The role of spatial location in irrelevant speech revisited: A registered replication of Jones and Macken (1995)

Mitra Hassanzadeh[1], Florian Kattner[1], & Wolfgang Ellermeier[2]

[1] Health and Medical University

[2] Technical University of Darmstadt

Author note

This is a State-1 Report for Scheduled Review. Authors contributed equally.

Correspondence concerning this article should be addressed to Florian Kattner, Institute for Mind, Brain and Behavior, Department of Psychology, Schiffbauergasse 14, 14467 Potsdam, Germany. E-mail: florian.kattner@health-and-medical-university.de

Abstract

The goal of the present investigation is to perform a preregistered replication of Jones and Macken's (1995b) study, which showed that the segregation of a sequence of sounds to distinct spatial locations reduced the detrimental effects of irrelevant speech on short-term memory. Thereby, it postulated an intriguing connection between the psycho-acoustical concept of spatial auditory stream segregation and the cognitive mechanism underlying the irrelevant speech effect. Specifically, it was found that spoken utterances (e.g., "V-J-X") were less disruptive in a "stereo" condition in which each auditory event (each letter) could be allocated to a separate location (right ear, left ear, center), compared to when the sounds were played in "mono", which does not allow for such streaming-by-location. It is important to replicate this influential study with enhanced statistical power, due to its relevance for probing the classic as well as more recent accounts of the irrelevant speech effect, but also since the results were somewhat equivocal, in that the stereo condition produced slightly more disruption compared to silence and compared to the respective steady-state conditions with a single repeated letter ("J-J-J"). Jones and Macken's study, which has never been replicated by a different laboratory, to our knowledge, is significant both theoretically, and from an applied perspective, since it postulates a role for the spatial distribution of sound to modulate auditory distraction with relevance, for example, for the acoustic design of the workplace.

Keywords: irrelevant speech effect, changing-state effect, streaming-by-location, auditory distraction

Word count: 3783

The role of spatial location in irrelevant speech revisited: A registered replication of Jones and Macken (1995)

The irrelevant speech effect

The irrelevant speech effect refers to the observation that working memory is strongly disrupted by task-irrelevant auditory input, particularly speech, even though it is presented at moderate levels and participants are told to ignore it. Starting from the first work on this effect (e.g., Colle & Welsh, 1976; Salamé & Baddeley, 1982), the study of auditory distraction has developed into a research domain of its own encompassing well-defined experimental paradigms and resulting in hundreds of scientific publications (for reviews, see Banbury et al., 2001; Ellermeier & Zimmer, 2014; Hughes, 2014; Marsh et al., 2021).

While the earliest theoretical explanation as to why irrelevant speech effects occur, focused on interference-by-content between to-be-remembered verbal materials encoded in the phonological store and the automatically encoded phonological elements of irrelevant speech in working memory (Baddeley & Hitch, 1974; Salamé & Baddeley, 1982), the most influential account to date assumes interference-by-process. Originally termed the object-oriented episodic record (O-OER) model, the account was proposed by Jones and co-workers (Jones et al., 1996; Jones & Macken, 1993) and postulates that disruption occurs as a result of automatically processed order information inherent to the auditory sequence (within the process of auditory scene analysis, Bregman, 1990). Specifically, a sequence of distinct objects that are perceived in irrelevant sound is assumed to interfere with the order information of to-be-remembered items during a serial recall task. The empirical basis of this account is the changing-state effect, that is, the observation that an irrelevant sound sequence consisting of multiple discriminable sounds (such as articulations of the letters "V-J-X") produces more disruption in a serial recall task (but not in non-serial memory tasks) than a sequence of repeated (steady-state) sounds (e.g. repetitions of the same letter/object, such as "J-J-J"). The idea is that changing-state sound gives rise to the formation of order cues during auditory scene analysis (i.e., pointers linking consecutive auditory objects, see Bregman, 1990), which then interfere with serial-order processing, whereas no order information is generated by steady-state sound.

In contrast, an attentional account (e.g., Bell et al., 2019; Cowan, 1995; Röer et al., 2015), provides an alternative explanation for the changing-state effect. The theory assumes that all incoming auditory stimuli receive a minimal amount of attention, which requires cognitive resources that are drawn away from the focal memory task. The degree of attentional capture, however, varies with features of the irrelevant sound. For instance, sounds that deviate from the listener's predictive model based on previous stimulation (e.g., an unexpected change in the variability of a melody; Röer et al., 2014) or sounds that are meaningful to the listener's current goals (e.g., their own name; Röer et al., 2013) are expected to capture more attention than regular or meaningless sounds. Hence, according to the attentional account, the changing-state effect may arise because a sequence of changing sounds is less predictable than a steady-state sequence. The less predictable an auditory sequence is (such as free-running speech), the larger the resource demand. If, by contrast, a steady-state stream (e.g., a repetition of identical utterances of a single letter) is presented as an irrelevant background, the attentional system will compare the incoming stimulus with the recently processed stimuli, and – due to the high predictability of the sequence - will require only a very basic amount of attentional resources for this kind of monitoring (Bell et al., 2019, p. 501). Note that, in contrast to the changing-state hypothesis, an attentional theory of auditory distraction postulates that steady-state effects should be measurable and significant in the irrelevant speech paradigm.

This theoretical controversy is relevant when reconsidering an early and influential study in the history of research on the irrelevant speech effect that claimed to demonstrate a simple manipulation to turn a changing-state sequence of letters into what perceptually amounts to three steady-state sequences by means of spatial panning (shifting the stereo signal between the left and right headphones). Thus, multiple empirical tests of the changing-state hypothesis were performed by Jones and Macken (1995b). Their study is both crucial and elegantly conceived, in that (a) it shows that principles of perceptual organization (Bregman, 1990; Handel, 1989) are involved in determining whether a changing-state effect will be observed or not, (b) it is the first to provide evidence that the spatial arrangement of the irrelevant-speech scenario may play a role, and (c) it provides a crucial and frequently cited example of contrasting a steady-state with a changing-state effect, the difference of which is caused by a small manipulation in the way an identical sequence of irrelevant sounds is presented via headphones.

Jones and Macken's (1995b) classic experiment

Jones and Macken (1995b) devised a typical irrelevant speech experiment with participants having to remember a sequence of visually presented letters (F, K, L, M, Q, R, Y in a different random order on each trial) while irrelevant speech was played back over headphones, with silence constituting a control condition. The irrelevant speech was composed of three different spoken letters being played back in a repeating loop (e.g., "V-J-X" in their Experiment 1b). The crucial manipulation, however, with respect to the spatial sound configuration at issue was that a "mono" version of the irrelevant letter sequence was contrasted with a "stereo" version. In the mono version, the three letters were presented in fast succession (200 ms per letter) and monophonically, i.e. diotically (same signal in both ears, see Figure 1, top left) delivered via headphones. In the stereo version, one letter ("V") was played on the right channel only, followed by a different letter ("J") on the left channel only, and the third letter ("X") was simultaneously played on both stereo channels (see Figure 1, top right). Due to auditory streaming with stereophonic headphone presentation, a sequence of "V"s should be heard in the right ear, a sequence of "J"s in the left ear, and an "X" sequence in the center of the head (see Figure 1, bottom right).

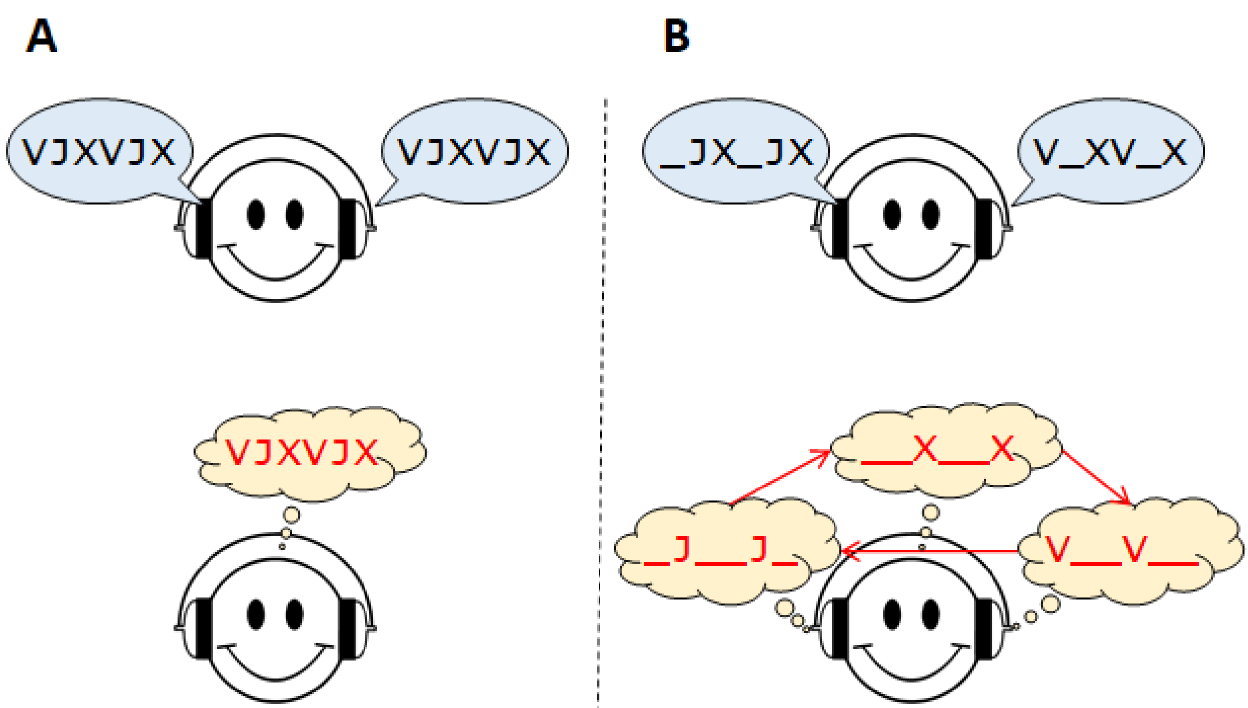

Figure 1: Schematic illustration of two irrelevant speech conditions used by Jones & Macken (1995b). (A) Monophonic changing-state presentation (top) leading to the perception of a changing-state stream of auditory objects (bottom). (B) Stereophonic changing-state presentation

(top), leading to the perception of three spatially separated steady-state streams in the left ear, right ear and the center of the head (bottom).

More specifically, in line with the idea of auditory scene analysis (a concept very influential at the time due to the work by Bregman, 1990), the left-ear, right-ear, and diotic (both-ear) stimuli are perceptually split into three different auditory sequences (or streams) emanating from three distinct locations. If such streaming-by-location occurs (Barsz, 1991; Hartmann & Johnson, 1991), what the listener perceives in the stereo condition (Figure 1, bottom right) is actually three distinct, localized non-changing letter sequences (a unique letter being repeated at regular intervals in each stream), which amounts to three steady-state percepts, rather than a changing-state percept as with monophonic presentation (Figure 1, bottom left).

Consequently, the crucial prediction made by Jones and Macken (1995b) is that the changing-state mono condition should be significantly more disruptive to serial recall than the stereo condition in which the auditory scene is organized perceptually into three separate steady-state streams. This prediction is largely borne out by the data, in that three independent experiments (Experiments 1a, 1b, and 1c, distinguished by slight variations in the letters used for the irrelevant sound, or in the respective control conditions implemented) all show the mono versions to be significantly more disruptive to serial recall than the (spatially streamed) stereophonic sound conditions.

Furthermore, two control conditions implemented in Jones and Macken's (1995b) Experiment 1c are of particular interest: They consisted of steady-state sequences "J-J-J" (not contained in Experiments 1a and 1b) either presented in mono by means of diotic headphone presentation, or in stereo, the latter resulting in three perceptual steady-state streams of repetitions of the letter "J" alternating between right, left, and center-of-the head localization. Both of these conditions produced no disruption at all when compared with the silent control. In our view, these steady-state control conditions from Experiment 1c constitute a true steady-state reference and allow to tease apart the irrelevant speech effect (comparison with silence) into a changing-state effect (changing condition vs. steady condition) and a steady-state effect (steady condition vs. silence). Note that this kind of reference condition is needed to be able to assess whether the release from interference caused by spatial streaming (in the "V-J-X" stereo condition) actually reduces to a perceptual steady-state condition (or three steady-state streams, according to Jones and Macken's reasoning). Of course, as Jones and Macken (1995b) argue, observing a small disruptive effect of the streamed steady-state condition may be due to the fact that the residual disruptive effects of three steady-state streams simply add up (a speculation not supported by the results of Experiment 1c), or, alternatively, that the perceptual streaming is unstable (as reasoned by Jones et al., 1999). Since the steady-state control conditions were not presented in Experiments 1a and 1b, any conclusion about these conjectures hinges on the equivocal outcome of Experiment 1c (where, for example, the steady-state control conditions came out slightly better than silence).

Another repetition of the original experiment was reported four years later (Jones et al., 1999): They replicated Experiment 1b with essentially the same outcome: The stereo condition where the three letters "stream out" to three distinct spatial locations (right column of Figure 1) produced significantly less memory disruption than the mono condition consisting of a single stream looping through the same set of letters (left column of Figure 1), but again there was no contrast with the corresponding steady-state conditions in mono and stereo.

This result is (1) significant in the web of empirical results determining what factors modulate memory disruption by irrelevant speech, particularly, since, among the many (psycho)acoustical factors modulating the irrelevant speech effect (such as pitch changes or spectral detail; see Ellermeier & Zimmer, 2014), it is the first to demonstrate an effect of the spatial layout of the irrelevant sound background in affecting the amount of auditory distraction. It is further (2) theoretically significant, in that it supports the Object-Oriented Episodic Record (O-OER) model proposed at the time (Jones et al., 1996; Jones & Macken, 1993) which has since become integrated into a more general *interference-by-process* account of auditory distraction (Jones & Tremblay, 2000; Marsh et al., 2009). It does so by invoking the concept of perceptual organization (Bregman, 1990) or auditory stream segregation as a vehicle to switch between steady-state and changing-state percepts. Specifically, following this rationale, the stereo condition implemented by Jones and Macken (1995b) turned a changing-state sequence into three perceptual steady-state sequences. Note, that the idea that perceptual organization or streaming plays a significant role in modulating irrelevant sound effects is supported by other studies in which increasing the frequency separation between tones constituting the irrelevant-sound sequence (pitch streaming: Jones et al., 1999) or speeding up the distracting tone sequence (streaming by tempo: Macken et al., 2003) both reduced the magnitude of the irrelevant sound effect observed. The demonstration that streaming-by-location is suited to attenuate irrelevant sound effects, however, largely hinges on the evidence provided by the work to be replicated here.

Robustness of the original finding

In our view, the crucial experiment of the classic study by (Jones & Macken, 1995b) should be replicated for three reasons:

First, it is the first study manipulating the spatial arrangement of distractor sources in the irrelevant speech effect. Though it does so with only minimal, highly artificial means to spatialize the sound, i.e. by using monotic (left, right) vs. diotic (center) headphone presentation, resulting in a lateralized or central sound image in the listener's head, Jones and Macken's (1995b) study remains the uncontended reference as to spatial sound perception modulating the irrelevant speech effect. Its only competitor, though with quite a different rationale, is a study by the same authors, published in the same year (Jones & Macken, 1995a) similarly exploiting the spatial properties of the irrelevant speech to further explore both the interference-by-process account and its relation to auditory stream segregation. They placed the listeners in an environment of "babble speech": While a mix of 6 speakers talking (the standard babble condition) produced only moderate memory impairment, disruption increased significantly when each of the 6 human speakers was played back by one of 6 loudspeakers arranged in a circle around the listener (the streamed babble condition). Just like the outcome of Jones and Macken's (1995b) original study, this is consistent both with the interference-by-process account, and the idea that auditory streaming (by location) will exert an influence, since the babble becomes more noise-like (and thus less segmented) the more speakers are added to the audio track, whereas the streamed babble allows the listener to process the much more segmented utterances of individual speakers originating from different directions in space.

Second, Jones and Macken's (1995b) study is statistically underpowered: Even though they obtain a spatial-streaming effect (statistically significant difference between their mono and stereo conditions) in three separate experiments (1A-C), each of them is based on data from twenty participants only. However, a simulation-based power analysis of the effects reported in Exp. 1C revealed that considerably more participants are needed to reliably demonstrate the crucial streaming effects in a single experiment (N = 20 provides about 51% power for the crucial contrast; see below). Given the theoretical impact of these effects, the aim of the present study is to provide an independent well-powered replication of Jones and Macken's (1995b) Exp. 1C in a different laboratory.

Third, the results are not as clear-cut as Jones and Macken's (1995b) interpretation suggests: Ideally, the critical stereo condition affording spatial streaming into three steady-state sources should reduce the changing-state effect to a residual steady-state effect. Nevertheless, the memory disruption produced by the stereo condition in all three experiments is still substantial, with mean performance levels falling roughly midway between the silent control and the most disruptive monophonic changing-state condition (see Figure 2). That might suggest that this is not just a steady-state effect (which typically is hardly measurable with so few participants, see Bell et al., 2019) but a residual changing-state effect, which may be due to either imperfect or unstable streaming (Jones et al., 1999) or the spatial switching between locations (although there is little support for the latter; see Jones & Macken, 1995b, Exp. 1C, Fig. 3). To address this issue, a true steady-state control condition like the one implemented in Experiment 1c (see the two rightmost bars in Figure 2) is required for comparison; provided, of course, it is equipped with sufficient statistical power to potentially distinguish performance from that in the quiet control.

Thus, a replication of Jones and Macken (1995b) might - from a theoretical perspective not available at the time - also permit us to assess whether the kind of attentional capture postulated by the attentional account might better explain the results than the changing-state hypothesis. If a call for attention is issued whenever, and to the extent that a new stimulus deviates from its predecessor, clearly, the monophonic condition in which different letters are presented in succession should be the most distracting, a prediction not differing from that made by the changing state hypothesis. If, on the other extreme, one and the same letter is presented repeatedly and in mono, a minimal but robust amount of distraction is expected by the attentional account due to the obligatory monitoring of the albeit constant auditory input. If the amount of distraction depends on the mismatch between the current distractor sound and a neural, predictive model built from the features of previous distractors (Escera et al., 1998), then a prediction for the crucial stereo condition in Jones and Macken (1995b) may be derived: Perceptually assigning the three different distractor letters (e.g. “V-J-X”) employed in the stereo condition to three in-the-head locations will certainly not diminish the degree of mismatch between successive distractors, on the contrary, that mismatch will increase, since the existing spectro-temporal differences between the letter sounds are supplemented by interaural intensity differences resulting in changing lateralizations. Note, however, that the procedure chosen by Jones and Macken (1995b), in which a repeating sequence of three fixed letters is played throughout each trial (“V-J-X-V-J-X…”) makes the upcoming utterances perfectly predictable after a few seconds. Hence, a predictability-based attentional capture account, would not predict much more disruption in the changing-state than in the steady-state conditions. Furthermore, the stereophonic control condition of Jones and Macken’s Experiment 1C, where one and the same letter alternates between three locations, might be thought to require slightly more attentional resources than the monophonic steady-state condition (repetitions of the same letter at a single location), since the changing spatial position of that letter (driven by the interaural level differences) constitutes a changing sound feature. Thus, though - admittedly - the attentional account is not as rigorously formulated as the interference-by-process account, different predictions may be derived tentatively from the two competing models.

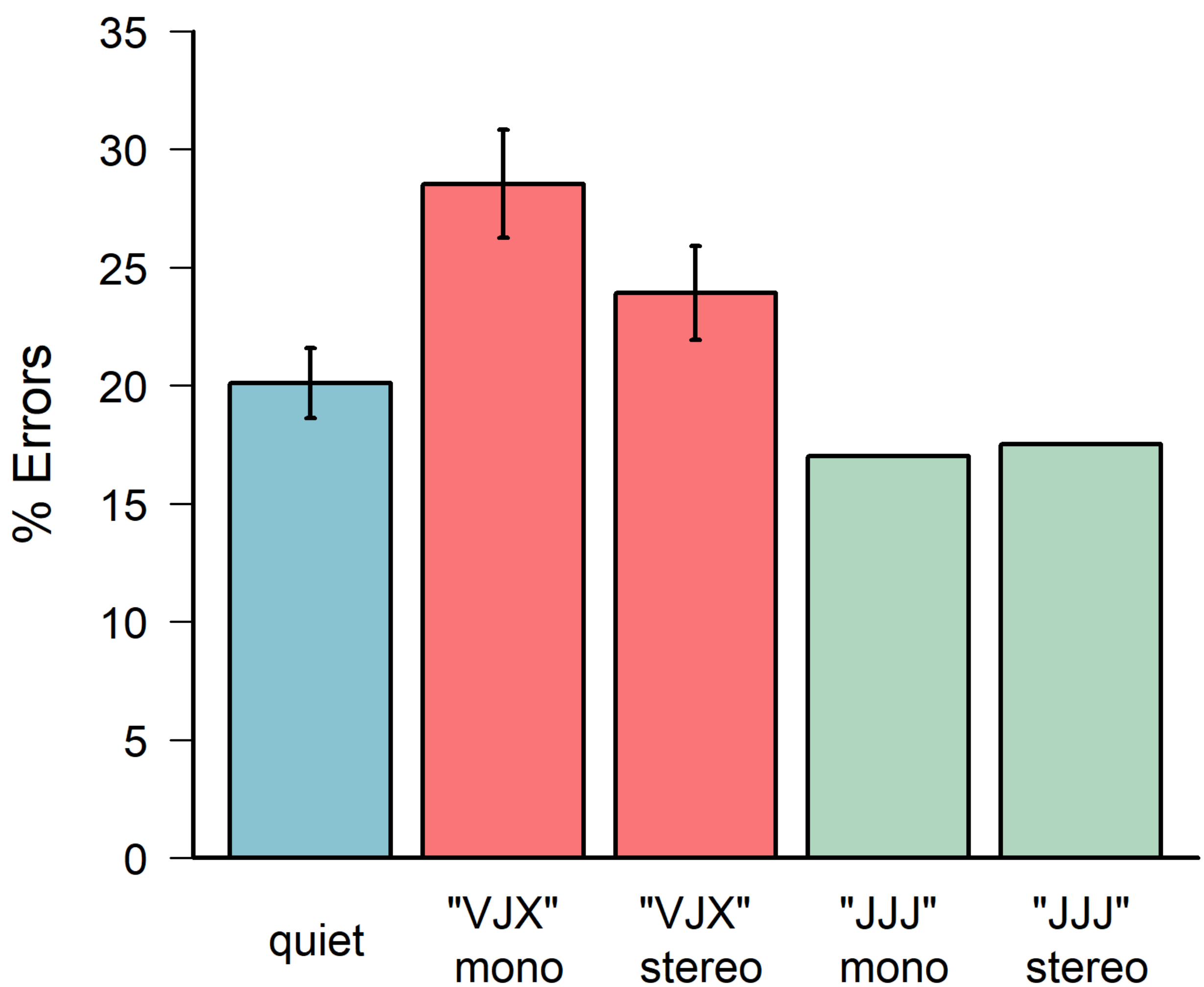

Figure 2: Summary of the results reported by Jones & Macken (1995b; Exp. 1A-1C) and Jones et al. (1999; Exp. 1). As can be seen, collapsed across four experiments the stereo “VJX” changing-state condition (perceptually a steady-state condition) appears to produce more recall errors

compared to the real steady-state conditions (“JJJ” in terms of stimulus presentation; which were only contained in Exp. 1C). Error bars represent standard errors of the means computed across serial positions and all four experiments (not available for the two control conditions that were exclusively contained in Experiment 1C).

The criticisms of low statistical power and ambiguous results, incidentally, also apply to the earlier replication by Jones and colleagues (1999). Therefore, Jones and Macken’s (1995b) Experiment 1c, which contains all relevant control conditions including and beyond those used in Experiments 1a and 1b (and Jones et al., 1999b), is to be replicated as closely as possible (making it a ‘direct’ replication, see Zwaan et al, 2018) and with sufficient statistical power.

Hypotheses for the present replication

Two specific hypotheses are tested with the present replication of the original study (Jones & Macken, 1995b): (1) The stereophonic changing-state condition, by virtue of spatial streaming, should induce significantly fewer serial recall errors than the corresponding changing-state mono condition. (2) If, in fact, spatial streaming reduces the changing-state effect by producing nothing but a steady-state effect in the stereophonic changing-state condition, then the disruption should be equivalent, or not significantly different, from the corresponding stereophonic steady-state condition.

Method

Participants

We note that this replication project is motivated partially by the fact that the original study (Jones and Macken, 1995b; Exp. 1C) may have been underpowered with only 20 participants. To determine the statistical power, simulations of the replication experiment were conducted in R to determine the minimum sample size required to observe both the main effect of auditory condition (5-level factor in a repeated-measures ANOVA) and the crucial contrast between the monophonic and stereophonic changing-state conditions (using corrections for multiple comparisons, Holm, 1979) with a power of 95% ($\alpha = .05$). For the simulations, we conservatively assumed population means and standard errors of recall errors as depicted by Jones and Macken (1995b, Fig. 3, M_{quiet} = 18.0; $M_{\text{changing-state/mono}}$ = 22.7; $M_{\text{changing-state/stereo}}$ = 19.0; $M_{\text{steay-state/mono}}$ = 17.0 vs $M_{\text{steady-state/stereo}}$ = 17.5; *SD* = 5.6; d_z = 0.33 for the crucial mono/stereo contrast with changing-state speech). The simulation was conducted with 2000 iterations for each sample size ranging from 20 and 80 participants (in steps of 5) and it revealed both a significant main effect of auditory condition and the crucial contrast (stereo/mono changing-state) in 95% of the simulated experiments ($1 - \beta = .95$) with a sample size of $N = 60$ participants or more (see Fig. 3; the code to run the power simulations is available at https://osf.io/hvp58).

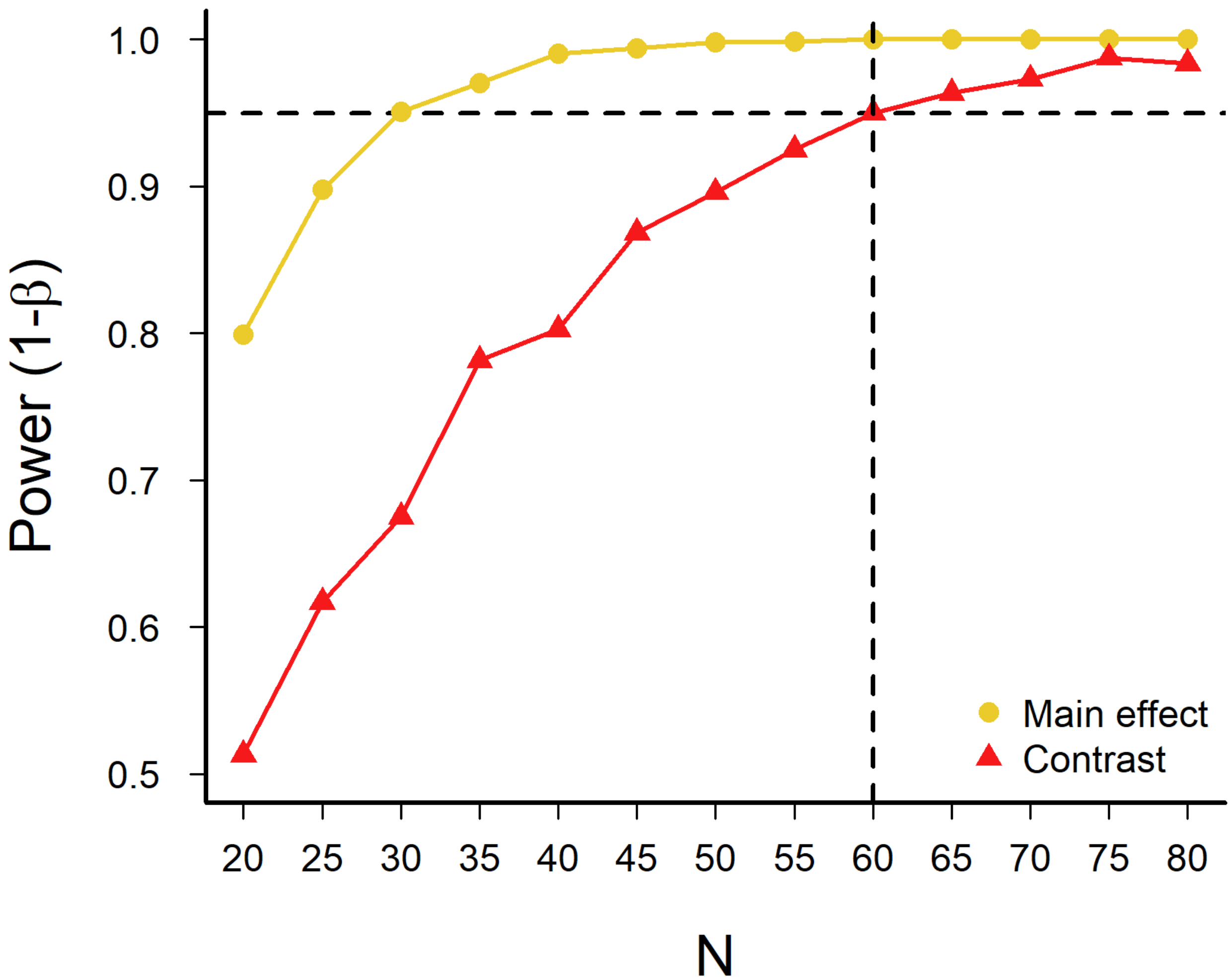

Figure 3: Statistical power to observe the main effect of auditory condition as well as the crucial contrast between mono and stereo changing-state speech as a function of the sample size (2000 iterations each, based on the means and standard deviations as reported by Jones & Macken, 1995b, Exp. 1C).

The 60 participants will be recruited at Health and Medical University, and will be compensated with partial course credit. Only participants who reported no hearing loss and normal or corrected-to-normal vision will be included for this study.

Apparatus and Stimuli

The experiment was conducted individually in a sound-attenuated listening booth. The experimental routines were programmed in PsychoPy (Peirce, 2007; Peirce et al., 2019). The to-be-remembered items consisted of the letters “F”, “K”, “L”, “M”, “Q”, “R”, and “Y”. The to-be-ignored auditory stimuli consisted of recordings of the three letters “V”, “J”, and “X”, all pronounced in UK English by a female computer-generated voice (using the text-to-speech generator “naturalreaders.com” with the voice “Libby” producing utterances at a rate of 360 words/min, 44.1 kHz sampling rate). Each recording had a duration of 200 ms. Four types of auditory materials were prepared. 17-s sequences of steady-state and changing-state speech (mono) were generated by repeating the letter triplets “J-J-J” and “V-J-X”, respectively (i.e., 85 letters per sequence) and recording them for diotic headphone presentation. In addition to the mono conditions, two additional stereo conditions were created by presenting the spoken letters dichotically, allocating each letter of the triplet to a unique location (first letter to the left channel, second letter to the right channel, and third letter to both channels, as in Figure 1B, top). Hence, in the stereo changing-state condition, a sequence of the letter “V” was heard only in the left ear, whereas a sequence of the letter “J” was heard in the right ear and the “X” sequence was heard in the center of the head (see Figure 1B, bottom). In the stereo steady-state condition, spatially streamed sequences of the letter “V” should be heard at each location (at the same rate). Silence was used as a control condition. All sounds were played with an average sound pressure level of about 65 dB(A), as measured in the headphones.

Experimental Design and Procedure

A 7 (serial position) x 5 (auditory condition: quiet, mono steady-state, mono-changing-state, stereo steady-state, stereo-changing-state) repeated-measures design was implemented. Each auditory condition was repeated 16 times, resulting in a total of 80 trials.

The experiment started with five practice trials in quiet, followed by 80 experimental trials with the five auditory conditions presented in full random order. The participant started each trial by pressing the spacebar. The sound sequence was then started immediately, and after 1 s the seven to-be-remembered letters were presented sequentially on the screen. The visual letters were presented in random order at a rate of 1/s, with each letter being shown on the screen for 800 ms, followed by a 200-ms inter-stimulus interval. After the seventh letter, a text message ("Bitte warten" [Please wait]) was presented on the screen for the remaining 10 s (the retention interval), indicating that participants should wait for the response screen. The irrelevant sound was presented both during visual letter presentation and the retention interval, accumulating to a total sound duration of 17 s. A response matrix with eight boxes showing all seven letters and a question mark was then presented on the screen and participants were asked to click the letters in the order they were presented. Participants could use the question mark if they did not recall a specific letter, but there was no option to correct their responses. The entered digits were shown at the top of the screen. No feedback on accuracy was provided and the next trial started immediately after the last digit was entered.

Data Analysis

As in the original article, a 7 (serial position) x 5 (auditory condition: quiet, steady mono, steady stereo, changing mono, changing stereo) repeated-measures ANOVA of the proportion of recall errors in the serial recall task will be conducted to demonstrate significant main effects of serial position and auditory condition. No interaction is expected (in line with Jones & Macken, 1995b; Exp. 1C). To disentangle the disruptive effects of the different sound conditions, Holm (1979) corrected contrasts will be tested between quiet and all other auditory conditions, with the only significant contrast being expected for the "changing mono" condition (as reported by Jones & Macken, 1995b; p. 196). In addition, the crucial contrast between "changing mono" and "changing stereo" will be tested to demonstrate that the stereo presentation mode significantly reduced the number of recall errors with changing-state sound (i.e., through generating the perception of three spatially segregated steady-state streams). Finally, in addition to the tests reported by Jones and Macken (1995b), the changing-state effects (i.e., the difference in recall errors between the changing- and steady-state conditions) are to be contrasted between mono and stereo presentation modes, in order to demonstrate that the disruptive effect of changing-state sound was significantly reduced in case of stereo only.

Acknowledgment

The R package {papaja} (Aust & Barth, 2022) was used to prepare this manuscript.

References

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://github.com/crsh/papaja

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). Academic Press. https://doi.org/10.1016/j.cub.2009.12.014

Banbury, S., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, *43*, 12–29. https://doi.org/10.1518/001872001775992462

Barsz, K. (1991). Auditory pattern perception: The effect of tone location on the discrimination of tonal sequences. *Perception & Psychophysics*, *50*(3), 290–296. https://doi.org/10.3758/BF03206752

Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology*, *47*(6), 1182–1191. https://doi.org/10.1111/j.1469-8986.2010.01029.x

Bell, R., Röer, J. P., Dentale, S., & Buchner, A. (2012). Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(6), 1542–1557. https://doi.org/10.1037/a0028459

Bell, R., Röer, J. P., Lang, A. G., & Buchner, A. (2019). Distraction by steady-state sounds: Evidence for a graded attentional model of auditory distraction. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(4), 500–512. https://doi.org/10.1037/xhp0000623

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

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

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195119107.001.0001

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

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

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.

Hammershøi, D., & Moller, H. (2005). Binaural technique - Basic methods for recording, synthesis, and reproduction. In *Communication acoustics* (pp. 223–254). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27437-5_9

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. MIT Press. https://doi.org/10.2307/898748

Hartmann, W. M., & Johnson, D. (1991). Stream Segregation and Peripheral Channeling. *Music Perception*, *9*(2), 155–183. https://doi.org/10.2307/40285527

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. https://www.jstor.org/stable/4615733

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*(1), 30–41. https://doi.org/10.1002/pchj.44

Jones, D. M., Beaman, P., & Macken, W. J. (1996). The object-oriented episodic record model. In *Models of short term memory* (pp. 209–240). Psychology Press.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 369–381. https://doi.org/10.1037/0278-7393.19.2.369

Jones, D. M., & Macken, W. J. (1995a). Auditory Babble and Cognitive Efficiency: Role of Number of Voices and Their Location. *Journal of Experimental Psychology: Applied*, *1*(3), 216–226. https://doi.org/10.1037/1076-898X.1.3.216

Jones, D. M., & Macken, W. J. (1995b). 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., Saint-Aubin, J., & Tremblay, S. (1999). Modulation of the Irrelevant Sound Effect by Organizational Factors: Further Evidence from Streaming by Location. *The Quarterly Journal of Experimental Psychology Section A*, *52*(3), 545–554. https://doi.org/10.1080/713755832

Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin and Review*, *7*(3), 550–558. https://doi.org/10.3758/BF03214370

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., Kattner, F., & Ruhnau, P. (2021). Research Collection: On Theoretical Advancement in Auditory Distraction Research. *Auditory Perception & Cognition*, *4*(3-4), 133–138. https://doi.org/10.1080/25742442.2022.2036524

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

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

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41,* e120. https://doi.org/10.1017/S0140525X17001972

Question	**Hypothesis**	**Sampling plan**	**Analysis Plan**	**Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis**	**Interpretation given different outcomes**	**Theory that could be shown wrong by the outcomes**
Is Jones and Macken's (1995) study replicable, in that the auditory separation of a changing sound sequence into three non-changing perceptual "streams" will reduce auditory distraction?	A "changing-state stereo" condition, by virtue of spatial streaming, should induce significantly fewer serial recall errors than the corresponding "changing-state mono" condition.	A power analysis based on the data reported by Jones and Macken (1995) revealed that *N* = 54 participants are required to demonstrate not only the general main effect of auditory condition, but the crucial contrast between mono and stereo versions of changing-state speech (*f* = .25) with a statistical power of 95% (α = .05).	A repeated-measures ANOVA with a planned contrasts analysis will be conducted to demonstrate a significant contrast between mono- and stereophonic versions of changing-state speech.	The relevant effect sizes were determined from the statistics reported in the study to be replicated (see sampling plan).	If the spatially separated distractor condition does not reduce serial recall errors, the "auditory streaming" interpretation of the interference-by-process account is challenged. Consequently, either other factors must be operating to produce changing-state effects, or the entire concept must be abandoned in favor of, e.g., an attentional capture interpretation.	Jones et al. (1996) object-oriented episodic record (O-OER) model and ensuing, more recent "interference-by-process" accounts.
Does the streaming-by-location effect observed in Jones and Macken's (1995) study reduce interference-by-process	Spatial streaming should reduce serial recall errors in the stereophonic "changing-state" condition to	Based on the Experiment 1C reported by Jones and Macken (1995), an effect size of *f* = .25 was estimated for the contrast in changing-state	The changing-state effects (i.e., the difference in recall errors between the changing- and steady-state conditions) will	Steady-state control conditions were only included in Jones and Macken's (1995) Exp. 1C, and the remaining experiments suggest that there is residual "changing-state"	If the spatial separation of sound sources (via stereo) does not reduce distraction to the level of "steady-state" speech, the interpretation given by Jones and Macken (1995) is challenged,	

completely (i.e., to the level of steady-state distractor sequence)?	the level of the corresponding stereophonic “steady-state” condition.	effects between mono and stereo conditions, requiring N = 54 participants.	be contrasted between mono and stereo presentation modes.	effect in the “stereo” condition. The effect size for the contrast was estimated based on the figures.	and the residual interference will have to be explained..	

Guidance Notes

- **Question**: articulate each research question being addressed in one sentence.
- **Hypothesis**: where applicable, a prediction arising from the research question, stated in terms of specific variables rather than concepts. Where the testability of one or more hypotheses depends on the verification of auxiliary assumptions (such as positive controls, tests of intervention fidelity, manipulation checks, or any other quality checks), any tests of such assumptions should be listed as hypotheses. Stage 1 proposals that do not seek to test hypotheses can ignore or delete this column.
- **Sampling plan**: For proposals using inferential statistics, the details of the statistical sampling plan for the specific hypothesis (e.g power analysis, Bayes Factor Design Analysis, ROPE etc). For proposals that do not use inferential statistics, include a description and justification of the sample size.
- **Analysis plan**: For hypothesis-driven studies, the specific test(s) that will confirm or disconfirm the hypothesis. For non-hypothesis-driven studies, the test(s) that will answer the research question.
- **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis**: For hypothesis-driven studies that employ inferential statistics, an explanation of how the authors determined a relevant effect size for statistical power analysis, equivalence testing, Bayes factors, or other approach.
- **Interpretation given different outcomes**: A prospective interpretation of different potential outcomes, making clear which outcomes would confirm or disconfirm the hypothesis.
- **Theory that could be shown wrong by the outcomes**: Where the proposal is testing a theory, make clear what theory could be shown to be wrong, incomplete, or otherwise inadequate by the outcomes of the research.