Irrelevant speech disrupts memory more when spoken by a familiar talker

Jens Kreitewolf*[1], Malte Wöstmann[1], Sarah Tune[1], Michael Plöchl[1], Jonas Obleser*[1]

[1]*Department of Psychology, University of Lübeck*

**Correspondence should be addressed to:*

Jens Kreitewolf and Jonas Obleser

Department of Psychology, University of Lübeck, Maria-Goeppert-Str. 9a, 23562 Lübeck, Germany

Phone: +49 451 3101 3624

Fax: +49 451 3101 3604

E-mail: jens.kreitewolf@uni-luebeck.de; jonas.obleser@uni-luebeck.de

Abstract

When listening, familiarity with an attended talker's voice improves speech comprehension. Here, we instead investigated the effect of familiarity with a *distracting* talker. In an irrelevant-speech task, we assessed listeners' working memory for the serial order of spoken digits when a task-irrelevant, distracting sentence was produced by either a familiar or an unfamiliar talker. We tested two groups of listeners using the same experimental procedure. The first group were undergraduate psychology students (N=66) who had attended an introductory statistics course. Critically, each student had been taught by one of two course instructors, whose voices served as familiar and unfamiliar task-irrelevant talkers. The students made *more* errors when the task-irrelevant sentence was produced by the familiar compared to the unfamiliar talker. There was evidence for the absence of this effect in a second group of listeners — that is, family members and friends (N=20) who had known either one of the two talkers for more than ten years. While familiarity with an attended talker benefits speech comprehension, our findings indicate that familiarity with an ignored talker deteriorates working memory for target speech. The absence of this effect for family members and friends suggests that the degree of talker familiarity modulates distraction by irrelevant speech.

Keywords: talker familiarity, working memory, irrelevant-speech task, attention

Introduction

In natural situations, listeners are whelmed by a multitude of sounds that compete for attention. To comprehend a talker's speech message in the presence of competing distractors, both bottom-up voice characteristics, such as the talker's pitch, timbre, and articulatory style (reviewed by Diehl et al., 2004; Mathias and von Kriegstein, 2014), and top-down listening goals, such as selective attention to and working memory of the talker's speech (reviewed by Fritz et al., 2007; Shinn-Cunningham, 2008), guide listeners to focus on relevant sounds and to ignore irrelevant distractors.

One beneficial factor for speech comprehension that highlights the interplay of bottom-up features and top-down processes is familiarity with the talker's voice. Several studies have shown that listeners are better at comprehending target speech when it is produced by a familiar compared to an unfamiliar talker (Johnsrude et al., 2013; Kreitewolf et al., 2017; Levi et al., 2011; Newman and Evers, 2007; Nygaard and Pisoni, 1998, Souza et al., 2013). The so-called *familiarity benefit* is thought to rely on listeners' top-down knowledge about the bottom-up acoustics of the familiar talker's voice which can be exploited to direct selective attention in the sound mixture (Bressler et al., 2014; Kreitewolf et al., submitted). A conceivable implication of the familiarity benefit is that listeners do not only benefit from talker familiarity when their goal is to attend to the familiar talker's speech but also when they want to ignore it. In other words, if the familiarity benefit is based on prior knowledge about the familiar talker's vocal characteristics, then this knowledge might also help listeners to filter out irrelevant speech produced by the familiar talker.

Only few studies have investigated whether talker familiarity helps listeners to ignore distracting, task-irrelevant speech. Johnsrude et al. (2013) presented listeners with two concurrent spoken sentences and asked them to report key words from the target

sentence. Critically, the authors manipulated talker familiarity in these sentences such that either the attended (target), the unattended (masker), or none of the two was spoken by a highly familiar talker (i.e., the listeners' spouses). Listeners reported more key words from the target sentence when either the target or the masker sentence was spoken by a familiar compared to an unfamiliar talker, suggesting that talker familiarity facilitates both attending to target speech and ignoring distracting speech. In an earlier study, Newman and Evers (2007) asked listeners to attend to and immediately repeat (i.e., to shadow) speech from a target talker while a distracting talker was presented in the background. Listeners differed in their familiarity with one of the two concurrent talkers and whether they were told that they would hear a familiar voice (i.e., explicit knowledge) or not (i.e., implicit knowledge). In this study, talker familiarity was ensured by presenting speech produced by the listeners' university professor. The results showed that listeners with explicit knowledge about talker familiarity made fewer shadowing errors than listeners who only had implicit knowledge and listeners who were not familiar with the talker at all. Yet, this benefit was limited to familiarity with the target talker. Unlike Johnsrude et al. (2013), listeners' comprehension of target speech did not benefit from familiarity with the distracting background talker. Therefore, these two studies produced somewhat incongruent results with regard to the question of whether talker familiarity helps listeners to ignore distracting, task-irrelevant speech.

Here, we investigated the effect of talker familiarity on the distraction induced by task-irrelevant speech using a different, yet well-established experimental paradigm: the *irrelevant-speech task* (e.g., Colle and Welsh, 1976; Salamé and Baddeley, 1982). The irrelevant-speech task requires listeners to keep the serial order of to-be-attended target stimuli in working memory, while task-irrelevant, to-be-ignored speech is presented during memory retention. The number of incorrectly recalled targets is thought to increase

proportionally to the distraction by task-irrelevant speech, making the irrelevant-speech task an effective paradigm to study memory disruption by distracting speech. To modulate talker familiarity, we used an adaptation of the irrelevant-speech task in which a task-irrelevant, to-be-ignored sentence was either spoken by a familiar or an unfamiliar talker (Fig. 1).

The major objective of this study was to test whether familiarity with a task-irrelevant talker affects the serial recall of attended target speech. Of note, both directions of the familiarity effect are conceivable: on the one hand, one might expect that talker familiarity helps listeners to filter out irrelevant speech which would manifest in *fewer* recall errors when the task-irrelevant talker is familiar versus unfamiliar (in line with Johnsrude et al., 2013). On the other hand, distracting speech in an irrelevant-speech task has been shown to be more memory-disrupting when presented in a familiar (i.e., native) compared to an unfamiliar (i.e., foreign) language (Ellermeier et al., 2015). It is therefore possible that task-irrelevant speech is more memory-disrupting when produced by a familiar versus unfamiliar talker which would manifest in *more* recall errors when the task-irrelevant talker is familiar. Furthermore, we compared two groups of listeners that differed in the degree to which they were familiar with the task-irrelevant talkers: students who heard irrelevant speech produced by one of their course instructors (moderate familiarity) versus a group of family members and friends of the course instructors (high familiarity).

Methods

Participants

Two groups of listeners participated in this study. The first group were 66 undergraduate psychology students (59 females, mean age 23.11 yrs, age range 17-48 yrs; see Table I for

details) who had received classroom instructions by one of two talkers. Classroom teaching comprised a total of fourteen 90-minute sessions (of which students attended at least nine; see Table I). The second group of listeners were 20 family members and close friends of either one of the two talkers (8 females, mean age 39 yrs, age range 30-65 yrs; see Table I for details), who did not receive classroom instructions. All listeners were native German speakers. Students gained course credit for their participation, family members and friends were paid € 10. The experimental procedures were approved by the ethics committee of the University of Lübeck.

Table I. Details of listener groups.

Group 1: Students

familiar talker	*N*	*gender (females/males)*	*mean age (yrs)*	*age range (yrs)*	*mean familiarity (sessions)*	*familiarity range (sessions)*
JK	32	28/4	24.27	17-48	13.19	9-14
MP	34	31/3	22.09	18-30	12.94	10-14

Group 2: Family and friends

familiar talker	*N*	*gender (females/males)*	*mean age (yrs)*	*age range (yrs)*	*mean familiarity (yrs)*	*familiarity range (yrs)*
JK	10	3/7	39.10	30-65	23.10	14-34
MP	10	5/5	38.90	32-42	25.30	11-40

Stimuli

The to-be-attended speech stimuli were recordings of the German digits 1 to 9 which we had used in previous studies (Obleser et al., 2012; Wöstmann and Obleser, 2016; Wöstmann et al., 2017). All digits were spoken by a native German female talker (mean fundamental frequency, *f0*, 192 Hz). None of the listeners was familiar with the talker's voice. Digits had an average duration of 0.6 s (ranging from 0.5 to 0.7 s) and were concatenated with an onset-to-onset delay of 0.75 s. The resulting digit streams had an average duration of 6.6 s.

For the task-irrelevant speech, we used a German version of the speech-in-noise sentences (Erb et al., 2012) adopted from Kalikow et al. (1977). The same 50 sentences were recorded from two male talkers who were the authors JK and MP (both native German speakers). The mean *f0*, averaged across all sentences, was 93 Hz for talker JK, and 85 Hz for talker MP. Sentences produced by JK had an average duration of 2.17 s (ranging from 1.83 to 2.58 s); sentences produced by MP had an average duration of 2.19 s (ranging from 1.91 to 2.72 s). All sentences and digit streams were adjusted to the same root mean square (RMS) value using MATLAB (version 8.6, MathWorks, United States). On a given trial, the onset of the task-irrelevant sentence was delayed by 1409 ms (± 400 ms) so that, on average, sentences were centered in the middle of a 5-second memory retention phase (Fig. 1B).

Procedure

The listeners performed an online experiment implemented in Labvanced (Scicovery GmbH, Osnabrück, Germany) that used an adaptation of the irrelevant-speech paradigm (e.g., Colle and Walsh, 1976; Jones and Morris, 1992). The online experiment was executed in the browser in full-screen mode. Online experiments allow for the rapid collection of large datasets (e.g., Buhrmester et al., 2011) and have been shown to produce reliable data in several areas of behavioral research, including linguistics (e.g., Saunders et al., 2013) and psychoacoustics (e.g., McPherson and McDermott, 2018). Here, the online experiment had the additional advantage to prevent direct contact between the listeners and one of the two task-irrelevant talkers immediately before the start of the experiment which would have otherwise contaminated our manipulation of talker familiarity.

All listeners completed the experiment within one hour. On each trial, listeners heard the nine spoken digits in random order followed by a task-irrelevant sentence (Fig. 1B), while a fixation cross was presented on the computer screen. In about 17 % of trials (i.e., 20 out of

120 trials), silence was presented instead of task-irrelevant speech. Five seconds after the offset of the digit stream (i.e., at the end of the memory retention phase), a number pad consisting of the digits 1 to 9 was visually presented. Listeners were asked to select the digits in the order of their presentation. Each visually presented digit disappeared directly after it had been selected. This was done to avoid that the same digit could be selected more than once per trial. After the selection of the ninth digit, the next trial started with a delay of 500 ms. No feedback was given.

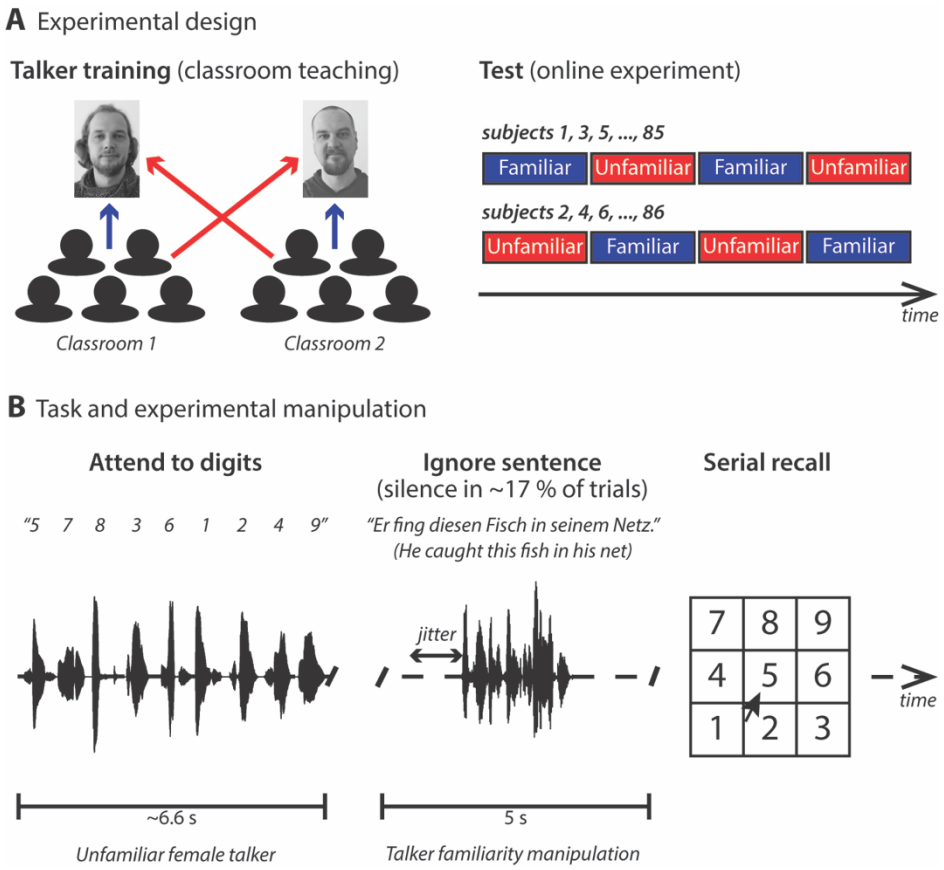

Figure 1. (**A**) Experimental design. Talker training in the students group was accomplished via classroom teaching. Students attended an introductory statistics course taught by one of two talkers. After the semester, the students as well as a group of family members and friends of the two talkers completed an online experiment. Familiar and unfamiliar talkers were presented in alternating blocks of trials. For half of the listeners, the experiment started with a familiar-talker block. (**B**) Task and experimental manipulation. In each trial of the online experiment, listeners heard the spoken digits 1 to 9 in random order followed by a task-irrelevant sentence. The onset of the sentence was jittered. In about 17 % of the trials, the sentence was omitted. The listeners' task was to select the digits in the order of presentation from a visually presented number pad. The same task-irrelevant sentences were spoken by both talkers. Digits were spoken by an unfamiliar female talker.

Listeners were asked to perform the online experiment in a quiet setting, to use a computer (no tablets, smart phones, etc.), and to listen to the sounds using headphones. Prior to the start of the experiment, listeners were instructed to silently rehearse the digit stream during the memory retention phase, to keep their eyes open and not to speak the digits aloud during a trial. Listeners could adjust the loudness of the sounds to a comfortable level. They were asked not to change the sound level during the experiment.

The experiment comprised four blocks (Fig. 1A, 'Test'). Each block consisted of 30 trials; 25 trials with a distracting, task-irrelevant sentence (distractor trials) and 5 trials with silence in the memory retention phase (no-distractor trials). In each block, no-distractor trials were pseudo-randomly interspersed with distractor trials with the restriction that the first no-distractor trial within a block could not occur within the first five trials and that two no-distractor trials could not occur in succession. The task-irrelevant familiar and unfamiliar talker were presented in alternating blocks of trials. Half of the listeners started with a familiar-talker block; the other half started with an unfamiliar-talker block (Fig. 1A, 'Test'). Note that no-distractor trials were acoustically identical for the task-irrelevant familiar and unfamiliar talker. However, the blocked presentation of the familiar and the unfamiliar talker allowed us to test whether the *infrequent* presentation of silence in the memory retention phase would affect performance differently when a familiar compared to an unfamiliar talker's voice was expected.

In the first and second half of the experiment (consisting of one familiar- and one unfamiliar-talker block), the exact same trials were presented; that is, the same combinations of digit streams and task-irrelevant sentences. This was done to ensure that differences in performance between blocks were due to familiarity with the task-irrelevant talker and not due to differences in the memorability of digit streams or the distractibility

of task-irrelevant sentences. To reduce item-specific learning, we ensured that the trial order was always different in familiar- and unfamiliar-talker blocks.

Statistical Analyses

To assess listeners' memory of the serial order of digits, we considered digits recalled at their respective position of presentation as *correct* and all other responses as *incorrect*. As a measure of distraction by task-irrelevant speech, we counted the number of errors per trial (0–9). All statistical analyses were carried out in R (R Core Team, 2017) using RStudio (version 1.1.383).

To overcome problems related to the unequal number of trials with and without a distractor in the memory retention phase, we fitted generalized linear mixed-effects models as implemented in the *lme4* package (Bates et al., 2015) to the number of errors per trial using Poisson regression (log link function; treating the number of errors as count data).

We followed an iterative model fitting procedure: starting with a minimal model that only included the random intercepts for subjects, we first added fixed-effects and then random-effects terms in a stepwise fashion. Fixed-effects terms were added in the order of their conceptual importance (i.e., Talker Familiarity, Distractor, Listener Group, and interactions between these factors; see below). Random-effects terms included the random intercepts for sentences and the subject- and sentence-specific random slopes for all significant main factors and interactions. After each step, we fitted the model using maximum-likelihood estimation, and assessed the change in model fit using likelihood-ratio tests. Model terms that significantly improved the model fit were kept in the model, non-significant terms were dropped (unless they were involved in higher-order interactions), resulting in a best-fitting model.

To investigate the potential effects of *Talker Familiarity* (familiar vs unfamiliar task-irrelevant talker), *Distractor* (distractor vs no-distractor trial), and *Listener Group* (students vs family and friends) on the number of errors, we modeled these predictors as fixed effects using deviation coding. To explore significant interaction terms, we performed post-hoc comparisons using Tukey's range tests as implemented in the *lsmeans* package (Lenth, 2016). We report unstandardized coefficients *b* to provide an estimate of effect size for fixed effects. Note that Poisson regression operates on a log transform of the dependent measure. Coefficients are therefore given in log-scale units. For significant random-effects terms, we report the likelihood-ratio test comparing the more complex model that includes the random-effects term with the simpler model excluding the term.

To enhance the interpretability of non-significant effects in particular, we calculated Bayes Factors (BFs) using the *brms* package (Bürkner, 2016). When comparing two statistical models, the BF indicates how many times more likely the observed data are under the more complex model (including a particular model term of interest) compared to the simpler model (excluding the model term of interest). In accordance with Jeffreys (1939/1961), a BF of 0.33 or smaller is interpreted as providing evidence in favor of the null hypothesis, while a BF of 3 or larger is interpreted as evidence against it.

Results

Figure 2 (A) shows the average proportions of errors as a function of digit position for distractor and no-distractor trials. Several basic observations can be made based on the data shown in this figure. Descriptively, listeners made fewer errors for digits presented at initial and final positions which is likely due to primacy and recency effects (e.g., Jones and Macken, 1993; Röer et al., 2013; Salamé and Baddeley, 1982; Schlittmeier et al., 2011; Wöstmann and Obleser, 2016). Interestingly, differences between distractor and no-

distractor trials were most pronounced for digits presented in the second half of the digit stream (i.e., between the fifth and eighth digit), and digits presented at these positions were generally more difficult to recall. Another observation is that, averaged across digit positions, listeners made more errors in distractor compared to no-distractor trials (Fig. 2B).

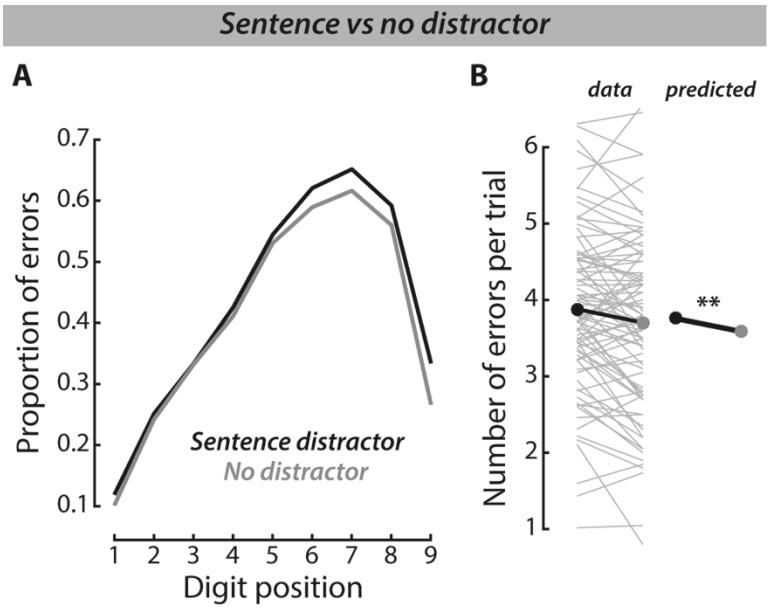

Figure 2. Distraction by task-irrelevant speech. (**A**) Average proportions of errors in sentence-distractor (black) and no-distractor trials (gray) as a function of digit position. (**B**) Left: number of errors per trial averaged across digit positions for sentence-distractor (left; black) and no-distractor trials (right; gray). Individual listeners' data are shown as gray lines. Across-listener averages are shown by dots and are connected by the thick black line. Right: number of errors per trial as predicted by the best-fitting model (i.e., controlling for the effects of all other model terms; see Table II for a summary of model terms). Significant effects are denoted by asterisks: ** $p < 0.01$.

Task-irrelevant speech distracts memory

To test the effect of Distractor, amongst other things, on the number of errors per trial, we fitted linear mixed-effects models. In sum, the best-fitting model included the three main factors Distractor, Talker Familiarity, and Listener Group as well as the interaction between the factors Talker Familiarity and Listener Group as fixed-effects terms (see Table II for a summary of fixed-effects terms). The random-effects terms included the random intercepts

for subjects and sentences as well as the subject-specific random slopes for Talker Familiarity.

Table II. Fixed-effects terms included in the best-fitting model.

Fixed effect	***Estimate b***	***Standard error SE***	***p-value***
(Intercept)	1.30057	0.03974	< 2e-16
Talker Familiarity (familiar)	–0.04718	0.02383	0.04772
Listener Group (students)	–0.01155	0.07803	0.88236
Distractor (distractor trial)	–0.04704	0.01676	0.00501
Talker Familiarity × Listener Group	0.09499	0.04737	0.04491

Note that the results are in log-scale units. Reference condition is given in brackets.

The inclusion of Distractor in the best-fitting model ($Z = -2.81$; $p = 0.005$; $b = -0.0470$; $BF = 7.6$) demonstrated *the irrelevant-speech effect*: presentation of a task-irrelevant sentence within the memory retention phase was indeed more distracting than a silence period. The sentence-specific random intercepts were also included in the best-fitting model ($X^2_{(1)} = 9.75$; $p = 0.002$; $BF = 4.09$), suggesting that the task-irrelevant sentences differed in distractibility.

Familiar talkers' speech is more memory-disrupting

The main aim of the present study was to investigate the effect of familiarity with a task-irrelevant talker on the serial recall of target digits. Figure 3 shows how the average proportion of errors evolves over digit positions in familiar- and unfamiliar-talker blocks (Fig. 3A) and the number of errors per trial averaged across digit positions for task-irrelevant familiar and unfamiliar talkers (Fig. 3B).

The best-fitting model included the factor Talker Familiarity ($Z = -1.98$; $p = 0.048$; $b = -0.0472$; $BF = 1.58$) and the interaction between the factors Talker Familiarity and Listener Group ($Z = 2.01$; $p = 0.045$). Post-hoc comparisons revealed that there was a significant

effect of talker familiarity in the students group ($Z = -4.07$; $p < 0.001$; $b = -0.0947$; $BF = 13.71$) but not in the group of family members and friends ($Z = 0.01$; $p = 0.994$; $b = 0.0003$; $BF = 0.04$). This means that students (Fig. 3C) but not family members and friends (Fig. 3D) made more errors when they were familiar with the task-irrelevant talker than when they were not.

Notably, the sample size was much smaller in the group of family and friends (N=20) than in the students group (N=66). However, the small Bayes Factor of 0.04 provides strong evidence for the absence of the effect of familiarity for family and friends. It is therefore unlikely that the comparably small sample was responsible for the lack of a familiarity effect in this group of listeners.

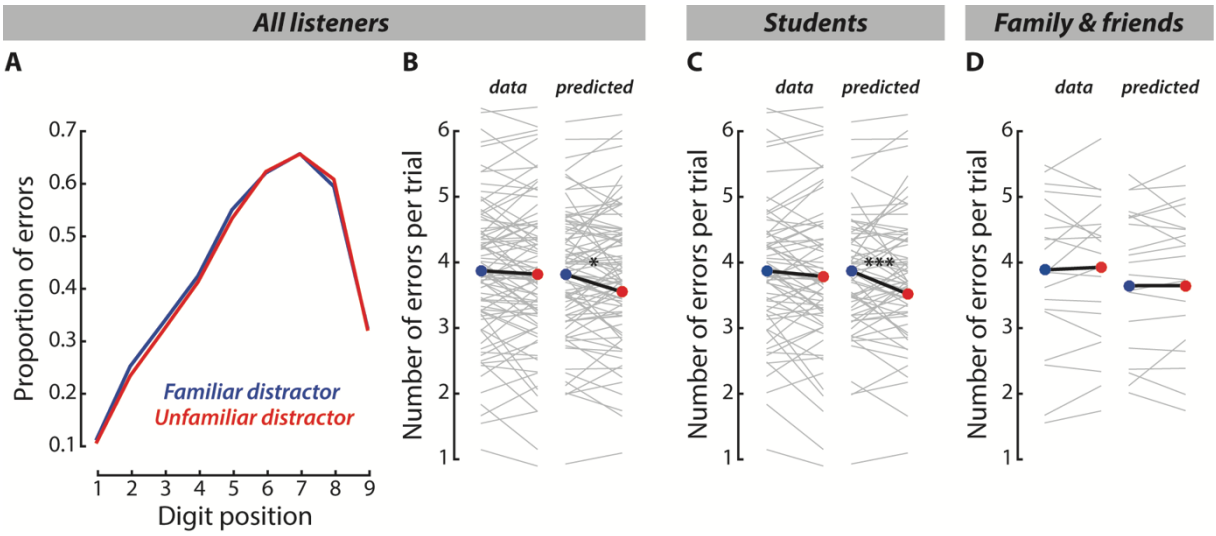

Figure 3. Effect of talker familiarity in all listeners (A, B) as well as in students (C) and in family members and friends (D). (**A**) Average proportions of errors in familiar- (blue) and unfamiliar-talker blocks (red) as a function of digit position. (**B, C, D**) Left: number of errors per trial averaged across digit positions for familiar- (left; blue) and unfamiliar-talker blocks (right; red). Individual listeners' data are shown as gray lines. Across-listener averages are shown by colored dots and are connected by the thick black line. Right: number of errors per trial as predicted by the best-fitting model. Gray lines show the pairs between the random intercepts for subjects and the subject-specific random slopes for talker familiarity. Significant effects are denoted by asterisks: * $p < 0.05$; *** $p < 0.001$.

The best-fitting model also included the subject-specific random slopes for Talker Familiarity ($X^2_{(2)} = 64.38$ $p < 0.001$, $BF = 17.72$), suggesting that the effect of talker familiarity

differed across listeners. Interestingly, however, the best-fitting model did not include the interaction between the factors Talker Familiarity and Distractor: compared to the simpler model, the inclusion of the fixed-effects term for the talker familiarity-by-distractor interaction did not significantly improve the model fit ($X^2_{(1)} = 0.61$; $p = 0.434$; BF < 0.001). Thus, the effect of talker familiarity (higher distraction in familiar- than unfamiliar-talker blocks) was not modulated by the presentation of a task-irrelevant sentence within the memory retention phase.

Note that the identity of the familiar talker was balanced across the entire group of listeners; that is, a similar number of listeners was familiar with JK's voice (N=42) or MP's voice (N=44) (see Table I for details). It thus seems unlikely that the effect of talker familiarity was mediated by differences in the distractibility of speech produced by the two talkers. Furthermore, compared to the best-fitting model, the inclusion of the talker-specific random intercepts did not significantly improve the model fit ($X^2_{(1)} = 0.00$; $p = 0.999$; BF = 0.02). This result indicated that task-irrelevant speech produced by the two talkers was similarly distracting.

No effect of listener group and block order

In total, students did not differ from family members and friends in their serial recall of digits. Compared to the simpler model, the inclusion of the factor Listener Group did not significantly improve the model fit ($X^2_{(1)} = 0.02$; $p = 0.888$; BF = 1.11). Note, however, that the BF shows no evidence for either the absence or presence of the effect of Listener Group. It is thus possible that the non-significant difference between the two listener groups was due to the rather small sample of family members and friends.

To investigate potential effects of the presentation order of familiar and unfamiliar talkers (Fig 1A, 'Test'), we carried out a control analysis in which we added the factor *Block Order*

(familiar first vs unfamiliar first) as a fixed-effects term to the best-fitting model. Compared to the respective simpler model, neither the main effect of Block Order ($X^2_{(1)} = 0.04$; $p = 0.843$; BF = 0.06) nor the interaction term between the factors Block Order and Talker Familiarity ($X^2_{(4)} = 5.69$; $p = 0.223$; BF = 0.05) significantly improved the model fit. This means that neither the overall performance nor the effect of talker familiarity depended on whether listeners started the experiment with a familiar- or an unfamiliar-talker block.

Discussion

The present study used a variant of the irrelevant-speech task to investigate the effect of familiarity with a task-irrelevant talker on the serial recall of target speech. The most important finding was that task-irrelevant speech is more distracting when produced by a familiar compared to an unfamiliar talker. This effect critically depends on the degree to which listeners are familiar with the task-irrelevant talker. Students who were moderately familiar with a task-irrelevant talker produced more recall errors when task-irrelevant speech was spoken by the familiar talker, while talker familiarity had no effect on the serial recall of target speech in a group of family members and friends who were highly familiar with the task-irrelevant talker.

Degree of familiarity modulates distraction by irrelevant speech

One previous study (Newman and Evers, 2007) showed that listeners' comprehension of target speech does not benefit from moderate familiarity with a distracting background talker (i.e., the listeners' university professor). However, Johnsrude et al. (2013) found that high familiarity with a distracting talker (i.e., the listener's spouse) improves comprehension of target speech — a finding that has been recently extended to familiarity with accented speech (Senior and Babel, 2018). It might thus seem surprising that in the

group of students in the present study, the familiar talker was more distracting than the unfamiliar one.

Together, these previous studies suggest that familiarity with a distracting talker can be beneficial for listeners' comprehension of target speech, but that a high degree of familiarity with the talker is needed for this benefit to occur. Talker familiarity in the present study is comparable to the two previous studies both in terms of the magnitude of familiarity (students who acquired familiarity with their course instructor during one semester were compared to family members and friends who had years of experience with the familiar talker's voice) and in terms of the social setting in which talker familiarity was instantiated (classroom teaching for students versus various settings of personal life for family members and friends). Despite these apparent similarities between the present and previous studies, our results follow the exact opposite pattern: first, familiarity with a distracting talker was not beneficial but rather harmful for the recall of target speech from memory, and second, moderate instead of high talker familiarity was necessary for the familiarity effect to occur.

Talker familiarity disrupts working memory for target speech

To succeed in the irrelevant-speech task, selective attention to items in working memory is needed (for a review on the interaction of attention and working memory, see Awh et al., 2006). Notably, the present study used an irrelevant-speech task to investigate how irrelevant speech disrupts memory for target speech, while previous studies by Johnsrude et al. (2013) and Newman and Evers (2007) investigated how irrelevant speech disrupts comprehension of target speech.

Furthermore, these previous studies examined the distraction by a familiar talker using concurrent speech. For example, Johnsrude et al. (2013) presented listeners with two

concurrent sentences (from the Coordinate Response Measure corpus; Bolia et al., 2000) of the form "Ready [call sign], go to [color] [number] now" (e.g., "Ready Baron, go to green six now"). The listeners' task was to report the color and number word from the target sentence indicated by the call sign. A conceivable strategy in this task is to divide attention between target and masker sentences and retrospectively select the two key words from the target sentence. This strategy creates low working-memory load because listeners only need to remember four words (i.e., two from the target and two from the masker sentence). In contrast, the irrelevant-speech task used in the present study loads high on working memory because listeners had to remember the serial order of nine digits while ignoring a task-irrelevant sentence. The familiarity effect in the present study therefore indicates that listeners' working memory of target speech is more disrupted when task-irrelevant speech is produced by a familiar talker.

Why is a familiar talker more memory-disrupting?

We consider three possible factors that contribute to the disruption of memory by a familiar talker. First, our results are in line with previous work suggesting that the intelligibility of task-irrelevant speech modulates memory disruption in the irrelevant-speech task. Specifically, it has been shown that listeners' working memory of target speech is more disrupted when task-irrelevant speech is spectrally intact and thus more intelligible (Ellermeier et al., 2015; Wöstmann and Obleser, 2016; Wöstmann et al., 2017). Spectral degradation deteriorates speech intelligibility (e.g., Shannon et al., 1995) which can explain why listeners' working memory suffered more from high than low spectral detail in task-irrelevant speech. Talker familiarity constitutes another determinant of speech intelligibility as several studies have shown that listeners are better at comprehending speech from a familiar compared to an unfamiliar talker (Johnsrude et al.,

2013; Kreitewolf et al., 2017; Levi et al., 2011; Newman and Evers, 2007; Nygaard and Pisoni, 1998, Souza et al., 2013). For example, Johnsrude et al. (2013) found a 10–15 % improvement in speech intelligibility by talker familiarity. The stronger disruption of working memory by the familiar talker's irrelevant speech can be therefore explained by the higher intelligibility of the familiar talker which renders it more difficult to ignore.

Second, compared to an unfamiliar talker, task-irrelevant speech produced by a familiar talker might capture attention and draw it away from items in memory. For example, it has been shown that speech items of high familiarity, such as the listener's own name (Röer et al., 2013) and the listeners' native language (Ellermeier et al., 2015), enhance memory-disruption, possibly by drawing attention resources away from the serial memory of target items in the irrelevant-speech task.

Third, the familiarity effect in the present study was unaffected by the presence versus absence of a task-irrelevant sentence. In other words, listeners produced more recall errors in blocks of trials with a task-irrelevant familiar talker, irrespective of whether a task-irrelevant sentence was presented (distractor trial) or not (no-distractor trial). Although no-distractor trials were relatively rare (about 17 % of trials), this finding is consistent with the view that the sheer expectation of a familiar talker's voice disrupts working memory for target speech, lending credence to the notion that memory disruption in the irrelevant-speech task is modulated by top-down predictions on the distractor.

Notably, the present results extend previous findings by showing that stronger memory disruption by task-irrelevant speech is not only driven by bottom-up stimulus features, such as acoustic detail, but also by prior experience of the listener; that is, talker familiarity. Thus, our results have implications for a prominent theory in irrelevant-speech research — the changing-state hypothesis (e.g., Jones et al., 1992) — which holds that the disruption

of working memory depends on the variability of the task-irrelevant acoustic structure. The familiarity effect found in the present study is difficult to reconcile with the changing-state hypothesis because, across the entire group of listeners, there was no acoustic difference in task-irrelevant sentences produced by the familiar and unfamiliar talkers: the same task-irrelevant sentences were produced by both talkers and the identity of the familiar talker was balanced across listeners. Instead of the variability of acoustic input, our manipulation of talker familiarity determined whether or not listeners had prior knowledge about the acoustic characteristics of the task-irrelevant talker.

Benefits and costs of talker familiarity

While previous work showed that familiarity with an attended talker is beneficial for speech comprehension, our findings showed that familiarity with an ignored talker can be detrimental for working memory of target speech. One of our original hypotheses was that listeners might actually be better at ignoring a familiar task-irrelevant talker because they are better at filtering out task-irrelevant speech produced by the familiar talker. The absence of a familiarity effect for family members and friends provides at least indirect evidence for this hypothesis: listeners who are highly familiar with a talker's voice might, on the one hand, experience higher distraction by the familiar talker's speech, but on the other hand, form a more efficient perceptual filter for the voice characteristics of the familiar talker. These opposing effects might annul an effect of talker familiarity.

The familiarity benefit for an attended talker's speech requires rather little experience with the talker's voice. For example, we showed in a previous study that even incidental exposure to a talker's voice in the laboratory benefits speech comprehension (Kreitewolf et al., 2017). In contrast, the benefit from familiarity with a distracting talker, might require a much higher degree of talker familiarity as previous studies suggest that comprehension

of target speech only benefits from high (Johnsrude et al., 2013) but not moderate (Newman and Evers, 2007) familiarity with a distracting talker. Hence, the degree of talker familiarity modulates the degree to which a listener benefits from familiarity with a distracting talker, potentially because the effectiveness of the filter for the familiar talker's voice depends on high familiarity.

Conclusions

Here, we demonstrated that familiarity with a distracting talker's voice deteriorates working memory of target speech. Our results speak to the more general view that the goal of a listener determines whether a specific feature of the speech signal is beneficial or detrimental: whatever helps listeners' performance when they are attending to relevant speech, such as acoustic detail (e.g., Wöstmann and Obleser, 2016) and talker familiarity (e.g., Kreitewolf et al., 2017), can hinder them when their goal is to ignore irrelevant speech.

References

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201-208.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America, 107*(2), 1065-1066.

Bressler, S., Masud, S., Bharadwaj, H., & Shinn-Cunningham, B. (2014). Bottom-up influences of voice continuity in focusing selective auditory attention. *Psychological research, 78*(3), 349-360.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science, 6*(1), 3-5.

Bürkner, P. C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1-28.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of verbal learning and verbal behavior, 15*(1), 17-31.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol., 55*, 149-179.

Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K., & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America, 138*(3), 1561-1569.

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia, 50*(9), 2154-2164.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current opinion in neurobiology, 17*(4), 437-455.

Jeffreys, H. (1939/1961). The Theory of Probability. 1st/3rd Edn. Oxford: Oxford University Press.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science, 24*(10), 1995-2004.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 369.

Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology Section A, 44*(4), 645-669.

Jones, D., & Morris, N. (1992). Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scandinavian Journal of Psychology, 33*(3), 212-229.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the acoustical society of America, 61*, 1337–1351.

Kreitewolf, J., Mathias, S. R., Trapeau, R., Obleser, J., & Schönwiesner, M. (submitted). Voice-feature continuity helps perceptual grouping in the cocktail party.

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit Talker Training Improves Comprehension of Auditory Speech in Noise. *Frontiers in psychology, 8*, 1584.

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software, 69*(1), 1-33.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible?. *The Journal of the Acoustical Society of America, 130*(6), 4053-4062.

Mathias, S. R., & von Kriegstein, K. (2014). How do we recognise who is speaking. *Front Biosci (Schol Ed), 6*, 92-109.

McPherson, M. J., & McDermott, J. H. (2018). Diversity in pitch perception revealed by task dependence. *Nature Human Behaviour, 2*(1), 52.

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics, 35*(1), 85-103.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & psychophysics, 60*(3), 355-376.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience, 32*(36), 12376-12383.

R Core Team (2017). R: A Language and Environment for Statistical Computing. *Vienna: The R Foundation for Statistical Computing*.

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

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of verbal learning and verbal behavior, 21*(2), 150-164.

Saunders, D. R., Bex, P. J., & Woods, R. L. (2013). Crowdsourcing a normative natural language dataset: A comparison of Amazon Mechanical Turk and in-lab data collection. *Journal of medical Internet research, 15*(5).

Schlittmeier, S. J., Weisz, N., & Bertrand, O. (2011). What characterizes changing-state speech in affecting short-term memory? An EEG study on the irrelevant sound effect. *Psychophysiology, 48*(12), 1669-1680.

Senior, B., & Babel, M. (2018). The role of unfamiliar accents in competing speech. *The Journal of the Acoustical Society of America, 143*(2), 931-942.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in cognitive sciences, 12*(5), 182-186.

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology, 24*(8), 689-700.

Wöstmann, M., & Obleser, J. (2016). Acoustic detail but not predictability of task-irrelevant speech disrupts working memory. *Frontiers in human neuroscience, 10*, 538.

Wöstmann, M., Lim, S. J., & Obleser, J. (2017). The human neural alpha response to speech is a proxy of attentional control. *Cerebral Cortex*, *27*(6), 3307-3317.