Journal of Experimental Psychology: Learning, Memory, and Cognition

Changing-State Irrelevant Speech Disrupts Visual–Verbal but Not Visual–Spatial Serial Recall

John E. Marsh, Mark J. Hurlstone, Alexandre Marois, Linden J. Ball, Stuart B. Moore, François Vachon, Sabine J. Schlittmeier, Jan Philipp Röer, Axel Buchner, Frederik Aust, and Raoul Bell

Online First Publication, June 24, 2024. https://dx.doi.org/10.1037/xlm0001360

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001360

Changing-State Irrelevant Speech Disrupts Visual–Verbal but Not Visual–Spatial Serial Recall

John E. Marsh[1, 2], Mark J. Hurlstone[3], Alexandre Marois[4], Linden J. Ball[1], Stuart B. Moore[5], François Vachon[4], Sabine J. Schlittmeier[6], Jan Philipp Röer[7, 8], Axel Buchner[8], Frederik Aust[9, 10], and Raoul Bell[8]

[1] School of Psychology and Humanities, University of Central Lancashire
[2] Department of Health, Learning and Technology, Luleå University of Technology
[3] Department of Psychology, Lancaster University
[4] École de Psychologie, Université Laval
[5] School of Psychology, Keele University
[6] Work and Engineering Psychology, Institute of Psychology, RWTH Aachen University
[7] Department of Psychology and Psychotherapy, Witten/Herdecke University
[8] Department of Experimental Psychology, Heinrich Heine University Düsseldorf
[9] Department of Psychology, University of Cologne
[10] Department of Psychology, University of Amsterdam

In an influential article, Jones et al. (1995) provide evidence that auditory distraction by changing relative to repetitive auditory distracters (the changing-state effect) did not differ between a visual–verbal and visual–spatial serial recall task, providing evidence for an amodal mechanism for the representation of serial order in short-term memory that transcends modalities. This finding has been highly influential for theories of short-term memory and auditory distraction. However, evidence vis-à-vis the robustness of this result is sorely lacking. Here, two high-powered replications of Jones et al.'s (1995) crucial Experiment 4 were undertaken. In the first partial replication ($n = 64$), a fully within-participants design was adopted, wherein participants undertook both the visual–verbal and visual–spatial serial recall tasks under different irrelevant sound conditions, without a retention period. The second near-identical replication ($n = 128$), incorporated a retention period and implemented the task-modality manipulation as a between-participants factor, as per the original Jones et al. (1995; Experiment 4) study. In both experiments, the changing-state effect was observed for visual–verbal serial recall but not for visual–spatial serial recall. The results are consistent with modular and interference-based accounts of distraction and challenge some aspects of functional equivalence accounts.

Keywords: auditory distraction, functional equivalence, modularity, serial order, short-term memory

Erik M. Altmann served as action editor.

John E. Marsh https://orcid.org/0000-0002-9494-1287
Linden J. Ball https://orcid.org/0000-0002-5099-0124

John E. Marsh's contribution to this article was supported by grants received from the Bial Foundation (Grant 201/20) and the German Academic Exchange Service (Deutscher Akademischer Austauschdienst [Grant 91875827]).

Open Access funding provided by the University of Central Lancashire:

John E. Marsh served as lead for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, validation, writing–original draft, and writing–review and editing, contributed equally to supervision, and served in a supporting role for software. Mark J. Hurlstone served as lead for software and contributed to the formal analysis, investigation, methodology, writing–original draft, and writing–review and editing. Alexandre Marois served in a supporting role for formal analysis and writing–review and editing. Linden J. Ball served in a supporting role for resources and writing–review and editing. Stuart B. Moore served in a supporting role for data curation, formal analysis, and visualization. François Vachon served in a supporting role for writing–review and editing. Sabine J. Schlittmeier served in a supporting role for writing–review and editing. Jan Philipp Röer served in a supporting role for data curation and writing–review and editing. Axel Buchner served in a supporting role for formal analysis, software and writing–review and editing. Frederik Aust served in a supporting role for formal analysis, software, and writing–review and editing. Raoul Bell contributed equally to methodology and served in a supporting role for data curation and writing–review and editing. Jan Philipp Röer and Raoul Bell contributed equally to investigation.

Correspondence concerning this article should be addressed to John E. Marsh, School of Psychology and Humanities, University of Central Lancashire, Darwin Building, Marsh Lane, Preston, Lancashire, PR1 2HE, United Kingdom. Email: jemarsh@uclan.ac.uk

The notion of modularity within working memory (WM) is pervasive, perhaps in part because of its intuitive appeal—there can be little dispute that the effector systems involved in fulfilling goal-directed-behavior in verbal and spatial tasks are distinct (e.g., Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006). For example, the nature of rehearsal closely resembles the stimulus input—inner speech for sequential verbal information and eye movements for sequential visual–spatial information (Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006; but see Awh & Jonides, 2001). It seems plausible that a task's susceptibility to disruption by secondary, or task-irrelevant information should be dictated, at least in part, by the different processes required to rehearse verbal and spatial material. In an influential study, Jones et al. (1995) compared the disruptive impact of active and passive secondary tasks on visual–verbal and visual–spatial tests of serial short-term memory. They reported that active secondary tasks involving spatial processing (manual spatial tapping) and verbal processing (articulatory suppression) disrupted visual–verbal and visual–spatial serial recall alike if they involved changing sequences of actions, as compared with a single repeating action. Furthermore, and of direct relevance to the current article, Jones et al. (1995) reported that passive exposure to irrelevant speech comprising changing verbal items produced more disruption than one repeatedly presented verbal item for both visual–verbal and visual–spatial serial recall.

The study of Jones et al. (1995) has been very influential. It has been used to argue against the modular nature of WM, and it has helped shape and constrain theories of auditory distraction (Jones et al., 1996; Jones & Tremblay, 2000; Neath, 2000; Norris et al., 2004). However, more recent studies have called into question the results of Jones et al. (1995) regarding the effects of actively performed secondary tasks (e.g., manual spatial tapping and articulatory suppression) on visual–verbal and visual–spatial short-term memory (e.g., Guérard & Tremblay, 2008; Guitard & Saint-Aubin, 2015; Meiser & Klauer, 1999). However, to date, there has been little attempt to examine the reproducibility of the Jones et al. (1995) finding that visual–spatial serial recall is disrupted to the same degree as visual–verbal serial recall by irrelevant speech containing auditory changes (but see Tremblay et al., 2001). The purpose of the present study is to examine the robustness of this central finding by providing two high-powered replications of the original experiment from which it was derived.

Modularity of Memory

On the classic WM model (see Baddeley, 1986, 2000; Baddeley & Hitch, 1974), modularity is assumed because of the purported existence of separate short-term memory subsystems for storing verbal and visual–spatial information—the phonological loop and the visuospatial sketchpad, respectively. The phonological loop comprises two subcomponents—a phonological store and an articulatory control process (Baddeley, 1986). The phonological store holds speech-based, phonological, representations of verbal items that are subject to loss because of decay. The articulatory control process—which is analogous to subvocal speech—can be used to rehearse and refresh the contents of the store, in order to offset this decay process. For auditory input, entry into the phonological store is automatic and obligatory, whereas for visual input, entry depends on visual–verbal information being converted into phonological form via the articulatory control process. The articulatory control process can be impeded by articulatory suppression, which involves the repeated utterance of an irrelevant verbal token (e.g., "the," "the," "the"). This obstructs the articulatory control process from being used to rehearse the decay-prone contents of the phonological store (it also prevents visual–verbal input from being converted into the phonological form necessary to access the phonological store). Serial recall performance is disrupted by the concurrent or subsequent presentation of irrelevant speech (Colle, 1980; Jones, 1993; Jones & Macken, 1993). On the WM model, this arises because the sound is thought to gain obligatory access to the phonological store, wherein it disrupts the store's contents (Salamé & Baddeley, 1982).

The mechanism within the WM model concerned with the storage and manipulation of visual–spatial information is the visual–spatial sketchpad (Baddeley & Hitch, 1974). Although this component stores both visual and spatial information, it has been suggested that there are distinct subcomponents for each type of input. According to Logie (1995), visual information (e.g., color and form) is stored in a visual cache that, like the phonological store, is prone to decay, whereas spatial information is processed by an inner scribe that can also be used to rehearse information held in the visual cache to offset the decay process. The visual–spatial sketchpad can be disrupted by manual spatial tapping (tapping a sequence of different keys), which is thought to impede the operation of the inner scribe component (Logie, 1995). There are two dominant competing views on how spatial information is rehearsed in short-term memory—one assumes that rehearsal involves covert shifts of spatial selective attention (e.g., Awh et al., 2006; Awh & Jonides, 2001; Postle et al., 2003), whereas the other assumes that rehearsal involves eye movements (e.g., Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006).

The WM model predicts that because verbal and spatial information are stored in separate bespoke short-term memory subsystems, the degree of interference observed on verbal and spatial short-term memory primary tasks from active secondary activities (e.g., manual spatial tapping and articulatory suppression) or passive exposure to to-be-ignored material (e.g., irrelevant speech) should depend on whether they draw on the same or different subsystems. This assumption has been corroborated by double dissociations observed under dual-task conditions in which secondary tasks that are thought to draw on verbal short-term memory resources produce greater interference on verbal than spatial short-term memory primary tasks, whereas the converse pattern of interference is observed with secondary tasks that are thought to draw on spatial short-term memory resources (e.g., Farmer et al., 1986; Guérard & Tremblay, 2008; Lange, 2005; Logie et al., 1990; Meiser & Klauer, 1999). The assumption of separate subsystems for verbal and spatial information also receives support from double dissociations observed under neuroimaging conditions (Awh et al., 1996; Smith et al., 1996; Smith & Jonides, 1997), between neuropsychological case studies (De Renzi & Nichelli, 1975; Hanley et al., 1991; Vallar & Baddeley, 1984), and different patient groups (Wang & Bellugi, 1994).

However, the assumption of separate subsystems for verbal and spatial information has been challenged based on findings indicating functional equivalence between verbal and spatial short-term memory (e.g., Jones et al., 1995, 1996).

Functional Equivalence

The notion of functional equivalence between codes implies the existence of an "amodal" mechanism for the representation of serial order in short-term memory that transcends domains and modalities (e.g., verbal, visual, spatial; see, e.g., Jones et al., 1995, 1996). Evidence for such a mechanism has been obtained from studies showing that order recall across different domains and modalities exhibits similar general characteristics. For example, order recall of verbal and visual–spatial materials exhibits similar accuracy serial position curves (Avons, 2007; Cortis et al., 2015; Farrand et al., 2001; Guérard & Tremblay, 2008; Jones et al., 1995; Smyth et al., 2005; Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006; Ward et al., 2005); latency serial position curves (Hurlstone & Hitch, 2015, 2018; Parmentier et al., 2005, 2006); effects of sequence length (Jones et al., 1995; Smyth et al., 2005; Smyth & Scholey, 1996); distributions of item and order errors (Guérard & Tremblay, 2008); transposition gradients and latency-displacement functions (Hurlstone & Hitch, 2015, 2018); effects of temporal grouping (Hurlstone, 2019; Hurlstone & Hitch, 2015, 2018; Parmentier et al., 2006); and effects of Hebb repetition learning (Couture & Tremblay, 2006; Horton et al., 2008). Additionally, there is some evidence to suggest that verbal and visual–spatial serial recall may rely on a unitary memory system. For example, secondary tasks that require order memory—as compared to item memory—produce more disruption to tasks that also require serial order (Depoorter & Vandierendonck, 2009), even if presented in different modalities (Vandierendonck, 2016). Taken together, this evidence suggests that the same mechanism may be involved in the representation of serial order across different domains and modalities. Moreover, the evidence suggests that verbal and visual–spatial short-term memory may rely on a common resource.

In their influential article, Jones et al. (1995) reported a series of findings that cast doubt on the modularity of WM. In the critical experiments of their article, the authors compared visual–verbal with visual–spatial serial recall tasks that were equated in terms of the number of stimuli presented and the presentation and recall procedures. In the visual–verbal task, participants studied sequences containing random orderings of seven letters (F, K, L, M, Q, R, Y) presented one at a time on a computer display. In the recall phase, the letters were presented in a jumbled vertical array and participants had to reconstruct the order of the sequence by mouse-clicking on the letters in turn. In the visual–spatial task, participants studied sequences containing dots presented one at a time in random locations on a computer display. In the recall phase, the dots were presented in their original locations and participants had to reconstruct the order of the sequence by mouse-clicking on the locations in turn.

Jones et al. (1995) investigated the effects of verbal (articulatory suppression and irrelevant speech) and spatial (manual spatial tapping) distractors/activities on visual–verbal and visual–spatial serial recall performance. As we have seen, the WM model predicts that interference between primary and secondary tasks should occur to the extent that they engage the same putative short-term memory sub-system. Accordingly, articulatory suppression and irrelevant speech should disrupt visual–verbal serial recall performance more so than visual–spatial serial recall performance, whereas the converse should be true for manual spatial tapping.

However, Jones et al. (1995) sought to contrast this WM prediction with a competing one based on a unitary memory system. In their earlier work on the irrelevant speech effect, Jones and colleagues (Jones & Macken, 1993, 1995; Macken & Jones, 1995) had shown that the degree of disruption of visual–verbal serial recall performance by to-be-ignored irrelevant speech is based on the principle of changing state. Specifically, when the background speech contains sounds that change from one item to the next (e.g., "a," "b," "c," "d"; a so-called changing-state sequence), the degree of disruption of visual–verbal serial recall performance is stronger than when the background speech contains the same repeating sound (e.g., "b," "b," "b," "b"; a so-called steady-state sequence). In later work, Jones and colleagues showed that the disruption of visual–verbal serial recall performance by articulatory suppression also exhibits this so-called changing-state effect (Macken & Jones, 1995). This led Jones et al. (1995) to speculate that the changing-state effect might be a general feature of interference in serial short-term memory. In turn, this yielded the prediction that disruption of visual–verbal and visual–spatial serial recall should be based not on the verbal or spatial content of the primary and secondary task—and therefore the extent to which they draw on the same putative short-term memory sub-system—but rather the degree to which the secondary task conforms to the principle of changing state.

In three experiments, Jones et al. (1995; Experiments 2–4) found that the degree of disruption of visual–verbal and visual–spatial serial recall was consistent with their competing view. Accordingly, in Experiment 2 changing-state manual spatial tapping (tapping a sequence of 12 keys) was more disruptive than steady-state manual spatial tapping (tapping a single key repeatedly); in Experiment 3, changing-state articulatory suppression (mouthing the alphabetic sequence "a," "b," "c," "d," "e," "f," and "g") was more disruptive than steady-state articulatory suppression (mouthing the syllable "bee" repeatedly); whilst in Experiment 4 changing-state irrelevant speech (a voice speaking the alphabetic sequence "a," "b," "c," "d," "e," "f," and "g") was more disruptive than steady-state irrelevant speech (a voice speaking the syllable "ah" repeatedly). However, critically, the magnitude of the changing-state effect in each of the three experiments was the same for both the visual–verbal and visual–spatial tasks. This pattern of cross-modal interference is clearly at variance with the expectations under the WM model. In light of these findings, Jones et al. (1995) concluded that a common order mechanism exists across different types of items. This assumption has been embodied in their object-oriented episodic record (O-OER) model (Jones, 1993; Jones et al., 1996), which assumes the existence of a unitary memory system served by a common serial ordering mechanism. According to the model, streams of items within memory are represented on an episodic surface (blackboard) common to materials from different sensory origins. Interference occurs between any to-be-remembered and to-be-ignored materials (or streams) providing both comprise a serial order element (i.e., they conform to the principle of changing state).

The results of Jones et al. (1995) observed using actively performed verbal and spatial secondary tasks (viz., articulatory suppression and manual spatial tapping) have been the subject of numerous failed replication attempts. For example, in two studies (Guérard & Tremblay, 2008; Meiser & Klauer, 1999) changing-state manual spatial tapping disrupted visual–spatial serial recall, whereas changing-state articulatory suppression did not. By contrast, changing-state manual spatial tapping and articulatory suppression both disrupted visual–verbal serial recall, but the degree of disruption was smaller with the former than with the latter secondary task. In a close replication of the original

Jones et al. (1995) experiments, Guitard and Saint-Aubin (2015) witnessed cross-modal interference effects of changing-state manual spatial tapping and articulatory suppression on visual–verbal and visual–spatial serial recall but these effects were weaker than when the primary and secondary tasks originated from the same modality. Finally, Alloway et al. (2010) observed no cross-modal changing-state effects in their experiments—changing-state manual spatial tapping disrupted visual–spatial serial recall but not visual–verbal serial recall, whereas the converse was true with respect to changing-state articulatory suppression.

These results with actively performed secondary tasks are clearly at odds with the pattern of findings from Jones et al. (1995). However, the cross-modal changing-state interference of passively heard to-be-ignored irrelevant speech observed in that study has been replicated by Tremblay et al. (2001) who found that changing-state broadband noise was more disruptive than steady-state broadband noise on both visual–verbal and visual–spatial serial recall. Moreover, the degree of disruption of visual–verbal and visual–spatial serial recall by changing-state broadband noise was roughly comparable in magnitude. More recently, in a partial replication, Kvetnaya (2018) repeated the spatial condition of Jones et al. (1995; Experiment 4) and failed to observe a reliable changing-state interference effect, but a reliable changing-state interference effect was observed in a second partial replication wherein the effect of irrelevant sound was examined on the verbal condition of Jones et al. (1995; Experiment 4). On the face of it, the results from the partial replications of Kvetnaya (Kvetnaya, 2018; Kvetnaya et al., 2019) are inconsistent with the notion of functional equivalence of Jones et al. (1995) and are at the same time consistent with the notion of modularity of WM. These recent findings underscore the need to address the replicability of the findings of Jones et al. (1995; Experiment 4).

Current Study

Given that the findings of Jones et al. (1995; Experiment 4) have been so influential that central assumptions of current theories of auditory distraction (Jones et al., 1996; Norris et al., 2004) build upon them, it is important to evaluate their robustness. Accordingly, in this article, we sought to test the reproducibility of this canonical finding through two high-powered replications. One experiment was a partial replication of the original Jones et al. (1995) Experiment 4 that involved some minor changes to the design and procedure (Experiment 1), whereas the second was a near-identical replication that was faithful to the design and procedure of the original experiment (Experiment 2). To foreshadow the main results, across both experiments, we observed a changing-state effect of irrelevant speech in the verbal domain but not in the spatial domain, consistent with the WM model but at odds with a unitary account of short-term memory. These results were corroborated by a Bayesian meta-analysis of the results of our experiments. In the General Discussion, we consider the implications of our findings for the WM and O-OER accounts and then entertain other prominent theories of auditory distraction and short-term memory.

Experiment 1: Partial Replication

The first experiment was a partial replication of Experiment 4 of Jones et al. (1995). Their original experiment used a mixed design, whereby irrelevant sound condition (quiet vs. steady state vs. changing state; where "quiet" was a no-sound control condition) was a within-participants factor, whereas task modality (visual–verbal vs. visual–spatial) was a between-participants factor. In the original study, there were 36 participants in total, 18 participants per task-modality condition. In the present experiment, we increased the total sample size from 36 to 64, almost doubling the sample size of the original experiment. We also chose to deploy a fully within-participants design, as it is both a more economical and a more powerful option. We made one additional change to the experimental protocol of the original experiment. Specifically, Jones et al. (1995) included a 10-s retention period from the offset of the last study item to the onset of the recall phase. In the conditions involving to-be-ignored irrelevant sound, the sound was delivered during both the encoding phase and retention period of the serial reconstruction tasks. Here, we opted to remove the 10-s retention period, as a long retention period has been observed to reduce auditory distraction in previous studies (Elliott & Cowan, 2005; Körner et al., 2019).

In its original formulation, the WM model predicts that steady state and changing-state irrelevant speech should disrupt visual–verbal, but not visual–spatial, serial recall compared to a quiet control condition. Thus, any interference effects should be confined to the visual–verbal serial recall task. The model does not predict a changing-state effect, as it contains no explicit mechanism for accounting for this finding. An ad hoc explanation could be that changing-state sequences are more disruptive of visual–verbal serial recall than steady-state sequences because changing sequences contain more phonological variation, thus producing greater interference with the phonological representations of to-be-remembered items in the phonological store. In this revised account, the WM model predicts an interaction between sound condition (steady state vs. changing state) and task modality (visual–verbal vs. visual–spatial). By contrast, the functional equivalence account based on the O-OER model (Jones et al., 1996) predicts that the changing-state effect should be observed in both tasks.[1] Accordingly, the functional equivalence view predicts a main effect of sound condition (steady state vs. changing state), in conjunction with the absence of a sound condition (steady state vs. changing state) by task modality (visual–verbal vs. visual–spatial) interaction.

Method

Participants

Sixty-four (34 female, 30 male, $M_{age} = 21.98$; $SD = 3.78$) participants were recruited from the participant panel at the University of Central Lancashire. All participants reported normal or corrected-to-normal visual acuity and normal hearing.

[1] In the extreme case, the O-OER model predicts that the magnitude of the changing-state effect will be the same in the visual–verbal and visual–spatial modality, thus giving rise to the absence of an interaction between task modality and irrelevant sound, which is what Jones et al. (1995) find in their Experiment 4. However, the model might also predict a two-way interaction whereby the magnitude of the changing-state effect is larger in one modality compared to the other. For simplicity, we attribute to the O-OER model the prediction that there should be no two-way interaction between modality and irrelevant sound as this was the pattern observed across all four experiments in Jones et al. (1995).

The original Jones et al. (1995) Experiment 4 had a sample size of $N = 36$ ($n = 18$ per task-modality condition). Given $\alpha = \beta = .05$, they were only able to detect differences of size $f = 0.62$ between the visual–spatial and the visual–verbal task conditions in the changing-state effect. Thus, the sensitivity was very low in their experiment, enabling them to detect the critical effect only if it were much larger than what Cohen (1988) defined as "large" ($f = 0.4$). We took two steps to arrive at a larger sensitivity for the present experiment. First, we used a within-participant manipulation of the modality of the task instead of a between-participants manipulation. Second, we made sure we had a larger sample. A sensitivity analysis showed that with $N = 64$, $\alpha = \beta = .05$, and a correlation of the changing-state effect between the visual–spatial and the visual–verbal condition of $\rho = 0.3$, it is possible to detect effects of size $f = 0.27$ (close to what Cohen, 1988, defined as a "medium" effect). Note that assuming $\rho = 0.3$ can be considered conservative, in that the correlation among the levels of the repeated measures variable might well be larger. In that case, the sensitivity would be even higher than what we report for the current set of assumptions. The sensitivity analyses were performed using G*Power (Faul et al., 2007).

Short-Term Memory Tasks

The visual–verbal and visual–spatial serial recall tasks were executed on a personal computer running an E-Prime (Psychology Software Tools, Pittsburgh, Pennsylvania, United States) program that controlled stimulus presentation and collected all responses. The E-Prime programs for Experiment 1 are available from https://doi.org/10.17030/uclan.data.00000321 (Marsh et al., 2023).

Visual–Verbal Serial Recall Task. The stimuli for the visual–verbal task were sequences containing random orderings of the seven letters, F, K, L, M, Q, R, and Y. Each letter was presented visually in the central screen position in black 30-point Geneva-bold uppercase font on a white background. Participants initiated each trial by mouse-clicking on a "begin trial" button located in the central screen position after which a sequence of letters was presented. Each letter was displayed for 1 s followed by a 1 s blank delay (Figure 1). Immediately after the presentation of the sequence, all seven letters simultaneously appeared on screen, each within its box, organized horizontally from left to right in a jumbled order. Participants were required to click on the letters in their original presentation order using a mouse-driven pointer. Once a letter had been clicked on, its shade changed to denote that it had been selected. A selected letter could not be de-selected or re-selected again, and all seven letters had to be selected before progressing to the next trial. Therefore, repetition and omission errors were not permitted.

Visual–Spatial Serial Recall Task. The stimuli for the visual–spatial task were sequences comprising seven black dots, each with a diameter of 0.81 cm, presented in quasi-random positions within a 16.5 × 16.5 cm invisible white matrix encased by a black border. The coordinates for the dots were randomly generated subject to the constraint that the centers of any sequential pair of dots were separated by at least 2.86 cm on either axis of the matrix. The maximum distance between any two sequential dots was 10.07 cm. Furthermore, none of the dots appeared within 2.86 cm of the center of the presentation screen. Participants initiated each trial by mouse-clicking on a "begin trial" button located in the central screen position after which a sequence of dots was presented. Each dot was displayed for 1 s followed by a 1 s blank delay (Figure 1). Immediately after the presentation of the sequence, all seven dots simultaneously reappeared on the screen in their original spatial locations. Participants were required to click on the dots in their original presentation order using a mouse-driven pointer. Once a dot had been clicked on, its shade changed to denote that it had been selected. Once selected, it could not be de-selected or re-selected again, and all seven dots had to be selected before progressing to the next trial. Thus, as per the visual–verbal serial recall task, repetition and omission errors were not permitted.

Irrelevant Sounds

The syllable "Ah" and the letters "A" through "G" were digitally recorded in an even-pitched voice in 16-bit resolution at a 44.1 kHz sampling rate using Sony Sound Forge 8.0 software. Each spoken letter, recorded by a British male, was edited to 250 ms and concatenated into a sequence, wherein there was a 250 ms temporal gap between each item. Irrelevant sequences were thus spoken at a rate of two items per second. Steady-state sequences comprised repetitions of the syllable "Ah," whereas changing-state sequences consisted of the looped letter sequence "A" through "G," but for a given changing-state sound trial the starting point of the sequence was random. The onset of the sequences was contemporaneous with the onset of the first to-be-remembered letter/dot and offset at the onset of the response phase. Thus, the auditory sequences were presented only during the encoding of the visual sequence. The auditory stimuli were presented at approximately 60 dB(A) via over-the-ear headphones (Sennheiser HD-202) that the participants wore throughout the study.

Design and Procedure

The experiment employed a 3 (sound condition: quiet vs. steady state vs. changing state) × 2 (task modality: visual–verbal vs. visual–spatial) × 7 (serial position: 1–7) within-participants design.

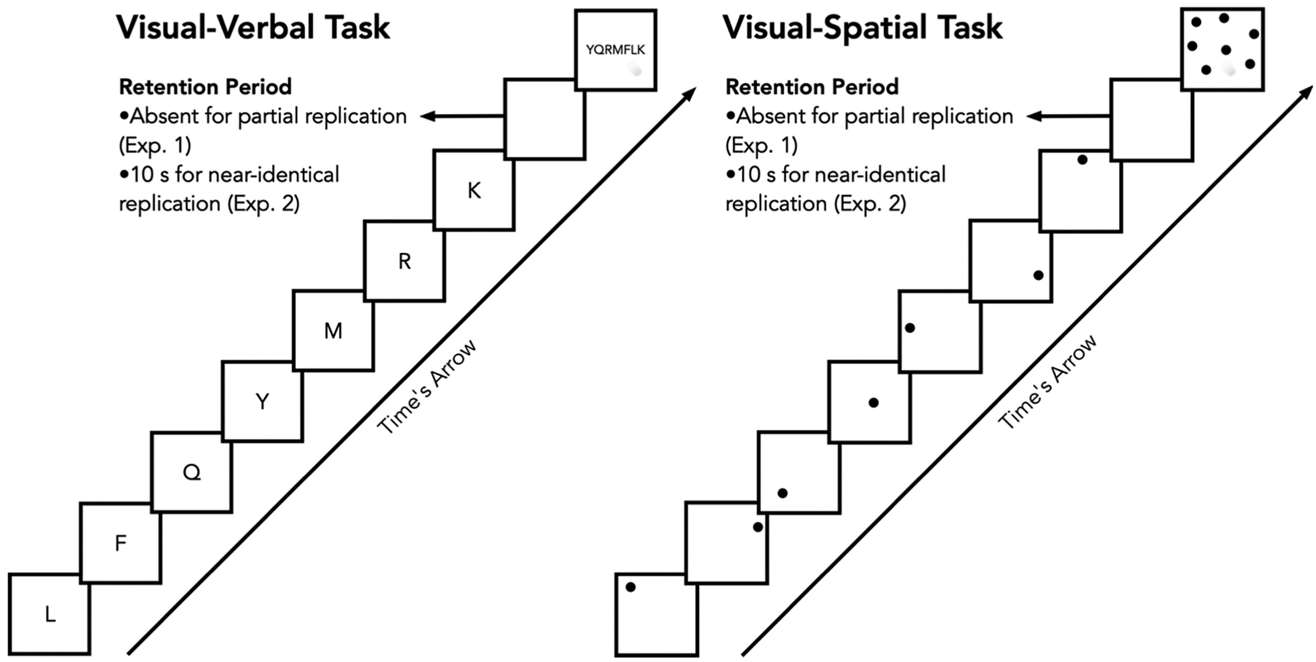
Participants read standardized instructions that told them to recall the order of presentation of the seven letters in the visual–verbal task or dots in the visual–spatial task. Participants knew that once a letter/dot had been selected, they would be unable to alter their response. They were informed that sounds would be presented over their headphones but that they were irrelevant to the recall task and that they should ignore them as best as they could. The experiment contained two blocks of trials, one for the visual–verbal serial recall task and one for the visual–spatial serial recall task. Each block began with three practice trials (one quiet, one steady state, and one changing state) prior to participants receiving a block of 48 experimental trials (16 quiet, 16 steady state, and 16 changing state). The order of administration of the visual–verbal and visual–spatial serial recall tasks was counterbalanced across participants, whereas the order of the three sound conditions was randomized from trial to trial. A brief, optional pause was offered between tasks. While Experiment 4 of Jones et al. (1995) comprised a "quiet" condition wherein participants were exposed to 50 dB(A) air conditioning noise (prior to attenuation by the headphone cups), Experiment 1 was undertaken in a quiet laboratory with no air-conditioning noise.

Results and Discussion

Data for Experiment 1 are available from https://doi.org/10.17030/uclan.data.00000412 (Marsh et al., 2023). Analyses were undertaken

Figure 1

Schematic of the Trial Structure in the Visual–Verbal (Left) and Visual–Spatial (Right) Serial Recall Tasks Used in the Experiments



Note. The recall phase commenced immediately after the presentation phase in the partial replication (Experiment 1), whereas there was a 10-s retention period separating the presentation and recall phases in the near-identical replication (Experiment 2). The illustration is not to scale. Exp. = experiment.

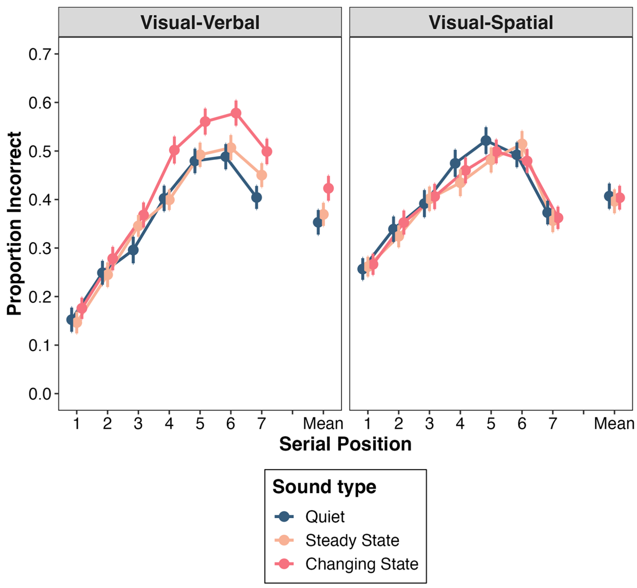
on the mean number of errors as a function of sound condition, task modality, and serial position. The results were scored in accordance with a strict serial recall criterion: an item was only recorded as correct if it was recalled in its original presentation position. The data are shown in Figure 2 from which it can be seen that primacy and recency effects can be observed in the shape of the serial position curves for both the visual–verbal (left panel) and visual–spatial (right panel) serial recall tasks, regardless of sound condition. In addition to reporting conventional analysis of variance (ANOVA, Greenhouse–Geisser-corrected) tests, we also report the Bayes factors (BF_{10}) for each main effect and interaction effect. Analyses were conducted using R (V4.2.0) in RStudio (V2022.02.03) using the packages tidyverse (Wickham et al., 2019), afex (V1.3-0, Singmann et al., 2022), and BayesFactor (V0.9.12-4.6, R. D. Morey & Rouder, 2023). For the Bayesian analyses, we used the model specification recommended by van den Bergh et al. (2022) and default prior settings ($r = .5$ for fixed effects, $r = 1$ for random effects). In frequentist and Bayesian analyses, we performed Type-III model comparisons.

An initial analysis was undertaken to determine whether there was any effect on the order in which the visual–verbal and visual–spatial serial recall tasks were completed. A 3 (sound condition: quiet vs. steady state vs. changing state) × 2 (task modality: visual–verbal vs. visual–spatial) × 7 (serial position: 1–7) × 2 (task order: visual–verbal → visual–spatial vs. visual–spatial → visual–verbal) mixed ANOVA was conducted on the mean number of errors. There was no significant main effect of the task order, $F(1, 62) = 0.79$, $MSE = 226.50$, $p = .379$, $\eta_p^2 = .013$, $BF_{10} = 0.44$, nor any significant two-way interactions with Sound Condition, $F(1.97, 122.20) = 0.14$, $MSE = 8.19$, $p = .869$, $\eta_p^2 = .002$, $BF_{10} = 0.03$, or task modality, $F(1, 62) = 0.24$, $MSE = 52.80$, $p = .627$, $\eta_p^2 = .004$, $BF_{10} = 0.26$. There was a significant interaction between task order and serial position, $F(2.53, 156.55) = 3.16$, $MSE = 15.14$, $p = .034$, $\eta_p^2 = .048$, $BF_{10} = 2.20$, but this did not relate to any of the hypotheses and thus was not considered further. The three-way interaction between sound condition, task modality, and task order was not significant, $F(1.94, 120.51) = 0.33$, $MSE = 12.06$, $p = .716$, $\eta_p^2 = .005$, $BF_{10} = 0.09$. This is important as it demonstrates that there were no transfer effects—that is, undertaking the visual–verbal serial recall task first, did not result in a change in the susceptibility of the visual–spatial serial recall task to disruption via the presence of sound compared to quiet or changing-state compared to steady-state distracters.

Since our initial analysis did not indicate any effect of the order in which the visual–verbal and visual–spatial tasks were completed, we now concentrate on the results of the 3 (sound condition) × 2 (task modality) × 7 (serial position) repeated-measures ANOVA. This ANOVA revealed a significant main effect of sound condition, $F(1.97, 124.12) = 11.51$, $MSE = 8.08$, $p < .001$, $\eta_p^2 = .154$, $BF_{10} = 351.87$, and a significant main effect of serial position, $F(2.52, 158.52) = 180.16$, $MSE = 15.72$, $p < .001$, $\eta_p^2 = .741$, $BF_{10} = 7.24 \times 10^{103}$. However, there was no significant main effect of task modality, $F(1, 63) = 1.38$, $MSE = 52.16$, $p = .244$, $\eta_p^2 = .021$, $BF_{10} = 0.37$. There were significant interactions between sound condition and serial position, $F(8.63, 543.99) = 3.48$, $MSE = 2.95$, $p < .001$, $\eta_p^2 = .052$, $BF_{10} = 5.52$, and between task modality and serial position, $F(4.12, 259.80) = 30.45$, $MSE = 5.36$, $p < .001$, $\eta_p^2 = .326$, $BF_{10} = 3.40 \times 10^{26}$. However, the three-way interaction between sound condition, task modality, and serial position was not significant, $F(8.14, 513.04) = 1.05$, $MSE = 3.50$, $p = .394$, $\eta_p^2 = .016$, $BF_{10} = 0.004$. Thus far, the pattern of data is relatively consistent with that reported by Jones et al. (1995; Experiment 4).

Figure 2

Probability of Serial Recall Error as a Function of Irrelevant Sound Condition for the Visual–Verbal (Left) and Visual–Spatial (Right) Serial Recall Tasks in Experiment 1



Note. Error bars show standard errors of the means. See the online article for the color version of this figure.

However, critically, there was a significant interaction between sound condition and task modality, $F(1.94,\ 122.28) = 5.68$, $MSE = 11.94$, $p = .005$, $\eta_p^2 = .083$, $BF_{10} = 6.39$, which is inconsistent with the results of Jones et al. (1995; Experiment 4).

To test directly the competing predictions of the WM and O-OER models, we removed the quiet condition from the Sound Condition factor—permitting a focused assessment of the changing-state effect—and performed a 2 (sound condition: steady state vs. changing state) × 2 (task modality: visual–verbal vs. visual–spatial) ANOVA. There was a significant main effect of sound condition, $F(1, 63) = 13.40$, $MSE = 1.26$, $p < .001$, $\eta_p^2 = .175$, $BF_{10} = 15.72$, but no significant main effect of task modality, $F(1, 63) = 0.08$, $MSE = 5.61$, $p = .772$, $\eta_p^2 = .001$, $BF_{10} = 0.26$. Critically, the interaction between the two factors was significant, $F(1,\ 63) = 5.45$, $MSE = 1.40$, $p = .023$, $\eta_p^2 = .080$, $BF_{10} = 3.40$. The interaction materialized because the changing-state effect was present in the visual–verbal task, $F(1,\ 63) = 18.82$, $MSE = 1.26$, $p < .001$, $\eta_p^2 = .230$, $BF_{10} = 354.31$, whereas it was absent in the visualspatial task, $F(1, 63) = 0.66$, $MSE = 1.41$, $p = .422$, $\eta_p^2 = .010$, $BF_{10} = 0.25$.

Thus, with a larger sample size and more power than Jones et al. (1995; Experiment 4) owing to the inclusion of modality as a within-participants factor, the changing-state effect was completely absent for visual–spatial as compared to visual–verbal serial recall, consistent with the WM model but at variance with the functional equivalence account based on the O-OER model.

Experiment 2: Near-Identical Replication

Although the results of Experiment 1 are contrary to those reported by Jones et al. (1995; Experiment 4), we have already noted that there were two methodological differences between their experiment and our partial replication that could potentially explain the discrepant findings. First, rather than deploying the task-modality manipulation as a between-participants factor, as Jones et al. (1995) did, we instead opted to deploy this as a within-participants factor to benefit from the increased statistical power this affords. Second, we removed the 10-s retention period included in the original Jones et al. (1995) experiment. Although these changes were made to increase the power to detect a cross-modal changing-state effect and reduce the duration of testing time, it is possible that they had a counterproductive effect. Accordingly, to rule out this possibility, in Experiment 2 we conducted a near-identical replication of their original experiment in which the task-modality manipulation was implemented as a between-participants factor, and the 10-s retention period was reinstated. We replicated Jones et al.'s (1995) Experiment 4 as faithfully as possible but because we used a sample of German participants, we replaced the English with German auditory distracters. The E-prime programs for Experiment 2 are available from: https://doi.org/10.17030/uclan.data.00000321 (Marsh et al., 2023).

Method

Participants, Short-Term Memory Tasks, and Irrelevant Sounds

The methods closely followed those of Experiment 1, with the following exceptions. One hundred and twenty-eight participants (88 females, 40 males; $M_{age} = 23.48$; $SD = 4.28$) were recruited from Heinrich-Heine University and were randomly assigned to either the visual–verbal or visual–spatial serial recall task. Participants reported normal or corrected-to-normal visual acuity and normal hearing. The short-term memory tasks were the same as those used in Experiment 1 except that a 10-s blank retention period was inserted between the offset of the final study item and the onset of the recall phase of the visual–verbal and visual–spatial serial recall tasks (Figure 1). For the irrelevant sounds, the syllable "Ah" and the letters "A" through "G" were spoken and digitally recorded in an even-pitched German male voice in 16-bit resolution at a 44.1 kHz sampling rate using Sony Sound Forge 8.0 software. These German distractor letters are one-syllable letters and phonologically similar to their English counterparts. Each spoken letter was edited in the same manner as for Experiment 1. The sequences were played throughout the presentation phase and retention period of the short-term memory tasks and stopped with the onset of the recall phase. The onset of the auditory sequence co-occurred with the onset of the first to-be-remembered letter/location and offset with the onset of the serial order reconstruction screen. Auditory stimuli were presented at approximately 60 dB(A) via over-the-ear headphones with high-insulation hearing protection covers (Beyerdynamic DT-150) plugged directly into Apple iMac computers with 3.4 GHz Intel Core i5 processors and Radeon Pro 560 (4,096 MB) graphics boards. Visual stimuli were displayed on 21.5-inch thin-film-transistor liquid-crystal display displays. Participants were seated at a distance of approximately 45 cm from the screen. Experiment 2 was undertaken in a quiet laboratory with no air-conditioning noise.

Using G*power (Faul et al., 2007), we performed a sensitivity analysis for the critical test of whether the changing-state effect differs between the visual–spatial and the visual–verbal task. With 64 participants in the visual–spatial condition, 64 participants in the visual–verbal condition, and given $\alpha = \beta = .05$, it was possible to detect differences of size $f = 0.32$ between the visual–spatial and the visual–verbal condition in the changing-state effect. Thus, the sensitivity of the critical comparison in the present experiment is much larger than that of Experiment 4 by Jones et al. (1995) who could only detect huge effects of size $f = 0.62$ (see the Participants section of Experiment 1).

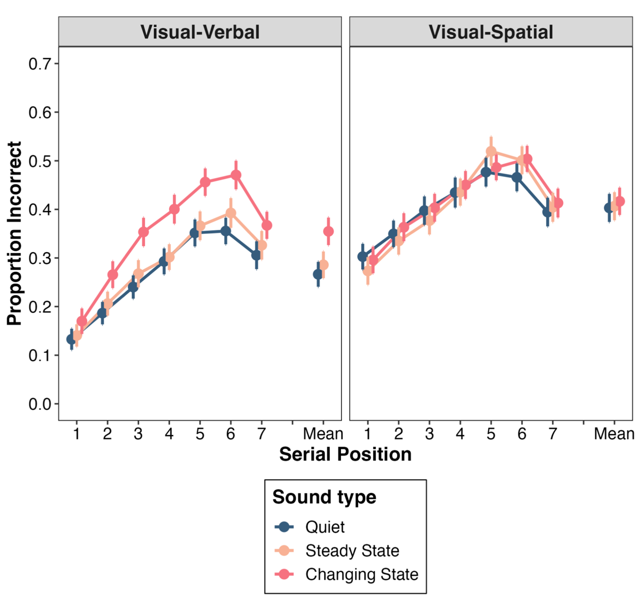
Results and Discussion

Data for Experiment 2 are available from: https://doi.org/10.17030/uclan.data.00000412 (Marsh et al., 2023). Figure 3 shows mean error rates as a function of serial position and sound condition for the visual–verbal serial recall task (left panel) and the visual–spatial serial recall task (right panel). As per Experiment 1, primacy and recency effects are apparent in both the visual–verbal and visual–spatial serial recall tasks, irrespective of sound condition.

A 3 (sound condition: quiet vs. steady state vs. changing state) × 2 (task modality: visual–verbal vs. visual–spatial) × 7 (serial position: 1–7) mixed ANOVA revealed a significant main effect of sound condition, $F(1.93, 243.45) = 18.42$, $MSE = 9.20$, $p < .001$, $\eta_p^2 = .128$, $BF_{10} = 276{,}234.78$, a significant main effect of task modality, $F(1, 126) = 9.81$, $MSE = 198.30$, $p = .002$, $\eta_p^2 = .072$, $BF_{10} = 15.63$, and a significant main effect of serial position, $F(2.54, 319.72) = 148.91$, $MSE = 10.71$, $p < .001$, $\eta_p^2 = .542$, $BF_{10} = 6.88 \times 10^{120}$. There were significant interactions between sound condition and serial position, $F(8.57, 1{,}079.96) = 2.21$, $MSE = 2.90$, $p = .021$, $\eta_p^2 = .017$, $BF_{10} = 0.06$,[2] and between serial position and task modality, $F(2.54, 319.72) = 3.31$, $MSE = 10.71$, $p = .027$, $\eta_p^2 = .026$, $BF_{10} = 2.41$. The three-way interaction between sound condition, serial position, and task modality was not significant, $F(8.57, 1{,}079.96) = 1.87$, $MSE = 2.07$, $p = .055$, $\eta_p^2 = .015$, $BF_{10} = 0.085$.

Figure 3

Probability of Serial Recall Error as a Function of Irrelevant Sound Condition for the Visual–Verbal (Left) and Visual–Spatial (Right) Serial Recall Tasks in Experiment 2



Note. Error bars show standard errors of the means. See the online article for the color version of this figure.

Again, the data pattern thus far appears generally consistent with that of Jones et al. (1995; Experiment 4). However, at odds with Jones et al. (1995; Experiment 4), but consistent with our partial replication (Experiment 1), the interaction between sound condition and task modality was significant, $F(1.93, 243.45) = 10.01$, $MSE = 9.20$, $p < .001$, $\eta_p^2 = .074$, $BF_{10} = 281.83$. As before, the interaction materialized because the changing-state effect was present for the visual–verbal serial recall task, $F(1, 63) = 34.22$, $MSE = 1.14$, $p < .001$, $\eta_p^2 = .352$, $BF_{10} = 56{,}219.66$, but absent for the visual–spatial serial recall task, $F(1, 63) = 0.73$, $MSE = 1.14$, $p = .397$, $\eta_p^2 = .011$, $BF_{10} = 0.26$.

As per Experiment 1, to test the changing-state effect directly, we repeated the previous analysis with the quiet condition removed and excluding serial position as a factor. There was a significant main effect of sound condition, $F(1, 126) = 22.47$ $MSE = 1.14$, $p < .001$, $\eta_p^2 = .151$, $BF_{10} = 2{,}439.40$, and a significant main effect of task modality, $F(1, 126) = 6.85$, $MSE = 19.92$, $p = .010$, $\eta_p^2 = .052$, $BF_{10} = 4.32$. Critically, the interaction between the two factors was also significant, $F(1, 126) = 12.49$, $MSE = 1.14$, $p < .001$, $\eta_p^2 = .090$, $BF_{10} = 45.21$. The interaction materialized because the changing-state effect was present in the visual–verbal task, $F(1, 63) = 34.22$, $MSE = 1.14$, $p < .001$, $\eta_p^2 = .352$, $BF_{10} = 56{,}219.66$, whereas it was absent in the visual–spatial task, $F(1, 63) = 0.73$, $MSE = 1.14$, $p = .397$, $\eta_p^2 = .011$, $BF_{10} = 0.26$.

In brief, the results of our near-identical replication preclude the possibility that the discrepant findings observed in our partial replication were a consequence of (a) our deployment of the task-modality manipulation as a within-participants, rather than a between-participants factor; and/or (b) our decision to remove the 10-s retention period. Our observation of a reliable interaction between task modality and auditory distraction may be because of a lack of statistical power in the study of Jones et al. (1995) to detect this interaction effect. The results fall squarely in line with the predictions of the WM model and call into question the functional equivalence account based on the O-OER model.

Bayesian Meta-Analysis

To examine the broader evidence in opposition to, or support of, the existence of a cross-modal changing-state effect of irrelevant speech on visual–spatial serial recall, we conducted a Bayesian meta-analysis. To facilitate the analysis, we conducted a systematic review of the literature for studies that examined the impact of changing-state irrelevant speech on visual–verbal and visual–spatial serial recall.

A systematic search of the Web of Science was conducted using the search string: (verbal OR spatial) AND serial recall AND

[2] The frequentist test detects an effect, but the Bayesian evidence is in favor of the null hypothesis. Results like these are referred to as the Jeffrey–Lindley paradox and may occur when the sample size is large, and effects are small relative to the prior distribution. The paradox can be resolved either by a more stringent α-level as sample size increases or by specifying a narrower prior distribution. In the Bayesian analysis, we used default settings for the prior distribution. Because of this, we lean on the results from the frequentist analyses here. It is important to note this result is not central to adjudicating between the contrasting theoretical approaches entertained within Experiments 1 and 2.

(irrelevant speech OR irrelevant sound) AND changing-state effect AND (short-term memory OR serial memory). The search was executed on August 10, 2022, and confined to the period 1995–2022. This revealed a total of 34 unique records. Of these records, only two were relevant, namely the original study of Jones et al. (1995) and the study of Tremblay et al. (2001).[3] To these records, we added the two partial attempts to replicate the Jones et al. (1995) findings by Kvetnaya (Kvetnaya, 2018; Kvetnaya et al., 2019), which did not feature in our systematic search results, but were instead identified by an earlier nonsystematic search of the literature.

We opted to exclude the original study (i.e., Experiment 4) from Jones et al. (1995). This decision was made on two grounds. First, it was not possible to extract the mean difference and standard error of the mean difference for their visual–verbal and visual–spatial conditions. This is because the tasks were analyzed together and the changing-state effect size was reported for both tasks, rather than each task separately. Second, since it is often the case that effect sizes are inflated in the original studies, perhaps in part because of publication bias (e.g., see Etz & Vandekerckhove, 2016), it is potentially misleading to include the original finding in the meta-analysis. This point is reinforced by the goal of a replication study being to test whether the original effect can be trusted. Finally, we also opted to exclude from our meta-analysis the study by Tremblay et al. (2001), since their study used bursts of nonspeech noise, rather than irrelevant letter stimuli, as used in Jones et al. (1995; Experiment 4). Bursts of noise may be particularly likely to capture attention (e.g., Leiva et al., 2015) and give rise to disruption that is qualitatively distinct from the changing-state effect.

With these exclusions, the analysis was undertaken using our own partial (Experiment 1) and near-identical (Experiment 2) replications, and the partial replications of Kvetnaya (Kvetnaya, 2018; Kvetnaya et al., 2019). Table 1 summarizes the test statistics and effect sizes of the studies included in the final analysis.

Method

We determined our prior mean difference and prior standard error of the mean difference from the partial replication studies of Kvetnaya (Kvetnaya, 2018; Kvetnaya et al., 2019). Bayesian techniques can be used to determine the relative support for the changing-state hypothesis over the null hypothesis. BF provides a continuous measure of how probable the data are under the changing-state hypothesis, as compared to how probable the data are under the null hypothesis. The BF calculations were undertaken using existing software (Dienes, 2008, 2011, 2014). Using this software, a null hypothesis is assumed by default, whereby the true population value is equal to zero.

The Bayesian approach requires specificity about the hypothesis to be contrasted with the null. We assumed that the changing-state effect would vary in size between zero and an upper limit set by the typically observed magnitude of the changing-state effect. We based our prediction on a half-normal distribution wherein predicting smaller effect sizes is more likely than large effect sizes. Since we were unable to obtain the relevant data from the visual–verbal and visual–spatial conditions of Jones et al. (1995; Experiment 4), the estimate of the standard deviation of the p(population value|theory) was computed as an average of the mean difference between the steady-state and changing-state conditions (7.67; $SE = 1.57$) from the varied versus repeated sound conditions of Experiments 1, 2, and 4 of LeCompte (1995). The rationale for using these studies to determine effect size was that LeCompte (1995) used a single repeated letter against varying letters, as in Jones et al. (1995; from which the relevant data were not retrievable), and the replications (Kvetnaya, 2018; Kvetnaya et al., 2019; and the partial and near-identical replications reported here). Further, as the author (LeCompte) has not published with any of the authors of the current article, nor with any of the authors of the target study, then impartiality is exercised for the purpose of replication. It should be noted that the experiments of LeCompte (1995) differed from those of Jones et al. (1995; Experiment 4), Kvetnaya (2018), and the near-identical replication of the current study because the exact letter sounds were different (for Experiments 2 and 4, the steady-state condition involved repeated presentations of the same letter that differed across trials and a retention period was not deployed). Because the magnitude of the changing-state effect, according to Jones et al. (1995; Experiment 4), should not differ as a function of task modality, the same estimate of the standard deviation of the p(population value|theory) was used for the Bayes meta-analysis for both visual–verbal and visual–spatial task modalities.

Results

The visual–verbal data from Kvetnaya et al. (2019) and the partial and near-identical replication data were combined in a meta-analysis using the mean and standard deviation from Kvetnaya et al. (2019) as the prior mean and prior standard deviation and the mean from the partial replication as the likelihood mean and likelihood's standard deviation to calculate the posterior mean and posterior standard deviation. Once these were obtained, they were entered as the new prior mean and standard deviation, and the mean and standard deviation from the near-identical replication were then used as the likelihood's mean and standard deviation. Following this stepwise procedure (Dienes, 2008), a final BF was computed representing the combined data. The final BF was 601397356231.48. Therefore, the results indicate "extreme evidence" for the changing-state hypothesis over the null hypothesis in the verbal domain (see Jeffreys, 1961).

The visual–spatial data from Kvetnaya (2018) were used as the prior mean and prior standard deviation using the same Bayesian meta-analytic procedure as for the visual–verbal condition. The final BF was 0.76. Therefore, the results indicate "anecdotal

[3] Our literature search also uncovered a study by Tremblay et al. (2012) wherein the disruptive impact of irrelevant sound (air traffic speech) was contrasted with a quiet control condition on a task that involved recalling either verbal or spatial stimuli (seven letters presented in different spatial locations). Tremblay et al. (2012) found that regardless of the task requirement (verbal or spatial serial recall), irrelevant sound impaired performance. Furthermore, this pattern was unchanged regardless of whether participants knew in advance which task they were required to perform on a given trial. We excluded this study from our analysis on three grounds. First, it is very likely that the requirement to encode one dimension (letters) over-spilled to the second dimension (spatial), resulting in a process-impure measure of "spatial" serial recall. In this regard, it is not at all surprising that the spatial serial recall task was disrupted by changing-state irrelevant sound—we would be surprised if it was not. Second, the isolated semantically meaningful phrases presented as irrelevant sounds may have generated intrigue for the participants, thereby promoting a qualitatively distinct effect of attentional diversion (see Hughes & Marsh, 2020). Thus, it may be that the disruption was due to attentional capture rather than the changing-state effect. Third, the study did not include a steady-state comparison condition to evaluate the effects of changing-state irrelevant sound.

Table 1

Test Statistics and Effect Sizes of the Study Included in the Bayesian Meta-Analysis to Test the Changing-State Effect on the Visual–Verbal and Visual–Spatial Serial Recall Tasks

	Main effects		Task Type × Sound Type interaction	
Statistics	Sound type	SS versus CS difference	With quiet	Without quiet
Visual–verbal serial recall task				
Experiment 1: partial replication ($n = 64$)				
Effect size	$\eta_p^2 = .232$	$\eta_p^2 = .230$		
Test statistic	$F(1.97, 124.34) = 19.057$, $p < .001$	$F(1, 63) = 18.82$, $p < .001$		
Experiment 2: near-identical replication ($n = 64$)				
Effect size	$\eta_p^2 = .337$	$\eta_p^2 = .352$		
Test statistic	$F(1.95, 122.94) = 32.077$, $p < .001$	$F(1, 63) = 34.215$, $p < .001$		
Kvetnaya et al. (2019; $n = 80$)				
Effect size	$\eta_p^2 = .350$	$\eta_p^2 = .150$		
Test statistic	$F(2, 78) = 20.660$, $p < .001$	$t(39) = 2.60$, $p = .013$		
Visual–spatial serial recall task				
Experiment 1: partial replication ($n = 64$)				
Effect size	$\eta_p^2 = .005$	$\eta_p^2 = .010$		
Test statistic	$F(1.91, 120.24) = 0.296$, $p = .734$	$F(1, 63) = 0.655$, $p = .422$		
Experiment 2: near-identical replication ($n = 64$)				
Effect size	$\eta_p^2 = .009$	$\eta_p^2 = .011$		
Test statistic	$F(1.87, 117.89) = 0.562$, $p = .560$	$F(1, 63) = 0.728$, $p = .397$		
Kvetnaya (2018; $n = 40$)				
Effect size	$\eta_p^2 = .020$	$\eta_p^2 = .037$		
Test statistic	$F(2, 78) = 0.807$, $p = .450$	$t(39) = -1.222$, $p = .229$		
Visual–verbal and visual–spatial serial recall tasks				
Experiment 1: partial replication ($n = 64$)				
Effect size			$\eta_p^2 = .083$	$\eta_p^2 = .080$
Test statistic			$F(1.94, 122.28) = 5.678$, $p = .005$	$F(1, 63) = 5.45$, $p = .023$
Experiment 2: near-identical replication ($n = 64$)				
Effect size			$\eta_p^2 = .074$	$\eta_p^2 = .090$
Test statistic			$F(1.93, 243.45) = 10.01$, $p < .001$	$F(1, 126) = 12.49$, $p < .001$
Kvetnaya (2018) and Kvetnaya et al. (2019)				
Effect size			$\eta_p^2 = .100$	$\eta_p^2 = .09$
Test statistic			$F(2, 156) = 8.96$, $p < .001$	$F(1, 78) = 7.29$, $p = .008$

Note. SS = steady state; CS = changing state.

evidence" for the null over the changing-state hypothesis in the spatial domain (see Jeffreys, 1961).

General Discussion

Consistent with much previous research, our experiments demonstrated pronounced bow-shaped serial position curves exhibiting primacy and recency effects in both the visual–verbal and visual–spatial serial recall tasks (e.g., Guérard & Tremblay, 2008; Jones et al., 1995; Tremblay et al., 2001). However, critically, our results revealed a pronounced vulnerability of visual–verbal serial recall to disruption via changing-state against steady-state sounds but a lack of sensitivity of visual–spatial serial recall to this changing-state effect. As such, the study demonstrates a reliable interaction between task modality and auditory distraction, thereby failing to replicate the lack of interaction originally reported by Jones et al. (1995). This result was observed regardless of whether the task-modality manipulation was implemented as a within-participants factor (Experiment 1) or as a between-participants factor (Experiment 2), and whether the recall phase occurred immediately after the presentation phase (Experiment 1) or a retention period was inserted between the presentation and recall phases (Experiment 2). Moreover, the results were observed with sample sizes that were at least twice that of the original Jones et al. (1995; Experiment 4) study.

It could be argued that the German stimuli adopted in our near-identical replication (Experiment 2) produced a discrepancy between the results of that experiment and those of Jones et al. (1995, Experiment 4). We consider this extremely unlikely given that the to-be-ignored stimuli (A B C D E F G) are one-syllable letters in German and are phonologically similar to their English counterparts. Similarly, the to-be-recalled letters (F K L M Q R Y) are also very similar in pronunciation to their English counterparts, each containing a single syllable, apart from "Y" (Ypsilon), which contains three syllables. Critically, we observed a robust changing-state effect with these to-be-remembered and to-be-ignored materials in the visual–verbal serial recall task. Indeed, the partial eta squared for the changing-state effect in the visual–verbal serial recall task was 0.35, which is more than double that reported in the original study by Jones et al. (1995, Experiment 4; 0.16). Therefore, a hypothesis that the German stimulus material was not sufficiently distracting to foster a changing-state effect on the visual–spatial serial recall task is untenable. Moreover, the results of Experiments 1 and 2 lead to the same conclusions, regardless of language. This demonstrates that the key findings do not depend on whether German or English stimulus materials are adopted.

A Bayesian meta-analysis, pooling the effect sizes of prior studies into one overall effect, revealed extreme evidence for the changing-state hypothesis over the null hypothesis for visual–verbal serial recall but, in stark contrast, anecdotal evidence for the null hypothesis over the changing-state hypothesis for visual–spatial serial recall. The results are inconsistent with those of Jones et al. (1995; Experiment 4), wherein the changing-state effect manifested with the same order of magnitude for visual–spatial serial recall as it did for visual–verbal serial recall.

A wide array of conceptual and exact replications of Jones et al.'s (1995) experiments (i.e., Experiments 2–4) have now failed to support the original findings (Alloway et al., 2010; Guérard & Tremblay, 2008; Guitard & Saint-Aubin, 2015; Kvetnaya, 2018; Kvetnaya et al., 2019; Meiser & Klauer, 1999; Experiments 1 and 2 of the current article). One can only speculate about the causes of this lack of reproducibility. Given the substantial productivity of the laboratory during the time period within which the data were collected and the probability that participants were taking part in many visual–verbal serial recall tasks during this era, it cannot be ruled out that the results were attributable to transfer effects between experiments. For example, given a history of taking part in visual–verbal serial recall tasks, participants may have brought to bear a verbal encoding strategy for visual–spatial stimuli. Such verbal recoding would render the visual–spatial serial recall task susceptible to disruption via distractor activities in the same way as visual–verbal serial recall, thereby preventing an interaction. It is by now clear, based on the balance of evidence from our results and the results of others, that the cross-modal interference effects reported by Jones et al. (1995) do not replicate.

In what follows, we consider the implications of the failure to replicate the findings of Jones et al. (1995, Experiment 4) for the WM and O-OER models presented at the outset, before considering the implications for other prominent theories of short-term memory and auditory distraction.

WM Model

The selective influence of irrelevant changing-state sound on visual–verbal serial recall, but not its visual–spatial counterpart, is consistent with the WM model (Baddeley, 1986; Baddeley & Hitch, 1974). As noted at the outset, given its modular architecture, the WM model predicts that the degree to which active or passive secondary activities impede performance on primary short-term memory tasks is based on whether they draw on the resources of the same short-term memory subsystem. Irrelevant speech is thought to gain obligatory access to the phonological store of the phonological loop, wherein it disrupts the store's contents (Salamé & Baddeley, 1982). Accordingly, irrelevant speech should interfere with verbal short-term memory tasks that also recruit the phonological loop, but it should not interfere with visual–spatial short-term memory tasks that recruit the visuospatial sketchpad. The results of the current experiments are consistent with this expectation.

Nevertheless, the phonological loop account of the precise manner by which irrelevant speech disrupts visual–verbal serial recall is incorrect for several reasons. First, the phonological loop account does not acknowledge or explain the fact that the magnitude of disruption produced by irrelevant speech on visual–verbal serial recall is based on the degree to which it conforms to the property of changing state (Jones & Macken, 1993, 1995; Macken & Jones, 1995). Second, so long as the principle of changing state is satisfied, both "speech" and "nonspeech" sounds are equipotent in their ability to disrupt visual–verbal serial recall (Jones & Macken, 1993), whereas on the phonological loop account, it is assumed that only "speech" sounds gain access to, and can disrupt the contents of, the phonological store.

To explain the changing-state effect and interference by nonspeech sounds, the WM model requires refinement. To resolve this problem, Page and Norris (2003) present a revised account of the irrelevant sound effect in their computational model of the phonological loop, known as the primacy model (Page & Norris, 1998). In the primacy model, the order of a sequence of items is represented in the phonological store by a primacy gradient of activation strength, with the first item activated strongest and the last item activated weakest. According to the model, changing-state, but not steady state, irrelevant sound is represented by a separate primacy gradient that draws on the same pool of limited resources used to construct the primacy gradient over to-be-remembered items in the phonological store. The consequence of this competition is that compared to a no-sound scenario: (a) to-be-remembered items are stored with lower activation and (b) the relative difference in item activation levels is reduced. When random Gaussian noise is applied to item activation levels to simulate errors, the implication of (a) is that items will be more likely to drop beneath a response threshold for selection, causing an increase in omission errors, whereas the implication of (b) is that it will be easier for a nearby item to the target item to-be-recalled to assume a larger activation level (because of the smaller difference in item activation levels, which can be more readily bridged by noise), triggering an increase in transposition errors. Both an increase in omission and transposition errors are characteristic features of interference from changing-state irrelevant sound, and Page and Norris (2003) show that the primacy model can reproduce these key empirical characteristics.[4]

What then about the findings of functional equivalence between visual–verbal and visual–spatial short-term memory in serial order phenomena, such as serial position curves and error patterns (Hurlstone, 2024; Hurlstone et al., 2014)? How can the WM model account for these functional similarities? The WM model explains these similarities by recourse to the assumption that the phonological loop and visuospatial sketchpad are structured in similar ways

[4] What the phonological loop account (Baddeley, 1986) and indeed any other account cannot explain is that syncopated physical or spatial tapping exerts the same effects as articulatory suppression and irrelevant speech in reducing the magnitude of the phonological similarity effect (Guérard et al., 2009; Larsen & Baddeley, 2003; Saito, 1994, 1997; Surprenant et al., 2008). According to the WM model (Baddeley, 1986), for example, syncopated tapping should not affect the phonological loop.

(cf. Baddeley, 1986; Logie, 1995) and rely on common mechanisms and representational principles (Hurlstone, 2024; Hurlstone & Hitch, 2015, 2018; Hurlstone et al., 2014). Hurlstone et al. (2014) propose that the phonological loop and visuospatial sketchpad both operate as competitive queuing, sequence planning and control systems (Grossberg, 1978; Houghton, 1990), wherein items are simultaneously activated in parallel, according to an activation gradient that dictates their serial order, and a scanning mechanism iteratively selects the item with the strongest activation level for output. Evidence from behavioral and computational analyses of the dynamics of transposition errors (Farrell & Lewandowsky, 2004; Hurlstone & Hitch, 2015, 2018) in visual–verbal and visual–spatial serial recall tasks suggests that within the verbal and spatial short-term memory competitive queuing systems, serial order is represented by three common principles, namely, a primacy gradient (items are encoded with gradually decreasing strength), position marking (items are associated with some representation of their sequence position that is later used to drive recall), and response suppression (items are temporarily suppressed in memory once they have been emitted). Accordingly, functional similarities in serial order phenomena across the visual–verbal and visual–spatial domains reflect the fact that the phonological loop and visuospatial sketchpad rely on the same mechanisms and principles for representing and generating serial order.

O-OER Model

According to the O-OER model (Jones, 1993; Jones et al., 1996), a common order mechanism operates across different types of items. From this perspective, the presentation of materials from different sensory origins results in streams of items represented on an episodic surface (blackboard) within memory. Interference occurs to the extent that to-be-remembered and to-be-ignored materials, regardless of their domain of origin, possess an element of serial order. Thus, according to the O-OER account, the changing-state effect arises because of an interference-by-process (e.g., Jones & Tremblay, 2000). Specifically, the order of the auditorily presented changing-state distracters is automatically processed, which interferes with the voluntary processing of order information in the focal memory task, regardless of whether the order of verbal or spatial information must be retained. However, at odds with the O-OER model (and the interference-by-process view) is that, unlike Jones et al. (1995), the current study found a changing-state effect in the context of short-term order memory for visual–verbal but not visual–spatial material. The failure to find a changing-state effect in the context of the visual–spatial task argues against the notion that a common memory system is associated with both visual–spatial and visual–verbal information (cf. Jones et al., 1995). That is, the changing-state effect does not generalize across modalities (e.g., visual–verbal vs. visual–spatial). In recent years the O-OER model (e.g., Jones et al., 1996) and the interference-by-process view (Jones & Tremblay, 2000) have been subsumed within the perceptual-gestural account (see, Hughes et al., 2011, 2016; Hughes & Marsh, 2017; Jones et al., 2004, 2006, 2007; Macken & Jones, 2003) and the duplex mechanism account (Hughes, 2014; Hughes & Marsh, 2019; Hughes et al., 2007).

Perceptual-Gestural Account

According to the perceptual-gestural account, short-term memory performance is a product of the functioning of general-purpose perceptual and motor processes, not of dedicated mnemonic structures or mechanisms (Hughes et al., 2009, 2016; Hughes & Marsh, 2017; Jones et al., 2004, 2006, 2007; Maidment & Macken, 2012). Visual–verbal serial short-term memory is underpinned by vocal-motor processes: a vocal-motor plan is assembled opportunistically to graft constraints onto a verbal sequence where such information is lacking. This motor planning, however, is inherently "open" such that it is susceptible to interference from sequences of changing-state items that are passively organized into streams as a result of auditory scene analysis (Bregman, 1990). The changing-state effect arises because changes in successive sound elements give rise to cues pertaining to the order of those sounds. The involuntary, preattentive processing gives rise to extraneous order information that interferes with the voluntary, goal-driven process of serially rehearsing the to-be-remembered items which support that sequential output (e.g., Jones & Macken, 1993). The interference-by-process component of the perceptual-gestural account suggests that the vulnerability of a given focal task to disruption via the mere presence of a task-irrelevant sound sequence is dictated by a clash between the particular processes involved in the cognitive task at hand, and those applied automatically to the task-irrelevant sound: only when a task requires the planning of deliberate, coherent motor-actions (e.g., serial rehearsal) will it be susceptible to disruption via the perceptual organization of the auditory environmental input into objects. Thus, tasks that do not invoke serial rehearsal are immune to the changing-state effect (Beaman & Jones, 1997; Jones & Macken, 1993; Stokes & Arnell, 2012). This account can better accommodate the interaction between interference and task modality observed in the current study than the O-OER model (Jones et al., 1995) from which it evolved. This is because irrelevant speech competes strongly for motor-planning processes involved in vocal-articulatory planning and only weakly for the motor-planning processes underpinning visual–spatial serial recall.

In the context of visual–spatial serial recall, the notion is that eye movements are adopted as an effective rehearsal strategy as indicated by eye-tracking measures (Guérard et al., 2009; Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006). Participants engaging in a greater quantity of eye movements, as indexed by fixating pairs of dots in the same order as the to-be-remembered sequence, have better visual–spatial serial recall performance (Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006). Moreover, preventing the use of eye movements by requiring participants to engage in irrelevant saccadic eye movements reduces serial recall performance (Tremblay, Parmentier, et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006). For visual–spatial serial recall, eye movement and oculomotor control are co-opted to meet the goal of retaining and reproducing the to-be-remembered sequence. The perceptual-gestural view extended to the visual–spatial serial recall task must assume that the oculomotor sequence planning is also "open" to the population via the perceptual organization of the auditory environmental input into objects. The changing-state effect therefore transcends the motor-action means of rehearsal (oculomotor vs. vocal-motor).

The results of the current study could imply that the mapping between the perceptual organization of sound and motor planning functions in different ways depending on whether vocal-motor or oculomotor planning is required (cf. Hughes & Marsh, 2017, p. 545). In addition, it could be argued that the current results shed light on the locus where the interference with motor planning occurs. That is, changing-state sound may interfere with vocal-articulatory planning rather than operating at more general/"earlier" stages of motor planning. Thus, the finding that visual–spatial serial recall in comparison to visual–verbal serial recall is invulnerable to the changing-state effect may help refine what processes in the context of the interference-by-process account are being disrupted. In this way, the current results do not necessarily undermine the central tenet of the interference-by-process account that motor processes are disrupted rather than "short-term storage."

Alternatively, the results could be taken as consistent with the notion that distinct memory systems or mechanisms are associated with serial order retention of information from different domains of origin, but that the general principles of these systems or mechanisms are highly similar (cf. Logie, 1995). The problem with the latter suggestion is a burgeoning body of work demonstrating that many short-term memory phenomena can be explicated in terms of motor planning, perceptual organization, and the mapping between them, without appealing to dedicated storage systems or mechanisms (Hughes et al., 2011; Jones et al., 2004, 2006).

Further empirical work has undermined the body of evidence supporting the hypothetical existence of the phonological store (Jones et al., 2004, 2006). Cited as key evidence for the phonological store is a three-way interaction between irrelevant sound, articulatory suppression, and presentation modality. With visually presented memoranda, the irrelevant sound effect disappears when opportunities for articulatory (e.g., vocal-motor) processes are reduced by articulatory suppression. It is argued that under articulatory suppression, visual–verbal items cannot be converted into phonological form and access the phonological store. Consequently, item representations from auditory origin cannot interfere with the representations of visual target items within the store. Coherent with this view is that the irrelevant sound effect is abolished by articulatory suppression for visual–verbal presentation (Hanley, 1997; Salamé & Baddeley, 1982; Schlittmeier et al., 2008). However, the irrelevant sound effect survives for auditory presentation of memoranda under conditions of articulatory suppression (Hanley & Broadbent, 1987; Schlittmeier et al., 2008), even when to-be-ignored sound is presented during a retention interval thereby ruling out auditory masking explanations (Hanley & Bakopoulou, 2003). The WM model accounts for this persistence of the irrelevant sound effect with the auditory presentation of memoranda because auditory-verbal target items gain direct, automatic access to the phonological store because they are already in a phonological form. Therein, representations of to-be-ignored items of auditory origin can interfere with auditory-verbal target items. Thus, according to the WM model (e.g., Baddeley & Larsen, 2007; Norris et al., 2018; Salamé & Baddeley, 1982), the impact of irrelevant sound on serial recall arises because sound interferes with the representations of target items in a postcategorical phonological store.

However, several studies have now shown that the survival of the irrelevant sound effect under articulatory suppression is restricted primarily to recency (Jones et al., 2004), which suggests the effect is driven by the modality effect (or auditory recency) whereby there is a recall advantage for the final one or two items with an auditory compared to visually-presented sequence (Conrad & Hull, 1968). Jones et al. (2004) found that the presence of a redundant to-be-ignored spoken item that immediately followed the auditory list presentation removed the irrelevant sound effect. Thus, irrelevant sound produces a suffix effect (e.g., Crowder, 1967; Nicholls & Jones, 2002) for auditory-verbal presentation. This finding has been conceptually replicated several times by Hanley and colleagues (Hanley & Bourgaize, 2018; Hanley & Hayes, 2012; Hanley & Shah, 2012). Problematic with the WM model is that it proposes the modality effect arises from acoustic against phonological factors (Nicholls & Jones, 2002) and is therefore "peripheral to the working memory system" (Baddeley, 1986, p. 95; Hurlstone et al., 2014). The suffix effect, however, is attributable to acoustic factors: To elaborate, a suffix, when acoustically similar to the to-be-remembered sequence is subject to an auditory grouping process that means it becomes perceptually integrated with the to-be-remembered sequence. Thus, the distinctiveness of the end-boundary position that the auditory list-final items would have (which governs their superior recall), now belongs to the suffix which disrupts the order encoding of the final list items. The acoustic-based irrelevant sound effect that survives under suppression with auditory presentation is consistent with the view that the vocal-motor processes that are necessary for the expression of the irrelevant sound effect transcend presentation modality. Consequently, proponents of the perceptual-gestural view (Jones et al., 2004, 2006) propose that the irrelevant sound effect does not have its locus within the phonological store, but rather an articulatory planning process.

The Duplex Mechanism Account

The duplex-mechanism account of auditory distraction (e.g., Hughes, 2014) postulates that interference-by-process is one of two qualitatively distinct forms of distraction. The changing-state effect is attributed to order interference while the disruption of serial recall by deviant or intrinsically interesting distracters (Hughes & Marsh, 2020; Marsh et al., 2018, 2020; Röer et al., 2013, 2017; Vachon et al., 2012, 2017) is attributed to a task-unspecific attentional diversion/capture. In the current study, the visual–spatial serial recall task was invulnerable to disruption via the changing-state effect. At first glance, one might consider visual–spatial tasks to be insensitive to disruption via any task-irrelevant auditory stimulation. However, a growing number of studies reveal that the visual–spatial serial recall task is susceptible to the auditory deviation effect: unexpected auditory events within an otherwise predictable sequence of events (deviants) reliably disrupt visual–spatial serial recall as they do visual–verbal serial recall (Marsh et al., 2017; C. C. Morey & Miron, 2016; Vachon et al., 2017; see also Vachon et al., 2020; but see Lange, 2005). Combined with the current results, these studies demonstrate that the visual–spatial task dissociates changing-state and attention-based auditory deviation effects, which does not cohere with an "attentional account" (discussed below), according to which both forms of distraction are underpinned by the same attentional capture mechanism and should therefore manifest in both visual–verbal and visual–spatial serial recall.

Similarity-Based Interference Accounts

Another class of theoretical accounts that lend themselves well to explaining verbal and spatial short-term memory dissociations observed in the presence of secondary tasks or task-irrelevant information are interference-based models (Farrell, 2006; Lewandowsky & Farrell, 2008; Nairne, 1988, 1990; Neath, 2000; Oberauer et al., 2012). According to these accounts, the to-be-remembered items in a short-term memory primary task are represented in terms of stimulus features. For a verbal short-term memory primary task, these features might correspond to the phonological or semantic properties of words, whereas for a spatial short-term memory primary task, they might correspond to the orientation or spatial position of objects. The content of actively performed secondary tasks or passive to-be-ignored information is also assumed to be represented in terms of stimulus features, and these features can be adopted by—or superimposed upon—the features corresponding to the to-be-remembered items, thus causing them to be overwritten. The degree of overwriting or interference depends upon the extent to which the short-term memory primary task and active secondary task or passive to-be-ignored information possess similar stimulus features—that is, interference is similarity-based. Thus, articulatory suppression or irrelevant speech, because of their verbal character, will share more stimulus features with a verbal short-term memory primary task than a spatial short-term memory primary task; by contrast, spatial tapping, because of its spatial character, will share more stimulus features with a spatial short-term memory primary task than a verbal short-term memory primary task.

Poirier et al. (2019) reported a dual-task experiment in which participants completed the verbal and spatial versions of the Brooks matrix task in the presence of either articulatory suppression or manual spatial tapping. They found that the verbal version of the Brooks matrix task was disrupted more by articulatory suppression than by manual spatial tapping, whereas the spatial version of the Brooks matrix task was disrupted more by manual spatial tapping than by articulatory suppression. They subsequently fitted the feature model (Nairne, 1988, 1990; Neath, 2000)—an interference-based model of memory—to the data from their experiment and found that it was successfully able to reproduce the observed dual-task dissociation. In a further simulation study, Poirier et al. (2019) showed that the feature model could also reproduce the dual-task dissociation between visual–verbal and visual–spatial serial recall under conditions of articulatory suppression and manual spatial tapping reported in the study by Guérard and Tremblay (2008; see earlier). This provides a proof-of-principle that interference-based accounts can reproduce the dissociations between verbal and spatial short-term memory tasks under dual-task conditions. It suggests such dissociations may be more parsimoniously explained in terms of the degree to which primary and secondary tasks share similar features, rather than the same bespoke short-term memory store, thus calling into question the modularity assumption of the WM model.

A key assumption of the feature model account of the irrelevant speech effect is that the locus of the effect is not the disruption of serial order processing in the primary memory task, but rather the feature overwriting of item information in that task by task-irrelevant distracters (Neath, 2000). This distinguishes the model from the O-OER account (Jones et al., 1996). How then does the model explain the changing-state effect? In the feature model, it is assumed that secondary tasks or passive to-be-ignored information will deplete attentional resources from the short-term memory primary memory task. Active secondary tasks, like articulatory suppression and spatial tapping, are assumed to deplete attentional resources more so than passive to-be-ignored information like irrelevant speech, because of their production and monitoring requirements. The model explains the changing-state effect by assuming that changing-state irrelevant speech draws more attentional resources away from the short-term memory primary task than steady-state irrelevant speech does. Consistent with our results, the model would also predict that irrelevant speech should not impair visual–spatial serial recall performance because irrelevant speech should not share any overlapping features with spatial locations.

There are some potential issues with the feature model account of the irrelevant speech effect, however. First, the feature model assumes that it is not a necessary precondition for the short-term memory primary task to require the retention of serial order for the changing-state irrelevant speech effect to manifest. However, this runs counter to results demonstrating that tasks that require the processing of items but not order information are not disrupted (Hughes et al., 2007; Jones & Macken, 1993)—or are disrupted markedly less (Henson et al., 2003)—by changing- compared with steady-state speech or sound than tasks that require the retention of serial order information. This suggests that changing-state speech does have a specific disruptive effect on the retention of serial order information, contrary to the feature model. Second, the model's assumption that changing-state speech is more disruptive than steady-state speech because it is more attention-demanding is questionable. For example, Hughes et al. (2013) found that visually degrading the letter stimuli in a visual–verbal serial recall task—a manipulation that should increase focused visual attention to the primary memory task—did not reduce disruption by changing-state speech, compared to a condition involving nondegraded letter stimuli. However, the same manipulation abolished the disruptive effect of an auditory deviant—a form of auditory distraction that is known to be caused by attentional capture (Hughes, 2014; Jones et al., 2010). Kindred dissociations between the two forms of auditory distraction under conditions of high versus low primary task encoding or attentional load have been reported in other studies (Hughes & Marsh, 2019; Marsh et al., 2020).

Notwithstanding these limitations, interference-based accounts such as the feature model are attractive because they can explain dissociations between verbal and spatial short-term memory in the presence of secondary tasks or passive to-be-ignored information without appealing to bespoke memory systems for verbal and spatial information. Such accounts are arguably more parsimonious therefore than modular accounts such as the WM model.

Attentional Accounts

The attentional (diversion) account (e.g., Cowan, 1995; Röer et al., 2015) is an alternative to the perceptual-gestural model and the duplex-mechanism account (e.g., Hughes, 2014). Based on the functional, process-oriented, embedded-processes model of WM (Cowan, 1995), the attentional account postulates that auditory changes attract the focus of attention, which can no longer be used to refresh the memory traces of the to-be-remembered items. According to this account, attention is drawn away from a focal task by a sound that differs from its immediate predecessor. Thus, when a sound is not a repeat of its predecessor its likelihood of diverting attention is much greater. The graded attentional model (Bell et al., 2019a, 2019b) is a specific example of the attentional account. In this view, each stimulus, regardless of whether it mismatches its predecessor, elicits a basic call for attention and is compared to the preceding pattern of stimulation. Because of this basic call for attention, some resources are already diverted away from the primary task even before the system matches the incoming stimulus to previous representations and thereafter denies the call of attention (if a match is found) or answers the call for attention (if a mismatch is detected). The graded attentional account assumes that reliable steady-state effects and changing-state effects should be observable on a variety of tasks, including those that are not seriation-based, providing that they are attentionally demanding. Therefore, the graded attentional model (Bell et al., 2019a, 2019b) assumes that a changing-state effect should be observed in the context of both visual–verbal and visual–spatial short-term memory tasks. This model thus leads to the same prediction as the interference-by-process account that the changing-state effect should generalize to a visual–spatial serial recall task. Furthermore, like the feature model, the graded attentional model fails to account for the process sensitivity of the changing-state effect (Beaman & Jones, 1997; Jones & Macken, 1993), or its insensitivity to modulation by high focal-task load (Hughes et al., 2013).

The immunity of the visual–spatial task to the changing-state effect is thus at odds with the attentional account. Some modifications of the attentional account are therefore necessary to explain the invulnerability of visual–spatial serial recall to disruption by changing-state sounds. It is worth noting here that the attentional account subsumed within the embedded-process model (Cowan, 1995) does not strictly negate the existence of similarity-based (or task-modality-determined) interference:

> Every stimulus has the potential of activating a number of types of memory features corresponding to that stimulus (…). Interference with activated features in memory would come from any subsequent stimulus or thought process that elicited the activation of similar types of features. (Cowan, 1995, p. 36)

Therefore, according to this model, visual–spatial serial recall might be immune to the changing-state effect because of less interference between activated features within memory arising from the stimuli and thought processes involved in maintaining items for serial recall. However, on this attentional account, the specifics of the stimuli, thought processes, and the characteristics of the features that both activate, require elaboration. The results of the current study suggest that the similarity between the to-be-remembered and irrelevant input is an important determinant of irrelevant-sound disruption. However, a wealth of previous studies demonstrate that overlap during encoding/maintenance between to-be-ignored and to-be-remembered items in their phonological (e.g., Jones & Macken, 1993; but see Eagan & Chein, 2012) or semantic (e.g., Marsh et al., 2008, 2009; but see Neely & LeCompte, 1999) features are not determinants of the magnitude of disruption of serial short-term recall, regardless of their activation (Röer et al., 2017; Vachon et al., 2020). Therefore, it remains unclear how the attentional account might accommodate both the present evidence, and other evidence showing the magnitude of disruption to serial short-term memory is unaffected by the feature overlap between to-be-ignored and to-be-remembered items.

Conclusions

Current theories of auditory distraction were strongly influenced by the finding that visual–verbal and visual–spatial short-term serial recall are disrupted by changing-state auditory material to the same extent (Jones et al., 1995; Experiment 4). However, until now this finding has received little further investigation. Here, we demonstrate that the notion of functional equivalence supported by these data is also no longer tenable. In our study, the changing-state effect of auditory distraction emerges for visual–verbal, but not visual–spatial, serial recall tasks. This asymmetry in disruption from changing-state auditory distracters in the context of visual–verbal and visual–spatial recall should, in addition to the plethora of findings pertaining to short-term memory and distraction, be incorporated into general models of WM as well as those pertaining to the vulnerability of cognitive processing to the passive processing of sound (e.g., Bell et al., 2019a, 2019b; Hughes, 2014; Hughes et al., 2013).

References

Alloway, T. P., Kerr, I., & Langheinrich, T. (2010). The effect of articulatory suppression and manual tapping on serial recall. *European Journal of Cognitive Psychology*, *22*(2), 297–305. https://doi.org/10.1080/09541440902793731

Avons, S. E. (2007). Spatial span under translation: A study of reference frames. *Memory & Cognition*, *35*(3), 402–417. https://doi.org/10.2758/BF03193281

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, *7*(1), 25–31. https://doi.org/10.1111/j.1467-9280.1996.tb00662.x

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208. https://doi.org/10.1016/j.neuroscience.2005.08.023

Baddeley, A. D. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Baddeley, A. D. (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

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 47–89). Academic Press.

Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop: Some answers and some questions. *Quarterly Journal of Experimental Psychology*, *60*(4), 512–518. https://doi.org/10.1080/17470210601147663

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, & Cognition*, *23*(2), 459–471. https://doi.org/10.1037/0278-7393.23.2.459

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, & 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

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

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

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, *10*(4), 135–136. https://doi.org/10.3758/bf03331446

Cortis, C., Dent, K., Kennett, S., & Ward, G. (2015). First things first: Similar list length and output order effects for verbal and nonverbal stimuli.

Journal of Experimental Psychology: Learning, Memory, & Cognition, *41*(4), 1179–1214. https://doi.org/10.1037/xlm0000086

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. *Memory & Cognition*, *34*(8), 1720–1729. https://doi.org/10.3758/BF03195933

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

Crowder, R. G. (1967). Prefix effects in immediate memory. *Canadian Journal of Psychology*, *21*(5), 450–461. https://doi.org/10.1037/h0082997

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *Quarterly Journal of Experimental Psychology*, *62*(3), 531–549. https://doi.org/10.1080/17470210801995002

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex*, *11*(4), 341–354. https://doi.org/10.1016/S0010-9452(75)80026-8

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on?. *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781), 1–17. https://doi.org/10.3389/fpsyg.2014.00781

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, & Cognition*, *38*(2), 473–481. https://doi.org/10.1037/a0025368

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

Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *PLoS ONE*, *11*(2), Article e0149794. https://doi.org/10.1371/journal.pone.0149794

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *Quarterly Journal of Experimental Psychology*, *38*(4), 675–688. https://doi.org/10.1080/14640748608401620

Farrand, P. A., Parmentier, F. B. R., & Jones, D. M. (2001). Temporal-spatial memory: Retrieval of spatial information does not reduce recency. *Acta Psychologica*, *106*(3), 285–301. https://doi.org/10.1016/S0001-6918(00)00054-8

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, *55*(4), 587–600. https://doi.org/10.1016/j.jml.2006.06.002

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language*, *51*(1), 115–135. https://doi.org/10.1016/j.jml.2004.03.007

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*(2), 175–191. https://doi.org/10.3758/BF03193146

Grossberg, S. (1978). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, *17*(3), 199–219. https://doi.org/10.1016/0022-2496(78)90016-0

Guérard, K., Jalbert, A., Neath, I., Surprenant, A. M., & Bireta, T. J. (2009). Irrelevant tapping and the acoustic confusion effect: The effect of spatial complexity. *Experimental Psychology*, *56*(5), 367–374. https://doi.org/10.1027/1618-3169.56.5.367

Guérard, K., & Tremblay, S. (2008). Revisiting evidence for modularity and functional equivalence across verbal and spatial domains in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 556–569. https://doi.org/10.1037/0278-7393.34.3.556

Guitard, D., & Saint-Aubin, J. (2015). A replication of "Functional equivalence of verbal and spatial information in serial short-term memory (1995; Experiments 2 and 3)". *Quantitative Methods for Psychology*, *11*(2), r4–r7. https://doi.org/10.20982/tqmp.11.2.r004

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

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review*, *10*(2), 435–444. https://doi.org/10.3758/bf03196503

Hanley, J. R., & Bourgaize, J. (2018). Similarities between the irrelevant sound effect and the suffix effect. *Memory & Cognition*, *46*(6), 841–848. https://doi.org/10.3758/s13421-018-0806-8

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

Hanley, J. R., & Hayes, A. (2012). The irrelevant sound effect under articulatory suppression: Is it a suffix effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*(2), 482–487. https://doi.org/10.1037/a0025600

Hanley, J. R., & Shah, N. (2012). The irrelevant sound effect under articulatory suppression is a suffix effect even with five-item lists. *Memory*, *20*(5), 415–419. https://doi.org/10.1080/09658211.2012.670249

Hanley, J. R., Young, A. W., & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *The Quarterly Journal of Experimental Psychology Section A*, *43*(1), 101–125. https://doi.org/10.1080/14640749108401001

Henson, R., Hartley, T., Burgess, N., Hitch, G. J., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology*, *56*(8), 1307–1334. https://doi.org/10.1080/02724980244000747

Horton, N., Hay, D. C., & Smyth, M. M. (2008). Hebb repetition effects in visual memory: The roles of verbal rehearsal and distinctiveness. *Quarterly Journal of Experimental Psychology*, *61*(12), 1769–1777. https://doi.org/10.1080/17470210802168674

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). Academic Press.

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., Chamberland, C., Tremblay, S., & Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal short-term memory. *Journal of Memory & Language*, *90*(10), 126–146. https://doi.org/10.1016/j.jml.2016.04.006

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., & 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, & Cognition*, *43*(4), 537–551. https://doi.org/10.1037/xlm0000325

Hughes, R. W., & Marsh, J. E. (2019). Dissociating two forms of auditory distraction in a novel Stroop serial recall experiment. *Auditory Perception & Cognition*, *2*(3), 129–142. https://doi.org/10.1080/25742442.2020.1760757

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., Marsh, J. E., & Jones, D. M. (2009). Perceptual–gestural (mis)mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1411–1425. https://doi.org/10.1037/a0017008

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2011). Role of serial order in the impact of talker variability on short-term memory: Testing a perceptual organization-based account. *Memory & Cognition*, *39*(8), 1435–1447. https://doi.org/10.3758/s13421-011-0116-x

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

Hurlstone, M. J. (2019). Functional similarities and differences between the coding of positional information in verbal and spatial short-term order memory. *Memory*, *27*(2), 147–162. https://doi.org/10.1080/09658211.2018.1495235

Hurlstone, M. J. (2024). Serial recall. In M. Kahana & A. Wagner (Eds.) *The Oxford handbook of human memory*. Oxford University Press.

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *41*(2), 295–324. https://doi.org/10.1037/a0038223

Hurlstone, M. J., & Hitch, G. J. (2018). How is the serial order of a visual sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *44*(2), 167–192. https://doi.org/10.1037/xlm0000440

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

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

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. A., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*(4), 1008–1018. https://doi.org/10.1037/0278-7393.21.4.1008

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory & Language*, *54*(2), 265–281. https://doi.org/10.1016/j.jml.2005.10.006

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology*, *60*(4), 505–511. https://doi.org/10.1080/17470210601147598

Jones, D. M., Hughes, R. W., & Macken, W. J. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise and Health*, *12*(49), Article 201. 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, & Cognition*, *19*(2), 369–381. https://doi.org/10.1037/0278-7393.19.2.369

Jones, D. M., & Macken, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition*, *23*(2), 192–200. https://doi.org/10.3758/BF03197221

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*(3), 656–674. https://doi.org/10.1037/0278-7393.30.3.656

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

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

Kvetnaya, T., Schopf, K., & Wickelmaier, F. (2019). *Irrelevant background speech disrupts serial short-term memory for verbal but not for spatial information: A pre-registered replication study* [conference paper]. DAGA, Rostock.

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory & Language*, *53*(4), 513–531. https://doi.org/10.1016/j.jml.2005.07.002

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1249–1268. https://doi.org/10.1080/02724980244000765

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

Leiva, A., Parmentier, F. B. R., & Andrés, P. (2015). Distraction by deviance: Comparing the effects of auditory and visual deviant stimuli on auditory and visual target processing. *Experimental Psychology*, *62*(1), 54–65. https://doi.org/10.1027/1618-3169/a000273

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 49, pp. 1–48). Academic Press. https://doi.org/10.1016/S0079-7421(08)00001-7

Logie, R. H. (1995). *Visuo-spatial working memory. Essays in cognitive psychology*. Lawrence Erlbaum Associates.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*(1), 55–74. https://doi.org/10.1016/0001-6918(90)90066-O

Macken, W. J., & Jones, D. M. (1995). Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 436–448. https://doi.org/10.1037/0278-7393.21.2.436

Macken, W. J., & Jones, D. M. (2003). Reification of phonological storage. *Quarterly Journal of Experimental Psychology*, *56*(8), 1279–1288. https://doi.org/10.1080/02724980245000052

Maidment, D. W., & Macken, W. J. (2012). The ineluctable modality of the audible: Perceptual determinants of auditory verbal short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, *38*(4), 989–997. https://doi.org/10.1037/a0027884

Marsh, J. E., Campbell, T. A., Vachon, F., Taylor, P. J., & Hughes, R. W. (2020). How the deployment of visual attention modulates auditory distraction. *Attention, Perception, & Psychophysics*, *82*, 350–362. https://doi.org/10.3758/s13414-019-01800-w

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory & Language*, *58*(3), 682–700. https://doi.org/10.1016/j.jml.2007.05.002

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. (2023). *SPSS dataset and programs to accompany submission entitled: Changing-state Irrelevant Speech Disrupts Visual-Verbal but not Visual-Spatial Serial Recall*. https://doi.org/10.17030/uclan.data.00000321.

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., Vachon, F., & Sörqvist, P. (2017). Increased distractibility in schizotypy: Independent of individual differences in working memory capacity?. *Quarterly Journal of Experimental Psychology*, *70*(3), 565–578. https://doi.org/10.1080/17470218.2016.1172094

Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*(5), 1272–1299. https://doi.org/10.1037/0278-7393.25.5.1272

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, & Cognition*, *42*(12), 1907–1918. https://doi.org/10.1037/xlm0000280

Morey, R. D., & Rouder, J. N. (2023). *BayesFactor: Computation of Bayes factors for common designs.* https://richarddmorey.github.io/BayesFactor/

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*(4), 343–352. https://doi.org/10.3758/bf03197045

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

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

Neely, C. B., & Lecompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & cognition*, *27*, 37–44. https://doi.org/10.1037/e537272012-173

Nicholls, A. P., & Jones, D. M. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 12–28. https://doi.org/10.1037/0278-7393.28.1.12

Norris, D., Baddeley, A. D., & Page, M. (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

Norris, D., Butterfield, S., Hall, J., & Page, M. P. (2018). Phonological recoding under articulatory suppression. *Memory & Cognition*, *46*(2), 173–180. https://doi.org/10.3758/s13421-017-0754-8

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

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 Section A*, *56*(8), 1289–1300. https://doi.org/10.1080/02724980343000233

Parmentier, F. B. R., Andrés, P., Elford, G., & Jones, D. M. (2006). Organization of visuo-spatial serial memory: Interaction of temporal order with spatial and temporal grouping. *Psychological Research*, *70*(3), 200–217. https://doi.org/10.1007/s00426-004-0212-7

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, & Cognition*, *31*(3), 412–427. https://doi.org/10.1037/0278-7393.31.3.412

Poirier, M., Yearsley, J. M., Saint-Aubin, J., Fortin, C., Gallant, G., & Guitard, 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

Postle, B. R., Druzgal, T. J., & D'Esposito, M. (2003). Seeking the neural substrates of visual working memory storage. *Cortex*, *39*(4–5), 927–946. https://doi.org/10.1016/S0010-9452(08)70871-2

Röer, J. P., Bell, R., & Buchner, A. (2013). Self-relevance increases the irrelevant sound effect: Attentional disruption by one's own name. *Journal of Cognitive Psychology*, *25*(8), 925–931. https://doi.org/10.1080/20445911.2013.828063

Röer, J. P., Bell, R., & Buchner, A. (2015). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception & Performance*, *41*(3), 692–702. https://doi.org/10.1037/xhp0000028

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

Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect?. *Memory & Cognition*, *22*(2), 181–187. https://doi.org/10.3758/BF03208889

Saito, S. (1997). When articulatory suppression does not suppress the activity of the phonological loop. *British Journal of Psychology*, *88*(4), 565–578. https://doi.org/10.1111/j.2044-8295.1997.tb02658.x

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 & Verbal Behavior*, *21*(2), 150–164. https://doi.org/10.1016/S0022-5371(82)90521-7

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

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2022). *Afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive psychology*, *33*(1), 5–42. https://doi.org/10.1006/cogp.1997.0658

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, *6*(1), 11–20. https://doi.org/10.1093/cercor/6.1.11

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual-spatial domain: Reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology*, *58*(5), 909–930. https://doi.org/10.1080/02724980443000412

Smyth, M. M., & Scholey, K. A. (1996). The relationship between articulation time and memory performance in verbal and visuo-spatial tasks. *British Journal of Psychology*, *87*(2), 179–191. https://doi.org/10.1111/j.2044-8295.1996.tb02584.x

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

Surprenant, A. M., Neath, I., Bireta, T. J., & Allbritton, D. W. (2008). Directly assessing the relationship between irrelevant speech and irrelevant tapping. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *62*(3), 141–149. https://doi.org/10.1037/1196-1961.62.3.141

Tremblay, S., Macken, W. J., & Jones, D. M. (2001). The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. *Memory*, *9*(4–6), 323–331. https://doi.org/10.1080/09658210143000010

Tremblay, S., Parmentier, F. B. R., Guérard, K., Nicholls, A. P., & Jones, D. M. (2006). A spatial modality effect in serial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*(5), 1208–1215. https://doi.org/10.1037/0278-7393.32.5.1208

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movements. *Psychonomic Bulletin & Review*, *13*(3), 452–457. https://doi.org/10.3758/BF03193869

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), 164–177. https://doi.org/10.1037/a0025054

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, & Cognition*, *43*(4), 622–634. https://doi.org/10.1037/xlm0000330

Vachon, F., Marsh, J. E., & Labonté, K. (2020). The automaticity of semantic processing revisited: Auditory distraction by a categorical deviation. *Journal of Experimental Psychology: General*, *149*(7), 1360–1397. https://doi.org/10.1037/xge0000714

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology*, *1*(2), 121–141. https://doi.org/10.1080/02643298408252018

van den Bergh, D., Wagenmakers, E., & Aust, F. (2022). *Bayesian repeated-measures ANOVA: An updated methodology implemented in JASP* [pre-print]. PsyArXiv. https://doi.org/10.31234/osf.io/fb8zn

Vandierendonck, A. (2016). A working memory system with distributed executive control. *Perspectives on Psychological Science*, *11*(1), 74–100. https://doi.org/10.1177/1745691615596790

Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology*, *16*(2), 317–322. https://doi.org/10.1080/01688639408402641

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory*, *13*(3–4), 308–317. https://doi.org/10.1080/09658210344000279

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Received November 30, 2021
Revision received March 5, 2024
Accepted March 8, 2024 ■