

SCIENTIFIC REPORTS



OPEN

Effects of Auditory Distraction on Face Memory

Raoul Bell¹ , Laura Mieth¹, Jan Philipp Röer² & Axel Buchner¹

Effects of auditory distraction by task-irrelevant background speech on the immediate serial recall of verbal material are well established. Less is known about the influence of background speech on memory for visual configural information. A recent study demonstrated that face learning is disrupted by joyful music relative to soothing violin music and quiet. This pattern is parallel to findings in the serial-recall paradigm showing that auditory distraction is primarily caused by auditory changes. Here we connect these two streams of research by testing whether face learning is impaired by irrelevant speech. Participants learned faces either in quiet or while ignoring auditory changing-state sequences (sentential speech) or steady-state sequences (word repetitions). Face recognition was impaired by irrelevant speech relative to quiet. Furthermore, changing-state speech disrupted performance more than steady-state speech. The results were replicated in a second study using reversed speech, suggesting that the disruptive potential of the background speech does not depend on its semantic content. These findings thus demonstrate robust effects of auditory distraction on face learning. Theoretical explanations and applied implications are discussed.

People usually prefer a quiet environment to study because they are concerned that learning may be impaired by background sounds. Research has indeed confirmed that important working-memory functions such as maintenance¹, updating², problem solving³, and binding⁴ suffer when task-irrelevant speech has to be ignored. As yet, research has mainly focused on the effects of auditory distraction on short-term memory for verbal information (e.g.⁵). Here, we test whether the disruptive effects of background speech generalize to face learning. This is relevant for the evaluation of theories about auditory distraction but also for applied contexts such as eyewitness testimonies^{6,7}.

The distracting effect of task-irrelevant background sounds relative to quiet is often referred to as *irrelevant sound effect*^{8–12}. The disruption of working-memory processes is mainly determined by the degree to which the auditory signal is changing abruptly. One of the key signatures of auditory distraction is the *changing-state effect*, which refers to the observation that auditory distractor sequences that contain many different auditory objects (such as words or tones) disrupt cognitive performance more than continuous or repetitive sounds^{9,13,14}. Most research focuses on the serial-recall paradigm in which short lists of verbal items (words, digits, or consonants) have to be recalled in the correct order, either immediately after presentation or after a short retention interval⁵. The typical finding is that changing-state sequences that are composed of different one-syllable words^{15–17} or sentential speech^{1,18,19} disrupt serial recall more than so-called steady-state sequences that consist of regular repetitions of the same distractor word. The changing-state effect is also thought to be responsible for the surprising finding that multiple voices disrupt performance less than a single voice because when listening to multiple voices the changes in each speech signal are masked by the other voices so that the resulting signal is smoothed out²⁰. The disruptive effect of irrelevant speech on working-memory processes cannot be solely attributed to its lexical, syntactic, or semantic properties as speech has a strong impact on performance even when it is incomprehensible because it is reversed^{19,21–23}. Compared to changing speech, steady-state speech consisting of repeated speech utterances such as consonants or one-syllable words cause less disruption^{9,14,15}. Until recently, it was widely believed that steady-state stimuli cause no reliable disruption at all (e.g.¹¹). However, the disruptive effect of steady-state speech is simply smaller than that of changing-state speech so that it is not always detected in statistical tests²⁴.

Auditory distraction is not limited to speech^{19,25–30}. Klatte, Kilcher, and Hellbrück³¹ showed that a musical piece played *staccato* disrupted performance whereas the same piece played *legato* did not disrupt performance relative to a quiet control condition (see also²⁷). Furthermore, sequences consisting of glissandi (continuous pitch

¹Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. ²Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany. Correspondence and requests for materials should be addressed to R.B. (email: raoul.bell@hhu.de)

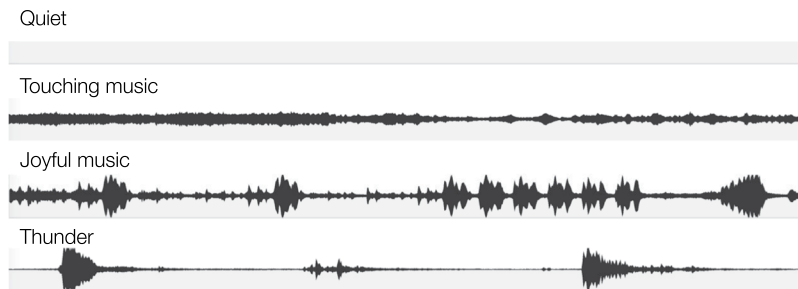


Figure 1. Examples for the stimulus material used in the study of Proverbio *et al.*³⁴: A quiet control condition, emotionally touching music (Arvo Pärt: Cantus in memoriam Benjamin Britten), joyful music (Paul Hindemith: Kammermusik No. 1, 1st movement), sound of rain with bursts of thunder (available at <https://youtu.be/WvRv-243Cmk>).

glides) did not cause distraction compared to a quiet control condition. By contrast, the same auditory sequences disrupt performance when the glissandi are interrupted by periods of silence, rendering the sequence into distinct auditory objects with abrupt onsets³². Furthermore, changing tones disrupt immediate memory more than repeated tones¹³. The disruptive effect of irrelevant non-speech sound is thus characterized by a changing-state effect, just as that of irrelevant speech^{13,14}.

While the effect of task-irrelevant sounds on working-memory processes is comparatively well understood, less is known about how auditory distractors affect the processing of complex visual stimuli (cf.³³). In particular, the question of whether face learning is affected by task-irrelevant background speech is interesting from a theoretical as well as from an applied perspective. In forensic settings, for instance, it is important to know whether the identification of a perpetrator face is affected by background noise^{6,7}.

The present study was inspired by a recent study of Proverbio *et al.*³⁴ who examined how background music affects face learning. In their experiment 300 faces were presented while participants listened to joyful music, touching music, the sound of rain, or quiet. Face recognition was better when participants had listened to touching music at encoding than when they had listened to joyful music or rain. Proverbio *et al.* discussed their results in light of an arousal-and-mood hypothesis suggesting that the “binding of facial properties with auditory and emotionally charged information (music) ... may result in deeper memory encoding” (p. 1). However, they also mentioned several aspects of their results that are more consistent with an interpretation in terms of auditory distraction. First, Proverbio *et al.*'s measures of emotional arousal (blood pressure and heart rate) were numerically higher for both joyful and touching music in comparison to rain and quiet and thus did not follow the same pattern as face recognition which was disrupted by joyful but not by touching music. Second, there was no beneficial effect of touching music on face recognition compared to the quiet control condition. Instead, face recognition was impaired when the faces were encoded while listening to joyful music or rain in comparison to quiet. This raises a theoretically interesting question: Could these effects be due to auditory distraction?

The finding that joyful music caused a large amount of interference while the emotionally touching music did not affect performance relative to quiet is consistent with a changing-state effect. In the study of Proverbio *et al.*³⁴, the joyful music contained a large number of changes in quick succession (see Fig. 1) while the emotionally touching music consisted almost entirely of glissandi. The finding that the joyful music disrupted performance more than the emotionally touching music can be explained with the general principle that auditory sequences with many abrupt changes disrupt memory more than sequences containing no such changes. At first glance, it seems surprising that the sound of rain should interfere with memory because rain usually causes a regular sound pattern that does not contain any abrupt changes. However, in the sound file used by Proverbio *et al.*, this regular pattern was interrupted by sudden bursts of thunder (see Fig. 1). Therefore, the disruption of memory in this condition in comparison to the quiet control condition is in line with the common finding that sudden deviations from regular patterns capture attention and cause distraction^{2,35–39}.

The results of Proverbio *et al.*³⁴ may thus provide evidence for an effect of auditory distraction on face learning. This finding is important because as yet evidence for a disruption of face processing by irrelevant sound is not well documented, due to a lack of systematic research on the influence of auditory distraction on visual memory (cf.³³). There are, however, a few pioneering studies^{6,7} that provide preliminary evidence for an irrelevant speech effect on face processing. In these studies, eyewitnesses were required to create a facial composite of a perpetrator face with the help of a computer program. The quality of the facial composites was affected by background speech. In the first study⁶, interference depended on whether the content of the irrelevant speech was incongruent with the target face. The verbal description of another face with different facial features caused a reduced likeness of the created facial composites while the congruent description of the face of the perpetrator did not affect the quality of the facial composites. In a follow-up study⁷, irrelevant speech, per se, did not negatively influence the quality of the facial composites and the identification of the perpetrator face in a photo lineup. Meaningful halfalogues (overhearing one side of a cell-phone conversation), by contrast, disrupted both the quality of the facial composites as well as the identification of the perpetrator face in the lineup. This finding was explained by attentional diversion. Halfalogues contain less information than dialogues but are more difficult to predict. The unpredictability of the auditory input thus seems to be a more important determinant of auditory distraction than the amount of information that has to be ignored.

While first evidence thus suggests that special properties of speech can distort the memory representations of faces⁶ and cause attentional diversion⁷, it is as yet unclear whether face learning is affected by the mere presence of irrelevant speech. In fact, irrelevant speech, per se, did not seem to have a reliable effect on face processing relative to a quiet control condition in both aforementioned studies^{6,7}. However, Marsh *et al.*⁷ refer to an unpublished study providing evidence in favor of a changing-state effect on face learning. Lineup identification accuracy was worse when a changing-state sequence (made up of different letters) had to be ignored during the encoding of the perpetrator face than when a steady-state sequence (made up of a repeated letter) had to be ignored. While the two studies providing evidence against an irrelevant speech effect^{6,7} focused on the *incidental* learning of faces, the unpublished study providing evidence in favor of such an effect focused on the *intentional* learning of faces, just as the study of Proverbio *et al.*³⁴. This suggests that intentional face learning may indeed be affected by irrelevant speech. However, given that this assumption as yet relies only on a single unpublished study, more research is needed to determine the robustness of the irrelevant speech effect on intentional face learning. The present study focuses on the distraction by sentential speech. As the central medium of human communication, sentential speech is a ubiquitous part of many acoustic environments. It is therefore important to understand its disruptive power on memory for theoretical as well as for applied reasons.

The present study aims at testing whether intentional face learning is impaired by irrelevant speech. The study was modeled after that of Proverbio *et al.*³⁴ in that it involved an intentional learning procedure but speech instead of music was played during face presentation. As in our previous studies^{38,40–42}, we contrasted changing-state sequences consisting of sentential speech with so-called steady-state sequences in which the same word was repeatedly presented. To arrive at clear conclusions about the vulnerability of face processing to auditory distraction, we averaged over a large number of faces and used large sample sizes of over 100 participants in each study to decrease the impact of unsystematic variance and to obtain sufficient statistical power.

Study 1

Method. *Participants.* The final sample consisted of 103 participants who were recruited at Heinrich Heine University Düsseldorf (74 female; mean age of 23, $SD = 5$). Three datasets had to be removed because three students had participated twice (the data sets from their first participation were retained).

Materials. For each participant, color photographs (256 × 384 pixels) of 60 female and 60 male faces were randomly drawn from a pool of 100 adult female and 100 adult male faces from the FERET database⁴³. Each participant saw a different, randomly selected set of faces. For the encoding phase, 15 sequences of four faces were created by randomly drawing faces from the pool (without replacement) with the restriction that each sequence contained two male and two female faces in random order. Each face was shown in frontal view at the center of the computer screen. Sixty other faces (30 female, 30 male) were randomly drawn from the same pool to serve as new faces in the old-new recognition test.

For the auditory distractor sequences, we used the same stimulus material as in the study of Röer *et al.*⁴⁰. In the changing-state condition, sentential speech was used (e.g., “Peel and quarter the onions and slice them into thin pieces, then add the tomatoes, then simmer at medium heat”; translated from German). The sentences were taken from eight different categories (weather forecast, prose text, cooking recipe, scientific textbook, poem, operating manual, road message, aphorism). For each steady-state sequence, a monosyllabic word (from the sentence) was randomly selected and repeated 18 times. This corresponded to the mean number of words in the changing-state sequences. The auditory sequences were spoken by a male voice and lasted 8 s each. With this material, a robust changing-state effect on serial recall has been obtained in several previous studies^{30,38,40–42,44}. The sounds were played binaurally at about 65 dB (A) using headphones with high-insulation hearing protection covers that were plugged directly into the Apple iMac computer which controlled the experiment.

Procedure. A standalone application built with LiveCode [<https://livecode.com>] controlled the experiment. Participants received standard written instructions on the computer screen informing them that they should concentrate only on the faces which have to be recognized later and ignore the speech which is completely irrelevant for the task.

For each participant, the sequences of faces were randomly assigned to the three conditions (quiet, steady state, changing state). Participants started each trial by pressing a “continue” button. This initiated the presentation of the auditory distractor sequences. After 1200 ms, the first face was shown. The four faces of a sequence were presented one after another, each for 700 ms with a 1200 ms inter-stimulus interval. Immediately after the offset of the last face, the “continue” button appeared with which participants started the next trial. Pilot testing had shown that it was necessary to repeat the presentation of the faces to obtain above-floor memory performance when presenting a large number of faces in quick succession, as in the study of Proverbio *et al.*³⁴. Therefore, all of the 15 face sequences (each consisting of 4 faces) were presented three times, each time in a new, randomly determined order. All 60 faces were shown before being repeated. To prevent habituation to the auditory sequences (cf.¹⁹), the faces assigned to the distractor conditions were accompanied by different changing-state or steady-state sequences each time they were presented. All encoding trials were completed prior to the test phase.

In the test phase, the 60 old faces (20 of each auditory condition) were intermixed with 60 new faces. The participants’ task was to classify each face as old or new. For each participant, the faces were presented in random order. The order of the faces at test was thus entirely unrelated to the order with which the faces were presented at encoding. In each trial of the recognition test, a face was presented in the middle of the screen along with the headline “Was this face old or new?”, and two buttons (labeled “old” and “new”). Upon clicking one of the buttons, a “continue” button appeared. Upon clicking this button, the next face was shown.

Written informed consent was obtained from all participants. Ethical approval had been obtained from the ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf.

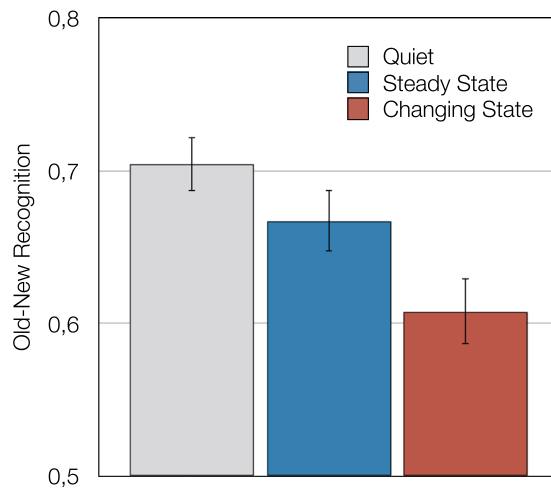


Figure 2. Old-new face recognition in terms of the sensitivity measure of the two-high threshold model⁴⁵—measured by hit rate minus false alarm rate—as a function of distractor condition (quiet, steady state, changing state) in Study 1. All distractors were played in forward direction. The error bars represent the standard errors of the means.

The study involved standard procedures in accordance with the ethical guidelines published by the American Psychological Association and the Declaration of Helsinki. The data are available at <https://osf.io/ksqx9/>.

Results. Figure 2 shows old-new recognition as a function of distractor condition in terms of P_r . The sensitivity measure of the two-high threshold model was chosen because it has been positively evaluated in validation studies (e.g.⁴⁵) and avoids the problem of undefined values that comes with using d' . The measure is often referred to as “corrected hit rate” because it is calculated by subtracting the false alarm rate from the hit rate. There was a main effect of auditory distraction on face recognition, $F(2,101) = 20.38$, $p < 0.001$, $\eta_p^2 = 0.29$. Orthogonal contrasts showed that performance in the quiet control condition was better than performance in the two distractor conditions, $F(1,102) = 33.12$, $p < 0.001$, $\eta_p^2 = 0.25$, which provides evidence of an irrelevant speech effect. Furthermore, changing-state sequences disrupted performance more than steady-state sequences, $F(1,102) = 15.93$, $p < 0.001$, $\eta_p^2 = 0.14$, which represents evidence of a changing-state effect. Note that an analysis of d' leads to the same statistical conclusions.

Discussion. The results of Study 1 confirm that intentional face learning is disrupted by distractor speech. Furthermore, the pattern of findings is consistent with a changing-state effect. Changing-state sequences consisting of sentential speech disrupted face learning more than steady-state sequences consisting of word repetitions. According to the changing-state explanation, the increased distraction caused by sentential speech in comparison to word repetitions is primarily due to auditory changes in the distractor material.

However, the changing-state sequences did not only comprise more changes, they also contained more semantic information than the steady-state sequences. In principle, the data pattern of Study 1 could be explained by assuming that the semantic processing of the sentences diverted resources away from the face-learning task. This semantic-interference hypothesis has the disadvantage that it cannot account equally well for the disruption of face learning by sequences of non-speech sounds observed in the study of Proverbio *et al.*³⁴. Furthermore, the literature on auditory distraction suggests that semantic content of the irrelevant speech has no influence on distraction unless the primary task strongly depends on semantic processing⁴⁶ or unless the speech captures attention because it is emotionally salient⁴⁷ or violates expectations⁴⁸. Furthermore, it is well established that the semantic processing of speech is not a necessary condition for obtaining auditory distraction effects on memory as such effects have often been observed with non-speech sounds such as instrumental music^{14,49} as well.

Nevertheless, to substantiate that the semantic processing of distractor speech is not a necessary condition for obtaining a disruptive effect on face learning, it is important to replicate the changing-state effect with distractor material that does not lend itself equally well to semantic processing as forward speech. Within the literature on auditory distraction, this is typically done by using reversed distractor speech that has been shown to disrupt working-memory processing in several studies^{19,21–23}. In Study 2, we used the same stimulus material as in Study 1 but the distractor sequences were played in reverse. Acoustically, reversed speech has similar properties (i.e., contains a similar amount of acoustic changes) as forward speech. The changing-state hypothesis therefore implies that both the irrelevant speech effect and the changing-state effect should be observed with reversed speech as well.

Study 2

Method. *Participants.* The sample consisted of 106 participants who were recruited at Heinrich Heine University Düsseldorf (70 female; mean age of 24, $SD = 4$).

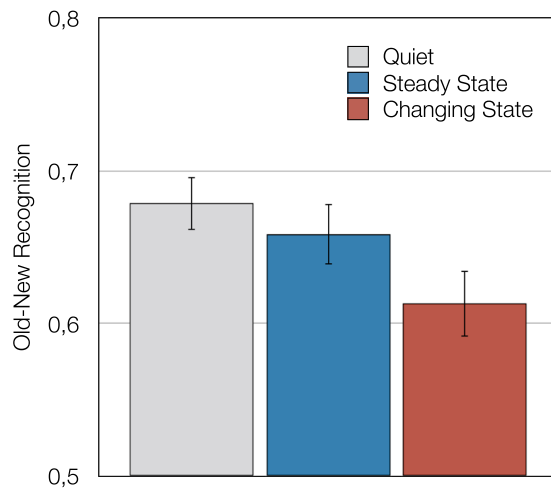


Figure 3. Old-new face recognition in terms of the sensitivity measure of the two-high threshold model⁴⁵—measured by hit rate minus false alarm rate—as a function of distractor condition (quiet, steady state, changing state) in Study 2. All distractors were played in backward direction. The error bars represent the standard errors of the means.

Materials and procedure. Materials and procedure were identical to those of Study 1 with the following exceptions. The same sound files were used as in Study 1, but the sound files for the changing-state sequences as well as the sound files for the steady-state sequences were reversed. For each participant, one steady-state sequence was randomly selected. This sequence was played in all steady-state trials throughout the experiment.

Results. Figure 3 shows face recognition as a function of distractor condition in terms of P_r . There was a main effect of auditory distraction on face recognition, $F(2,104) = 12.13$, $p < 0.001$, $\eta_p^2 = 0.19$. Orthogonal contrasts showed that performance in the quiet control condition was superior to that in the two distractor conditions, $F(1,105) = 14.26$, $p < 0.001$, $\eta_p^2 = 0.12$, which provides evidence of an irrelevant speech effect. As in Study 1, changing-state sequences disrupted performance more than steady-state sequences, $F(1,105) = 10.06$, $p = 0.002$, $\eta_p^2 = 0.09$, which represents evidence of a changing-state effect. As in Study 1, an analysis of d' leads to the same statistical conclusions.

Discussion. In Study 2, the effect of auditory distraction on intentional face learning was successfully replicated. As in Study 1, face recognition was better when the faces were encoded in quiet than when distractor speech had to be ignored. This finding generalizes the irrelevant speech effect to face learning. Furthermore, there was also a significant changing-state effect in that sentential speech disrupted face learning more than the repeated presentation of the same auditory distractor word. These results were obtained even though the speech was reversed suggesting auditory distraction effects on face learning do not depend on semantic processing, consistent with the finding that music disrupted face learning as well³⁴.

This finding is parallel to previously observed effects of auditory distraction on working memory. Semantic processing of comprehensible speech is no necessary precondition for observing effects of auditory distraction on working-memory processes as disruptive effects are commonly observed with reversed speech as well^{19,21–23}. Furthermore, effects of auditory distraction on working-memory processing are not restricted to speech, but are robustly observed with non-speech sounds such as music^{19,25–30} or environmental sounds^{49,50}.

General Discussion

Two studies consistently show that intentional face learning is disrupted by irrelevant speech. The disruptive effect of changing-state sequences consisting of sentential speech was stronger than that of steady-state sequences consisting of the repeated presentation of a single distractor word. The results thus suggest that auditory distraction by changing-state sounds generalizes to face learning. Here, we were able to demonstrate this with changing-state sequences whose disruptive effects on serial recall have been established in several previous studies^{38,40–42}. The changing-state effect was obtained regardless of whether the speech was reversed (Study 2) or not (Study 1), suggesting that the disruptive effect of speech on face learning is not restricted to meaningful speech. A changing-state effect could also be responsible for the disruptive effect of joyful music on face learning observed in the study of Proverbio *et al.*³⁴. They observed that joyful music with many abrupt changes disrupted face learning more than touching violin music with smooth transitions from one tone to another. This is parallel to the disruptive effects of staccato music on serial recall that have been attributed to a changing-state effect^{27,31}.

How can the robust disruption of face learning by auditory distractors be explained? Given that in the present study speech was only played at encoding but not at retrieval, the retrieval context differs from the encoding context. At first glance, *context-dependent memory*⁵¹ seems to be a candidate explanation for the effect of auditory background sound on learning. This account is based on the assumption that memory improves with increasing similarity between the retrieval and the encoding context. The basic idea behind context-dependent memory is

that task-irrelevant context details become automatically associated with the focal stimuli. Retrieval can benefit from the reinstatement of the encoding context because contextual details can be used as retrieval cues. For example, when the same music is played at retrieval and encoding, this could have beneficial effects on memory because the music serves as a retrieval cue⁵². However, this account does not provide a satisfactory explanation for the present findings. First, it is debatable whether the *absence* of contextual cues should benefit retrieval of faces encoded in quiet. Secondly, the encoding-retrieval match is known to have only a minor or no effect on recognition⁵³. Moreover, the context dependency of learning cannot explain the main finding of the present study that changing-state sound is more disruptive than steady-state sound because both distractor conditions involve a context change between encoding and test. It also does not offer any explanation for the finding of Proverbio *et al.*³⁴ that joyful music and sounds of thunder at encoding disrupted performance while touching music with smooth transitions between tones did not. Particularly troubling for the context-dependency account is that the psychophysiological measures (blood pressure and heart rate) in the Proverbio *et al.* study indicated that participants responded sensitively to the touching music that nevertheless did not have any effect on face recognition. This pattern of findings suggests that a different explanation has to be found for the disruption of face learning by irrelevant sounds.

Auditory distraction seems to provide a better explanation. Two broad classes of theories are available to explain the effects of auditory distraction on cognitive processing: The interference-by-process account (e.g.^{46,54}) and the attentional account (e.g.²⁴).

The first account is based on *similarity-based interference*. In general, the more similar two processes are, the more likely they are to interfere with each other⁵⁵. The interference-by-process account of auditory distraction^{46,54} assumes that interference occurs because certain properties of the auditory distractors are automatically processed and come into conflict with processes needed to encode and to retain the to-be-remembered material. Based on the similarity-based interference principle, the finding of auditory distraction in the present paradigm is surprising given that the to-be-learned stimuli are quite dissimilar to the to-be-ignored stimuli. Not only were the faces presented in a different modality than the to-be-ignored speech, but it is also often assumed that the processing of faces requires visual-configural processing⁵⁶ and relies on specialized processing modules in the ventral visual processing stream⁵⁷. A priori, there is no reason to assume that similarity-based interference with background speech should occur.

To explain the present results, the interference-by-process account could invoke the auxiliary assumption that participants verbally recode the faces to encode them. The verbal recoding of faces may seem plausible when familiar faces with well-known names are used⁵⁸, when long encoding periods are used⁶, and when the verbal description of the target face is explicitly required⁵⁹. However, in the present study unfamiliar faces were presented briefly (700 ms with a 1200 ms inter-stimulus interval) for later visual face recognition. Furthermore, it has been shown that the verbal recoding of unfamiliar faces does not necessarily improve face-recognition performance. The verbal overshadowing effect^{59,60} refers to the observation that face recognition suffers when people are required to verbally recode a face because they have to provide a verbal description of its facial features. An explanation of this effect is that a verbal code is inappropriate for face recognition that lends itself better to non-verbalizable configural processing. According to the transfer-inappropriate processing account of the verbal overshadowing effect⁶¹, “verbalization induces inappropriate processing operations, which (...) are incommensurate with the processes required for successful recognition performance” (p. 992). If irrelevant speech discouraged such inappropriate verbalization attempts, this should not necessarily lead to an impairment of face recognition. Another problem of the speech-based explanation is that it cannot easily account for the finding of Proverbio *et al.*³⁴ that the distraction effect generalizes to non-speech sounds. These factors make an explanation based on the interference between verbal codes unattractive. Study 2 further suggests that lexical, syntactical or semantic processing of the distractor speech can also be ruled out as causes for interference.

In the working-memory literature, the changing-state effect is often attributed to the interference between two types of order information⁶². According to this explanation, changes lead to the segmentation of the auditory stream into distinct objects, the order of which is automatically processed. The involuntary seriation of the auditory stream disrupts the voluntary seriation of to-be remembered material. An implication of this account is that the changing-state effect should only be found in tasks that rely on the processing of serial order such as serial recall⁶³. In the present study, however, the order in which the faces were presented in the recognition test was unrelated to their order at encoding. It therefore seems unclear why face recognition should have relied on the processing of serial order.

By contrast, the *attentional account* explains the changing-state effect by assuming that abrupt changes in the auditory modality capture attention. Based on literature on attentional orienting^{11,64–68}, the graded attentional account²⁴ postulates that changes in the auditory modality elicit a call for attention. Steady-state stimuli are less disruptive than changing-state stimuli because the system adapts to repetitive stimulus patterns so that repeated steady-state distractors cause gradually less distraction²⁴. Changing stimuli, by contrast, require some processing before the focal attention can be reoriented to the visual modality so that they impair performance in tasks that benefit from focal attention. Sentential speech is particularly disruptive because it contains many auditory changes¹⁹ that are efficiently detected due to the importance of speech processing for human communication³². When attention is captured by irrelevant speech, less attention is available for ongoing performance. The attentional explanation thus does not require distractors and targets to share a common representational code or to rely on the same processing modules. Instead, it assumes that less focal attention is available for modulating the domain-specific processing routines when it is involuntarily captured by auditory changes. An implication of this model is that the changing-state effect should not be strictly confined to the processing of serial order but should generalize to different kinds of encoding processes that rely on focal attention such as binding⁴ and updating².

The attentional explanation of the changing-state effect on face learning thus relies on the assumption that intentional face learning is influenced by the amount of attentional resources devoted to the faces at encoding.

Given the widespread belief that face processing is largely automatic, this assumption may seem surprising at first glance. However, there is substantial evidence that the processing of faces is affected by attentional load. Markers of the neuronal processing of faces such as the N170 in the event-related potential depend on selective attention^{69,70}. Accordingly, there is convincing evidence that a reduction of attentional resources at encoding disrupts face learning⁷¹. The attentional explanation of the present results thus implies that intentional face learning is impaired when attention is captured by changes in the auditory modality. A potential implication of the attentional-diversion view is that the effect of changing irrelevant speech should be more robust in intentional face learning when participants voluntarily devote attention to the faces than when they do not attend to the faces in which case face processing may be largely automatic. Two recent studies suggest that this may be the case. When the encoding of the faces was incidental, irrelevant speech had no detectable effect on face processing relative to a quiet control condition^{6,7}. By contrast, in the present study and the previous study providing evidence of changing-state disruption³⁴, participants were instructed to encode the faces intentionally. The available literature thus seems to support the assumption that irrelevant speech interferes with the voluntary direction of attention to the faces rather than with the more automatic components of incidental face processing (cf.⁷).

Here we report studies showing that intentional face learning is negatively affected by irrelevant speech. This finding suggests that auditory distraction by changing-state sounds is more general and thus more relevant than previously assumed. Changing auditory distractors do not only interfere with the retention of sequences of digits (e.g., telephone numbers), but also with the learning of complex visual stimuli such as faces. This disruptive effect is not restricted to comprehensible speech as it is found with reversed speech as well. Our findings are not only relevant for the evaluation of models of auditory distraction, they are interesting from an applied perspective as well. A practical conclusion from this research is that exposure to background speech and music should be reduced when a task requires accurate face recognition (e.g., in security screenings). In eyewitness testimony, accurate recognition of the perpetrator face is crucial. Due to the lack of research, one may incorrectly assume that only visual factors—such as the view on the target, viewing distance, and light conditions—are relevant for evaluating the accuracy of eyewitness testimonies. However, the present results—along with those of two previous studies by Marsh and colleagues^{6,7}—suggest that auditory distraction can have a significant detrimental effect on face-recognition accuracy. These findings thus advance our knowledge about the factors that influence face recognition in eyewitness testimony as well as in other contexts in which face recognition accuracy is of critical importance.

References

- Miles, C., Jones, D. M. & Madden, C. Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **17**, 578–584, <https://doi.org/10.1037/0278-7393.17.3.578> (1991).
- Körner, U., Röer, J. P., Buchner, A. & Bell, R. Time of presentation affects auditory distraction: Changing-state and deviant sounds disrupt similar working memory processes. *Quarterly Journal of Experimental Psychology* **72**, 457–471, <https://doi.org/10.1177/1747021818758239> (2019).
- Threadgold, E., Marsh, J. E., McLatchie, N. & Ball, L. J. Background music stints creativity: Evidence from compound remote associate tasks. *Applied Cognitive Psychology*, <https://doi.org/10.1002/acp.3532> (in press).
- Bell, R., Röer, J. P. & Buchner, A. Irrelevant speech disrupts item-context binding. *Experimental Psychology* **60**, 376–384, <https://doi.org/10.1027/1618-3169/a000212> (2013).
- Ellermeier, W. & Zimmer, K. The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology* **35**, 10–16, <https://doi.org/10.1250/ast.35.10> (2014).
- Marsh, J. E. *et al.* The impact of irrelevant auditory facial descriptions on memory for target faces: Implications for eyewitness memory. *The Journal of Forensic Practice* **17**, 271–280, <https://doi.org/10.1108/JFP-08-2014-0029> (2015).
- Marsh, J. E. *et al.* Chatting in the face of the eyewitness: The impact of extraneous cell-phone conversation on memory for a perpetrator. *Canadian Journal of Experimental Psychology* **71**, 183–190, <https://doi.org/10.1037/cep0000101> (2017).
- Tremblay, S., Nicholls, A. P., Alford, D. & Jones, D. M. The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 1750–1754, <https://doi.org/10.1037/0278-7393.26.6.1750> (2000).
- Campbell, T., Beaman, C. P. & Berry, D. C. Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory* **10**, 199–214, <https://doi.org/10.1080/09658210143000335> (2002).
- Beaman, C. P. The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 1106–1118, <https://doi.org/10.1037/0278-7393.30.5.1106> (2004).
- Bell, R., Dentale, S., Buchner, A. & Mayr, S. ERP correlates of the irrelevant sound effect. *Psychophysiology* **47**, 1182–1191, <https://doi.org/10.1111/j.1469-8986.2010.01029.x> (2010).
- Klatte, M., Lachmann, T., Schlittmeier, S. & Hellbrück, J. The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology* **22**, 1168–1191, <https://doi.org/10.1080/09541440903378250> (2010).
- Jones, D. M. & Macken, W. J. 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> (1993).
- Bell, R., Röer, J. P., Lang, A.-G. & Buchner, A. Reassessing the token set size effect on serial recall: Implications for theories of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **45**, 1432–1440, <https://doi.org/10.1037/xlm0000658> (2019).
- Jones, D. M., Madden, C. & Miles, C. Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **44**, 645–669, <https://doi.org/10.1080/14640749208401304> (1992).
- Elliott, E. M. & Cowan, N. Coherence of the irrelevant sound effect: Individual profiles of short-term memory and susceptibility to task-irrelevant materials. *Memory & Cognition* **33**, 664–675, <https://doi.org/10.3758/bf03195333> (2005).
- Röer, J. P., Bell, R., Dentale, S. & Buchner, A. The role of habituation and attentional orienting in the disruption of short-term memory performance. *Memory & Cognition* **39**, 839–850, <https://doi.org/10.3758/s13421-010-0070-z> (2011).
- Ellermeier, W. & Zimmer, K. Individual differences in susceptibility to the “irrelevant speech effect”. *Journal of the Acoustical Society of America* **102**, 2191–2199, <https://doi.org/10.1121/1.419596> (1997).
- Röer, J. P., Bell, R. & Buchner, A. Evidence for habituation of the irrelevant sound effect on serial recall. *Memory & Cognition* **42**, 609–621, <https://doi.org/10.3758/s13421-013-0381-y> (2014).
- Jones, D. M. & Macken, W. J. Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied* **1**, 216–226, <https://doi.org/10.1037/1076-898X.1.3.216> (1995).

21. Jones, D. M., Miles, C. & Page, J. Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology* **4**, 89–108, <https://doi.org/10.1002/acp.2350040203> (1990).
22. Surprenant, A. M., Neath, I. & Bireta, T. J. Changing state and the irrelevant sound effect. *Canadian Acoustics* **35**, 86–87 (2007).
23. Röer, J. P., Körner, U., Buchner, A. & Bell, R. Semantic priming by irrelevant speech. *Psychonomic Bulletin & Review* **24**, 1205–1210, <https://doi.org/10.3758/s13423-016-1186-3> (2017).
24. Bell, R., Röer, J. P., Lang, A.-G. & Buchner, A. Distraction by steady-state sounds: Evidence for a graded attentional model of auditory distraction. *Journal of Experimental Psychology: Human Perception and Performance* **45**, 500–512, <https://doi.org/10.1037/xhp0000623> (2019).
25. Salamé, P. & Baddeley, A. D. Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **41**, 107–122, <https://doi.org/10.1080/14640748908402355> (1989).
26. Ellermeier, W. & Hellbrück, J. Is level irrelevant in “irrelevant speech”? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance* **24**, 1406–1414, <https://doi.org/10.1037/0096-1523.24.5.1406> (1998).
27. Schlittmeier, S. J., Hellbrück, J. & Klatte, M. Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology* **20**, 252–271, <https://doi.org/10.1080/09541440701427838> (2008).
28. Perham, N. & Vizard, J. Can preference for background music mediate the irrelevant sound effect? *Applied Cognitive Psychology* **25**, 625–631, <https://doi.org/10.1002/acp.1731> (2011).
29. Röer, J. P., Bell, R. & Buchner, A. What determines auditory distraction? On the roles of local auditory changes and expectation violations. *PLoS one* **9**, e84166, <https://doi.org/10.1371/journal.pone.0084166> (2014).
30. Röer, J. P., Rummel, J., Bell, R. & Buchner, A. Metacognition in auditory distraction: How expectations about distractibility influence the irrelevant sound effect. *Journal of Cognition* **1**, 2, <https://doi.org/10.5334/joc.3> (2017).
31. Klatte, M., Kilcher, H. & Hellbrück, J. Wirkungen der zeitlichen Struktur von Hintergrundschaall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen. *Zeitschrift für Experimentelle Psychologie* **42**, 517–544 (1995).
32. Jones, D. M., Macken, W. J. & Murray, A. C. Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition* **21**, 318–328, <https://doi.org/10.3758/BF03208264> (1993).
33. Wais, P. E. & Gazzaley, A. The impact of auditory distraction on retrieval of visual memories. *Psychonomic Bulletin & Review* **18**, 1090–1097, <https://doi.org/10.3758/s13423-011-0169-7> (2011).
34. Proverbio, A. M. *et al.* The effect of background music on episodic memory and autonomic responses: listening to emotionally touching music enhances facial memory capacity. *Scientific Reports* **5**, 15219, <https://doi.org/10.1038/srep15219> (2015).
35. Lange, E. Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language* **53**, 513–531, <https://doi.org/10.1016/j.jml.2005.07.002> (2005).
36. Hughes, R. W., Vachon, F. & Jones, D. M. Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **33**, 1050–1061, <https://doi.org/10.1037/0278-7393.33.6.1050> (2007).
37. Röer, J. P., Bell, R., Marsh, J. E. & Buchner, A. Age equivalence in auditory distraction by changing and deviant speech sounds. *Psychology and Aging* **30**, 849–855, <https://doi.org/10.1037/pag0000055> (2015).
38. Körner, U., Röer, J. P., Buchner, A. & Bell, R. Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. *Journal of Memory and Language* **96**, 122–137, <https://doi.org/10.1016/j.jml.2017.05.005> (2017).
39. Vachon, F., Labonté, K. & Marsh, J. E. Attentional capture by deviant sounds: A noncontingent form of auditory distraction? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **43**, 622–634, <https://doi.org/10.1037/xlm0000330> (2017).
40. Röer, J. P., Bell, R. & Buchner, A. Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception and Performance* **41**, 692–702, <https://doi.org/10.1037/xhp0000028> (2015).
41. Bell, R., Röer, J. P., Marsh, J. E., Storch, D. & Buchner, A. The effect of cognitive control on different types of auditory distraction: A preregistered study. *Experimental Psychology* **64**, 359–368, <https://doi.org/10.1027/1618-3169/a000372> (2017).
42. Körner, U., Röer, J. P., Buchner, A. & Bell, R. Time of presentation affects auditory distraction: Changing-state and deviant sounds disrupt similar working memory processes. *Quarterly Journal of Experimental Psychology* **72**, 457–471, <https://doi.org/10.1177/1747021818758239> (2019).
43. Phillips, P. J., Wechsler, H., Huang, J. & Rauss, P. J. The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing* **16**, 295–306, [https://doi.org/10.1016/S0262-8856\(97\)00070-X](https://doi.org/10.1016/S0262-8856(97)00070-X) (1998).
44. Bell, R., Röer, J. P., Dentale, S. & Buchner, A. Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **38**, 1542–1557, <https://doi.org/10.1037/a0028459> (2012).
45. Snodgrass, J. G. & Corwin, J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General* **117**, 34–50, <https://doi.org/10.1037/0096-3445.117.1.34> (1988).
46. Marsh, J. E., Hughes, R. W. & Jones, D. M. Interference by process, not content, determines semantic auditory distraction. *Cognition* **110**, 23–38, <https://doi.org/10.1016/j.cognition.2008.08.003> (2009).
47. Röer, J. P., Körner, U., Buchner, A. & Bell, R. Attentional capture by taboo words: A functional view of auditory distraction. *Emotion* **17**, 740–750, <https://doi.org/10.1037/emo0000274> (2017).
48. Röer, J. P., Bell, R., Körner, U. & Buchner, A. A semantic mismatch effect on serial recall: Evidence for interlexical processing of irrelevant speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **45**, 515–525, <https://doi.org/10.1037/xlm0000596> (2019).
49. Buchner, A., Bell, R., Rothermund, K. & Wentura, D. Sound source location modulates the irrelevant sound effect. *Memory & Cognition* **36**, 617–628, <https://doi.org/10.3758/MC.36.3.617> (2008).
50. Röer, J. P., Bell, R. & Buchner, A. Please silence your cell phone: Your ringtone captures other people’s attention. *Noise & Health* **16**, 34–39, <https://doi.org/10.4103/1463-1741.127852> (2014).
51. Smith, S. M. Mood is a component of mental context: Comment on Eich (1995). *Journal of Experimental Psychology: General* **124**, 309–310, <https://doi.org/10.1037/0096-3445.124.3.309> (1995).
52. Mead, K. M. L. & Ball, L. J. Music tonality and context-dependent recall: The influence of key change and mood mediation. *European Journal of Cognitive Psychology* **19**, 59–79, <https://doi.org/10.1080/09541440600591999> (2007).
53. Godden, D. & Baddeley, A. When does context influence recognition memory? *British Journal of Psychology* **71**, 99–104, <https://doi.org/10.1111/j.2044-8295.1980.tb02735.x> (1980).
54. Marsh, J. E., Hughes, R. & Jones, D. Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language* **58**, 682–700, <https://doi.org/10.1016/j.jml.2007.05.002> (2008).
55. Wickens, C. D. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* **3**, 159–177, <https://doi.org/10.1080/14639220210123806> (2002).
56. Richler, J. J., Cheung, O. S. & Gauthier, I. Holistic processing predicts face recognition. *Psychological Science* **22**, 464–471, <https://doi.org/10.1177/0956797611401753> (2011).
57. Collins, J. A. & Olson, I. R. Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia* **61**, 65–79, <https://doi.org/10.1016/j.neuropsychologia.2014.06.005> (2014).
58. Bruce, V. & Young, A. Understanding face recognition. *British Journal of Psychology* **77**, 305–327, <https://doi.org/10.1111/j.2044-8295.1986.tb02199.x> (1986).

59. Schooler, J. W. & Engstler-Schooler, T. Y. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology* **22**, 36–71, [https://doi.org/10.1016/0010-0285\(90\)90003-M2295225](https://doi.org/10.1016/0010-0285(90)90003-M2295225) (1990).
60. Alogna, V. K. *et al.* Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science* **9**, 556–578, <https://doi.org/10.1177/1745691614545653> (2014).
61. Schooler, J. W. Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology* **16**, 989–997, <https://doi.org/10.1002/acp.930> (2002).
62. Jones, D. M., Farrand, P., Stuart, G. & Morris, N. Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 1008–1018, <https://doi.org/10.1037/0278-7393.21.4.1008> (1995).
63. Beaman, C. P. & Jones, D. M. Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23**, 459–471, <https://doi.org/10.1037/0278-7393.23.2.459> (1997).
64. Näätänen, R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences* **13**, 201–288, <https://doi.org/10.1017/S0140525X00078407> (1990).
65. Schröger, E. & Wolff, C. Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. *Cognitive Brain Research* **7**, 71–87, [https://doi.org/10.1016/S0926-6410\(98\)00013-5](https://doi.org/10.1016/S0926-6410(98)00013-5) (1998).
66. Escera, C., Alho, K., Schröger, E. & Winkler, I. Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro Otology* **5**, 151–166, <https://doi.org/10.1159/000013877> (2000).
67. Campbell, T., Winkler, I., Kujala, T. & Näätänen, R. The N1 hypothesis and irrelevant sound: Evidence from token set size effects. *Cognitive Brain Research* **18**, 39–47, <https://doi.org/10.1016/j.cogbrainres.2003.09.001> (2003).
68. Bendixen, A., Roeber, U. & Schröger, E. Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience* **19**, 1664–1677, <https://doi.org/10.1162/jocn.2007.19.10.1664> (2007).
69. Eimer, M. Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology* **17**, 103–116, <https://doi.org/10.1080/026432900380517> (2000).
70. Mohamed, T. N., Neumann, M. F. & Schweinberger, S. R. Perceptual load manipulation reveals sensitivity of the face-selective N170 to attention. *Neuroreport* **20**, 782–787, <https://doi.org/10.1097/WNR.0b013e32832b7e24> (2009).
71. Jenkins, R., Lavie, N. & Driver, J. Recognition memory for distractor faces depends on attentional load at exposure. *Psychonomic Bulletin & Review* **12**, 314–320, <https://doi.org/10.3758/BF03196378> (2005).

Acknowledgements

The research reported in this article was funded by the Deutsche Forschungsgemeinschaft (BE 4311/3-1).

Author Contributions

R.B., L.M., J.P.R. and A.B., designed the study, collected and analyzed the data, and wrote the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

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



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

© The Author(s) 2019