Partially masking task-irrelevant speech has opposite effects on metacognitive judgments of distraction and objective distraction effects

Gesa Fee Komar, Axel Buchner, Laura Mieth & Raoul Bell

Metacognition and Learning (2025) 20:35
https://doi.org/10.1007/s11409-025-09436-y

Partially masking task-irrelevant speech has opposite effects on metacognitive judgments of distraction and objective distraction effects

Gesa Fee Komar[1] · Axel Buchner[1] · Laura Mieth[1] · Raoul Bell[1]

Received: 12 February 2025 / Accepted: 27 July 2025

Abstract

Two experiments served to test the hypothesis that partially masking speech with pink noise (Experiment 1) or speech babble (Experiment 2) induces particularly pronounced metacognitive illusions in judgments about the distracting effects of task-irrelevant speech on cognitive performance. We hypothesized that the experimental manipulations would have opposite effects on judgments of distraction and objective distraction effects. Specifically, masked speech should be perceived as being more difficult to listen to than pure speech, thereby evoking a subjective experience of relative disfluency. If people rely on a (dis)fluency heuristic, masked speech should be predicted to be more distracting than pure speech. However, given that pink noise and speech babble mask the auditory changes in the speech signal that drive auditory distraction, people should objectively be less distracted by masked speech than by pure speech. The findings of both experiments support this hypothesis. Masked speech evoked a subjective experience of relative disfluency and was predicted to be more distracting than pure speech. However, participants were objectively less distracted by masked speech than by pure speech in a serial-recall task. Even after multiple firsthand experiences of having ignored masked and pure speech during the serial-recall task, participants judged masked speech to have been more, but not less, distracting than pure speech. Partially masking speech thus had opposite effects on judgments of distraction and objective distraction effects. These findings provide evidence of particularly pronounced metacognitive illusions and support the hypothesis that people rely on (dis)fluency as a cue for predicting distraction.

Keywords Task-irrelevant speech · Partial masking · Auditory distraction · Judgments of distraction · Metacognitive illusions of distraction

✉ Gesa Fee Komar
gesa.komar@hhu.de

[1] Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Published online: 26 September 2025

With the widespread use of portable devices such as tablets and laptops, people often study and work in shared spaces, facing the challenge of performing cognitively demanding tasks in environments with task-irrelevant speech. This likely results in distraction, as task-irrelevant speech has been identified as one of the most distracting types of sound (Ellermeier & Zimmer, 2014; Schlittmeier et al., 2012). As a potential countermeasure against distraction, modern computer operating systems such as macOS Sequoia offer the possibility to play background noise to mask task-irrelevant speech. For instance, the playback of pink noise—referred to as "balanced noise"—is offered as an option to reduce distractions and facilitate concentration. Indeed, because pink noise provides a homogenous auditory background, embedding speech in pink noise masks the auditory changes in the speech signal that drive auditory distraction (e.g., Ellermeier & Hellbrück, 1998). Similarly, multiple speech streams in open-plan offices—provided they cannot be perceptually segregated based on distinct spatial locations (Jones & Macken, 1995)—may jointly form a more homogeneous auditory background than the speech of a single individual because auditory changes in each speech stream can be masked by the surrounding speech babble. In such cases, working in open-plan offices may be less distracting than working in a private office shared only with a telephoning colleague. These insights refer to the objective effects of embedding speech in pink noise or speech babble on cognitive performance. However, it is unclear whether people can correctly predict these effects. Indeed, there is evidence of metacognitive illusions as it has been shown that people's predictions about the distracting effects of sounds on cognitive performance do not align with objective distraction effects (Bell et al., 2023; Komar et al., 2024). To understand whether people are able to effectively choose and control their digital and physical study and work environments so as to minimize auditory distraction, it is important to examine how people arrive at metacognitive judgments of distraction (Marsh et al., 2024). Can people correctly predict that embedding speech in pink noise or speech babble causes less, but not more, distraction? This question is addressed in the two experiments reported here.

Whereas the present study is focused on task-*irrelevant* stimuli, most research on metacognition to date has been concerned with metacognitive judgments about task-*relevant* stimuli. For example, judgments of learning refer to people's predictions about how well they will remember specific stimuli, such as words, that they study (Begg et al., 1989; Besken & Mulligan, 2013, 2014; Frank & Kuhlmann, 2017; Koriat, 1997; Kornell et al., 2011; Mieth et al., 2021; Rhodes & Castel, 2008; Schaper et al., 2019a, b, 2023; Soderstrom et al., 2015; Undorf et al., 2017). The overall consensus is that learners lack direct metacognitive access to the mechanisms underlying their learning. Instead, they often rely on simple heuristics which may lead to correct judgments of learning but may also result in systematic biases. One of the most important heuristics is the (dis)fluency heuristic (Alter & Oppenheimer, 2009). The fundamental principle underlying the (dis)fluency-heuristic hypothesis is that people use their subjective experience of the difficulty or ease of processing as a cue for predicting learning. For instance, it has been proposed that people predict stimuli that are experienced as being more difficult to process to be harder to learn than stimuli that are experienced as being easier to process (e.g., Begg et al., 1989; Rhodes & Castel, 2008; Undorf & Erdfelder, 2011, 2015; Undorf et al., 2017). Relying on the (dis)fluency heuristic may lead to metacognitive judgments that do not align with objective performance measures—referred to as metacognitive illusions—because stimuli that are experienced as being more difficult to process can be learned better if they are elaborated more than stimuli that are experienced as being easier to process (Besken & Mulligan, 2013, 2014; Bjork & Bjork, 2011, 2020).

In recent years, there has been increasing interest in the metacognition of auditory distraction (Atienzar et al., in press; Beaman et al., 2014; Bell et al., 2023; Bell et al., 2022; Hanczakowski et al., 2017, 2018; Kattner & Bryce, 2022; Komar et al., 2024; Röer et al., 2017). Here, metacognitive judgments refer primarily to people's predictions about how specific task-irrelevant stimuli affect cognitive performance. In most studies examining the metacognition of auditory distraction, the serial-recall task has been used. This task is an essential component of an established experimental paradigm for studying auditory distraction (for a review, see Ellermeier & Zimmer, 2014). In a common variant of this task, sequences of digits have to be maintained for immediate serial recall. Serial recall is scored as correct if a digit is recalled in the serial position in which it was presented. The number of correctly recalled digits typically decreases when task-irrelevant sounds have to be ignored during digit presentation or retention compared to quiet (Macken et al., 1999). Objective distraction effects can be calculated by subtracting, from the mean number of correctly recalled digits in distractor trials with sounds, the mean number of correctly recalled digits in quiet control trials (Bell et al., 2023; Bell et al., 2022; Komar et al., 2024). It is known from decades of research that auditory distraction in this paradigm is driven by changes in the to-be-ignored auditory signal (Bell et al., 2019; Jones, 1993; Jones et al., 1993): The more pronounced the changes in the auditory signal, the larger the objective distraction effect.

Following procedures initially developed for measuring judgments of learning (Begg et al., 1989; Frank & Kuhlmann, 2017; Koriat, 1997; Soderstrom et al., 2015), research on the metacognition of auditory distraction has been focused on two variants of judgments of distraction (Bell et al., 2023; Bell et al., 2022; Komar et al., 2024): Prospective judgments and retrospective judgments, which will be explicated below.

Prospective judgments are predictions about future distraction. These judgments are typically assessed in a stimulus-specific way, that is, each prospective judgment is based on the immediate experience of a specific distractor sound. Specifically, participants are first introduced to the serial-recall task they will later perform. The sounds are then played one by one. After having listened to a sound, participants are asked to predict how distracting or helpful that specific sound would be, relative to quiet, for serial recall. Understanding prospective judgments is essential for addressing the real-world situation in which people must predict the effect of an auditory background on cognitive performance after having listened to the specific sound. For example, in a work environment, one may briefly listen to an auditory background—such as a colleague's telephone conversation—to predict how distracting or helpful the auditory background would be for a certain task and to decide whether or not it would be necessary to adjust the work environment.

Retrospective judgments are postdictions about past distraction. These judgments are typically based on multiple firsthand experiences of having ignored a specific type of distractor sound. Specifically, after having completed the serial-recall task, participants are asked to judge how distracting or helpful each type of sound has been, relative to quiet, for serial recall. Returning to the earlier example, after having worked several times in the presence of a specific auditory background—such as a colleague's telephone conversation—one may no longer need to listen to the background in order to be able to judge the background's effect on cognitive performance. Instead, one may form a global judgment based on multiple firsthand experiences. In the following, we will first develop a hypothesis regarding prospective judgments and we will then return to retrospective judgments at the end of the Introduction.

In line with research on judgments of learning (e.g., Begg et al., 1989; Rhodes & Castel, 2008; Undorf & Erdfelder, 2011, 2015; Undorf et al., 2017), it has been proposed that people rely on a *(dis)fluency heuristic* by using their subjective experience of the difficulty or ease of processing when predicting the distracting effects of sounds on cognitive performance (Bell et al., 2023; Komar et al., 2024). Specifically, it has been proposed that people predict sounds that are experienced as being more difficult to process to be harder to ignore than sounds that are experienced as being easier to process. While relying on the (dis)fluency heuristic may have some ecological validity (Herzog & Hertwig, 2013), leading to correct prospective judgments in some contexts (Bell et al., 2022; Kattner & Bryce, 2022), it can also result in metacognitive illusions in other contexts (Bell et al., 2023; Bell et al., 2024; Komar et al., 2024).

Such metacognitive illusions have been demonstrated in studies by Bell et al. (2023) and Komar et al. (2024). In these studies, prospective judgments about the distracting effects of music or speech on serial-recall performance were affected by experimental manipulations such as playback direction (backward versus forward) and language (foreign versus native), even though these manipulations did not affect the objective distraction effects. In the study by Bell et al. (2023), backward music was experienced as being more difficult to process than forward music, leading participants to predict backward music to be more distracting than forward music. However, the objective distraction effects of backward and forward music on serial-recall performance were comparable, which was anticipated given that backward and forward music contain a similar number of distinct changes in the auditory signal. In the study by Komar et al. (2024), metacognitive illusions have been demonstrated by comparing backward and forward speech, and foreign and native speech. Backward and foreign speech were experienced as being more difficult to process than forward and native speech, respectively, leading participants to predict backward and foreign speech to be more distracting than forward and native speech. However, the objective distraction effects of backward and forward speech, and foreign and native speech, on serial-recall performance were comparable, which was anticipated given that backward and forward speech, and foreign and native speech, contain a similar number of distinct changes in the auditory signal.

This pattern of results was expected based on theoretical considerations, which had led to the experimental manipulations that were specifically designed to affect prospective judgments of distraction but not objective distraction effects. As such, the results were interpreted as evidence for metacognitive illusions, supporting the hypothesis that people rely on the (dis)fluency heuristic when predicting the distracting effects of sounds on cognitive performance and demonstrating that people lack direct metacognitive access to the cognitive effects of sounds that determine auditory distraction. However, the same pattern of results—that the prospective judgments of distraction were affected by the experimental manipulations, while the objective distraction effects remained unaffected—could have been observed if the direct-rating measure used to assess the prospective judgments of distraction had been more sensitive to the experimental manipulations than the serial-recall measure used to assess the objective distraction effects. The present study addressed this issue by using experimental manipulations that were expected to have *opposite* effects on prospective judgments of distraction and objective distraction effects. Specifically, we compared speech embedded in pink noise (Experiment 1) and speech embedded in speech babble (Experiment 2) with pure speech. Speech that is partially masked with pink noise or speech babble (hereafter referred to as masked speech) should be experienced as being more difficult to process than pure speech, an assumption that has been confirmed in separate validation experiments (see below). If people rely on their subjective experience of relative disfluency when predicting the distracting effects of masked and pure speech on serial-recall performance, they should indicate in their prospective judgments that they predict masked speech to be *more* distracting than pure speech. However, given that pink noise (Ellermeier & Hellbrück, 1998; Schlittmeier & Hellbrück, 2009) and speech babble (Jones & Macken, 1995; Keus van de Poll et al., 2015; Zaglauer et al., 2017) mask the auditory changes in the speech signal that drive auditory distraction (Bell et al., 2019; Jones, 1993; Jones et al., 1993), people should objectively be *less* distracted by masked speech than by pure speech. Therefore, masking speech with pink noise or speech babble should have opposite effects on prospective judgments of distraction and objective distraction effects, leading to even more pronounced metacognitive illusions in prospective judgments than those previously observed (Bell et al., 2023; Komar et al., 2024).

In previous studies (Bell et al., 2023; Komar et al., 2024), multiple firsthand experiences of having ignored different types of music or speech during the serial-recall task led participants to arrive at retrospective judgments that were somewhat better aligned with the objective distraction effects than the prospective judgments. However, in most cases (i.e., in all but one experiment), there was a metacognitive illusion in the retrospective judgments. Ellermeier and Zimmer (1997) found that metacognitive judgments given after the serial-recall task were better aligned with objective distraction effects than those given before the serial-recall task. Beaman et al. (2014), using recognition tasks, found that auditory distraction reduced participants' confidence in their responses and their willingness to report them during task performance, suggesting that, when performing the task, "participants seem to be aware that auditory distraction is harmful for memory" (p. 11). The present study included retrospective judgments to test whether particularly pronounced metacognitive illusions—that is, effects of masking speech with pink noise or speech babble on judgments of distraction in the opposite direction of their objective distraction effects—would be observed in retrospective judgments.

Experiment 1

Method

Participants

Following a previous study on the metacognition of auditory distraction (Komar et al., 2024), the experiment was conducted online using SoSci Survey (https://www.soscisurvey.de/). To gain fast access to a large and diverse participant pool for data collection, we utilized the online-access panel provider Cint (https://www.cint.com/). This approach is validated by empirical evidence demonstrating that key effects of auditory distraction can be reliably replicated in online settings (Elliott et al., 2022). Participants were required to use a desktop or laptop computer. We aimed at collecting at least 300 valid data sets and stopped data collection at the end of the day during which this criterion was reached. The same exclusion criteria were applied as in a previous study on the metacognition of auditory distraction (Komar et al., 2024). Of the participants who had passed the audio check (see below), 73 participants did not complete the experiment, four participants withdrew their consent to the use of their data, one participant was under 18 years old and thus could not legally provide consent and 16 participants reported studying or having studied Psychology. The data sets of these participants were excluded from the analyses. Based on responses to the catch trial or the post-experiment questions (see below), 64 further data sets could have been excluded. However, following a recommendation by Elliott et al. (2022), these data sets were retained in the analyses. This decision had no impact on the statistical conclusions. The final sample, characterized by diverse levels of education, consisted of 309 participants (116 female, 192 male, 1 non-binary) with a mean age of 36 ($SD = 13$) years. All participants were proficient German speakers; 297 of them indicated that German was their native language. Participants were randomly assigned to either the prospective-judgment group ($n = 168$) or the control group ($n = 141$). A sensitivity analysis using G*Power (Faul et al., 2007) showed that, given a sample size of $N = 309$ and $\alpha = \beta = .05$ (and, thus, a statistical power of $1 - \beta = .95$), a main effect of speech type on the objective distraction effect of the size $\eta^2_p = .04$ could be detected.

Materials

The speech sequences were recorded and edited using Amadeus Pro 2.8.9 (https://www.hairersoft.com/). Thirty-two speech sequences similar to the speech sequences that have been shown to elicit pronounced auditory distraction in previous studies (Bell et al., 2021; Komar et al., 2024; Körner et al., 2017; Röer et al., 2014, 2015) were used (e.g., “On Tuesday, it will be mostly sunny, with only isolated rain showers possible. A weak to moderate northeast wind will be blowing.”; translated from German). The speech sequences were spoken by a male voice, recorded at a sampling rate of 44.1 kHz in a 16-bit format, edited to last 8 s and normalized to peak amplitude. Masked-speech sequences were generated by combining the speech sequences with pink noise at a speech-to-noise ratio of 0 dB(A) L_{eq}. The pink noise featured an amplitude envelope with attack and decay times of 250 ms to ensure a smooth onset and offset. Pure-speech sequences contained only speech and no pink noise (Fig. 1). Thus, each of the 32 speech sequences was available both as a masked-speech sequence and as a pure-speech sequence.

Validation experiment

A separate validation experiment with different participants was conducted to test the effectiveness of the experimental manipulation of processing (dis)fluency. After having applied

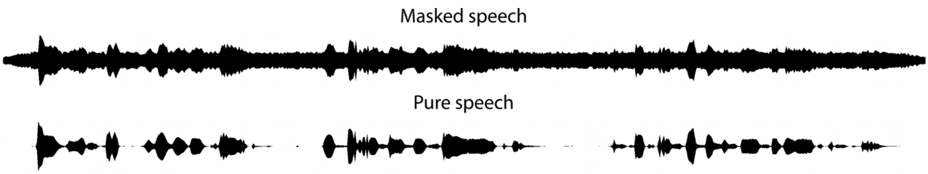

Fig. 1 Example waveforms of masked and pure speech used in Experiment 1

the same exclusion criteria as those reported above, data sets of 106 participants (35 female, 71 male) with a mean age of 40 ($SD=15$) years were included in the analysis. For each participant, 16 masked-speech sequences and 16 pure-speech sequences were randomly selected from the pool of 32 masked-speech sequences and 32 pure-speech sequences with the restriction that each speech sequence could be selected in only one of the two conditions. The subjective experience of processing (dis)fluency was measured on a single-item scale. Graf et al. (2018) have shown that a single-item scale captures the effects of various experimental manipulations of processing (dis)fluency—such as pronounceability, repeated exposure, visual contrast and typicality—as validly as a multiple-item scale (for an overview of successful single-item scales in various research areas, see Gräve et al., 2024). In each trial of the validation experiment, participants clicked a "Play" button to listen to one of the speech sequences. After the playback of the speech sequence, participants were asked: "How difficult or easy was it for you to listen to the sound?" They responded using the single-item scale of processing (dis)fluency ranging from −6 to +6. The scale contained verbal labels for the values of −6 (very difficult), −4 (difficult), −2 (somewhat difficult), 0 (neither nor), +2 (somewhat easy), +4 (easy) and +6 (very easy) to facilitate the interpretation of the scale (Maitland, 2009; Rohrmann, 1978). The order of the speech sequences was randomly determined for each participant. Using the speech sequences as the units of analysis, masked speech ($M=-2.72$, $SD=0.67$) was rated to be significantly more difficult to listen to than pure speech ($M=5.08$, $SD=0.24$), $F(1, 62)=3870.76$, $p<.001$, $\eta_p^2 = .98$. These results confirm that the experimental manipulation of processing (dis)fluency was effective.

Design and procedure

Design and procedure followed those of previous studies on the metacognition of auditory distraction (Bell et al., 2023; Komar et al., 2024). At the beginning of the experiment, participants were instructed to make sure they were alone in a quiet, distraction-free environment. They were asked to turn off their smartphone and to close any additional browser tabs or programs on their computer. Participants were informed that a browser supporting automatic sound playback was required. They could opt to view on-screen instructions on how to enable this feature in different browsers. Participants were asked to wear headphones for the entire duration of the experiment. An audio check was conducted to ensure that participants could hear the speech sequences. Participants were instructed to adjust the computer volume to ensure that they could hear all sounds well. In each trial of the audio check, one letter from the set {j, q, r, x, z} was randomly selected and masked with pink noise at a speech-to-noise ratio of 0 dB(A) L_{eq}. The letter was spoken by a male voice. The masked letter was played at the same intensity as the masked speech used later in the experiment. Participants were asked to type, from the letters j, q, r, x and z, the letter they had heard into a text field. They were asked to adjust the browser settings or switch browsers if they could not hear the spoken letter. To proceed to the experiment proper, participants had to correctly identify five letters in a row. After having completed the audio check, participants were instructed not to change the computer volume for the entire duration of the experiment.

Participants were randomly assigned to one of two groups (Fig. 2). In the prospective-judgment group, participants provided prospective judgments about the distracting effects of masked and pure speech on serial-recall performance before performing the serial-recall task in which the objective distraction effects were measured. To control for possible reactive effects of the prospective judgments, we included a control group. Participants in the control group did not provide any prospective judgments. Instead, they began immediately with the serial-recall task. This experimental design served to test whether providing prospective judgments about the distracting effects of masked and pure speech on serial-recall performance influenced the objective distraction effects. Such reactive effects could arise, for instance, if the act of providing prospective judgments prompted participants to invest more effort into the serial-recall task to compensate for the predicted distraction. After having experienced the distracting effects of masked and pure speech firsthand during the serial-recall task, all participants provided retrospective judgments about the distracting effects of the two types of speech on their serial-recall performance.

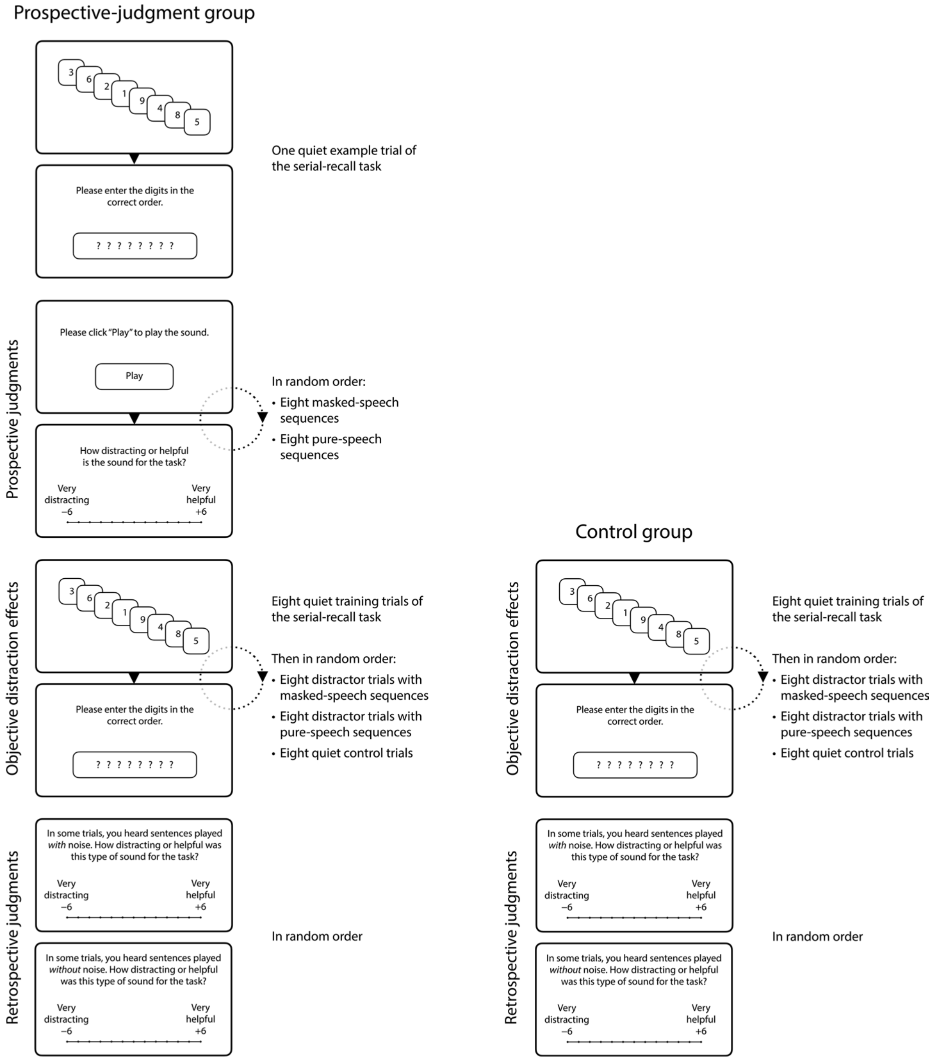

Fig. 2 Schematic illustration of the design and procedure (see main text for details)

Prospective judgments about the distracting effects of masked and pure speech on serial-recall performance were collected in the prospective-judgment group before the serial-recall task. In the prospective-judgment task, participants were asked to imagine performing the serial-recall task that was first described verbally and then illustrated with a quiet example trial (see below for details on the serial-recall task). After having completed the example trial of the serial-recall task, participants were informed that various sounds would be played to them and that they should imagine hearing the sounds while memorizing the digits in the serial-recall task that had just been described and illustrated.

In each trial of the prospective-judgment task, participants clicked a “Play” button, whereupon one of the speech sequences was presented. Participants were asked to predict how distracting or helpful the sound would be, relative to quiet, for serial recall, using a metacognition scale ranging from −6 to +6. The scale was presented 1 s after the end of the speech sequence and contained verbal labels for the values of −6 (very distracting), −4 (distracting), −2 (somewhat distracting), 0 (neither nor), +2 (somewhat helpful), +4 (helpful) and +6 (very helpful) to facilitate the interpretation of the scale (Maitland, 2009; Rohrmann, 1978). For each participant in the prospective-judgment group, eight masked-speech sequences and eight pure-speech sequences were randomly selected from the pool of 32 masked-speech sequences and 32 pure-speech sequences with the restriction that each speech sequence could be selected in only one of the two conditions.

Objective distraction effects of masked and pure speech on serial-recall performance were measured in the serial-recall task. Participants were informed that they would complete several trials in which they would have to memorize the order of digits without using external aids or speaking the digits aloud. They were instructed to guess any digits they could not remember. Participants were informed that they would occasionally hear sounds over their headphone and were instructed to ignore these sounds.

In each trial of the serial-recall task, a sequence of eight digits was generated by randomly drawing digits from the set {1, 2, …, 9} without replacement. Each digit was presented at the center of the browser window for 1 s. A text field with eight question marks appeared 1 s after all digits had been presented, prompting participants to enter the eight digits in the correct order.

The serial-recall task started with eight quiet training trials to familiarize participants with the task. The following 24 experimental trials consisted of eight distractor trials in which masked speech was played during digit presentation, eight distractor trials in which pure speech was played during digit presentation and eight quiet control trials. For each participant, eight masked-speech sequences and eight pure-speech sequences were randomly selected with the restriction that each speech sequence could be selected in only one of the two conditions. In the prospective-judgment group, speech sequences were selected from the 16 masked-speech sequences and the 16 pure-speech sequences that had not been presented in the prospective-judgment task. In the control group, speech sequences were selected from the complete pool of 32 masked-speech sequences and 32 pure-speech sequences. The order in which the experimental trials were presented was randomly determined for each participant.

After the serial-recall task, a catch trial was presented in which the spoken letter q could be heard. Participants were asked to type the letter they had heard into a text field.

Retrospective judgments about the distracting effects of masked and pure speech on the participants’ serial-recall performance were collected after the catch trial. Participants were informed about the type of speech whose effect on their serial-recall performance they had to judge ("In some trials, you heard sentences played with noise" or "In some trials, you heard sentences played without noise"). Participants were then asked to judge how distracting or helpful the type of sound had been, relative to quiet, for serial recall, using the metacognition scale ranging from −6 (very distracting) to +6 (very helpful). The order in which the two types of speech were judged was randomly determined for each participant.

At the end of the experiment, participants were thanked for their participation. They were then asked to provide honest responses to the post-experiment questions "Was all the information presented correctly?" and "Did you follow the instructions?". They were assured that responding "No" would not have any negative consequences for them (Rouse, 2015). Finally, participants were asked to confirm their consent to the use of their data, with the option to revoke their consent they had given at the beginning of the experiment by selecting "No, I withdraw the consent to the use of my data". The median duration of the experiment was 19 min in the prospective-judgment group and 14 min in the control group.

Results

Prospective judgments

A repeated-measures analysis of variance was calculated to test the effect of the within-subjects factor speech type (masked, pure) on the prospective judgments (left panel of Fig. 3). Participants in the prospective-judgment group predicted masked speech to be significantly more distracting than pure speech, $F(1, 167)=283.20$, $p<.001$, $\eta_p^2 = .63$. Relative to the neutral midpoint of the metacognition scale, masked speech was predicted to be distracting, $t(167)=-31.02$, $p < .001$, $d=-2.39$, while pure speech was predicted to be helpful, $t(167) = 2.79$, $p = .006$, $d=0.21$.

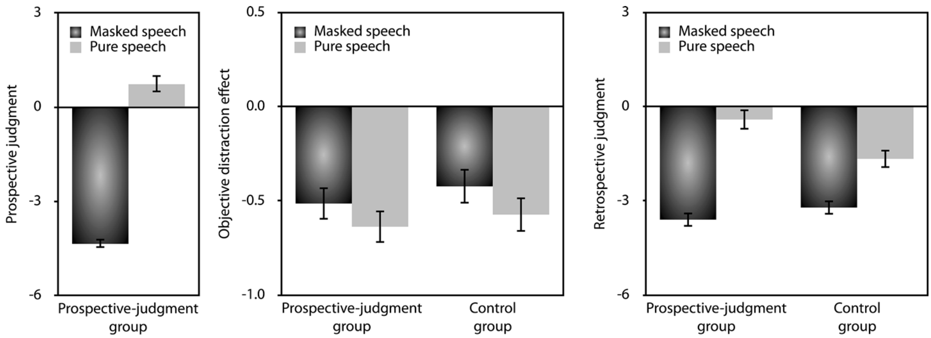

Fig. 3 Results of Experiment 1. The left panel shows the mean prospective judgment about the distracting effect of speech on serial-recall performance on a scale from −6 (very distracting) to +6 (very helpful) as a function of speech type (masked, pure). The middle panel shows the mean objective distraction effect of speech on serial-recall performance relative to quiet as a function of speech type (masked, pure) and group (prospective judgment, control). The right panel shows the mean retrospective judgment about the distracting effect of speech on the participants' serial-recall performance on a scale from −6 (very distracting) to +6 (very helpful) as a function of speech type (masked, pure) and group (prospective judgment, control). Error bars represent the standard errors of the means

Objective distraction effects

As in previous studies on the metacognition of auditory distraction (Bell et al., 2023; Bell et al., 2022; Komar et al., 2024), the objective distraction effects of masked and pure speech on serial-recall performance were calculated for each participant by subtracting, from the mean number of digits per trial that were recalled in the correct serial position in each of the two types of distractor trials, the mean number of digits per trial that were recalled in the correct serial position in the quiet control trials. Negative values indicate that participants recalled fewer digits per trial in the correct serial position in the distractor trials than in the quiet control trials and thus reflect a distracting effect of the speech (positive values would reflect a helpful effect). The serial-recall results that form the basis for calculating the objective distraction effects, averaged across participants in each group, are reported in Table 1.

A 2 × 2 mixed analysis of variance was calculated to test the effects of the within-subjects factor speech type (masked, pure) and the between-subjects factor group (prospective judgment, control) on the objective distraction effects (middle panel of Fig. 3). Participants were significantly less distracted by masked speech than by pure speech, $F(1, 307) = 6.50$, $p = .011$, $\eta_p^2 = .02$. Whether or not participants were asked to provide prospective judgments before the objective distraction effects were measured did not affect the objective distraction effects overall, $F(1, 307) = 0.47$, $p = .493$, $\eta_p^2 < .01$, and did not interact with speech type, $F(1, 307) = 0.06$, $p = .812$, $\eta_p^2 < .01$. Although masked speech disrupted serial recall significantly less than pure speech, both types of speech caused significant distraction relative to quiet. In the prospective-judgment group, participants were distracted by both masked speech, $t(167) = -5.93$, $p < .001$, $d = -0.46$, and pure speech, $t(167) = -7.38$, $p < .001$, $d = -0.57$. In the control group, participants were also distracted by both masked speech, $t(140) = -4.72$, $p < .001$, $d = -0.40$, and pure speech, $t(140) = -6.25$, $p < .001$, $d = -0.53$.

Retrospective judgments

A 2 × 2 mixed analysis of variance was calculated to test the effects of the within-subjects factor speech type (masked, pure) and the between-subjects factor group (prospective judgment, control) on the retrospective judgments (right panel of Fig. 3). Participants judged masked speech to have been significantly more distracting than pure speech, $F(1, 307) = 99.67$, $p < .001$, $\eta_p^2 = .25$. Whether or not participants were asked to provide prospective judgments before the objective distraction effects were measured did not affect the retrospective judgments overall, $F(1, 307) = 2.21$, $p = .138$, $\eta_p^2 = .01$, but interacted significantly with speech type, $F(1, 307) = 12.10$, $p < .001$, $\eta_p^2 = .04$. Both groups judged masked speech to have been more distracting than pure speech, but the main effect of speech type was larger in the prospective-judgment group, $F(1, 167) = 77.97$, $p < .001$, $\eta_p^2 = .32$, than

Table 1 Serial-recall results of Experiment 1

	Prospective-judgment group		Control group	
Masked speech	4.68	(0.16)	4.92	(0.17)
Pure speech	4.56	(0.17)	4.77	(0.18)
Quiet control	5.19	(0.16)	5.34	(0.16)

Note. The table shows the mean number of digits per trial that were recalled in the correct serial position in the distractor trials with masked or pure speech and in the quiet control trials. Values in parentheses represent the standard errors of the means in the control group, $F(1, 140) = 28.88$, $p < .001$, $\eta_p^2 = .17$. Relative to the neutral midpoint of the metacognition scale, participants in the prospective-judgment group judged masked speech to have been distracting, $t(167) = -16.69$, $p < .001$, $d = -1.29$, and pure speech to have been neither distracting nor helpful, $t(167) = -1.29$, $p = .199$, $d = -0.10$. Participants in the control group judged both masked speech, $t(140) = -14.80$, $p < .001$, $d = -1.25$, and pure speech, $t(140) = -5.72$, $p < .001$, $d = -0.48$, to have been distracting.

Discussion

A separate validation experiment has confirmed that speech masked with pink noise evokes a subjective experience of disfluency relative to pure speech. Consistent with the hypothesis that people rely on the (dis)fluency heuristic to arrive at prospective judgments about the distracting effects of sounds on cognitive performance (Bell et al., 2023; Komar et al., 2024), participants predicted masked speech to be more distracting than pure speech. In fact, they even predicted pure speech to be slightly helpful for serial recall. In sharp contrast to these prospective judgments but consistent with previous findings (Ellermeier & Hellbrück, 1998; Schlittmeier & Hellbrück, 2009), participants were objectively less distracted by masked speech than by pure speech. Given that masking speech with pink noise had opposite effects on prospective judgments of distraction and objective distraction effects, the present findings provide evidence of an even more pronounced metacognitive illusion in prospective judgments than those previously observed (Bell et al., 2023; Komar et al., 2024).

Even after multiple firsthand experiences of having ignored the two types of speech during the serial-recall task, participants judged masked speech to have been more, but not less, distracting than pure speech. This metacognitive illusion in retrospective judgments was more pronounced in participants who had previously provided prospective judgments than in those who had not. We will return to the retrospective judgments in the General discussion.

Experiment 2

Experiment 2 served to conceptually replicate the findings from Experiment 1, the main difference being that speech babble, rather than pink noise, was used to mask the speech (see also Jones & Macken, 1995; Keus van de Poll et al., 2015; Zaglauer et al., 2017). As in Experiment 1, masked speech should evoke a subjective experience of disfluency relative to pure speech, leading participants to predict masked speech to be more distracting than pure speech. However, given that speech babble masks the auditory changes in the speech signal that drive auditory distraction (Bell et al., 2019; Jones, 1993; Jones et al., 1993), participants should objectively be less distracted by masked speech than by pure speech.

Method

Participants

Data were collected as in Experiment 1. We aimed at collecting at least 750 valid data sets. Of the participants who had passed the audio check, 167 participants did not complete the experiment, 12 participants withdrew their consent to the use of their data, one participant was under 18 years old and 19 participants reported studying or having studied Psychology. The data sets of these participants were excluded from the analyses, together with the data sets of two participants who had participated twice (the data sets from their first participation were retained). Based on responses to the catch trial or the post-experiment questions, 140 further data sets could have been excluded. However, following a recommendation by Elliott et al. (2022), these data sets were retained in the analyses. This decision had no impact on the statistical conclusions. The final sample, characterized by diverse levels of education, consisted of 774 participants (388 female, 381 male, 5 non-binary) with a mean age of 37 ($SD = 12$) years. None of them had participated in Experiment 1. All participants were proficient German speakers; 730 of them indicated that German was their native language. Participants were randomly assigned to either the prospective-judgment group ($n = 383$) or the control group ($n = 391$). A sensitivity analysis using G*Power (Faul et al., 2007) showed that, given a sample size of $N = 774$ and $\alpha = \beta = .05$ (and, thus, a statistical power of $1 - \beta = .95$), a main effect of speech type on the objective distraction effect of the size $\eta^2_p = .02$ could be detected.

Materials

The same speech sequences as in Experiment 1 were used. A sequence of speech babble was generated by merging all the 32 speech sequences using Amadeus Pro 2.8.9 (https://www.hairersoft.com/). Masked-speech sequences were generated by combining the speech sequences with the sequence of speech babble at a speech-to-babble ratio of 0 dB(A) L_{eq}. The speech babble featured an amplitude envelope with attack and decay times of 250 ms to ensure a smooth onset and offset. Pure-speech sequences contained only the speech and no speech babble (Fig. 4). Thus, each of the 32 speech sequences was available both as a masked-speech sequence and as a pure-speech sequence.

Validation experiment

A separate validation experiment with different participants was conducted to test the effectiveness of the experimental manipulation of processing (dis)fluency. After having applied the same exclusion criteria as those reported above, data sets of 88 participants (32 female,

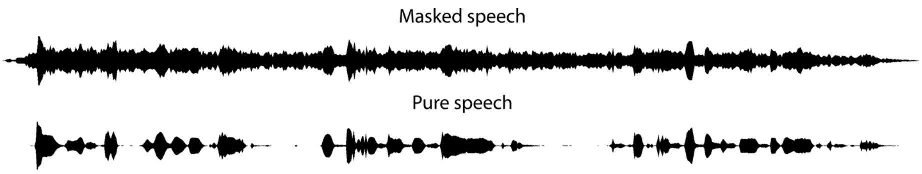

Fig. 4 Example waveforms of masked and pure speech used in Experiment 2

56 male) with a mean age of 39 (SD = 12) years were included in the analysis. The validation experiment for Experiment 2 was conducted as that for Experiment 1. Using the speech sequences as the units of analysis, masked speech (M = −2.99, SD = 0.81) was rated to be significantly more difficult to listen to than pure speech (M = 5.10, SD = 0.25), $F(1, 62) = 2926.60$, $p < .001$, $\eta_p^2 = .98$. These results confirm that the experimental manipulation of processing (dis)fluency was effective.

Design and procedure

Design and procedure were the same as those of Experiment 1. Questions and instructions were adapted to reflect the use of speech babble instead of pink noise (e.g., "In some trials, you heard sentences played with speech babble" or "In some trials, you heard sentences played without speech babble"). The median duration of the experiment was 19 min in the prospective-judgment group and 15 min in the control group.

Results

Prospective judgments

A repeated-measures analysis of variance was calculated to test the effect of the within-subjects factor speech type (masked, pure) on the prospective judgments (left panel of Fig. 5). Participants in the prospective-judgment group predicted masked speech to be significantly more distracting than pure speech, $F(1, 382) = 791.92$, $p < .001$, $\eta_p^2 = .67$. Relative to the neutral midpoint of the metacognition scale, masked speech was predicted to be distracting, $t(382) = -58.45$, $p < .001$, $d = -2.99$, while pure speech was predicted to be helpful, $t(382) = 5.67$, $p < .001$, $d = 0.29$.

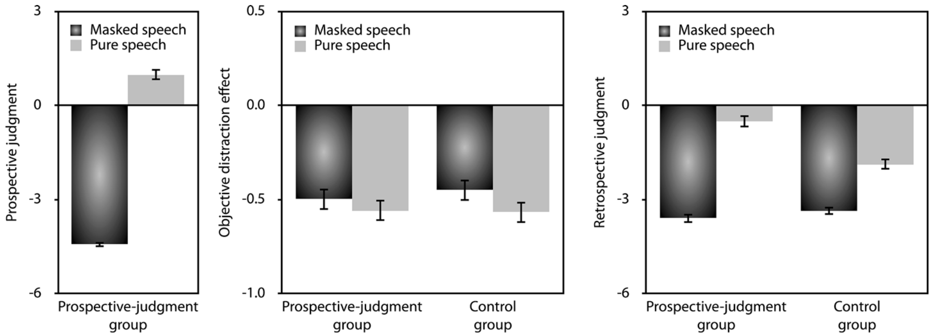

Fig. 5 Results of Experiment 2. The left panel shows the mean prospective judgment about the distracting effect of speech on serial-recall performance on a scale from −6 (very distracting) to +6 (very helpful) as a function of speech type (masked, pure). The middle panel shows the mean objective distraction effect of speech on serial-recall performance relative to quiet as a function of speech type (masked, pure) and group (prospective judgment, control). The right panel shows the mean retrospective judgment about the distracting effect of speech on the participants' serial-recall performance on a scale from −6 (very distracting) to +6 (very helpful) as a function of speech type (masked, pure) and group (prospective judgment, control). Error bars represent the standard errors of the means

Table 2 Serial-recall results of Experiment 2

	Prospective-judgment group		Control group	
Masked speech	4.65	(0.10)	4.81	(0.10)
Pure speech	4.59	(0.10)	4.69	(0.10)
Quiet control	5.15	(0.10)	5.26	(0.10)

Note. The table shows the mean number of digits per trial that were recalled in the correct serial position in the distractor trials with masked or pure speech and in the quiet control trials. Values in parentheses represent the standard errors of the means

Objective distraction effects

The objective distraction effects of masked and pure speech on serial-recall performance were calculated as in Experiment 1. The serial-recall results, averaged across participants in each group, are reported in Table 2.

A 2 × 2 mixed analysis of variance was calculated to test the effects of the within-subjects factor speech type (masked, pure) and the between-subjects factor group (prospective judgment, control) on the objective distraction effects (middle panel of Fig. 5). Participants were significantly less distracted by masked speech than by pure speech, $F(1, 772) = 5.63$, $p = .018$, $\eta_p^2 = .01$. Whether or not participants were asked to provide prospective judgments before the objective distraction effects were measured did not affect the objective distraction effects overall, $F(1, 772) = 0.09$, $p = .767$, $\eta_p^2 < .01$, and did not interact with speech type, $F(1, 772) = 0.55$, $p = .458$, $\eta_p^2 < .01$. Although masked speech disrupted serial recall significantly less than pure speech, both types of speech caused significant distraction relative to quiet. In the prospective-judgment group, participants were distracted by both masked speech, $t(382) = -8.56$, $p < .001$, $d = -0.44$, and pure speech, $t(382) = -10.05$, $p < .001$, $d = -0.51$. In the control group, participants were also distracted by both masked speech, $t(390) = -8.21$, $p < .001$, $d = -0.42$, and pure speech, $t(390) = -10.14$, $p < .001$, $d = -0.51$.

Retrospective judgments

A 2 × 2 mixed analysis of variance was calculated to test the effects of the within-subjects factor speech type (masked, pure) and the between-subjects factor group (prospective judgment, control) on the retrospective judgments (right panel of Fig. 5). Participants judged masked speech to have been significantly more distracting than pure speech, $F(1, 772) = 247.50$, $p < .001$, $\eta_p^2 = .24$. Overall, participants in the prospective-judgment group judged speech to have been significantly less distracting than participants in the control group, $F(1, 772) = 9.50$, $p = .002$, $\eta_p^2 = .01$. There was a significant interaction between speech type and group, $F(1, 772) = 30.84$, $p < .001$, $\eta_p^2 = .04$. Both groups judged masked speech to have been more distracting than pure speech, but the main effect of speech type was larger in the prospective-judgment group, $F(1, 382) = 184.08$, $p < .001$, $\eta_p^2 = .33$, than in the control group, $F(1, 390) = 66.56$, $p < .001$, $\eta_p^2 = .15$. Relative to the neutral midpoint of the metacognition scale, participants in the prospective-judgment group judged both masked speech, $t(382) = -24.70$, $p < .001$, $d = -1.26$, and pure speech, $t(382) = -2.55$, $p = .011$, $d = -0.13$, to have been distracting. Participants in the control group also judged both masked speech, $t(390) = -25.07$, $p < .001$, $d = -1.27$, and pure speech, $t(390) = -10.54$, $p < .001$, $d = -0.53$, to have been distracting.

Discussion

In Experiment 2, all findings from Experiment 1 were conceptually replicated. A separate validation experiment has confirmed that speech masked with speech babble evokes a subjective experience of disfluency relative to pure speech. Consistent with the hypothesis that people rely on the (dis)fluency heuristic to arrive at prospective judgments about the distracting effects of sounds on cognitive performance (Bell et al., 2023; Komar et al., 2024), participants predicted masked speech to be more distracting than pure speech. In fact, they even predicted pure speech to be slightly helpful for serial recall. In sharp contrast to these prospective judgments but consistent with previous findings (Jones & Macken, 1995; Keus van de Poll et al., 2015; Zaglauer et al., 2017), participants were objectively less distracted by masked speech than by pure speech. Given that masking speech with speech babble had opposite effects on prospective judgments of distraction and objective distraction effects, the present findings once again provide evidence of a particularly pronounced metacognitive illusion in prospective judgments.

Even after multiple firsthand experiences of having ignored the two types of speech during the serial-recall task, participants judged masked speech to have been more, but not less, distracting than pure speech. As in Experiment 1, this metacognitive illusion in retrospective judgments was more pronounced in participants who had previously provided prospective judgments than in those who had not.

General discussion

It has been proposed that people rely on the (dis)fluency heuristic when predicting the distracting effects of sounds on cognitive performance (Bell et al., 2023; Komar et al., 2024). The basic idea is that without direct metacognitive access, people rely on their subjective experience of the difficulty or ease of processing as a cue for predicting distraction. In the present study, we tested the hypothesis that if people use the (dis)fluency heuristic, then it should be possible to observe particularly pronounced metacognitive illusions in prospective judgments about the distracting effects of masked and pure speech on cognitive performance. Specifically, masking speech with pink noise (Experiment 1) or speech babble (Experiment 2) should have opposite effects on prospective judgments of distraction and objective distraction effects. The findings of both experiments support this hypothesis. Separate validation experiments have confirmed that masked speech evokes a subjective experience of disfluency relative to pure speech. Based on this subjective experience of relative disfluency, participants predicted masked speech to be more distracting than pure speech. In sharp contrast to these prospective judgments but consistent with previous findings (Ellermeier & Hellbrück, 1998; Jones & Macken, 1995; Keus van de Poll et al., 2015; Schlittmeier & Hellbrück, 2009; Zaglauer et al., 2017), participants were objectively less distracted by masked speech than by pure speech, likely because pink noise and speech babble masked the auditory changes in the speech signal that drive auditory distraction (Bell et al., 2019; Jones, 1993; Jones et al., 1993). Together, the present findings provide evidence of particularly pronounced metacognitive illusions in prospective judgments about the distracting effects of masked and pure speech on cognitive performance.

The present findings complement those of previous studies showing that people fail to correctly predict the distracting effects of music (Bell et al., 2023) and speech (Komar et al., 2024) on cognitive performance. In these studies, prospective judgments of distraction were affected by experimental manipulations such as playback direction (backward versus forward) and language (foreign versus native), even though these manipulations did not affect the objective distraction effects. However, the same pattern of results could have been observed if the direct-rating measure used to assess the prospective judgments of distraction had been more sensitive to the experimental manipulations than the serial-recall measure used to assess the objective distraction effects. The interpretation of the present findings is not complicated by such a limitation. Masking speech with pink noise or speech babble affected not only prospective judgments of distraction but also objective distraction effects, and it did so in opposite directions. The metacognitive illusions in prospective judgments reported here are thus even more pronounced than those reported in previous studies (Bell et al., 2023; Komar et al., 2024).

The hypothesis that people rely on the (dis)fluency heuristic when predicting the distracting effects of sounds on cognitive performance (Bell et al., 2023; Komar et al., 2024) implies that sounds that are experienced as being disfluent are predicted to be more distracting than sounds that are experienced as being fluent. In line with this hypothesis, participants predicted masked speech to be more distracting than pure speech. Surprisingly, they also predicted pure speech to be slightly helpful for serial recall. However, given that the effect was minuscule and has not been consistently observed across studies (cf. Komar et al., 2024), it is unclear whether this finding merits a substantive interpretation. That being said, it is possible that prospective judgments are influenced not only by the absolute (dis)fluency experience associated with a sound but also by the (dis)fluency experience relative to that associated with other sounds that are presented in the same context (cf. Wänke & Hansen, 2015). In the present study, the contrast between the (dis)fluency experiences associated with the two types of speech might have been particularly salient because masking speech with pink noise ($\eta_p^2 = .63$) or speech babble ($\eta_p^2 = .67$) influenced prospective judgments more than playing speech backward ($\eta_p^2 = .48$) or in a foreign language ($\eta_p^2 = .18$) in the previous study by Komar et al. (2024). This heightened contrast may have made the fluency experience associated with pure speech particularly salient.

Particularly pronounced metacognitive illusions were not only found in prospective judgments but also in retrospective judgments. Even though participants did not judge pure speech to have been helpful for serial recall, they judged masked speech to have been more, but not less, distracting than pure speech. This is in line with the finding that masking speech with pink noise or speech babble caused metacognitive illusions that were even more pronounced than those caused by the experimental manipulations used in previous studies (Bell et al., 2023; Komar et al., 2024). Notably, the metacognitive illusions in retrospective judgments were more pronounced in participants who had previously provided prospective judgments than in those who had not, suggesting that an initial focus on processing (dis)fluency when making prospective judgments amplified the metacognitive illusions in retrospective judgments. This pattern mirrors findings from research on self-regulated learning, which have shown that learners often continue to rely on ineffective learning strategies even after having experienced their limited effectiveness (Karpicke & Roediger, 2008; Kornell & Bjork, 2007, 2008). The fact that there was a metacognitive illusion in retrospective judgments may be viewed as contrasting with the findings from Beaman et al. (2014) showing that auditory distraction reduced participants' confidence in their responses and their willingness to report them during task performance. The differences between these two sets of results could potentially arise from differences in the memory tasks, stimuli, and timing and nature of the metacognitive judgments. More research is needed to understand why, in some situations, retrospective judgments remain remarkably prone to metacognitive illusions even when participants have experienced disconfirming evidence firsthand.

In practice, the metacognitive illusions reported here may lead people to make suboptimal decisions when choosing and controlling their study and work environments so as to optimize performance. For instance, they may fail to acknowledge the benefits of measures designed to reduce auditory distraction, such as masking distracting sounds with background noise or, under specific acoustic conditions, multi-talker environments like open-plan offices. However, a limitation of the present study is that this link between judgments of distraction and real-world decision-making has not been directly demonstrated. Therefore, a further aim for future research is to examine whether these judgments translate into choices about study and work environments and, ultimately, affect performance outcomes.

To summarize, the two experiments reported here demonstrate that masking speech with pink noise or speech babble causes the speech to be experienced as being more difficult to process than pure speech, leading participants to predict masked speech to be more distracting than pure speech. In sharp contrast to these prospective judgments, participants were objectively less distracted by masked speech than by pure speech. Even after multiple firsthand experiences of having ignored the two types of speech during the serial-recall task, masked speech was judged to have been more, but not less, distracting than pure speech. Masking thus had opposite effects on judgments of distraction and objective distraction effects. These findings provide evidence of particularly pronounced metacognitive illusions and support the hypothesis that people rely on (dis)fluency as a cue for predicting distraction.

Author contributions The study was part of a research project led by Raoul Bell. The study conception and design were developed collaboratively with contributions from all authors. Material preparation, data collection and analyses were performed by Gesa Fee Komar. The first draft of the manuscript was written by Gesa Fee Komar with subsequent input from Axel Buchner, Laura Mieth and Raoul Bell. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. The research reported here was funded by the German Research Foundation (BE 4311/6−1).

Declarations

Ethical approval In the two experiments reported here, participants gave written informed consent prior to participation. The local ethics committee has approved a series of experiments on the metacognition of auditory distraction which includes both experiments reported here. Both experiments were conducted in line with the Declaration of Helsinki.

Open practices statement The data of both experiments are available at the project page of the Open Science Framework, https://osf.io/jg8pq/.

Competing interests The authors declare that they have no competing interests.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. Personality and Social Psychology Review, 13(3), 219–235. https://doi.org/10.1177/1088868309341564

Atienzar, T. O., Pilgrim, L. K., Sio, U. N., & Marsh, J. E. (in press). Replicating and extending hemispheric asymmetries in auditory distraction: No metacognitive awareness for the left-ear disadvantage for changing-state sounds. *Journal of Cognitive Psychology*. Advance online publication. https://doi.org/10.1080/20445911.2024.2319268

Beaman, C. P., Hanczakowski, M., & Jones, D. M. (2014). The effects of distraction on metacognition and metacognition on distraction: Evidence from recognition memory. *Frontiers in Psychology*, *5*, 439. https://doi.org/10.3389/fpsyg.2014.00439

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632. https://doi.org/10.1016/0749-596X(89)90016-8

Bell, R., Komar, G. F., Mieth, L., & Buchner, A. (2023). Evidence of a metacognitive illusion in judgments about the effects of music on cognitive performance. *Scientific Reports*, *13*, 18750. https://doi.org/10.1038/s41598-023-46169-x

Bell, R., Mieth, L., Buchner, A., & Röer, J. P. (2021). Monetary incentives have only limited effects on auditory distraction: Evidence for the automaticity of cross-modal attention capture. *Psychological Research*, *85*(8), 2997–3009. https://doi.org/10.1007/s00426-020-01455-5

Bell, R., Mieth, L., Röer, J. P., & Buchner, A. (2022). The metacognition of auditory distraction: Judgments about the effects of deviating and changing auditory distractors on cognitive performance. *Memory & Cognition*, *50*(1), 160–173. https://doi.org/10.3758/s13421-021-01200-2

Bell, R., Mieth, L., Röer, J. P., & Buchner, A. (2024). The reverse Mozart effect: Music disrupts verbal working memory irrespective of whether you like it or not. *Journal of Cognitive Psychology*, *31*(1), 8–27. https://doi.org/10.1080/20445911.2023.2216919

Bell, R., Röer, J. P., Lang, A.-G., & Buchner, A. (2019). Reassessing the token set size effect on serial recall: Implications for theories of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(8), 1432–1440. https://doi.org/10.1037/xlm0000658

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*(6), 897–903. https://doi.org/10.3758/s13421-013-0307-8

Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 429–440. https://doi.org/10.1037/a0034407

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 56–64). Worth.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in "irrelevant speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1406–1414. https://doi.org/10.1037/0096-1523.24.5.1406

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect". *The Journal of the Acoustical Society of America*, *102*(4), 2191–2199. https://doi.org/10.1121/1.419596

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

Elliott, E. M., Bell, R., Gorin, S., Robinson, N., & Marsh, J. E. (2022). Auditory distraction can be studied online! A direct comparison between in-person and online experimentation. *Journal of Cognitive Psychology*, *34*(3), 307–324. https://doi.org/10.1080/20445911.2021.2021924

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Graf, L. K. M., Mayer, S., & Landwehr, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology*, *28*(3), 393–411. https://doi.org/10.1002/jcpy.1021

Gräve, E., Bell, R., & Buchner, A. (2024). Verbal and pictorial single-item scales are as good as their 10-item counterparts for measuring perceived usability. *Ergonomics*, *67*(12), 2096–2111. https://doi.org/10.1080/00140139.2024.2371061

Hanczakowski, M., Beaman, C. P., & Jones, D. M. (2017). When distraction benefits memory through semantic similarity. *Journal of Memory and Language*, *94*, 61–74. https://doi.org/10.1016/j.jml.2016.11.005

Hanczakowski, M., Beaman, C. P., & Jones, D. M. (2018). Learning through clamor: The allocation and perception of study time in noise. *Journal of Experimental Psychology: General*, *147*(7), 1005–1022. https://doi.org/10.1037/xge0000449

Herzog, S. M., & Hertwig, R. (2013). The ecological validity of fluency. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behavior* (pp. 190–219). Psychology Press. https://doi.org/10.4324/9780203078938

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

Jones, D. M., & Macken, W. J. (1995). Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied*, *1*(3), 216–226. https://doi.org/10.1037/1076-898X.1.3.216

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*(3), 318–328. https://doi.org/10.3758/BF03208264

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kattner, F., & Bryce, D. (2022). Attentional control and metacognitive monitoring of the effects of different types of task-irrelevant sound on serial recall. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(2), 139–158. https://doi.org/10.1037/xhp0000982

Keus van de Poll, M., Carlsson, J., Marsh, J. E., Ljung, R., Odelius, J., Schlittmeier, S. J., Sundin, G., & Sörqvist, P. (2015). Unmasking the effects of masking on performance: The potential of multiple-voice masking in the office environment. *The Journal of the Acoustical Society of America*, *138*(2), 807–816. https://doi.org/10.1121/1.4926904

Komar, G. F., Buchner, A., Mieth, L., van de Vijver, R., & Bell, R. (2024). Evidence of a metacognitive illusion in stimulus-specific prospective judgments of distraction by background speech. *Scientific Reports*, *14*, 24111. https://doi.org/10.1038/s41598-024-74719-4

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219–224. https://doi.org/10.3758/BF03194055

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Körner, U., Röer, J. P., Buchner, A., & Bell, R. (2017). Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. *Journal of Memory and Language*, *96*, 122–137. https://doi.org/10.1016/j.jml.2017.05.005

Macken, W. J., Mosdell, N., & Jones, D. M. (1999). Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 810–814. https://doi.org/10.1037/0278-7393.25.3.810

Maitland, A. (2009). Should I label all scale points or just the end points for attitudinal questions? *Survey Practice*, *2*(4), 1–4. https://doi.org/10.29115/SP-2009-0014

Marsh, J. E., Bell, R., Röer, J. P., & Hodgetts, H. M. (2024). Emerging perspectives on distraction and task interruptions: Metacognition, cognitive control and facilitation - part I. *Journal of Cognitive Psychology*, *36*(1), 1–7. https://doi.org/10.1080/20445911.2024.2314974

Mieth, L., Schaper, M. L., Kuhlmann, B. G., & Bell, R. (2021). Memory and metamemory for social interactions: Evidence for a metamemory expectancy illusion. *Memory & Cognition*, *49*(1), 14–31. https://doi.org/10.3758/s13421-020-01071-z

Röer, J. P., Bell, R., & Buchner, A. (2014). Evidence for habituation of the irrelevant-sound effect on serial recall. *Memory & Cognition*, *42*(4), 609–621. https://doi.org/10.3758/s13421-013-0381-y

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

Röer, J. P., Rummel, J., Bell, R., & Buchner, A. (2017). Metacognition in auditory distraction: How expectations about distractibility influence the irrelevant sound effect. *Journal of Cognition*, *1*(1), 2. https://doi.org/10.5334/joc.3

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung (Empirical studies on the development of rating scales for social-science research). *Zeitschrift für Sozialpsychologie*, *9*(3), 222–245.

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, *43*, 304–307. https://doi.org/10.1016/j.chb.2014.11.004

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019a). Metacognitive expectancy effects in source monitoring: Beliefs, in-the-moment experiences, or both? *Journal of Memory and Language*, *107*, 95–110. https://doi.org/10.1016/j.jml.2019.03.009

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019b). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 470–496. https://doi.org/10.1037/xlm0000602

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2023). Metacognitive differentiation of item memory and source memory in schema-based source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(5), 743–765. https://doi.org/10.1037/xlm0001207

Schlittmeier, S. J., & Hellbrück, J. (2009). Background music as noise abatement in open-plan offices: A laboratory study on performance effects and subjective preferences. *Applied Cognitive Psychology*, *23*(5), 684–697. https://doi.org/10.1002/acp.1498

Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention Perception & Psychophysics*, *74*(1), 194–203. https://doi.org/10.3758/s13414-011-0230-7

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 553–558. https://doi.org/10.1037/a0038388

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1264–1269. https://doi.org/10.1037/a0023719

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science*, *24*(3), 195–199. https://doi.org/10.1177/0963721414561766

Zaglauer, M., Drotleff, H., & Liebl, A. (2017). Background babble in open-plan offices: A natural masker of disruptive speech? *Applied Acoustics*, *118*, 1–7. https://doi.org/10.1016/j.apacoust.2016.11.004