise was played in the background. Although the correlational analyses of P3a did not yield significant results, MMN amplitude and latency for vowel changes measured in silence were both correlated with increased susceptibility to changing-state sounds in the serial recall task. Specifically, lower performance during changing-state speech in the ISE paradigm was linked to a reduced MMN amplitude and delayed peak.

These findings suggest that pre-attentive neural responses related to phoneme discrimination, particularly for vowel changes, are closely associated with vulnerability to background speech disruption in serial order retention. This may manifest as delayed neural processing and increased activation. The automatic processing of background speech during phonemic discrimination likely results in: 1) greater neural activation, as reflected by a more pronounced negative deflection for vowel changes, and 2) delayed processing, indicated by prolonged latencies—both of which may contribute to impaired working memory performance.

This finding has potential implications for theoretical accounts of the ISE, particularly in relation to the specific interference based approaches. The significant correlation between sound-induced disruption in serial recall and the MMN supports the changing-state hypothesis, which posits that cognitive performance is negatively affected by auditory stimuli characterized by frequent changes or variability in sound patterns. Such stimuli are automatically organized into streams of ordered objects through pre-attentive auditory streaming processes. The finding that lower performance is associated with both reduced MMN amplitude and delayed neural responses may support the changing-state hypothesis. A delayed MMN peak suggests that individuals who are more susceptible to background speech may take longer to process phonemic changes, resulting in greater interference and poorer task performance.

The correlation between reduced MMN amplitude, delayed peak latency, and lower performance during changing-state speech suggests that phoneme discrimination, especially for vowel changes, contributes significantly to susceptibility to the ISE. This supports the idea that the ISE arises from interference within the phonological loop. The reduced MMN amplitude could reflect weaker or less efficient encoding of phonological information, making it more difficult to maintain accurate verbal recall in the presence of changing-state sounds.

Future research could expand upon these findings by employing different methodologies to explore the MMN, such as the multi-feature paradigm (Pakarinen et al., 2009). This approach would facilitate a more comprehensive examination of the MMN and N1 responses, enabling the simultaneous investigation of multiple auditory attributes, such as speech sounds and tones, and their impact on serial recall performance. This potentially allows

3 General Discussion

to investigate the differential mechanisms proposed by the Duplex Mechanism Account (Hughes, 2014). This account suggests that auditory distraction arises from two distinct mechanisms, namely specific interference and attentional capture. By comparing an oddball paradigm with a multifeature paradigm, one could draw associations between the deviant effect (attentional capture) and the changing-state effect (specific interference). Such studies could further elucidate the relationship between different theoretical explanations underlying the ISE and offer deeper insights into the neural complexities of auditory distraction, thus providing neural evidence for the account's dual-process theory. To strengthen our understanding, it is essential to examine how specific ERP components, like the N1 and P3a, relate to distinct cognitive demands in recall tasks.

Drawing conclusions from neural indices, such as ERP components, to behavioral performance can be insightful, but this must be done cautiously for several reasons. First, multiple factors, such as individual differences, task difficulty, and cognitive strategies, can influence both neural responses and behavioral results, sometimes independently. For example, two participants might show similar neural responses (e.g., comparable N1 amplitudes) but differ in their behavioral outcomes due to factors like experience, strategy, or motivation. Second, neural data, being time-locked, offer a precise temporal mapping of cognitive processes that behavioral measures, like overall accuracy scores, cannot provide. Behavioral performance captures an outcome (e.g., correct recall or response accuracy) without revealing the specific temporal or phonological processes underlying task execution. The absence of a direct correspondence between neural signals and behavioral outcomes highlights the complexity of cognitive processes and underscores the importance of not assuming a straightforward relationship between neural activity and performance. By contrast, ERPs allow us to track the specific timing and sequence of cognitive events, enabling a deeper examination of the phonological and attentional processes involved in tasks like serial recall. This time-sensitive approach may reveal subtle neural dynamics—such as early sensory encoding or attention shifts—that are crucial for understanding the underlying mechanisms of the ISE beyond what behavioral data alone can show.

There are considerable practical implications of understanding children's high noise sensitivity. Given that working memory is crucial for learning and recalling information, background noise can hinder children's ability to process and retain what they learn, affecting their academic performance and overall cognitive development. Since children often underestimate the impact of irrelevant sounds, they might study or complete homework in noisy environments, diminishing their efforts. This highlights the urgent need for improved acoustic conditions in educational settings. By ensuring that classrooms and learning environments are conducive to concentration and clear communication, we can better support children in developing essential skills that will serve them throughout their education and beyond.

The majority of findings presented in this thesis emphasize the significant role of phonological processing in the ISE. Current research has increasingly focused on the unique

influences of semantic properties and the distinct effects of both the deviant and changing-state phenomena. I believe that further exploration of the assumption that irrelevant speech disrupts order-based performance through the integration of phonological representations would benefit the field, especially when considering children. Studies consistently show that irrelevant speech disrupts tasks involving the recognition of sublexical units, such as phonemes. For instance, Klatte et al. (2007) demonstrated that task-irrelevant speech significantly impaired children's performance on a phonological short-term memory task, in which they were asked to determine whether two non-words were the same or different. Similarly, Pröhl (2014) found interference effects in both primary school children and adults during an odd-one-out task (Bradley & Bryant, 1983), in which participants identified the "odd" word based on initial sound. Additionally, Smith et al. (1995) revealed that irrelevant speech severely disrupted word recognition from sequences of letters (e.g., M-T for "empty"). These studies collectively highlight the detrimental impact of irrelevant speech on phonological processing and emphasize the need for further investigation into the mechanisms driving this disruption.

In conclusion, none of the existing theoretical accounts of the ISE fully explain the complete pattern of findings presented in this thesis. The results are best understood by considering both specific interference with visual-verbal serial recall and more general attentional effects, as contributing factors to the observed outcomes.

4 Summary and Conclusion

The ISE is the robust finding that task-irrelevant background speech and specific non-speech sounds impair serial order retention for visual-verbal items. This thesis thoroughly examined how attention capture, serial order retention, and phonological processing contribute to the ISE, drawing from different theoretical frameworks. Across six studies, I explored how various sounds (speech, non-speech, environmental sounds, and classroom noise), age-related differences, habituation, sound characteristics, and neural correlates influence task performance of verbal and spatial recall tasks.

Study 1 showed that changing-state speech impaired verbal but not spatial serial recall, with children being more affected than adults, and no habituation observed. Study 2 found both narrative speech and non-speech sounds impacted verbal recall, but only younger children's spatial recall was affected, with no evidence of habituation. Study 3 demonstrated that phonological and changing-state factors independently disrupted verbal serial recall in adults, while the spatial task remained unaffected (with the exception of sinewave changing-state speech). Study 4 revealed no additional impact from binaural presentation of classroom-like noise compared to monaural, affecting verbal recall in both children and adults. Both age groups were equally impaired. Study 5 provided evidence of early attentional capture (N1 modulation) by irrelevant speech, without significant effects on later cognitive processing (P3a). Study 6 supports the notion that increased susceptibility to task-irrelevant sounds for visual-verbal serial recall is associated with heightened MMN amplitude and delayed neural responses to vowel changes.

The results of this thesis provide compelling evidence that seriation is not a sufficient criterion for the elicitation of the ISE due to a non-impairment of serial order retention of spatial items. Task-irrelevant speech and specific non-speech sounds reliably disrupt verbal serial recall. Depending on sound characteristics and task demands, differential effects have been found for children's serial recall performance, with most of the children being equally vulnerable to interference than adults. Importantly, no habituation to irrelevant sounds was observed, indicating a stable interference effect throughout tasks. Spatial task performance was mostly unaffected, underscoring the specificity of verbal disruption. Furthermore, the phonology and changing-state characteristics contributed independently to the sound-induced disruption in adults. The role of pre-attentive processes, as indicated by MMN, and heightened processing demands, as reflected by N1 highlight the complexity of the neural underpinnings underlying the susceptibility of sound-induced performance decrements.

These findings carry important theoretical implications, favoring the phonological loop account over competing theories like the attention capture or changing-state models for

4 Summary and Conclusion

verbal tasks. However, none of the theoretical accounts of the ISE adequately integrate the complete pattern of results. The results are best explained by an interplay between specific interference and attentional processes, with attentional mechanisms likely playing a crucial role in eliciting the ISE in children, depending on task characteristics.

Bibliography

- Alikadic, L., & Röer, J. P. (2022). Loud auditory distractors are more difficult to ignore after all. *Experimental Psychology*, 69(3), 163–171. <https://doi.org/10.1027/1618-3169/a000554>
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77(6), 1698–1716. <https://doi.org/10.1111/j.1467-8624.2006.00968.x>
- Antonietti, A., Capri, T., Fabio, R. A., Stuart, G., Towey, G., Pugliese, A., & Martino, G. (2018). The irrelevant sound effect: Testing the psychological effects of sequence predictability. *Atti della Accademia Peloritana dei Pericolanti - Classe di Scienze Medico-Biologiche*, Vol 106, No 1 (2018). <https://doi.org/10.6092/1828-6550/APMB.106.1.2018.OS1>
- Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195, Vol. 2). New York: Academic Press. [https://doi.org/10.1016/s0079-7421\(08\)60422-3](https://doi.org/10.1016/s0079-7421(08)60422-3)
- AuBuchon, A. M., McGill, C. I., & Elliott, E. M. (2019). Auditory distraction does more than disrupt rehearsal processes in children’s serial recall. *Memory & Cognition*, 47(4), 738–748. <https://doi.org/10.3758/s13421-018-0879-4>
- AuBuchon, A. M., McGill, C. I., & Elliott, E. M. (2020). Decomposing the role of rehearsal in auditory distraction during serial recall. *Auditory Perception & Cognition*, 3(1–2), 18–32. <https://doi.org/10.1080/25742442.2020.1842996>
- Baddeley, A. D. (2000a). The episodic buffer: A new component of working memory? *Trends in cognitive sciences*, 4(11), 417–423.
- Baddeley, A. D. (2000b). The phonological loop and the irrelevant speech effect: Some comments on neath (2000). *Psychonomic Bulletin & Review*, 7(3), 544–549. <https://doi.org/10.3758/bf03214369>
- Baddeley, A. D. (2021). Developing the concept of working memory: The role of neuropsychology. *Archives of Clinical Neuropsychology*, 36(6), 861–873.
- Baddeley, A. D., Hitch, G., & Allen, R. (2021). A multicomponent model of working memory. *Working memory: State of the science*, 10–43.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (pp. 47–89). Elsevier. [https://doi.org/10.1016/s0079-7421\(08\)60452-1](https://doi.org/10.1016/s0079-7421(08)60452-1)
- Baddeley, A. D., & Salamé, P. (1986). The unattended speech effect: Perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(4), 525–529. <https://doi.org/10.1037/0278-7393.12.4.525>

Bibliography

- Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 459–471. <https://doi.org/10.1037/0278-7393.23.2.459>
- Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology Section A*, 51(3), 615–636. <https://doi.org/10.1080/713755774>
- Beaman, C. P., & Roeer, J. (2009). Learning and failing to learn in immediate memory. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31(31).
- Beck, A.-K., Berti, S., Czernochowski, D., & Lachmann, T. (2021). Do categorical representations modulate early automatic visual processing? a visual mismatch-negativity study. *Biological Psychology*, 163, 108139. <https://doi.org/10.1016/j.biopsycho.2021.108139>
- Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). Erp correlates of the irrelevant sound effect. *Psychophysiology*, 47(6), 1182–1191. <https://doi.org/https://doi.org/10.1111/j.1469-8986.2010.01029.x>
- Bell, R., Mund, I., & Buchner, A. (2011). Disruption of short-term memory by distractor speech: Does content matter? *Quarterly Journal of Experimental Psychology*, 64(1), 146–168. <https://doi.org/10.1080/17470218.2010.483769>
- Bell, R., Röer, J. P., 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(6), 1542–1557. <https://doi.org/10.1037/a0028459>
- Bell, R., Röer, J. P., Marsh, J. E., Storch, D., & Buchner, A. (2017). The effect of cognitive control on different types of auditory distraction: A preregistered study. *Experimental Psychology*, 64(5), 359–368. <https://doi.org/10.1027/1618-3169/a000372>
- Bergström, K., Lachmann, T., & Klatte, M. (2012). Wann stört lärm das geistige arbeiten? einfluss von aufgaben- und geräuschcharakteristiken bei der wirkung moderaten lärms auf arbeitsgedächtnisleistungen. *Fortschritte der Akustik. Beiträge zur 38. Jahrestagung für Akustik, DAGA*.
- Blauert, J., & Braasch, J. (2008). Räumliches hören. In S. Weinzierl (Ed.), *Handbuch der audiotchnik* (pp. 87–121). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-34301-1_3
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. *Nature*, 301(5899), 419–421. <https://doi.org/10.1038/301419a0>
- 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>

- Broadbent, D. (1983). Recent advances in understanding performance in noise. In G. Rossi (Ed.), *Noise as a public health problem: Proceedings of the international congress (4th) held at turin (italy) on june 21-25, 1983. volume 2.* (pp. 719–738).
- Buchner, A., Bell, R., Rothermund, K., & Wentura, D. (2008). Sound source location modulates the irrelevant-sound effect. *Memory & Cognition*, 36(3), 617–628. <https://doi.org/10.3758/mc.36.3.617>
- Buchner, A., Irmen, L., & Erdfelder, E. (1996). On the irrelevance of semantic information for the “irrelevant speech” effect. *The Quarterly Journal of Experimental Psychology Section A*, 49(3), 765–779. <https://doi.org/10.1080/713755633>
- Campbell, T., Winkler, I., & Kujala, T. (2007). N1 and the mismatch negativity are spatiotemporally distinct ERP components: Disruption of immediate memory by auditory distraction can be related to n1. *Psychophysiology*, 44(4), 530–540. <https://doi.org/10.1111/j.1469-8986.2007.00529.x>
- Campbell, T., Winkler, I., Kujala, T., & Näätänen, R. (2003). The n1 hypothesis and irrelevant sound: Evidence from token set size effects. *Cognitive Brain Research*, 18(1), 39–47. <https://doi.org/10.1016/j.cogbrainres.2003.09.001>
- Campos, I. S., Almeida, L. S., Ferreira, A. I., & Martinez, L. F. (2013). Working memory as separable subsystems: A study with portuguese primary school children. *The Spanish Journal of Psychology*, 16. <https://doi.org/10.1017/sjp.2013.6>
- Cole, R., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (n.d.). The contribution of consonants versus vowels to word recognition in fluent speech. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. <https://doi.org/10.1109/icassp.1996.543255>
- Colle, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behavior*, 19(6), 722–735. [https://doi.org/10.1016/s0022-5371\(80\)90403-x](https://doi.org/10.1016/s0022-5371(80)90403-x)
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of verbal learning and verbal behavior*, 15(1), 17–31. [https://doi.org/https://doi.org/10.1016/S0022-5371\(76\)90003-7](https://doi.org/https://doi.org/10.1016/S0022-5371(76)90003-7)
- Conley, E. (1999). The n100 auditory cortical evoked potential indexes scanning of auditory short-term memory. *Clinical Neurophysiology*, 110(12), 2086–2093. [https://doi.org/10.1016/s1388-2457\(99\)00183-2](https://doi.org/10.1016/s1388-2457(99)00183-2)
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2), 163–191. <https://doi.org/10.1037/0033-2909.104.2.163>
- Cowan, N. (1995, February). *Attention and memory*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195119107.001.0001>
- Cowan, N. (1999, April). An embedded-processes model of working memory. In *Models of working memory* (pp. 62–101). Cambridge University Press. <https://doi.org/10.1017/cbo9781139174909.006>

Bibliography

- Cowan, N., Morey, C. C., & Naveh-Benjamin, M. (2020, November). An embedded-processes approach to working memory: How is it distinct from other approaches, and to what ends? In *Working memory* (pp. 44–84). Oxford University Press. <https://doi.org/10.1093/oso/9780198842286.003.0003>
- Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for p2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, 115(4), 732–744. <https://doi.org/10.1016/j.clinph.2003.11.021>
- Debruille, J. B., Touzel, M., Segal, J., Snidal, C., & Renoult, L. (2019). A central component of the n1 event-related brain potential could index the early and automatic inhibition of the actions systematically activated by objects. *Frontiers in Behavioral Neuroscience*, 13. <https://doi.org/10.3389/fnbeh.2019.00095>
- DEGA, A. L. (2023). Hörbeispiele [<https://www.ald-laerm.de/publikationen/hoerbeispiele/einzelquellen/strassenverkehrs-laerm-mit-grosser-entfernungsvariation>] [Accessed: 2022-01-10].
- Divin, W., Coyle, K., & James, D. T. T. (2001). The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *British Journal of Psychology*, 92(4), 593–616. <https://doi.org/10.1348/000712601162365>
- Ebbinghaus, H. (1885). *Über das gedächtnis: Untersuchungen zur experimentellen psychologie*. Duncker & Humblot.
- Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in "irrelevant speech"? effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), 1406–1414. <https://doi.org/10.1037/0096-1523.24.5.1406>
- Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the “irrelevant speech effect”. *The Journal of the Acoustical Society of America*, 102(4), 2191–2199. <https://doi.org/https://doi.org/10.1121/1.419596>
- Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, 30(3), 478–487. <https://doi.org/10.3758/bf03194948>
- Elliott, E. M., Bhagat, S. P., & Lynn, S. D. (2007). Can children with (central) auditory processing disorders ignore irrelevant sounds? *Research in Developmental Disabilities*, 28(5), 506–517. <https://doi.org/10.1016/j.ridd.2006.06.005>
- Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, 140(1), 64–74. <https://doi.org/10.1016/j.actpsy.2012.02.009>
- Elliott, E. M., & Cowan, N. (2005). Coherence of the irrelevant-sound effect: Individual profiles of short-term memory and susceptibility to task-irrelevant materials. *Memory & Cognition*, 33(4), 664–675. <https://doi.org/10.3758/bf03195333>
- Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., & Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental

- differences. *Journal of Memory and Language*, 88, 39–50. <https://doi.org/10.1016/j.jml.2015.12.008>
- Endrass, T., Mohr, B., & Pulvermüller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, 19(6), 1653–1660. <https://doi.org/10.1111/j.1460-9568.2004.03247.x>
- Escera, C., Alho, K., Schröger, E., & Winkler, I. W. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology*, 5(3–4), 151–166. <https://doi.org/10.1159/000013877>
- Evans, G. W., Hygge, S., & Bullinger, M. (1995). Chronic noise and psychological stress. *Psychological Science*, 6(6), 333–338. <https://doi.org/10.1111/j.1467-9280.1995.tb00522.x>
- Fogerty, D., & Humes, L. E. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America*, 131(2), 1490–1501. <https://doi.org/10.1121/1.3676696>
- Fogerty, D., & Kewley-Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *The Journal of the Acoustical Society of America*, 126(2), 847–857. <https://doi.org/10.1121/1.3159302>
- García-Larrea, L., Lukaszewicz, A.-C., & Mauguière, F. (1992). Revisiting the oddball paradigm. non-target vs neutral stimuli and the evaluation of erp attentional effects. *Neuropsychologia*, 30(8), 723–741. [https://doi.org/10.1016/0028-3932\(92\)90042-k](https://doi.org/10.1016/0028-3932(92)90042-k)
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. (2014). Cognitive neuroscience. the biology of the mind.
- Georgi, M., Leist, L., Klatte, M., & Schlittmeier, S. J. (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*, 6(1-2), 52–71. <https://doi.org/10.1080/25742442.2022.2127988>
- Gisselgård, J., Petersson, K. M., & Ingvar, M. (2004). The irrelevant speech effect and working memory load. *NeuroImage*, 22(3), 1107–1116. <https://doi.org/10.1016/j.neuroimage.2004.02.031>
- Gumenyuk, V., Korzyukov, O., Alho, K., Escera, C., & Näätänen, R. (2004). Effects of auditory distraction on electrophysiological brain activity and performance in children aged 8-13 years. *Psychophysiology*, 41(1), 30–36. <https://doi.org/10.1111/1469-8986.00123>
- Hadlington, L., Bridges, A. M., & Darby, R. J. (2004). Auditory location in the irrelevant sound effect: The effects of presenting auditory stimuli to either the left ear, right ear or both ears. *Brain and Cognition*, 55(3), 545–557. <https://doi.org/10.1016/j.bandc.2004.04.001>
- Hanley, J. R. (1997). Does articulatory suppression remove the irrelevant speech effect? *Memory*, 5(3), 423–431. <https://doi.org/10.1080/741941394>

Bibliography

- Hanley, J. R., & Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology*, 78(3), 287–297. <https://doi.org/10.1111/j.2044-8295.1987.tb02247.x>
- Hellbrück, J., Kuwano, S., & Namba, S. (1996). Irrelevant background speech and human performance: Is there long-term habituation? *Journal of the Acoustical Society of Japan (E)*, 17(5), 239–247. <https://doi.org/https://doi.org/10.1250/ast.17.239>
- Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, 3(1), 30–41. <https://doi.org/https://doi.org/10.1002/pchj.44>
- Hughes, R. W., Hurlstone, M. J., Marsh, J. E., 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 and Performance*, 39(2), 539–553. <https://doi.org/10.1037/a0029064>
- Hughes, R. W., & Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise and Health*, 4(13), 51–70.
- Hughes, R. W., & Marsh, J. E. (2020). When is forewarned forearmed? predicting auditory distraction in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(3), 427–442. <https://doi.org/10.1037/xlm0000736>
- Hughes, R. W., Tremblay, S., & Jones, D. M. (2005). Disruption by speech of serial short-term memory: The role of changing-state vowels. *Psychonomic Bulletin & Review*, 12(5), 886–890. <https://doi.org/10.3758/bf03196781>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2005). *Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model?* (Vol. 31). American Psychological Association (APA). <https://doi.org/10.1037/0278-7393.31.4.736>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6), 1050. <https://doi.org/10.1037/0278-7393.33.6.1050>
- James, W. (1890). *The principles of psychology*. Henry Holt; Company.
- Jasper, H. H. (1958). Ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol*, 10, 371–375.
- Jelelati, A., Leist, L., Lachmann, T., & Klatte, M. (2023). Effects of irrelevant speech on immediate serial recall: The role of phonological complexity. *Cognitive Psychology*, 22(8), 1168–1191.
- Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 464–473. <https://doi.org/10.1037/0278-7393.25.2.464>

- Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 1008–1018. <https://doi.org/10.1037/0278-7393.21.4.1008>
- Jones, D. M., Hughes, R. W., Macken, W. J., et al. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise and Health*, 12(49), 201. <https://doi.org/https://doi.org/10.4103/1463-1741.70497>
- 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. <https://doi.org/https://doi.org/10.1037/0278-7393.19.2.369>
- Jones, D. M., & Macken, W. J. (1995). Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied*, 1(3), 216–226. <https://doi.org/10.1037/1076-898x.1.3.216>
- 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/10.1111/j.2044-8295.1997.tb02657.x>
- Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology Section A*, 44(4), 645–669. <https://doi.org/10.1080/14640749208401304>
- Jones, D. M., Marsh, J. E., & Hughes, R. W. (2012). Retrieval from memory: Vulnerable or inviolable? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 905–922. <https://doi.org/10.1037/a0026781>
- 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/10.1002/acp.2350040203>
- 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>
- Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. *Models of short-term memory*, 209–238.
- Joseph, T. N., Hughes, R. W., Sörqvist, P., & Marsh, J. E. (2018). Differences in auditory distraction between adults and children: A duplex-mechanism approach. *Journal of Cognition*, 1(1). <https://doi.org/10.5334/joc.15>
- Kappenman, E. S., & Luck, S. J. (2011). Of brainwave recordings. *The Oxford handbook of event-related potential components*, 3.
- Kattner, F., & Ellermeier, W. (2020). Distraction at the cocktail party: Attenuation of the irrelevant speech effect after a training of auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 46(1), 10–20. <https://doi.org/10.1037/xhp0000695>

Bibliography

- Klatte, M., & Hellbrück, J. Effects of irrelevant speech on serial recall of verbal and spatial materials (A. Schick & M. Klatte, Eds.). In: *Contributions to psychological acoustics. results of the 7th oldenburg symposium on psychological acoustics* (A. Schick & M. Klatte, Eds.). Ed. by Schick, A., & Klatte, M. 1997, 531–538.
- Klatte, M., Hellbrück, J., Seidel, J., & Leistner, P. (2010). Effects of classroom acoustics on performance and well-being in elementary school children: A field study. *Environment and Behavior*, 42(5), 659–692. <https://doi.org/10.1177/0013916509336813>
- Klatte, M., Kilcher, H., & Hellbrück, J. (1995). Wirkungen der zeitlichen struktur von hintergrundschall auf das arbeitsgedächtnis und ihre theoretischen und praktischen implikationen. *Zeitschrift für Experimentelle Psychologie*, 42(4), 517–544.
- Klatte, M., Lachmann, T., Schlittmeier, S. J., & Hellbrück, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology*, 22(8), 1168–1191. <https://doi.org/10.1080/09541440903378250>
- Klatte, M., Meis, M., Sukowski, H., & Schick, A. (2007). Effects of irrelevant speech and traffic noise on speech perception and cognitive performance in elementary school children. *Noise and Health*, 9(36), 64. <https://doi.org/10.4103/1463-1741.36982>
- Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research*, 339, 40–49. <https://doi.org/10.1016/j.heares.2016.06.001>
- Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2017). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A p3 study. *Hearing Research*, 350, 58–67. <https://doi.org/10.1016/j.heares.2017.04.009>
- Kozou, H., Kujala, T., Shtyrov, Y., Toppila, E., Starck, J., Alku, P., & Näätänen, R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, 199(1–2), 31–39. <https://doi.org/10.1016/j.heares.2004.07.010>
- Kujala, T., & Brattico, E. (2009). Detrimental noise effects on brain's speech functions. *Biological Psychology*, 81(3), 135–143. <https://doi.org/10.1016/j.biopsycho.2009.03.010>
- Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74(1), 1–19. <https://doi.org/10.1016/j.biopsycho.2006.06.001>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the n400 component of the event-related brain potential (erp). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kvetnaya, T. (2018). Registered replication report: Testing disruptive effects of irrelevant speech on visual-spatial working memory. *Journal of European Psychology Students*, 9(1), 10–15. <https://doi.org/10.5334/jeps.450>

- Larsen, J. D., Baddeley, A. D., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory*, 8(3), 145–157. <https://doi.org/10.1080/096582100387579>
- LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1396–1408. <https://doi.org/10.1037/0278-7393.20.6.1396>
- LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: A new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1154–1165. <https://doi.org/10.1037/0278-7393.22.5.1154>
- LeCompte, D. C., Neely, C. B., & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 472–483. <https://doi.org/10.1037/0278-7393.23.2.472>
- Leist, L., Breuer, C., Yadav, M., Fremerey, S., Fels, J., Raake, A., Lachmann, T., Schlittmeier, S. J., & Klatte, M. (2022). Differential effects of task-irrelevant monaural and binaural classroom scenarios on children’s and adults’ speech perception, listening comprehension, and visual–verbal short-term memory. *International Journal of Environmental Research and Public Health*, 19(23), 15998. <https://doi.org/10.3390/ijerph192315998>
- Leist, L., Lachmann, T., & Klatte, M. (2025). Impact of irrelevant speech and non-speech sounds on serial recall of verbal and spatial items in children and adults. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-85855-w>
- Leist, L., Lachmann, T., Schlittmeier, S. J., Georgi, M., & Klatte, M. (2023). Irrelevant speech impairs serial recall of verbal, but not spatial items in children and adults. *Memory & Cognition*, 51(2), 307–320. <https://doi.org/10.3758/s13421-022-01359-2>
- Little, J. S., Martin, F. H., & Thomson, R. H. (2010). Speech versus non-speech as irrelevant sound: Controlling acoustic variation. *Biological Psychology*, 85(1), 62–70. <https://doi.org/10.1016/j.biopsycho.2010.05.004>
- 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*, 5(1–2), 1–24. <https://doi.org/10.1080/25742442.2022.2063609>
- Marois, A., Marsh, J. E., & Vachon, F. (2019). Is auditory distraction by changing-state and deviant sounds underpinned by the same mechanism? evidence from pupillometry. *Biological Psychology*, 141, 64–74. <https://doi.org/10.1016/j.biopsycho.2019.01.002>
- Marsh, J. E., Hanczakowski, M., Beaman, C. P., Meng, Z., & Jones, D. M. (2024). Thinking about meaning: Level-of-processing modulates semantic auditory distraction. *Journal of Cognitive Psychology*, 1–13. <https://doi.org/10.1080/20445911.2024.2381279>

Bibliography

- Marsh, J. E., & Jones, D. M. (2010). Cross-modal distraction by background speech: What role for meaning? *Noise and Health*, 12(49), 210. <https://doi.org/10.4103/1463-1741.70499>
- Marsh, J. E., Ljung, R., Jahncke, H., MacCutcheon, D., Pausch, F., Ball, L. J., & Vachon, F. (2018). Why are background telephone conversations distracting? *Journal of Experimental Psychology: Applied*, 24(2), 222–235. <https://doi.org/10.1037/xap0000170>
- 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(6), 882–897. <https://doi.org/10.1037/xlm0000492>
- Marsh, J. E., Hurlstone, M., Marois, A., Ball, L., Moore, S., Vachon, F., Schlittmeier, S. J., Röer, J. P., Buchner, A., Aust, F., et al. (2024). Changing-state irrelevant speech disrupts visual-verbal but not visual-spatial serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, 27(4), 382–398. [https://doi.org/10.1016/0749-596x\(88\)90063-0](https://doi.org/10.1016/0749-596x(88)90063-0)
- Martín-Loeches, M., Schweinberger, S. R., & Sommer, W. (1997). The phonological loop model of working memory: An erp study of irrelevant speech and phonological similarity effects. *Memory & Cognition*, 25(4), 471–483.
- Meinhardt-Injac, B., Imhof, M., Wetzell, N., Klatte, M., & Schlittmeier, S. J. (2022). The irrelevant sound effect on serial recall is independent of age and inhibitory control. *Auditory Perception & Cognition*, 1–21.
- Morey, C. C., & Miron, M. D. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(12), 1907–1918. <https://doi.org/10.1037/xlm0000280>
- 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>
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13(2), 201–233. <https://doi.org/10.1017/s0140525x00078407>
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (mmn) and its magnetic equivalent (mmnm). *Psychophysiology*, 38(1), 1–21. <https://doi.org/10.1111/1469-8986.3810001>
- Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (mmn): A review of the evidence. *Psychophysiology*, 42(1), 25–32. <https://doi.org/10.1111/j.1469-8986.2005.00256.x>

- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic bulletin & review*, 7(3), 403–423. <https://doi.org/10.3758/BF03214356>
- Novak, G., Ritter, W., & Vaughan, H. G. (1992). Mismatch detection and the latency of temporal judgments. *Psychophysiology*, 29(4), 398–411. <https://doi.org/10.1111/j.1469-8986.1992.tb01713.x>
- Oberauer, K. (2019). Working memory and attention – a conceptual analysis and review. *Journal of Cognition*, 2(1). <https://doi.org/10.5334/joc.58>
- Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (mmns) to phonetic and acoustic changes in speech sounds. *Biological Psychology*, 82(3), 219–226. <https://doi.org/10.1016/j.biopsycho.2009.07.008>
- Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *Journal of Neuroscience Methods*, 162(1-2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Pelletier, M.-F., Hodgetts, H. M., Lafleur, M. F., Vincent, A., & Tremblay, S. (2016). Vulnerability to the irrelevant sound effect in adult adhd. *Journal of Attention Disorders*, 20(4), 306–316. <https://doi.org/10.1177/1087054713492563>
- Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- Prölb, A. (2014). *Wirkungen akustischer umweltbedingungen auf arbeitsgedächtnisleistungen bei kindern und erwachsenen: Experimentelle untersuchungen zum "irrelevant sound effect"* [Doctoral dissertation, Kaiserslautern, Technische Universität Kaiserslautern, Diss., 2014].
- Rettie, L., Potter, R. F., Brewer, G., Degno, F., Vachon, F., Hughes, R. W., & Marsh, J. E. (2023). Warning—taboo words ahead! avoiding attentional capture by spoken taboo distractors. *Journal of Cognitive Psychology*, 1–17. <https://doi.org/10.1080/20445911.2023.2285860>
- Röer, J. P., Bell, R., & Buchner, A. (2013a). 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. (2013b). 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). Please silence your cell phone: Your ringtone captures other people's attention. *Noise and Health*, 16(68), 34. <https://doi.org/10.4103/1463-1741.127852>
- Röer, J. P., Bell, R., & Buchner, A. (2014b). What determines auditory distraction? on the roles of local auditory changes and expectation violations (M. S. Malmierca, Ed.). *PLoS ONE*, 9(1), e84166. <https://doi.org/10.1371/journal.pone.0084166>
- Röer, J. P., Bell, R., & Buchner, A. (2015). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception and Performance*, 41(3), 692–702. <https://doi.org/10.1037/xhp0000028>

Bibliography

- 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., Bell, R., Körner, U., & Buchner, A. (2018). Equivalent auditory distraction in children and adults. *Journal of Experimental Child Psychology*, 172, 41–58. <https://doi.org/10.1016/j.jecp.2018.02.005>
- Röer, J. P., Bell, R., Marsh, J. E., & Buchner, A. (2015). Age equivalence in auditory distraction by changing and deviant speech sounds. *Psychology and Aging*, 30(4), 849–855. <https://doi.org/10.1037/pag0000055>
- Röer, J. P., Buchner, A., & Bell, R. (2019). Auditory distraction in short-term memory: Stable effects of semantic mismatches on serial recall. *Auditory Perception & Cognition*, 2(3), 143–162. <https://doi.org/10.1080/25742442.2020.1722560>
- Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017). Attentional capture by taboo words: A functional view of auditory distraction. *Emotion*, 17(4), 740–750. <https://doi.org/10.1037/emo0000274>
- 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. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, 24(4), 263–265. <https://doi.org/10.3758/bf03330135>
- Salamé, P., & Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, 30(8), 1185–1194. <https://doi.org/10.1080/00140138708966007>
- 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>
- Salamé, P., & Baddeley, A. D. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society*, 28(6), 540–542. <https://doi.org/10.3758/bf03334073>
- Samper, J. R., Morrison, A., & Chein, J. (2021). Doubts about the role of rehearsal in the irrelevant sound effect. *Experimental Psychology*, 68(5), 229–242. <https://doi.org/10.1027/1618-3169/a000527>
- Schlittmeier, S. J., Hellbrück, J., Thaden, R., & Vorländer, M. (2008). The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance. *Ergonomics*, 51(5), 719–736. <https://doi.org/10.1080/00140130701745925>
- Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2011). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, & Psychophysics*, 74(1), 194–203. <https://doi.org/10.3758/s13414-011-0230-7>

- Schlittmeier, S. J., Weisz, N., & Bertrand, O. (2011). What characterizes changing-state speech in affecting short-term memory? an eeg study on the irrelevant sound effect. *Psychophysiology*, 48(12), 1669–1680. [https://doi.org/https://doi.org/10.1111/j.1469-8986.2011.01263.x](https://doi.org/10.1111/j.1469-8986.2011.01263.x)
- Schwarz, H., Schlittmeier, S. J., Otto, A., Persike, M., Klatte, M., Imhof, M., & Meinhardt-Injac, B. (2015). Age differences in the irrelevant sound effect: A serial recognition paradigm. *psihologija*, 48(1), 35–43. <https://doi.org/10.2298/PSI1501035S>
- Schweppe, J., & Knigge, J. (2020). Irrelevant music: How suprasegmental changes of a melody's tempo and mode affect the disruptive potential of music on serial recall. *Memory & Cognition*, 48(6), 982–993. <https://doi.org/10.3758/s13421-020-01037-1>
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33(11), 1433–1454. [https://doi.org/10.1016/0028-3932\(95\)00074-d](https://doi.org/10.1016/0028-3932(95)00074-d)
- Sokhadze, E. M., Casanova, M. F., Casanova, E. L., Lamina, E., Kelly, D. P., & Khachidze, I. (2017). Event-related potentials (erp) in cognitive neuroscience research and applications. *NeuroRegulation*, 4(1), 14–27. <https://doi.org/10.15540/nr.4.1.14>
- Sokolov, E. (1963). Perception and the conditioned reflex.
- Sörqvist, P. (2010). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition*, 38(5), 651–658. <https://doi.org/10.3758/mc.38.5.651>
- Stokes, K. A., & Arnell, K. M. (2012). New considerations for the cognitive locus of impairment in the irrelevant-sound effect. *Memory & Cognition*, 40(6), 918–931. <https://doi.org/10.3758/s13421-012-0194-4>
- Surprenant, A. M. (2007). Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging, Neuropsychology, and Cognition*, 14(2), 126–143. <https://doi.org/10.1080/13825580701217710>
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700), 1187–1188. <https://doi.org/10.1126/science.150.3700.1187>
- Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(3), 659–671. <https://doi.org/10.1037/0278-7393.24.3.659>
- Tremblay, S., & Jones, D. M. (1999). Change of intensity fails to produce an irrelevant sound effect: Implications for the representation of unattended sound. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 1005.
- Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1750–1754. <https://doi.org/10.1037/0278-7393.26.6.1750>

Bibliography

- Tremblay, S., Parmentier, F. B., Hodgetts, H. M., Hughes, R. W., & Jones, D. M. (2012). Disruption of verbal-spatial serial memory by extraneous air-traffic speech. *Journal of Applied Research in Memory and Cognition*, 1(2), 73–79. <https://doi.org/10.1016/j.jarmac.2012.04.004>
- Ueda, K., Nakajima, Y., Kattner, F., & Ellermeier, W. (2019). Irrelevant speech effects with locally time-reversed speech: Native vs non-native language. *The Journal of the Acoustical Society of America*, 145(6), 3686–3694. <https://doi.org/10.1121/1.5112774>
- 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.
- Valtonen, J., May, P., Mäkinen, V., & Tiitinen, H. (2003). Visual short-term memory load affects sensory processing of irrelevant sounds in human auditory cortex. *Cognitive Brain Research*, 17(2), 358–367.
- Viswanathan, N., Dorsi, J., & George, S. (2014). The role of speech-specific properties of the background in the irrelevant sound effect. *Quarterly Journal of Experimental Psychology*, 67(3), 581–589. <https://doi.org/10.1080/17470218.2013.821708>
- Weisz, N., & Schlittmeier, S. J. (2006). Detrimental effects of irrelevant speech on serial recall of visual items are reflected in reduced visual n1 and reduced theta activity. *Cerebral Cortex*, 16(8), 1097–1105. <https://doi.org/10.1093/cercor/bhj051>
- Wetzel, N., Scharf, F., & Widmann, A. (2018). Can't ignore—distraction by task-irrelevant sounds in early and middle childhood. *Child Development*, 90(6). <https://doi.org/10.1111/cdev.13109>
- Wetzel, N., Schröger, E., & Widmann, A. (2016). Distraction by novel and pitch-deviant sounds in children. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01949>
- Wetzel, N., Widmann, A., Berti, S., & Schröger, E. (2006). The development of involuntary and voluntary attention from childhood to adulthood: A combined behavioral and event-related potential study. *Clinical Neurophysiology*, 117(10), 2191–2203. <https://doi.org/10.1016/j.clinph.2006.06.717>
- Wetzel, N., Widmann, A., & Scharf, F. (2021). Distraction of attention by novel sounds in children declines fast. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-83528-y>
- Wetzel, N., Widmann, A., & Schröger, E. (2011). Processing of novel identifiability and duration in children and adults. *Biological Psychology*, 86(1), 39–49. <https://doi.org/10.1016/j.biopsycho.2010.10.005>
- Wilke, S. (2024). Indicator: Population exposure to traffic noise [Accessed : 2024-06-09]. <https://www.umweltbundesamt.de/en/data/environmental-indicators/indicator-population-exposure-to-traffic-noise#at-a-glance>
- Yadav, M., Georgi, M., Leist, L., Klatte, M., Schlittmeier, S. J., & Fels, J. (2023). Cognitive performance in open-plan office acoustic simulations: Effects of room acoustics

- and semantics but not spatial separation of sound sources. *Applied Acoustics*, 211, 109559. <https://doi.org/10.1016/j.apacoust.2023.109559>
- Zhang, Q., Williams, C., & Morgan, P. L. (2024). Partial habituation to disruption by irrelevant emotive speech—evidence for duplex-mechanism account. *Journal of Cognitive Psychology*, 1–19. <https://doi.org/10.1080/20445911.2024.2310881>

Appendix

Table A1. Studies investigating habituation underlying the ISE.

Publication	Exp.	N	Age groups	Trials per condition (in total)	Auditory conditions	Presentation duration	Analysis In blocks, single trial	Habituation effect
Morris & Jones, 1990	1	48	12 adults per auditory condition	4 blocks of 6 trials, 24 in total	Quiet, unfamiliar speech, Habituation phase between blocks, using 4 different auditory conditions, sth. Like preexposure	blocked	blocks	yes to familiar speech
Hellbrück, Kuwano, & Namba, 1996	1	12	6 German students (21-28 years) 6 Japanese employees (25-46 years)	3 blocks of 20 trials	Quiet control, unfamiliar speech, speech-simulating noise	Trial-by-trial variation	Single trial, averaged over 3 sessions and 3 blocks	no
Ellermeier & Zimmer, 1997	1	72	72 Adults (19-44 years)	2 or 3 blocks of 30 trials	Quiet, unfamiliar speech, pink noise	Trial-by-trial variation	averaged blockwise	no
Jones, Macken, & Mosdell, 1997	1	40	40 adults	60 trials	Quiet, steady-state words, changing-state words	Trial-by-trial	blocked	no
	2	30	30 adults	60 trials	Quiet, steady-state words, changing-state words	blocked	blocked	no
	3	60	Adults	12 trials per condition, 36 in total	Quiet, steady-state words, changing-state words	Trial-by-trial, with prior passive listening phase	blocked	no
Tremblay & Jones, 1998	1	24	Adults	15 trials per condition, 90 in total	Quiet, steady-state speech, four changing state speech conditions	trial-by-trial variation	Blocked (three blocks of 5 trials)	no
	2	48	adults	15 trials per condition, 90 in total	Quiet, steady-state tones and four changing-state tone conditions	trial-by-trial variation	Blocked (three blocks of 5 trials)	no
	3	80	20 adults per auditory condition	30 trials per condition	Quiet, steady-state speech and two changing-state speech conditions	trial-by-trial variation	blocked (five blocks of 6 trials)	no
	4	40	adults	15 trials per condition, 75 in total	Quiet, steady-state speech, three changing state speech conditions	blocked	blocked	no
	5	24	adults	15 trials per conditions, 90 in total	Quiet, steady-state speech, four changing state speech conditions	blocked	blocked	no
Elliott, 2002	1	128	32 second graders ($M=8;3y$, $SD=5m$) 32 third-fourth graders ($M=9;6y$, $SD=6m$)	60	Quiet, Steady-state words, Changing-state words, Steady -state tones, changing-state tones	blocked	blocked	yes, comparable across age groups

Appendix

32 fifth-sixt
graders
($M=11;8y$,
 $SD=7m$)
32 adults
($M=19;0y$,
 $SD=2y$)

Hughes, Vachon & Jones, 2005	1-4	34	Adults	15 trials per condition, 75 in total	Quiet, steady-state speech, steady-state speech + deviant, changing-state speech, changing-state speech + deviant	trial-by-trial variation	blocked	no
Beaman & Röer, 2009	2	18	Adults	10 trials per condition, 40 in total, over 4 days (160)	Quiet, changing-state words	n.a.	blocked	no
Röer et al., 2011	1	211	104 adults in auditory distractor group, 107 in quiet group ($M=26$ y)	90	Quiet, Changing-state speech	Blocked for auditory distractor group, distractors only played during introductory period and presentation phase	Blocked	no
	2	182	90 adults in auditory distractor group, 90 in quiet group ($M=26$ y)	90	Quiet, Changing-state speech	Blocked for auditory distractor group, distractors played throughout	Blocked	no
	3	146	73 adults in auditory distractor group, 73 in quiet group ($M=25$ y)	90	Quiet, Changing-state speech	Blocked for auditory distractor group, Four testing days	Blocked	no
Vachon et al., 2012	1	37	adults	18 blocks of 5 trials each (90)	Changing state letters, either in female or male voice	blocked	blocked	yes
	2	24	adults	14 blocks of 5 trials each (70)	Changing state letters, either in female or male voice	blocked	blocked	yes
Röer, Bell, & Buchner, 2013b	1	55	Adults ($M=24$ y)	8 per condition, 24 in total	Quiet, own name, other name	trial-by-trial variation	Trial-based	yes, but only to one's own name
Röer, Bell, & Buchner, 2013a	1	113	Adults ($M=23$ y)	8 trial per condition, 24 in total	Quiet, melody, speech	blocked	Trial-based	yes
	2	123	Adults ($M=23$ y)	8 trial per condition, 24 in total	Quiet, melody, speech	Trial-by-trial variation	Trial-based	yes
	3	204	103 adults in forward group	8 trial per condition, 24 in total	Quiet, Forward group: melody, speech	blocked	Trial-based	yes

			101 adults in backward group ($M=23$ y)		Backward group: reversed melody, reversed speech				
	4	118	Adults ($M=24$ y)	8 trial per condition, 24 in total	Melody or speech (only during 8s retention interval)	blocked	Trial-based	yes	
Röer, Bell, & Buchner, 2014b	1	98	Adults ($M=24$ y)	8 trial per condition, 24 in total	Quiet, steady-state ending, changing-state ending (melody)	Trial-by-trial variation with prior passive listening phase	Trial-based	yes (steady-state ending) no (changing-state ending)	
	2	59	Adults ($M=24$ y)	8 trial per condition, 24 in total	Quiet, steady-state ending, changing-state ending (speech)	Trial-by-trial variation with prior passive listening phase	Trial-based	no	
	3	69	35 adults in encoding group 34 adults in retention group ($M=24$ y)	8 trial per condition, 24 in total	Quiet, steady-state ending, changing-state ending (speech)	Trial-by-trial variation with prior passive listening phase	Trial-based	yes (steady-state ending) no (changing-state ending)	
	4	67	Adults ($M=23$ y)	8 trial per condition, 24 in total	Quiet, steady-state ending, changing-state ending (speech)	Trial-by-trial variation	Trial-based	yes (steady-state ending) no (changing-state ending)	
Röer, Bell, & Buchner, 2014a	1	51	Adults ($M=26$ y)	8 trial per condition, 24 in total	Quiet, own ringtone, other ringtone	Trial-by-trial variation	Trial-based	yes	
Röer, Bell, Marsh, & Buchner, 2015	1	260	128 older adults ($M=71$ y) 130 younger adults ($M=23$ y)	8 trial per condition, 24 in total	Quiet, steady-state speech, changing-state speech, steady-state + deviant	Trial-by-trial variation	Trial-by-trial variation	no	
Pelletier et al., 2016	1	64	32 Adults with ADHD 32 controls	25 per condition, 50 in total	Quiet, work environment	blocked	blocked	yes	
Röer et al., 2017	1	60	Adults ($M=24$ y)	10 trials per condition, 50 in total	Quiet, steady-state speech, steady-state speech + deviant, changing-state speech, changing-state speech + deviant	Trial-by-trial variation	Trial-based	no	
	2	74	Adults ($M=24$ y)	10 trials per condition, 50 in total	Quiet, steady-state neutral speech, steady-state + taboo deviant, changing-state neutral speech, changing-state speech + taboo deviant	Trial-by-trial variation	Trial-based	no	
	3	82	Adults ($M=25$ y)	10 trials per condition, 50 in total	Quiet, steady-state speech neutral, steady-state speech taboo, steady-state speech neutral + neutral deviant, steady-state speech taboo + taboo deviant	Trial-by-trial variation	Trial-based	no	
	4	84	Adults ($M=24$ y)	10 trials per condition, 50 in total	Quiet, steady-state speech neutral, steady-state speech taboo, steady-state speech neutral + neutral deviant, steady-state speech + taboo deviant	Trial-by-trial variation	Trial-based	no	

Appendix

Röer, et al., 2018	2	68	38 younger adults ($M=23.6$) 30 4 th -graders ($M=9.1$)	24	Changing-state words (familiar) Steady-state words (familiar) Deviant in Changing-state words (familiar)	Blocks of 4 trials	Trial-based	yes for changing-state and deviant condition, not for steady state condition comparable across age groups
Röer, et al., 2019	1	97	Adults ($M=24$ y)	16 trials per condition, 48 trials in total	Quiet, semantic match, semantic mismatch	Trial-by-trial variation	Trial-based	no
	2	90	Adults ($M=24$ y)	16 trials per condition, 64 trials in total	Semantic match and mismatch conditions, each with and without repetition	blocked	Trial-based	no
Littlefair et al., 2022	2	188	62 adults in acoustic deviant condition ($M=26$ y) 63 adults in categorical deviant condition ($M=25$ y) 65 adults combined condition ($M=24$ y)	18 deviant trials, 36 standard trials, 54 trials in total	Changing-state speech (standard), changing-state speech + deviant (either acoustic, categorical or combined)	n.a. deviant trials were presented at fixed position	blocked	yes, but only for acoustic deviant condition
Zhang et al., 2024	1	58	Adults ($M=20$ y)	12 trials per condition, 48 in total	Quiet, changing-state speech either neutral, positive or negative	Trial-by-trial variation	blocked	yes for negative speech
	2	78	Adults ($M=20$ y)	12 trials per condition, 48 in total	Quiet, changing-state speech either neutral, positive or negative	Trial-by-trial variation	blocked	yes for negative speech

Curriculum Vitae

Larissa Leist
Erwin-Schrödinger-Straße 57
67663 Kaiserslautern
lleist@rptu.de

Wissenschaftlicher Werdegang

Seit Februar 2018	Promotion RPTU Kaiserslautern-Landau
Okt. 2014 – Dez. 2017	Masterstudium (M. Sc.) Cognitive Science Technische Universität Kaiserslautern
Sep. 2011 – Sep. 2014	Bachelorstudium (B. A.) Integrative Sozialwissenschaft Technische Universität Kaiserslautern

Publications

Journal Publications and Proceedings

Leist, L., Lachmann, T., & Klatte, M. (2024). Impact of Irrelevant Speech and Non-speech Sounds on Serial Recall of Verbal and Spatial Items in Children and Adults (under revision).

Leist, L., Belke, C., Kuhlmann, J., Benz, S., Haba, A.-I., Schreckenberger, D., Spilski, J., Lachmann, T., & Klatte, M. (2024). The role of noise as part of neighborhood quality and its impact on well-being in German preschoolers: Results from the Equal-Life project.

Leist, L., Lachmann, T., Schlittmeier, S. J., Georgi, M., & Klatte, M. (2023). Irrelevant speech impairs serial recall of verbal but not spatial items in children and adults. *Memory & Cognition*, 51(2), 307-320. <https://doi.org/10.3758/s13421-022-01359-2>

Leist, L., Breuer, C., Yadav, M., Fremerey, S., Fels, J., Raake, A., ... & Klatte, M. (2022). Differential Effects of Task-Irrelevant Monaural and Binaural Classroom Scenarios on Children's and Adults' Speech Perception, Listening Comprehension, and Visual-Verbal Short-Term Memory. *International Journal of Environmental Research and Public Health*, 19(23), 15998. <https://doi.org/10.3390/ijerph192315998>

Leist, L., Reimers, C., Fremerey, S., Fels, J., Raake, A., & Klatte, M. (2023, February). Effects of binaural classroom noise scenarios on primary school children's speech perception and listening comprehension. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 265, No. 2, pp. 5199-5205). Institute of Noise Control Engineering.

Leist, L., Lachmann, T., & Klatte, M. (2021). Differential Effects of Irrelevant Speech and Environmental Sounds on Short-Term Memory in Children and Adults. 13th ICBEN Congress on Noise as a Public Health Problem. Stockholm, Sweden.

Breuer, C., Schmitt, R., Leist, L., Fremery, S., Raake, A., Klatte, M., & Fels, J. (2024). Exploring the Impact of Realistic Classroom Noise on Auditory Selective Attention. DAGA 2024 - 50. Jahrestagung für Akustik, DAGA 2024, 2024-03-18 - 2024-03-21, Hannover, Germany.

Fremerey, S., Breuer, C., Leist, L., Klatte, M., Fels, J., & Raake, A. (2024). AVT-ECoClass-VR: An open-source audiovisual 360° video and immersive CGI multi-talker dataset to evaluate cognitive performance. 2024 16th International Conference on Quality of Multimedia Experience (QoMEX).

Yadav, M., Georgi, M., Leist, L., Klatte, M., Schlittmeier, S. J., & Fels, J. (2023). Cognitive performance in open-plan office acoustic simulations: Effects of room acoustics and semantics but not spatial separation of sound sources. <https://doi.org/10.48550/arXiv.2306.08051>

Georgi, M., Leist, L., Klatte, M., & Schlittmeier, S. J. (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>

- Breuer, C., Loh, K., Leist, L., Fremerey, S., Raake, A., Klatte, M., & Fels, J. (2022). Examining the Auditory Selective Attention Switch in a Child-Suited Virtual Reality Classroom Environment. *International Journal of Environmental Research and Public Health*, 19 (24), 16569. MDPI AG. Retrieved from <https://doi.org/10.3390/ijerph192416569>
- Waye, K. P., Löve, J., Lercher, P., Dzhambov, A. M., Klatte, M., Schreckenberg, D., ... & Equal-life Scientific Team. (2023). Adopting a child perspective for exposome research on mental health and cognitive development-Conceptualisation and opportunities. *Environmental Research*, 239, 117279. <https://doi.org/10.1016/j.envres.2023.117279>
- Dzhambov, A. M., Lercher, P., Vincens, N., Waye, K. P., Klatte, M., Leist, L., Lachmann, T., ... & van Kamp, I. (2023). Protective effect of restorative possibilities on cognitive function and mental health in children and adolescents: A scoping review including the role of physical activity. *Environmental Research*, 116452.
- Breuer, C., Leist, L., Fremerey, S., Raake, A., Klatte, M., & Fels, J. (2023). Evaluating Cognitive Performance in Classroom Scenarios using Audiovisual Virtual Reality. Proceedings of the 1st AUDICTIVE Conference. Aachen, Germany.
- Reimers, C., Fremerey, S., Leist, L., Klatte, M., Raake, A., & Fels, J. (2023). Investigating the Influence of Visual Distractors on the Voluntary Switching of Auditory Selective Attention in Virtual Reality. 49. Jahrestagung für Akustik DAGA. Hamburg, Germany.
- Jelelati, A., Leist, L., Lachmann, T., & Klatte, M. (2023). Effects of irrelevant speech on immediate serial recall: The role of phonological complexity. 49. Jahrestagung für Akustik DAGA. Hamburg, Germany.
- Yadav, M., Georgi, M., Leist, L., Klatte, M., Schlittmeier, S.J., & Fels, J. (2023). Investigating the effect of realistic sound absorption variations in open-plan office simulations on short-term memory performance. 10th Convention of the European Acoustics Association. Turin, Italy.
- Breuer, C., Leist, L., Fremerey, S., Raake, A., Klatte, M., & Fels, J. (2023). Towards investigating listening comprehension in virtual reality. 10th Convention of the European Acoustics Association. Turin, Italy.
- Reimers, C., Loh, K., Leist, L., Fremerey, S., Raake, A., Klatte, M., & Fels, J. (2022). Investigating Different Cueing Methods for Auditory Selective Attention in Virtual Reality. 48. Jahrestagung für Akustik DAGA. Stuttgart, Germany.
- Belke, C., Benz, S., Kuhlmann, J., Leist, L., Klatte, M., & Schreckenberg, D. (2023). Walking interviews as a qualitative method of assessing environmental noise perception of preschool children. 14th ICBEN, Belgrade, Serbia.
- Fremerey, S., Reimers, C., Leist, L., Spilski, J., Klatte, M., Fels, J., & Raake, A. (2021). Generation of audiovisual immersive virtual environments to evaluate cognitive performance in classroom type scenarios. 47. Deutsche Jahrestagung für Akustik. Wien, Austria.

Acknowledgements

An dieser Stelle möchte ich meinen herzlichen Dank an all jene Menschen richten, die mich während meiner Promotionszeit unterstützt haben.

Besonderer Dank gilt Maria Klatte und Daniela Czernochowski. Maria, über die Jahre hast du mir stets dein Vertrauen geschenkt und mich mit deinem wertvollen Feedback begleitet. Ohne dich wären die zahlreichen Projekte, aus denen ich so viel lernen konnte, nicht möglich gewesen.

Daniela, du hast mir bereits während meines Masterstudiums Vertrauen in meine wissenschaftliche Arbeit entgegengebracht und mich ermutigt, mit der Promotion zu beginnen. Ich schätze es sehr, Teil deiner Arbeitsgruppe gewesen zu sein und dass du immer für Fragen, Diskussionen und wertvolle Ratschläge zur Verfügung standest.

Ein großes Dankeschön geht auch an Thomas Lachmann, der mir die Gelegenheit gab, meine vielen Ideen in die Realität umzusetzen, und für die Unterstützung in den vergangenen Jahren.

Ein besonderer Dank gebührt Ann-Kathrin Beck, die mir über die Jahre hinweg stets ein offenes Ohr geschenkt und mich mit ermutigenden Worten begleitet hat. Ich freue mich auf das nächste Kapitel. Ein großes Dankeschön geht auch an Kirstin Bergström, die meine Leidenschaft für diese Karriere geweckt hat und von der ich so viel lernen durfte.

Mein Dank gilt auch André Haese, Petra Ludowicy und Radha Meghanathan, die mir besonders zu Beginn der Promotion zur Seite standen.

Ein herzliches Dankeschön an Carolin Breuer, Arzu Arat und Sonja Jeram. Es hat mir große Freude bereitet, mit euch an spannenden Projekten zu arbeiten, inspirierende Gespräche zu führen und an zahlreichen Konferenzen teilzunehmen.

Ohne die zahlreichen Gespräche und die stetige Unterstützung meiner Freunde wäre diese Promotion nicht möglich gewesen. Ein besonderer Dank geht an Markus und Flo für einfach alles! Auch Lisa und Divi möchte ich für die schöne Zeit neben der Arbeit danken. Ein herzliches Dankeschön an all die anderen Freunde und Unterstützer, die nicht namentlich erwähnt wurden, für eure Motivation, eure ermutigenden Worte und euer Verständnis in allen Lebenslagen.

Declaration

Hiermit versichere ich,

- dass ich die vorgelegte Arbeit selbst angefertigt und alle benutzten Hilfsmittel in der Arbeit angegeben habe,
- dass ich diese Dissertation nicht schon als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, und
- dass weder die gleiche noch eine andere Abhandlung der Dissertation bei einer anderen Universität oder einem anderen Fachbereich der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau veröffentlicht wurde.

Kaiserslautern, den 28. Oktober 2024.

Larissa Leist