The Role of Talker Familiarity in Auditory Distraction

Brittan A. Barker and Emily M. Elliott

Louisiana State University

Author Note

Brittan A. Barker, Department of Communication Sciences and Disorders, Louisiana State University; Emily M. Elliott, Department of Psychology, Louisiana State University.

Brittan A. Barker is now at the Department of Communicative Disorders and Deaf Education, Utah State University.

The authors would like to thank Annie Schubert for her help with sound editing and E-Prime programming and Noelle Brown for recording the irrelevant speech stimuli for the study. Alicia Briganti assisted with data collection.

Correspondence concerning this article should be addressed to Brittan Barker, Department of Communicative Disorders and Deaf Education, Utah State University, Logan, UT 84342. E-mail: brittan.barker@usu.edu

Abstract

The current research employed a classic irrelevant sound effect paradigm and investigated the talker-specific content of the irrelevant speech. Specifically, we aimed to determine if the participants' familiarity with the irrelevant speech's talker affected the magnitude of the irrelevant sound effect. Experiment 1 was an exploration of talker familiarity established in a natural listening environment (i.e. a university classroom) in which we manipulated the participants' relationships with the talker. In Experiment 2 we manipulated the participants' familiarity with the talker via 4 days of controlled exposure to the target talker's audio recordings. For both Experiments 1 and 2, a robust effect of irrelevant speech was found; however, regardless of the talker manipulation, talker familiarity did not influence the size of the effect. We interpreted the results within the processing view of the auditory distraction effect and highlighted the notion that talker familiarity may be more vulnerable than once thought.

Keywords: irrelevant sound effect, irrelevant speech attention, working memory, talker familiarity

The Role of Talker Familiarity in Auditory Distraction

It is well established that background noise negatively affects peoples' capacity to immediately recall visually presented information (Colle & Welsh, 1976), and this auditory distraction phenomenon is often referred to as the irrelevant sound effect (ISE). The ISE is one example of an auditory distraction paradigm that provides researchers with an ecologically valid construct to explore theoretical aspects of cognition intertwined with short-term memory and selective attention, while simultaneously working towards an understanding of its underlying neurological mechanisms (for a review see Hughes & Jones, 2001). As a result, researchers explore the ISE from a number of perspectives via various manipulations, one of which is to investigate the acoustic and semantic content of the irrelevant speech in an attempt to determine which of its properties affect the magnitude of the ISE (e.g., Tremblay & Jones, 1999; Röer, Körner, Buchner, & Bell, 2017). Despite the fact that the research on this particular auditory distraction paradigm extends over 40 years, researchers do not agree on the precise mechanism(s) of the disruption to serial recall performance. The current research drew upon findings from the field of hearing science to understand the role of the talker-specific content of the irrelevant speech; thus, utilizing an interdisciplinary approach to gain a clearer understanding of the cause of this distraction.

Views of Auditory Distraction

While it is well accepted that the degree of acoustic variability of the irrelevant sounds is critical for the disruptive effect on serial recall to occur (i.e. the changing state effect; Ellermeier & Zimmer, 2014; Jones, Madden, & Miles, 1992), the role of attentional factors has been called into question. For example, an individual's working memory capacity does not have a relationship with the size of their ISE (see Beaman, 2004; Elliott & Briganti, 2012). This lack of

a relationship between the size of the ISE and working memory capacity is in contrast to research employing an auditory deviant paradigm. The serial recall version of the deviant paradigm is similar in design to the ISE; however, research participants are presented with irrelevant sounds that contain an unexpected and deviant item (e.g., a male voice changing to a female voice). The magnitude of the deviant effect in the auditory deviant paradigm was shown to correlate with working memory capacity (e.g., Sörqvist & Rönnerberg, 2014; although see Körner, Röer, Buchner, & Bell, 2017 for an opposing view).

Hughes (2014) built upon the divergence of these two signature findings, the *changing-state effect* (Jones et al., 1992) and the *auditory deviant effect* (Hughes, Vachon, & Jones, 2007) to support the duplex model of auditory distraction. Within this model, it is hypothesized that the degree of disruption in the changing-state effect is caused by the overlap in processing demands resulting from the obligatory order cues in to-be-ignored auditory stimuli, conflicting with the order cues of the to-be-remembered stimuli, especially when serial order recall is required (i.e. interference by process). The changing acoustic features of the to-be-ignored stimuli are referred to as a form of precategorical auditory distraction (Marsh et al., 2018). In contrast, postcategorical auditory distraction results from semantic or phonological features of the to-be-ignored stimuli that may result in attention capture (e.g., Röer et al., 2017).

Prior researchers have also examined the role of attentional factors in auditory distraction effects, and proposed a functional view of auditory distraction (e.g., Allport, 1993; Neumann, 1996; Hughes & Jones, 2003). Within this view, it is argued that people must maintain a balance of attentional selectivity. That is, they must be able to focus on processing the experimental task in the forefront (e.g., visual digit-span task), while also allowing for some monitoring or awareness of the irrelevant auditory channel's content. This internal monitoring promotes human

survival. The openness to incoming auditory information allows an individual to perceive and process something critical, like the sound of an emergency vehicle's siren or a building's fire alarm. Supporting this functional view, Röer et al. (2017) conducted experiments using an ISE serial recall paradigm with taboo words presented in the irrelevant channel. When a variety of taboo words were presented, as opposed to the repetition of the same taboo word or neutral words, the participants' serial recall was significantly more disrupted by the changing taboo words than changing neutral words.

The finding—changing taboo words yield the greatest disruption—is a critical piece of evidence that people process the semantic content of the irrelevant sounds. Because the participants in their study were instructed to ignore the auditory information, without a functional system of selectivity, the effect of the changing taboo words should have been no different than the changing neutral words. Röer et al. (2017; Experiment 2) further suggested that the disruptive effect of taboo words was not related to individual levels of working memory capacity. The authors' overall interpretation of the findings reflected upon the flexibility of the cognitive system, as opposed to viewing the disruption from the changing taboo words as a failure of the system. While researchers have shown a number of sounds cause disruption to focal task performance, such as tones or alarms, the authors discussed the potential importance of speech as a distracting stimulus. As stated by Röer et al., “it seems plausible that speech sounds (and sounds that resemble speech in their physical properties) play a special role” (p. 748). Examining the components of the speech sounds in the irrelevant channel may provide new insights into the processing of irrelevant information. This was the case in the present study, when we manipulated the research participants' familiarity with the speaker of the irrelevant speech.

The Role of Talker Familiarity

Generally speaking, speech is made up of two components: linguistic information and indexical or talker-specific information (Abercrombie, 1967; Pisoni, 1997). The linguistic information refers to the segmental speech components such as phonemes and morphemes while the talker-specific information includes components reflected in the speech signal that are particular to an individual talker's voice (e.g., accent, gender, age, etc...) as opposed to any linguistic content. Historically, many studies in the field of speech perception focused on the perception, encoding, processing, and retrieval of the linguistic information (e.g., Kuhl, 1993; Liberman, 1957; Nearey, 1990), while the talker-specific information in the speech signal was often unaccounted for or dismissed from the problem space. Over the past decades, however, researchers showed that the talker-specific information is neurologically coupled with its linguistic information (Chandrasekaran et al., 2011; Naoi, et al., 2012) and often affects the perception and processing of linguistic information across a variety of listeners and a number of laboratory tasks. In particular, when a listener is familiar with a talker's voice, the listener's ability to perceive and process the talker's speech is often better than if they are listening to a stranger (e.g., Nygaard & Pisoni, 1998; Souza, Gehani, Wright, & McCloy, 2013). Furthermore, the talker familiarity effect seems strongest when listeners are explicitly reminded of their relationship with the talker before the experimental task begins (Newman & Evers, 2007). Interestingly, despite past work examining talker variability in the to-be-ignored auditory stimuli (Hughes, Marsh, & Jones, 2009; 2011), no prior studies on the ISE have attempted to manipulate the relationship of the research participant's familiarity with the talker.

The Current Study

The work of Röer and colleagues suggested that certain factors increase the allocation of attentional resources towards the irrelevant channel (taboo words, Röer et al., 2017; one's own name, Röer, Bell, & Buchner, 2013), whereas an interference-by-process view instead places emphasis on the processes needed to complete the task at hand (Hughes, 2014) and the precategorical features of the irrelevant speech (Marsh et al., 2018). Based upon the findings of Röer and colleagues, in addition to the known effects of talker familiarity on speech perception (Nygaard & Pisoni, 1998; Souza et al., 2013), we predicted that the greater the degree of familiarity of the participant with the talker, the greater the resulting size of the disruption to serial recall performance. This talker familiarity effect would be further increased when the participant was made explicitly aware of their relationship with the talker, relative to being uninformed about the talker's identity, or having no familiarity at all. Alternatively, the interference-by-process view would only predict an effect of talker familiarity if the task required a specific interaction with the characteristics of the talker's voice, which is not necessary to complete a serial recall task. This prediction indicates that in a typical ISE task, the familiarity of the participant with the talker's voice producing the irrelevant speech would be processed as a postcategorical feature.

Experiment 1

In Experiment 1 we explored the role of talker familiarity established in a natural listening environment—a college classroom (see also Newman & Evers, 2007). We did so within the context of a classic ISE task and we recruited students from introductory psychology classes to serve as participants. One third of the participants were a part of the control group and were unfamiliar with the irrelevant talker's speech; the remaining participants were familiar with the

irrelevant talker's speech because it was recorded by their introductory psychology class instructor. Of these remaining participants—half of them were informed that they were familiar with the talker and half of them were not informed that they were familiar with the talker.

Method

Participants. 207 undergraduate students from a large, state university enrolled in introductory psychology classes participated in the experiment. We recruited participants through the university's Department of Psychology research participant pool, and compensated them with either course credit or extra credit in psychology classes. We quasi-randomly assigned participants to 1 of 3 groups based on their familiarity with the background talker. The final sample ($N = 179$) included $n = 96$ for the unfamiliar talker condition, $n = 40$ for the informed "familiar" talker condition, and $n = 43$ for the uninformed "familiar" talker condition. Participant inclusion criteria were as follows: normal or corrected-to-normal vision; hearing within-typical-limits (5 excluded); and native English speaker (7 excluded). Of the total number of original participants, 19 were removed from data analyses based on answers to the post-experiment questionnaire (see Results section for details).

Experimental design. We utilized a 2×3 mixed design with *auditory condition* (silent, speech) as a within-subjects factor and *talker familiarity* (unfamiliar, uniformed, informed) as a between-subjects factor.

Materials. We used *E-Prime* software 2.0 (2012) to conduct the experiment on *Dell* desktop computers equipped with 17-inch monitors and *Sony* circumaural headphones. The experimental setup was located in a sound-treated room. The *E-Prime* code and stimuli used for Experiment 1 and 2 can be secured by researchers for replication purposes by emailing the corresponding author.

Visual stimuli. Seven target digits (1-7) served as the visual stimuli. Digits were printed in black, typeface and were comfortably viewed by the participants. All digits were centered on the monitor and displayed on a white background.

Auditory stimuli. Talker details. A female instructor for an “Introduction to Psychology” course at the university recorded the distractor sentences.

We expected participants in the familiar condition to be familiar with this voice because the talker was their introductory psychology instructor. The participants were invited to assist with the study approximately mid-way through the semester, to ensure that they had ample time to become familiar with the instructor’s voice. The participants in the unfamiliar condition were enrolled in introductory psychology course sections other than the target talker’s.

Distractor sentences. Sentences from the *Revised List of Phonetically Balanced Sentences: Harvard Sentences* (IEEE, 1969) served as the distractor sentences for all participants. The sentences were combined and edited into sound files approximately 6s in length using the *Adobe Audition 2.0* (2005) sound editing software on a *Dell Optiplex 740* computer with a *Delta 101LT PCI* sound card. The recordings were edited and equated across root mean square (r.m.s.). These edited sentences served as the distractor sentences for all participants.

Procedure. Each participant was tested individually at a personal computer using a typical ISE paradigm. A researcher obtained informed consent prior to the experiment. The participants in the unfamiliar and uniformed groups were given no information about the irrelevant speech’s talker. However, for participants in the informed group, prior to the beginning of the experiment, “The voice you will hear is your PSYC 2000 professor, [INSTRUCTOR’S NAME], so you should be familiar with this voice.” appeared on the computer monitor.

The experiment began with 3 practice trials, followed by the 40 test trials presented randomly. The 7 target digits were presented visually on the computer monitor at a rate of 750 ms each, while the distractor sentences were simultaneously presented at a comfortable listening level via headphones worn by the participant. The total duration of stimuli presentation was approximately 6000 msec. Of the 40 experimental trials, 20 were silent trials and 20 were auditory distractor trials; the different trial types were intermixed across the 40 total trials. Each participant was instructed to remember the digits that they saw on the screen and to ignore the sentences they heard over the headphones. For each trial, after the digits were presented on the monitor, the participant was instructed to recall the digits in the order in which they were presented and use the number key pad to record their responses. After the completion of all trials, the participant answered follow-up, multiple-choice questions via the computer about the distractor talker. Follow-up questions can be found in Appendix A. Finally, the experimenter debriefed the participant and thanked them for their participation.

Results and discussion

Assessing talker familiarity. For our first step in data analysis we examined the participants' responses to the follow-up questionnaire. It was critical those participants in the informed condition, for example, could correctly recognize the talker as familiar and provide the correct name of their instructor (recall, they were told her name at the beginning of the experiment). From the original sample, 4 individuals in the informed group responded, "I don't know" to the question asking about the familiarity of the to-be-ignored voice. We removed those individuals from further data analyses since the voice was not "explicitly familiar" to these participants. Interestingly, in the unfamiliar condition, in which participants were recruited from a section of "Introduction to Psychology" that was *not* taught by the to-be-ignored talker, 15

participants reported that they were familiar with the voice. Given that these individuals may have tried to attach a label to the voice—because it is highly unlikely that they were actually familiar with the correct identity of the talker—we excluded these participants from further data analyses as well. Recall, the uniformed group was given no instructions regarding the identity of the speaker, and it was possible that they may have recognized her voice correctly, as she was their instructor.

Assessing the effect of talker familiarity on the irrelevant speech effect. With the final sample established ($N = 179$), we conducted a 2×3 mixed-model *ANOVA* with the within-participants factor of *auditory condition* (silent, speech) and the between-participants factor of *talker familiarity* (unfamiliar, uniformed, informed). A main effect of auditory condition was found, $F(1, 176) = 428.19$, $MSE = .007$, $p < .01$, $\eta_p^2 = .71$, with higher serial recall performance in the silent condition ($M = .79$) than in the presence of speech ($M = .59$). However, there was no main effect of talker familiarity ($p = .63$), and no interaction of talker familiarity with auditory condition ($p = .41$; see Figure 1).

We then collapsed the two familiarity conditions into one group ($N = 83$) and created a difference score to assess the size of the disruption created by the irrelevant speech. The familiar group ($M_{diff} = .20$) did not demonstrate a significantly larger difference score than the unfamiliar group ($M_{diff} = .19$) [$t(177) = .73$, $p = .46$].

It is possible that the talker—despite being the participants' professor and accurately identified by name via a post-experiment questionnaire—was not actually familiar to either the uniformed or informed “familiar” groups of listeners. An alternative possibility is that the demands of the task, ignoring the auditory information and performing serial recall on visually-presented digits, did not involve any overlap in processing between the irrelevant speech (which

was possibly spoken by a familiar talker) and the required response. Thus, in an attempt to differentiate between these two theoretical explanations of the results of Experiment 1, we conducted Experiment 2 and directly manipulated the familiarity of the talker's voice.

Experiment 2

Given the lack of a significant talker familiarity effect and the little experimental control available regarding listener familiarity in Experiment 1, we conducted Experiment 2 to address these concerns. We familiarized listeners with a specific talker's voice via four days of training that included 30 minutes/day of narrative exposure followed by an interactive voice detection task. We hypothesized that this voice training would provide controlled exposure to and interaction with the target talker's voice and subsequently affect the magnitude of the ISE, thus yielding a greater possibility of talker-familiarity effects that were not found in Experiment 1. By directly manipulating the participants' familiarity with the talker's voice, the likelihood of familiarity leading to an attention capture response (i.e. the strength of the postcategorical feature to disrupt serial recall performance) is increased. In contrast, if the precategorical features remain the dominant cause of disruption, then the familiarity of the talker's voice will not influence the size of the ISE.

Method

Participants. Sixty-three undergraduate students from a large, state university were recruited through the university's Department of Psychology research participant pool to participate in this experiment. We quasi-randomly assigned participants to either undergo or forgo talker familiarity training. The final sample included $n = 29$ talker training condition and $n = 31$ for the control condition ($N = 60$; 53 females, mean age = 23 years). Participant inclusion criteria were as follows: normal or corrected-to-normal vision; self-reported hearing within-

typical-limits; and a native English speaker. We compensated participants in the control group with either course or extra credit; we compensated participants undergoing talker training with \$25.

Experimental design. We employed a 2*2*2 mixed design with *auditory condition* (silent, speech) and *time of testing* (pre, post) as within-subjects factors; *talker training* (training, control) was the between-subjects factor.

Materials. Irrelevant sound task. We used the same materials and stimuli from Experiment 1 for Experiment 2's irrelevant sound task.

Talker training. We used *E-Prime* software 2.0 (2012) to conduct the talker-training portion of the experiment on *Dell* desktop computers equipped with 17-inch monitors and *Sony* circumaural headphones. The setup was located in a sound-treated room.

Talker exposure task's auditory stimuli. A total of 14 published short stories were recorded by the same, female talker who recorded the irrelevant sound task's distractor sentences. We recorded the stories in the same manner as the distractor sentences and used them as stimuli for the talker exposure task. Stories included such pieces as "Eleonora" by Edgar Allen Poe (1850) and "Experience" by Tessa Hadley (2013).

A researcher uploaded the stories' audio recordings to *Dell Optiplex 740* computer equipped with a *Delta 101LT PCI* sound card. They then edited and equated the recordings across total root mean square (r.m.s.) using *Adobe Audition 2.0* (2005) sound editing software. The stories were combined to yield four different, 30-minute blocks of story audio. Each block of story audio included a total of three or four stories and was used as stimuli for the 4-day-long talker exposure task.

Voice detection task. The same female talker who recorded the aforementioned distractor sentences and stories, also recorded 10 additional *Harvard sentences*. An additional 12 native English-speaking females recorded a total of 60 *Harvard sentences*. A researcher uploaded these audio recordings to same *Dell* computer. They again edited and equated across total root mean square (r.m.s.) using the *Adobe Audition 2.0* (2005) sound editing software for these stimuli. The sentences were utilized to yield four different sets of 15 voice detection trials. Each sentence set was used as stimuli for the voice detection task which served as the final component of the 4-day-long training protocol.

Procedure. Experiment 2's procedure consisted of three components: baseline, pre-training ISE task (pre-test); talker training; and post-training ISE task (post-test). The talker training component of this experiment had two phases 1) talker exposure and 2) voice detection task; talker training occurred four times over a series of four consecutive days. The pre-test and post-test components had one phase each (see Figure 2 for a schema).

We administered all components of the experiment in a sound-treated room at a personal computer equipped with headphones; we trained and tested all participants individually. All of the speech stimuli were presented at the listener's maximum comfort level. All components of the experiment were administered during 4 consecutive days. We counterbalanced talker-, digit-, and sentence-presentation across all irrelevant sound tasks and participants.

Irrelevant sound task. We used ISE task procedures identical to those of Experiment 1. A researcher obtained informed consent prior to the experiment. The participants in the control condition completed an ISE task on Day 1 and Day 4 of the four consecutive days (see Figure 2); we gave them no information about the irrelevant speech's talker. The participants in the training condition also completed an ISE task on Day 1 and Day 4 of the four consecutive days (see

Figure 2); we familiarized them with the irrelevant speech's talker over the four consecutive days of training.

Talker training. Participants in the talker training condition underwent talker training on Days 1-4. On Day 1 training occurred after the ISE task; on Day 4 training occurred before the ISE task (see Figure 2). Training began with talker exposure, followed by a voice detection task.

Talker exposure. Talker exposure began with the following message written across the computer's monitor: *Sally will read you 3 stories. Your job is to relax and listen.* The short story's title and its author then appeared printed across the screen and the story's audio began. The participant listened to a total of 30 minutes of stories—read by the talker who produced the irrelevant speech in the ISE tasks—via the computer's headphones. The stories were not repeated across the 4 days of training. The voice detection task began immediately after the participant completed their 30 minutes of talker exposure.

Voice detection task. The voice detection task began with the following instructions written across the computer's monitor: *For this next task, you will hear 3 spoken sentences. Sally may or may not speak one of the sentences. Your job is to indicate whether or not you heard Sally's voice...make your judgement as quickly and accurately as possible.* Then, the voice detection task began. If the participant heard Sally (i.e. the talker who read the stories during talker exposure) speak one of the three sentences, they pressed the letter "Y" on the computer's keyboard. If they participant did not detect Sally's voice, they pressed the letter "N" on the computer's keyboard. The voice detection task consisted of a total of 15 trials; the trials were never duplicated across the 4 days of training.

Results and discussion

Assessing the effect of talker familiarity on the irrelevant speech effect. We scored performance on the serial recall task using a proportion correct method, in which each item correctly recalled was counted independently of whether the entire list was correct or not. Strict serial recall scoring was used, such that the item had to be in the correct order to be counted as accurate. Descriptive statistics are illustrated in Figure 3. We utilized a $2 \times 2 \times 2$ mixed *ANOVA* with the between-subjects factor of participant group resulting in a non-significant comparison, $F(1, 58) = .161, p = .69$, and both of the within-subjects factors resulting in significant main effects: auditory condition, $F(1, 58) = 181.16, MSE = .01, p < .001, \eta_p^2 = .76$, and time of testing, $F(1, 58) = 33.76, MSE = .009, p < .001, \eta_p^2 = .37$. These main effects can be described as an overall detriment to recall in the presence of speech ($M = .69$), as compared to silence ($M = .86$), and as an improvement from pre-test performance ($M = .75$) to post-test performance ($M = .80$). In other words, in concert with the data from Experiment 1, these data demonstrated robust irrelevant sound and practice effects. However, despite the experimental manipulation of talker familiarity via talker training in Experiment 2, there was no effect of talker familiarity.

Talker training: voice detection data. Overall, voice detection performance was high ($M = .88, SD = .08$). The high level of accuracy suggests that participants encoded talker-specific, acoustic characteristics to memory and later used said information to determine whether or not the target talker's voice was present among the three audio recordings presented at test. We also examined the four, different story sets to test for story-specific effects on performance. Across the four days of talker training sessions, three participants each missed one session. We therefore excluded these participants from the repeated-measures *ANOVA* exploring the effect of story during talker exposure ($N = 26$). The use of the four different story sets across the four days of training did not result in any significant differences in voice detection accuracy, $F(3, 75) =$

1.55, $MSE = .01$, $p = .21$, $\eta^2_p = .06$. (See Appendix B for descriptive statistics for the voice detection task associated with each story set.)

Recall that we conducted Experiment 2 because we questioned whether or not the irrelevant talker (i.e. the listeners' professor) was truly "familiar" to the listeners in Experiment 1. Without said familiarity, the precategorical features of the auditory distractors would continue to drive the participants' performance without any influence from the postcategorical features. In Experiment 2 we included a 4-day-long talker training component (i.e. talker familiarization) prior to the ISE task. Training included 30 minutes of exposure to the ISE talker's voice via story narratives followed by a voice detection task including the target voice. Listening to the stories likely involved passive encoding of the talker-specific characteristics of the voice, while the voice detection task involved a more active, deeper level of processing on the listeners' part (Schneider & Shiffrin, 1977). Subsequently, we predicted that the irrelevant talker's voice would be familiar to the trained participants and affect the magnitude of their ISE. This was not the case. The size of the ISE was statistically indistinguishable in pre-testing and in post-testing, before and after voice detection training was completed. Familiarity with the irrelevant talker did not modulate auditory distraction in Experiment 2's ISE task, suggesting that the precategorical features remained the dominant contributor to the distraction effects we observed.

General Discussion

Across two experiments, participants' serial recall was significantly hindered in the presence of irrelevant background speech, compared to serial recall in quiet (ISE; Colle & Welsh, 1976; Elliott & Briganti, 2012). Furthermore, these data showed that a person's familiarity with the talker did not influence the size of the ISE, whether the participants were informed of the talker's identity (Experiment 1) or received four days of talker training that

required them to attend to and interact with talker-specific characteristics of the voice (Experiment 2). These findings are in contrast to our hypotheses and previous speech perception research comparing listeners' perception of words spoken by familiar and unfamiliar voices—familiarity with the talker's voice typically yields a perceptual advantage (e.g., Barker & Newman, 2004; Souza et al., 2013; Yonan & Summers, 2000). However, it is important to note, the majority of these aforementioned studies exploring talker familiarity utilized talkers that the listeners had long-standing, personal relationships with (e.g., a parent or spouse). It is probable that it is the degree of familiarity one has with a talker that enhances their ability to perceive speech.

For example, Newman and Evers (2007) conducted a study employing a perceptual shadowing task in which participants shadowed the speech of one talker while ignoring the speech of a second talker. Similar to our current study, Newman and Evers' familiar talker was the participants' Introductory Psychology instructor. In Experiment 1, the listeners were familiar with the voice of talker to be shadowed and the results showed a talker-familiarity effect (i.e. familiarity with the target talker significantly improved the listeners' shadowing accuracy). In Experiment 2, the listeners were familiar with the to-be-ignored talker's voice and familiarity did *not* affect the listeners' shadowing accuracy—just as the listeners' familiarity with the to-be-ignored talker did not affect the magnitude of the ISE in the present study.

The processing requirements of ignoring a voice while shadowing one of two speech streams is arguably similar to being asked to ignore a voice while remembering visually-presented digits, and task similarities may explain the lack of a talker familiarity effect. The emphasis on the overlap of the processing demands of the task and the to-be-ignored stimuli within an ISE paradigm—and a speech shadowing task—fits with the duplex-model account of

auditory distraction, and specifically, the interference-by-process view of the changing-state effect (Hughes, 2014). The current findings suggest that in these two experiments, the precategorical features of auditory distraction were the dominant cause of the distraction observed.

In this precategorical/postcategorical description of the duplex mechanism account (Marsh et al., 2018), the lack of an effect of talker familiarity is not necessarily problematic for the “functional” view of auditory distraction of Röer et al. (2017). The current stimuli were not emotionally arousing as a result of their semantic context or the talker-specific information. It is possible that if we increased the degree of familiarity our participants had with the to-be-ignored voice even further (e.g., if the talker was the participants’ close friend or family member), the talker-specific information may have affected the participants’ serial recall. Further manipulating the concept of familiarity, and thus strengthening the potency of the postcategorical features in future research, may ultimately give us both new insights into the “functional” view of auditory distraction and how listeners utilize talker familiarity cues during everyday communication in noisy listening environments.

These data also remind us that talker familiarity is a more vulnerable construct than once thought (Drouin, Monto, & Theodore, 2017). Some researchers argue that it’s the role of attention that seems to modulate said talker familiarity effects across a variety of speech perception paradigms. In other words, “when processing time during retrieval is decoupled from encoding factors, it fails to predict the emergence of talker-specificity effects. Rather attention during encoding appears to be the putative variable.” (Theodore, Blumstein, & Luthra, 2015, p. 1674). This is known as the *attention-driven specificity hypothesis* (Maibauer, Markis, Newell, & McLennan, 2014; Theodore et al., 2015). Our data from Experiment 1 support this

perspective. It seems probable that during their Introductory Psychology class, students were not attending to the instructor's identity and voice characteristics (e.g., gender, shimmer, quality, accent, etc.), rather their attentional resources were allocated to learning and understanding class material. Because the students were not explicitly attending to (and encoding) the talker characteristics of their instructor's voice during class, when they participated in Experiment 1—even those participants in the informed familiarity condition—they had few detailed, talker-specific exemplars available for recall and subsequent interference (Church & Schacter, 1994). Therefore, familiarity with the irrelevant speech's talker had no significant effect on the magnitude of the ISE. Our data from Experiment 2 counter this argument. In the future, if we employed a similar design and ISE methodology, but trained the student participants to document (i.e. actively attend to) specific vocal characteristics of their instructor during each class and prior to the experiment, we may gain further insight into whether or not attention allocation is driving talker-specific effects.

In summary, our study replicated the robust data already supporting the ISE but failed to show a talker-specific effect of familiarity in the context of ISE. These data suggest that talker familiarity is a more fragile construct than the speech perception literature originally suggested (Drouin et al., 2017), while also supporting the interference-by-process view of the changing-state effect (Hughes, 2014). Further research is needed to help resolve when talker-specific information is perceived and utilized by the listener. This knowledge will be key to confirming the theoretical explanations of the irrelevant speech effect and working toward a biologically plausible theory of speech perception that includes *both* linguistic and indexical information.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.
- Adobe Systems. (2005). Adobe Audition CC (Version 2) [Computer software]. San Jose, CA.
- Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In E. Meyer & S. Kornblum (Eds.), *Attention and performance XVI: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 182-218). Cambridge, MA: M.I.T. Press.
- Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, 94, B45-B53. doi: 10.1016/j.cognition.2004.06.001
- Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 5, 1106-1118. doi: 10.1037/0278-7393.30.5.1106
- Chandrasekaran, B., Chan, A. H. D., & Wong, P. C. M. (2011). Neural processing of what and who information in speech. *Journal of Cognitive Neuroscience*, 23, 2690–2700. doi: 10.1162/jocn.2011.21631
- Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20, 521-533.
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, 15, 17-31.
- Drouin, J. R., Monto, N. R., & Theodore, R. M. (2017). Talker-specificity effects in spoken language processing: Now you see them, now you don't. In A. Lahiri & S. Kotzer (Eds.),

- The speech processing lexicon: Neuro cognitive and behavioural approaches* (pp. 107-128). Berlin/Boston: Walter de Gruyter GmbH.
- Ellermeier, W., & Zimmer, K. (2014). The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology*, 35, 10-16. doi: 10.1250/ast.35.10
- Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, 140, 64-74. doi: 10.1016/j.actpsy.2012.02.009.
- Hadley, T. (2013, January 21). Experience. *The New Yorker*. NY, NY: Condé Nast.
- Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, 3, 30-41. doi: 10.1002/pchj.44
- Hughes, R. W., & Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise and Health*, 13, 55-74.
- Hughes, R. W., & Jones, D. M. (2003). Indispensable benefits and unavoidable costs of unattended sound for cognitive functioning. *Noise and Health*, 6, 63-76.
- Hughes, R. W., Marsh, J. E., & Jones, D. M. (2009). Perceptual-gestural (mis)mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35, 1411-1425. doi: 10.1037/a0017008
- Hughes, R. W., Marsh, J. E., & Jones, D. M. (2011). Role of serial order in the impact of talker variability on short-term memory: Testing a perceptual organization-based account. *Memory, & Cognition*, 39, 1435-1447. doi: 10.3758/s13421-011-0116-x
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory

- distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1050-1061. doi: 10.1037/0278-7393.33.6.1050
- IEEE. (1969). IEEE recommended practices for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17, 225-246.
- Jones, D. M., Madden, C. & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology Section A*, 44, 645-669. doi: 10.1080/14640749208401304
- 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. doi: 10.1016/j.jml.2017.05.005
- Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, 21, 125-139.
- Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117-123.
- Maibauer, A. M., Markis, T. A., Newell, J., & McLennan, C. T. (2014). Famous talker effects in spoken word recognition. *Attention, Perception, & Psychophysics*, 76, 11-18. doi: 10.3758/s13414-013-0600-4
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 306-321. doi: 10.1037/0278-7393.31.2.306
- Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J.-i., & Shozo, K. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *Neuroimage*, 59, 1735-1744. doi: 10.1016/j.neuroimage.2011.07.093

- Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, 18, 347-373.
- Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, 35, 85-103. doi: 10.1016/j.wocn.2005.10.004
- Neumann, O. (1996). Theories of attention. In O. Neumann & W. Prinz (Eds.), *Handbook of perception and action Vol. 3: Attention* (pp. 389-446). San Diego: Academic Press.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60, 355-3769.
- Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego, CA: Academic Press.
- Poe, E. A. (1842). Eleonora: A fable. *The gift: A Christmas and New Year's present for 1842* (pp. 17-322). Philadelphia, PA: Carey and Hart.
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from <http://www.pstnet.com>.
- Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017). Attentional capture by taboo words: A functional view of auditory distraction. *Emotion*, 17, 740-750. doi: 10.1037/emo0000274
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66. doi: 10.1037/0033-295X.84.1.1
- Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, 24, 689-700. doi: 10.3766/jaaa.24.8.6
- Sörqvist, P. & Rönnerberg, J. (2014). Individual differences in distractibility: An update and a model. *PsyCh Journal*, 3, 42-57. doi: 10.1002/pchj.47

Theodore, R. M., Blumstein, S. E., & Luthra, S. (2015). Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Attention, Perception, & Psychophysics*, 77, 1674-1684. doi: 10.3758/s13414-015-0854-0

Tremblay, S., & Jones, D. M. (1999). Change of intensity fails to produce an irrelevant sound effect: Implications for the representation of unattended sound. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1005-1015. doi: 10.1037//0096-1523.25.4.1005

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology & Aging*, 15, 88-99. doi: 10.1037/0882-7974.15.1.88

Figure 1. Mean proportion correct serial recall performance of the talker familiarity groups, by auditory condition. Error bars represent standard error of the mean. White bars represent trials in silence and grey bars represent trials presented in the presence of irrelevant speech.

Figure 2. Schema of three-component procedure for Experiment 2.

Figure 3. Mean proportion of correct serial recall for both groups of listeners (talker training, control) across time of testing, with standard error of the mean represented by the error bars.

White bars represent trials in silence and grey bars represent trials presented in the presence of irrelevant speech. Dotted bars on the left, represent listeners who underwent talker familiarity training.

Appendix A:**Follow-up questions used to access participants' familiarity with the spoken voice**

1. Was the voice speaking the sentences to you familiar?

A) Yes

B) No

2. If the voice was familiar, attempt to identify the voice by selecting the letter next to the person's last name.

A) Brown

B) Knapp

C) Munson

D) Toussaint

E) Vigna

F) The voice is not familiar to me

3. If you are taking PSYC 2000, how often a week do you attend class?

A) Once a week

B) Two times a week

C) Three times a week

Appendix B:**Descriptive Statistics for Experiment 2's Talker Training Voice Detection Data****from Each Story Session**

Training Session	n	M	SD	Min	Max
A	26	.85	.17	.27	1.00
B	27	.90	.09	.67	1.00
C	28	.90	.12	.47	1.00
D	28	.87	.15	.47	1.00