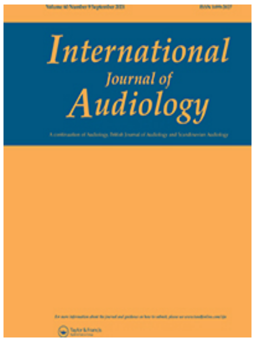

International Journal of Audiology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iija20

Spoken narrative comprehension for young adult listeners: effects of competing voices and noise

Peter A. Wasiuk, Gabriel A. Radvansky, Robert L. Greene & Lauren Calandruccio

To cite this article: Peter A. Wasiuk, Gabriel A. Radvansky, Robert L. Greene & Lauren Calandruccio (2021) Spoken narrative comprehension for young adult listeners: effects of competing voices and noise, International Journal of Audiology, 60:9, 711-722, DOI: 10.1080/14992027.2021.1878397

To link to this article: https://doi.org/10.1080/14992027.2021.1878397

INTERNATIONAL JOURNAL OF AUDIOLOGY
2021, VOL. 60, NO. 9, 711–722
https://doi.org/10.1080/14992027.2021.1878397

ORIGINAL ARTICLE

Spoken narrative comprehension for young adult listeners: effects of competing voices and noise

Peter A. Wasiuk[a], Gabriel A. Radvansky[b], Robert L. Greene[a] and Lauren Calandruccio[a]

[a]Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA; [b]Department of Psychology, University of Notre Dame, Notre Dame, IN, USA

ABSTRACT

Objective: To examine the influence of competing voices or noise on the comprehension of spoken narratives for young adults.
Design: First, an intelligibility assessment of the target narratives was conducted to establish a signal-to-noise ratio ensuring accurate initial speech recognition. Then, narrative comprehension for two target types (*fixed* and *varied target talker*) was measured in four listening conditions (*quiet, one-talker speech, speech babble, speech-shaped noise*). After hearing target narratives in each listening condition, participants completed a visual recognition memory task that assessed the comprehension of the narrative materials at three levels of representation (*surface form, propositional, event model*).
Study Sample: Seventy adults (18–32 years of age).
Results: Narrative comprehension results revealed a main effect of listening condition at the event model level, indicating poorer narrative memory of described situations for all noise conditions compared to quiet. Increased positive responses to thematically consistent but situationally "wrong" memory probes drove this effect. No other significant effects were observed.
Conclusion: Despite near-perfect speech recognition, background noise negatively influenced aspects of spoken narrative comprehension and memory. Specifically, noise did not disrupt memory for what was said (surface form and propositional memory), but only memory for what was talked about (event model memory).

ARTICLE HISTORY
Received 6 July 2020
Revised 7 January 2021
Accepted 13 January 2021

KEYWORDS
Spoken narrative comprehension; memory; background noise

Introduction

Background noise and other forms of acoustic signal degradation have a negative impact on the way that listeners recognise target speech, especially as signal-to-noise ratios (SNR) become less favourable (Kryter 1962; Mattys et al. 2012; Moore 2013). However, there may be more inconspicuous effects of acoustic challenge on spoken language processing, comprehension, and memory that go beyond the level of speech recognition per se (for a review, see Peelle 2018). Spoken communication in everyday life is marked by a variety of acoustic challenges that reduce the clarity of the speech signal available to the listener. Acoustic challenges can result from a variety of sources that degrade the clarity of the target speech signal, thereby reducing the amount of speech information available to the listener and requiring greater engagement of cognitive resources and/or listening effort for the listener. Speech signal degradations can be source-related (e.g. underarticulated conversational speech, foreign-accented speech, etc.), transmission-related (e.g. background noise, reverberation, bandwidth limitations, etc.), and/or receiver-related (e.g. hearing impairment, supra-threshold auditory processing difficulties, etc.) (Mattys et al. 2012). Successful spoken language comprehension is contingent upon the ability to rapidly integrate the incoming acoustic signal with information from stored representations in memory, even in cases when the speech signal is sparsely represented (for example, in situations of acoustic challenge). While accurate speech recognition is important for eliciting successful communication, empirical investigations of the downstream effects of acoustic challenge on memory encoding and retrieval, as well as language comprehension, have provoked serious questions about how people listen in the complex environments of the real world.

The presence of competing background noise during syllable (Surprenant 1999), word (Heinrich, Schneider, and Craik 2008; Kjellberg, Ljung, and Hallman 2008; Ljung, Israelsson, and Hygge 2013; Murphy et al. 2000; Pichora-Fuller, Schneider, and Daneman 1995; Rabbit 1968), and sentence (Koeritzer et al. 2018) processing impairs subsequent memory. The observation that challenges from acoustic signal degradation at early stages of speech processing leads to poorer episodic memory has also been shown in listeners with hearing loss (i.e. an intrinsic source of signal degradation). For example, McCoy et al. (2005) compared age-matched adults with mild-to-moderate sensorineural hearing loss with people having borderline normal hearing sensitivity (based on three-frequency pure-tone average (PTA); mean PTA of "better hearing" group = 21.4 dB, mean PTA of "hearing loss" group = 35.7 dB) on a memory task for running speech. In this paradigm, people listened to a running list of unrelated words that could stop at any moment. At that point, people were to repeat back the last three words they heard. Both groups did equally well for the most recent word (the final word heard before stoppage), indicating accurate speech recognition regardless of hearing status. However, the group with hearing loss had

CONTACT Peter A. Wasiuk peter.wasiuk@case.edu Department of Psychological Sciences, Case Western Reserve University, 11635 Euclid Avenue, Cleveland, OH 44106, USA

poorer recall than the group with better hearing for the middle and first word of the three-word sets. These results suggested that while reduced signal clarity (as a result of hearing loss, background noise, or both) did not produce errors in initial speech recognition, it did negatively influence subsequent memory for the speech materials (McCoy et al. 2005; Murphy et al. 2000; Rabbit 1968, 1991).

In everyday life, spoken communication typically consists of connected speech, rather than isolated sets of unrelated words or sentences. The comprehension of speech not only relies upon the successful recognition and understanding of the individual words and sentences, but also on the integration of meaning of these elements with general world knowledge to derive a larger understanding of what is being referred to as a whole (Kintsch 1988; van Dijk and Kintsch 1983). A number of studies have examined how varieties of acoustic challenge influence memory and comprehension of connected spoken discourse (Griffin, Poissant, and Freyman 2020; Piquado et al. 2012; Schneider et al. 2000; Sörqvist and Rönnberg 2012; Tye-Murray et al. 2008).

For instance, Ward et al. (2016) presented younger and older adults with short narratives (60–80 words each) degraded by two levels of noise vocoding (24 and 16 channel vocoding), or in an unaltered form (i.e. natural speech). Noise vocoding is a digital signal processing technique that alters the spectral detail of the speech signal (Shannon et al. 1995). The levels of vocoding reduced the fidelity of the speech signal without decreasing initial speech recognition for the listener (i.e. high levels of intelligibility across stimulus conditions). An intelligibility assessment revealed that all participants correctly recognised 100% of the speech materials in the natural and the two noise-vocoded conditions. After presenting the short stories to the listeners, free recall was used to assess any differences in narrative memory. Results revealed that, overall, older adults recalled less than the younger adults (Tye-Murray et al. 2008). Moreover, in both age groups, recall accuracy dropped in both noise vocoded speech conditions at the first level of narrative detail (which was scored based on participants recalling the subject and verb of a sentence) compared to natural speech, indicating that acoustic degradation negatively influenced narrative memory for the most prominent (i.e. the main) story ideas (Ward et al. 2016). This effect of signal clarity on recall of the main idea units of linguistic stimuli has also been shown in the visual domain. Specifically, recall performance of main ideas is negatively influenced by increased visual noise levels during reading (Gao, Levinthal, and Stine-Morrow 2012; Gao et al. 2011). Ward et al. (2016) also found that working memory capacity (WMC), as measured by a reading span task (Daneman and Carpenter 1980; McCabe et al. 2010), predicted narrative recall in the degraded listening conditions for both age groups. This aligns well with the Ease of Language Understanding (ELU) model (Rönnberg et al. 2013), and lends credence to the prediction that people with greater WMC are more adept at both recognising and remembering speech heard in challenging acoustic conditions.

In a similar study, Sörqvist and Rönnberg (2012) examined the comprehension of target narratives presented in distinct listening conditions meant to represent both *energetic* and *informational* masking. Energetic masking occurs as a result of spectrotemporal overlap of excitation patterns in the peripheral neural structures of the auditory system (i.e. the cochlea and auditory nerve), causing target speech information to be "covered up" by the masker, and degrading target representation (Yost 2013). Reductions in target speech recognition can also occur as a result of confusion or uncertainty in identifying and segregating target speech from masker competition, and/or if there is perceptual similarity between the target and masker. This type of interference is referred to as informational masking (Durlach et al. 2003; Kidd and Colburn 2017), and is often thought of as any masking that is "nonenergetic" in nature (Durlach et al. 2003; Watson 2005).

Sörqvist and Rönnberg (2012) presented target narratives to listeners with either an intelligible speech masker (a simultaneously presented narrative) or a spectrally rotated speech masker (the masker narrative spectrally rotated to be made unintelligible, but to approximate the same energetic characteristics as the natural speech), each at an SNR of 5 dB. Comprehension was indexed by having participants answer questions about the content of the target narratives after listening in each condition. Results indicated poorer comprehension accuracy after listening to target narratives in the intelligible speech masker compared to the spectrally rotated one. Sörqvist & Rönnberg interpreted this as an indication that informational masking may have an especially negative impact on speech comprehension over and above the impact of energetic masking (when energetic masking characteristics are similar between conditions). They also found that verbal WMC was a significant predictor of performance on the comprehension task (Sörqvist and Rönnberg 2012).

Modern spoken language processing frameworks support the idea of a meaningful relationship between acoustic challenge and the ease of comprehending and encoding speech material (e.g. effortfulness hypothesis: McCoy et al. 2005; ELU: Rönnberg et al. 2013; Framework for Understanding Effortful Listening: Pichora-Fuller et al. 2016). One foundational point of these theories is that there is a limited capacity of domain-general information processing resources that are engaged during challenging listening scenarios (assuming the listener is motivated to hear the target speech). When these resources are diverted towards the process of recognising target speech due to acoustic challenge (e.g. conditions with background noise and/or when the listener has hearing impairment) there are fewer resources remaining for speech comprehension and memory encoding. Therefore, target signal degradation likely causes people to allocate resources towards speech decoding and lexical access, taking resources away from memory and comprehension processes such as inference making, rehearsal, and elaborative association (Gao, Levinthal, and Stine-Morrow 2012).

Current experiment

The aim of the current experiment was to examine the potential effects that distinct forms of acoustic challenge (i.e. one-talker speech, speech babble, and speech-shaped noise) have on spoken narrative comprehension. The listening conditions were chosen to systematically explore the influence of a traditional *energetic* masking condition (i.e. speech-shaped noise) and acoustic challenge caused by competing talkers that often provokes *informational* masking (i.e. one-talker speech and speech babble) on comprehension. A principal objective was to present target narratives at an SNR that would ensure successful initial speech recognition, as well as be representative of the complex environments of everyday listening. By doing this, we could assess the influence of listening condition on narrative comprehension, rather than on speech recognition. Positive SNRs also provide greater ecological validity in simulating real-world listening scenarios in the noisy communication environments of everyday life (Smeds, Wolters, and Rung 2015; Weisser and Buchholz 2019).

Narratives that have been used successfully in the visual domain to study comprehension and memory (Fisher and Radvansky 2018; Radvansky et al. 2001; Zwaan and Radvansky 1998) were recorded and presented to listeners in four acoustic conditions. Listeners either heard target narratives spoken by the same talker across conditions (*fixed target talker*) or by four unique talkers (*varied target talker*). Memory and comprehension for the narratives was probed using a recognition memory task that evaluates three levels of discourse representation, namely the surface form, propositional, and event model levels. This was done using a well-established procedure in cognitive psychology (i.e. the Schmalhofer and Glavanov 1986 procedure). Based on frameworks for understanding the relationship between acoustic challenge, comprehension, and memory (McCoy et al. 2005; Pichora-Fuller et al. 2016; Rönnberg et al. 2013), we hypothesised that people would do more poorly after listening to narratives in the more challenging acoustic conditions (i.e. one-talker speech, speech babble, and speech-shaped noise) than in quiet, even with initial speech recognition being equivalent (and ensured) across the conditions. Furthermore, we predicted that the two speech masker conditions would degrade performance on the comprehension task to a greater extent than the speech-shaped noise masker (comprehension from best to worst: quiet > speech-shaped noise > speech babble > one-talker speech). This prediction is based on the idea that greater engagement of cognitive resources is required to maintain attention and segregation of the target speech in the presence of speech maskers, which in turn would negatively influence memory encoding and the robustness of narrative representations.

It is important to note that a substantial body of evidence has demonstrated that speech recognition is enhanced when words are presented in meaningful contexts, especially in the presence of background competition (Kalikow, Stevens, and Elliott 1977; Pichora-Fuller 2008; Pichora-Fuller, Schneider, and Daneman 1995; Winn 2016). That is, speech materials that have constraining contexts (e.g. meaningful sentences) improve speech recognition accuracy at the word and sentence levels, especially compared to those that are presented in isolation or in low-constraining contexts (Dubno, Ahlstrom, and Horwitz 2000; Kalikow, Stevens, and Elliott 1977; Sommers and Danielson 1999). Meaningful contexts likely reduce the listening effort and/or release the information processing resources that are required to recognise target speech at lower levels (Winn 2016). Therefore, it was also of interest whether the rich, coherent narrative speech materials used in this study would promote initial speech recognition in the challenging acoustic conditions enough to reduce/offset the increased engagement of resources required of the listener, thereby freeing up resources for comprehension and memory processes.

Levels of narrative representation

In modern theories of discourse comprehension, narrative memory can be divided into three levels: (a) surface form, (b) propositional, and (c) event model. The *surface form* is a verbatim memory of the exact words and syntax that were used. This type of memory is typically very short lasting and is often lost within a few minutes after initial encoding (Fisher and Radvansky 2018; Sachs 1967). The *propositional* level[1] is the memory for the idea units that were present in the discourse apart from the exact wording. At the propositional level, a paraphrased sentence conveying the same meaning as an actual utterance from a discourse would both map onto the same mental representation. Finally, an *event model* is a referential representation of the described situation, not of the language itself. It is a mental simulation of what is being described. Event models are created using information in the language itself along with inferences that listeners draw based on their world knowledge.

Thus, in other words, the surface form and the propositional levels capture what was said, and the event model captures what was talked about. The event model is the longest lasting in memory and can be thought of as being the most important for comprehension and knowledge acquisition (Kintsch 1988; Radvansky and Zacks 2014; Schmalhofer and Glavanov 1986; van Dijk and Kintsch 1983; Zwaan 1999; Zwaan and Radvansky 1998).

A common method used to separate out and examine these three levels of memory representation is the Schmalhofer and Glavanov (1986) procedure. This procedure, while often used in written narrative memory studies (Bohay et al. 2011; Fisher and Radvansky 2018; Fletcher and Chrysler 1990; Kintsch et al. 1990; Narvaez et al. 2011; Radvansky, Copeland, and von Hippel 2010; Radvansky, Copeland, and Zwaan 2003; Zwaan 1994), is novel in terms of its use in auditory science. The details of this approach are provided in the method section.

Materials and method

Spoken narrative materials

All stimuli for the intelligibility assessment and the *fixed target talker* narrative comprehension condition were recorded by two females who were native speakers of English (American dialect; referred to as Talkers A and B). The texts used for the target and masker speech are from Radvansky et al. (2001) and Fisher and Radvansky (2018). Four different narratives were recorded by talkers A and B (Talker A recorded the target narratives, while Talker B recorded the competing masker narratives). The four target narratives ranged in length from 614 to 681 words (M *length* = 650 words, $SD = 28$ words). The narrative recorded to be used as the one-talker competing speech masker was 680 words long. Both talkers used a natural speaking rate (M *speaking rate* = 4.7 syllables/second, $SD = 0.3$ syllables/second). Stimuli were recorded in a double-walled sound-isolated booth using 16-bit resolution and a 44.1 kHz sampling rate. Talkers stood six inches in front of a Shure KSM-42 omnidirectional cardioid condenser microphone with a pop-filter attached. The microphone was connected to an M-Audio M-Track 2 × 2 converter. Recorded narratives were spliced into individual WAV files and edited to minimise silences at the beginning and end of each sentence. The sentences were then root-mean-square (RMS) equalised to the same sound pressure level using Praat software (Boersma and Weenink 2017). Sentences were then concatenated with 200 ms of silence inserted between each pair to ensure equal pause duration throughout the narrative. All other pauses within the running speech were <200 ms in duration. The speaking rate and pause durations of our narrative materials sounded quite natural and were equated with great care across talkers and conditions (stimuli can be accessed via our OSF page; see Data Availability Statement). However, it should be noted that longer pauses may influence a listener's ability to recall information from passages (Wingfield et al. 1999), and including passages with different pause durations (Heldner and Edlund 2010; Winn 2016) should be considered in future work.

For the *varied target talker* narrative comprehension condition, target talker identity (i.e. the female telling the narrative) was varied randomly across the four narratives and four listening conditions for each participant. In everyday life, target talker identity changes across time as a result of conversational turn taking. Thus, this manipulation enhanced the ecological validity of the paradigm as well as increased the cognitive demand of the task by adding multiple talker-specific vocal characteristics. Original Talker A was one of the four target narrative voices. Three additional female talkers produced the three remaining narratives (the narratives were the same as the fixed target talker condition). All three females were native speakers of English (American dialect). All four talkers used a natural speaking rate that was nearly identical to the speaking rates of the fixed target talker recordings (M *speaking rate* = 4.9 syllables/second, $SD = 0.2$ syllables/second). All other procedures in producing the varied target talker narratives were identical to those used in the fixed target talker condition.

The two competing speech masker conditions included a one-talker stream and a speech babble. The speech babble masker was produced by overlaying four different masker narratives spoken by Talker B into a single audio file (subsequently producing a speech babble noise in which none of the speech streams are fully intelligible).[2] The steady-state, speech-shaped noise masker was generated by passing white noise through a filter shaped to match the long-term average spectrum of Talker B's narrative productions. The speech-shaped noise masker condition was used to test an energetic masker condition that was matched to the long-term frequency characteristics of the masker talker (but see Stone, Fullgrabe, and Moore 2012). This was done to evaluate the potential differences energetic and informational background competition may produce on the comprehension and memory of target discourse. Target narratives were intended to be fully intelligible to the listeners in every condition to allow us to assess the effect of acoustic challenge on narrative comprehension, rather than on initial speech recognition. To evaluate the intelligibility of the four target narratives in each masking condition (i.e. *one-talker speech, speech babble, speech-shaped noise*), an intelligibility assessment was conducted prior to the comprehension experiment to establish the appropriate SNR for stimulus presentation that would ensure accurate speech recognition. Finding an SNR that would ensure ceiling level speech intelligibility performance was crucial to this experimental design, in that we aimed to examine the influence of listening condition on narrative comprehension, rather than examine the influence of listening condition on speech recognition performance (which is far more common in the hearing science literature). However, it is important to note that even under conditions of equivalent speech intelligibility, there may be differences in listening effort and/or cognitive resource engagement that are occurring as a result of different acoustic backgrounds (Peelle 2018).

Materials for intelligibility assessment

The speech materials used in the intelligibility assessment[3] consisted of short segments of speech extracted from the target narratives (range of 2–11 words each; $M = 6$ words, $SD = 2$ words), to examine the intelligibility of these random windows of target speech at a range of SNRs (−3, −1, 1, 3, and 5 dB) in the three experimental noise conditions (*one-talker speech, speech babble, speech-shaped noise*). Random segments of speech were extracted from the final target narrative WAV files to minimise the contextual cues available to the listener, to allow local sound pressure levels to vary in the same manner that they would during the narrative comprehension experiment (i.e. fluctuating target speech sound pressure level associated with natural connected speech), and to ensure that segment lengths would not strain short-term memory. Segments were extracted via visual examination and listening cheques of the final target narrative WAV files to produce short WAV files that could be presented to listeners in a masked-speech recognition task. Segments extracted from the target narratives were coded with scoring words by multiple trained research assistants. The words chosen for scoring ensured that they lacked potential ambiguity resulting from coarticulation.

Participants for intelligibility assessment

Fifteen adult listeners ranging in age from 18 to 26 (10 female; M age = 20 years, $SD = 3$ years) participated in an intelligibility assessment to evaluate speech recognition accuracy for the target narrative stimuli in the three noise conditions. Participants were recruited from the Case Western Reserve University (CWRU) Department of Psychological Sciences research participant pool and received course credit for their involvement. All recruitment and testing methods were approved by the CWRU Institutional Review Board (IRB). Participants first signed an informed consent document approved by the CWRU IRB, and then completed a demographic questionnaire to confirm that they were native speakers of English (American dialect). Prior to testing, all participants had their hearing screened at all octave frequencies at and between 250 and 8 kHz bilaterally at 15 dB HL to ensure hearing sensitivity within the normal range. Each young adult participant passed this hearing screening.

Experimental procedures for intelligibility assessment

All participants in the intelligibility assessment listened to target narrative segments in three noise conditions (*one-talker speech, speech babble, speech-shaped noise*). Target speech segments were presented in each of these noise conditions at five different SNRs (−3, −1, 1, 3, 5 dB SNR) to examine proportion correct data for target speech recognition using our novel narrative stimuli in a masked-speech recognition task. A total of 15 listening conditions were presented to each participant (3 noise conditions x 5 SNRs), with listening condition blocked by SNR. The order of SNR block presentation was randomised across listeners, and the presentation of noise conditions was also randomised within each SNR block. Fifty-one segments of target speech were presented within each listening condition. Unique lists of target speech segments were produced to be randomly presented across noise conditions, with lists consisting of 95–110 scoring words each ($M = 98$ words, $SD = 5$ words). Overall, each listener heard a total of 765 target speech segments, and a total of 1468 scoreable words. Segments were extracted equally from each of the four target narratives (∼191 segments extracted from each narrative).

Stimuli were presented to listeners bilaterally over Sennheiser HDA 280 headphones at an overall presentation level of 70 dB SPL in a sound-treated room. Participants were told that they would be listening to a female target talker, and that the target talker would not change throughout the experiment. Participants were also told that other female talker(s) or noise would be present at the same time, and their task was to try and ignore the competing speech or noise and repeat back only what the target talker said. Participants were encouraged to guess after trials in which they were uncertain. Responses were spoken aloud and scored on-the-fly by an examiner seated across from the listener. The examiner had a clear view of the participant's face and heard the responses clearly, giving both visual and auditory cues for accurate scoring. Any variations of the keyword (including pluralisation and tense changes) were marked as incorrect. The examiner was naïve to the experimental hypotheses. Total testing time of the speech intelligibility assessment was approximately 45 min.

Participants for the narrative comprehension experiment

Fifty-five adults ranging in age from 18 to 32 (27 female; *M* age = 21 years, *SD* = 3 years) participated in the narrative comprehension experiment. The first 31 participants were assigned to the *fixed target talker* condition (age range 18–28; 14 female; *M* age = 19 years, *SD* = 2 years), while the next 24 participants were assigned to the *varied target talker* condition (age range 18–32; 13 female; *M* age = 22 years, *SD* = 4 years).[4] None of these adults had taken part in the intelligibility assessment. Participant recruitment, compensation, informed consent process, demographic questionnaire, and hearing screening procedures were identical to those used in the intelligibility assessment. All participants were native speakers of English (American dialect) and passed the hearing screening.

Experimental procedures for the narrative comprehension experiment

Working memory capacity (WMC) task

Prior to the narrative comprehension task, all participants completed a reading span task (Daneman and Carpenter 1980) using methods described in Klaus and Schriefers (2016). This span task has been shown to be an effective measure of complex verbal WMC, in that it requires an individual to process ongoing verbal information while simultaneously holding verbal material in memory. The task consists of two parts: a sentence processing component and a memory component. In the sentence processing component, the participant is required to evaluate the sensibility of a series of sentences. That is, in each experimental trial, a sentence (*M word length* = 12 words, *SD* = 2 words; 60 sentences total) is presented at the centre of a computer screen, and the task is to decide whether it is semantically sensible (via selecting "Y", plausible, or "N", implausible). An example of an implausible sentence is, "Every now and then I catch myself swimming blankly at the wall", whereas an example of a plausible sentence is, "After final exams are over, we'll be able to take a well-deserved break". For the memory component, the task is to hold a series of words in memory that are presented in between the sentences of the sentence processing component. After each sentence, a to-be-remembered noun is presented on the screen for 1200 ms, which is to be held in memory until the end of the sentence/word set. The nouns in each set are not phonologically, semantically, or associatively related to the sentences. At the end of each sentence/word set, the participant has to verbally report as many of the nouns from that set as possible to the experimenter (regardless of serial order). Randomised orders of two to six sentence/word combination sets are presented, and each set size is tested three times overall, resulting in 60 total trials. A participant's reading span score is the number of nouns accurately recalled.

Narrative comprehension task

Listening Phase. All participants listened to four target narratives in four acoustic conditions (i.e. *quiet, one-talker speech, speech babble, speech-shaped noise*). Participants listened to the spoken narratives in a double-walled sound isolated booth while seated in front of a computer screen and keyboard. Materials were presented bilaterally over Sennheiser HDA 280 headphones. Narratives were blocked by condition and the order of narratives was randomised across participants, as was the combination of target narrative and listening condition. Each narrative was presented at a fixed level of 70 dB SPL. A 5 dB SNR was used for all conditions that included acoustic competition (per the intelligibility assessment, please see Results section). This SNR provided near 100% target speech intelligibility for our young participants with normal hearing, enabling the examination of the potential effect of listening condition on narrative comprehension and memory, while ensuring accurate initial speech recognition.

Participants in the *fixed target talker* group were told that they would listen to a female talker tell a story, and that the title of this story would be displayed on the screen. They were told that the aim of the study was to remember the specific details of the story. Specifically, they would be asked whether they heard specific sentences from the story after it was over. They were also told to try to remember the exact wording of the sentences, as this would be important during the subsequent memory task. What would also make this task difficult is that there may be noise or other concurrent talkers speaking in the background while the main female talker was telling the story. The main female talker's voice would not change throughout the task, and would be the same for all four stories. Participants were told to try and ignore the competing speech or noise and listen to only what the main talker said.

For participants in the *varied target talker* group, instructions were moderately revised to indicate that the female talker's voice would change for each story, but that the main female talker would always be louder than the background noise and/or background voices (5 dB SNR), would start after the noise or background voices (1000 ms), and would always align with the title of the story displayed on the screen. The benefit of this experimental manipulation is that it allows us to evaluate the effect of talker variability on the narrative comprehension task. Negative effects of talker uncertainty have been shown in both speech recognition and memory tasks for isolated vowels, words, and sentences (Bradlow, Nygaard, and Pisoni 1999; Martin et al. 1989; Morton, Sommers, and Lulich 2015), and therefore, we hypothesised that this manipulation might negatively influence narrative comprehension. While the adaptation to a new target talker in varied target talker conditions appears to occur quite rapidly (even in response to a single isolated vowel; Morton, Sommers, and Lulich 2015), we were interested in whether this effect would hold true at the level of narrative stimuli, and/or if the different target talkers would require different degrees of listening effort or cognitive resource allocation that may influence narrative comprehension. Unique talkers vary significantly in their overall intelligibility (Bradlow, Torretta, and Pisoni 1996), and therefore may require different degrees of information processing resource allocation in initial speech processing. Thus, the *varied target talker* manipulation was used to explore these potential effects on narrative comprehension. One potential limitation of this was that only one of the four talkers was used in the *fixed target talker* condition. In future work, it will be beneficial to use each of these talkers in four separate *fixed target talker* conditions to examine the potential differences between the talkers on speech comprehension in a more direct manner. All other experimental procedures and instructions for the *fixed* and *varied target talker* groups were identical.

Memory Recognition Phase. Following the listening phase (all four narratives), participants completed a recognition memory task to probe target narrative memory using the Schmalhofer and Glavanov procedure (Fisher and Radvansky 2018; Schmalhofer and Glavanov 1986). The recognition memory task was presented visually on a computer screen to avoid confounds related to encoding specificity (i.e. to ensure people were not making memory judgements based on surface level attributes present at the time of encoding) (Tulving and Thomson 1973). At the beginning of each recognition phase, participants were given the title of the narrative to indicate which was being tested (the themes of the four narratives were unique). The order of the four recognition blocks was the same as the listening phase presentation order to equate retention intervals. Participants were told to indicate yes (via the "Y" key) or no (via the "N" key) regarding whether each probe sentence was heard earlier. They were told to only indicate yes if they were certain that the sentence had been said, word-for-word, in the original narrative, as there would be sentences presented that had not been said. Sixty-four probes were presented for each narrative (in a randomised order for each participant), with 256 probes presented in total.

Narrative comprehension analysis

Recognition memory was scored using the Schmalhofer and Glavanov procedure to index the surface form, propositional, and event model levels. This involves using signal detection analysis (*d'* sensitivity index: Hautus 1995; Stanislaw and Todorov 1999). Specifically, the 64 recognition sentences from each narrative were of four types: (a) a *verbatim* sentence that actually appeared in the narrative, (b) a *paraphrase* of the verbatim sentence, which was not heard but conveyed the same propositional idea units using different wording, (c) an *inference* sentence that conveyed an inference idea that was likely generated by the listener using their world knowledge, and (d) a *wrong* sentence that was thematically consistent with the narrative, but was inconsistent with the described events. An example *verbatim* sentence is, "The plot bitterly intensified government suspicions of farmers". An example *paraphrase* of that sentence is, "The plot greatly heightened government distrust of farmers". An *inference* example is, "The plot led to increased laws against farmers". While an example *wrong* sentence is, "After the plot, donations to the city government rose dramatically".

The surface form level was measured by treating "yes" responses to verbatim probes as hits and "yes" responses to paraphrases as false alarms. Both of these refer to idea units that were in the narrative, but only the verbatim probes were actually heard. The propositional level was measured by using "yes" to paraphrases as hits and "yes" responses to inferences as false alarms. Neither of these sentence types were actually heard, but only the paraphrases convey idea units actually present in the story. Finally, the event model level was measured by using "yes" responses for inference probes as hits and "yes" responses to wrong probes as false alarms. Neither of these probe types conveyed ideas that were actually present in the narratives, but the inferences were consistent with the described situations, while the wrong probes were not.

Recognition accuracy *d'* scores were calculated using this method. A log-linear correction rule was used to account for perfect performance, or perfectly inaccurate performance (Hautus 1995; and see, Koeritzer et al. 2018, for an example of this rule's usage in auditory science).[5]

Results

Intelligibility assessment

Proportion correct performance was at ceiling for all three masker conditions at 5 dB SNR. Basic unweighted psychometric functions using a general linear model and logit link function were fitted to group mean data for each listening condition across SNR (for a visualisation of these data, please visit our OSF page; see Data Availability Statement). As expected, listeners had more difficulty with target recognition in competing speech backgrounds (i.e. one-talker speech and speech babble) across the more challenging SNR conditions than in the steady-state noise masker (a common finding in the speech recognition literature, e.g. Carhart, Tillman, and Greetis 1969). This suggests potential inequality in the listening effort and/or cognitive resource engagement required of the listener to process target speech in these distinct masker conditions. Examination of mean proportion correct scores revealed that participants achieved near-ceiling speech recognition performance in all three conditions of acoustic challenge at 5 dB SNR (one-talker speech: $M = 0.98$, $SE = 0.004$, speech babble: $M = 0.99$, $SE = 0.003$, speech-shaped noise: $M = 0.99$, $SE = 0.003$). At 3 dB SNR, speech recognition performance was highest in the speech-shaped noise condition and lowest in the one-talker speech condition (one-talker speech: $M = 0.86$, $SE = 0.050$, speech babble: $M = 0.91$, $SE = 0.022$, speech-shaped noise: $M = 0.96$, $SE = 0.011$). As a result, a 5 dB SNR was selected for presenting target narrative stimuli in the noise conditions of the narrative comprehension experiment. While it is non-conventional to present target speech at an SNR resulting in near-ceiling level speech recognition, our research question aimed to examine the influence of listening condition on narrative comprehension, while attempting to control for (and ensure) initial speech recognition across conditions. It is also important to note that positive SNRs are more common in real world listening, making this particular (albeit non-conventional) method of using an advantageous SNR more ecologically valid (Smeds, Wolters, and Rung 2015; Weisser and Buchholz 2019).

Narrative comprehension experiment

Recognition accuracy for the three levels of memory is shown in Figure 1. The pattern of results is consistent with prior work in narrative memory in the visual domain. Specifically, surface form memory was worst (Mean *d'* range from 0.13 to 0.21), and event model memory was the best (Mean *d'* range from 0.45 to 0.75). As a performance check, *t*-tests indicated that mean *d'* scores were above chance for all three levels of memory, all p values $<$ 0.001. Prior work has consistently indicated that event model memory is more strongly encoded and retained (Fisher and Radvansky 2018; Radvansky et al. 2001; Schmalhofer and Glavanov 1986). To examine the influence of talker and listening condition on the levels of narrative memory, we consider each level separately (see Fisher and Radvansky 2018 for a similar approach). Four conditions are missing for four separate listeners due to experimenter error during data collection (one one-talker speech condition, one steady-state noise condition, and one speech babble condition in the fixed target talker group, and one quiet condition in the varied target talker group).

Linear mixed models, with participant as a random effect, were used to examine the main effects of talker condition (*fixed target talker* and *varied target talker*), listening condition (*quiet, one-talker speech, speech babble, steady-state noise*), and the

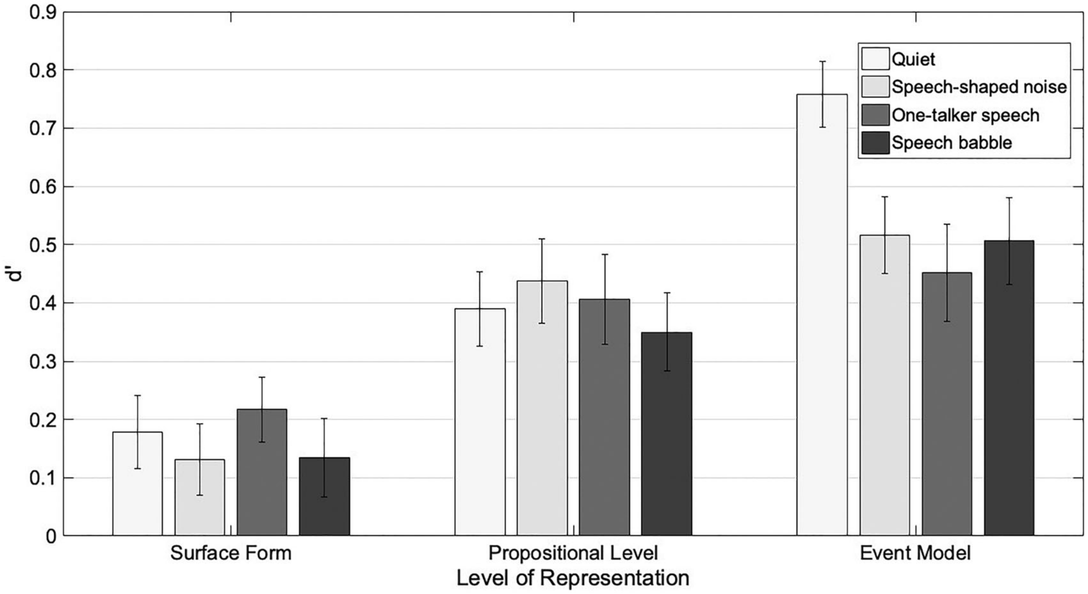

Figure 1. Narrative memory recognition accuracy (*d'*) for the narrative comprehension experiment (data are collapsed across talker condition). Mean recognition scores and standard errors are displayed via grouped bar graphs for surface form, propositional level, and event model narrative representations across each listening condition (*quiet, one-talker speech, speech babble, speech-shaped noise*).

interaction of these two effects (talker condition x listening condition), with WMC included as a covariate, for each of the three levels of memory. Restricted maximum likelihood (REML) estimation with unbounded variance components was used to produce unbiased estimates of variance and covariance parameters.

Results indicated no significant effects at either the surface form or propositional levels. For the surface form, there was no main effect of talker condition ($F(1, 51) = 1.30$, $p = 0.26$), or listening condition ($F(3, 156) = 0.66$, $p = 0.58$), and no interaction ($F(3, 156) = 1.46$, $p = 0.23$). Moreover, the covariate WMC was not significant ($F(1, 51) = 0.09$, $p = 0.76$). For the propositional level, there was no significant main effect of talker condition ($F(1, 51) = 0.99$, $p = 0.32$), or listening condition ($F(3, 156) = 0.31$, $p = 0.82$), nor the interaction ($F(3, 156) = 0.37$, $p = 0.78$). Here the covariate WMC was marginally significant ($F(1, 51) = 3.60$, $p = 0.06$). Thus, with more power, we might be able to detect an influence of WMC, but this influence would be small. It is also not surprising that there might be a relationship here because both the working memory span tests and propositional level memory are focussed on memory for language per se, rather than the referents of described situations in the language signal, which is measured via event model memory.

In contrast, at the event model level there was a main effect of listening condition ($F(3, 156) = 4.47$, $p = 0.005$), with performance being better in the quiet than the background masker conditions. There was no main effect of talker condition ($F(1, 51) = 0.03$, $p = 0.86$), nor a significant interaction ($F(3, 156) = 0.10$, $p = 0.96$). WMC was not a significant predictor of performance ($F(1, 51) = 0.45$, $p = 0.51$). This is consistent with prior work showing that traditional working memory span scores do not predict performance at the event model level (Radvansky and Copeland 2004).

Random effects covariance parameter estimates were used to calculate the intraclass correlation (ICC) for this model, which indicated significant clustering of observations at the level of participant (ICC = 0.16). For this reason, and due to missing data points (described above), a mixed model was used. Post-hoc Tukey-Kramer HSD all pairwise comparisons for the event model displayed a significant reduction in accuracy between the one-talker speech (least squares mean (LSM): $LSM = 0.45$, $SE = 0.07$) and quiet ($LSM = 0.76$, $SE = 0.07$) conditions ($t(53) = -3.33$, $p = 0.006$), between the speech babble ($LSM = 0.50$, $SE = 0.07$) and quiet ($LSM = 0.76$, $SE = 0.07$) conditions ($t(53) = -2.80$, $p = 0.029$), and between speech-shaped noise ($LSM = 0.51$, $SE = 0.07$) and quiet ($LSM = 0.76$, $SE = 0.07$) conditions ($t(53) = -2.70$, $p = 0.038$).

To further examine the main effect of listening condition on narrative comprehension, and to use a more standard analysis approach, positive response proportions to each recognition memory probe type were evaluated. These data are shown in Figure 2. Four separate linear mixed models were conducted to examine mean "yes" proportions to each memory recognition probe type (*verbatim, paraphrase, inference, wrong*), with participant as a random effect, and talker condition (*fixed target talker* and *varied target talker*), listening condition (*quiet, one-talker speech, speech babble, speech-shaped noise*), and the interaction of these two effects (talker condition x listening condition), as fixed effects. WMC was included in each model as a covariate.

REML estimation with unbounded variance components was utilised to produce unbiased estimates of variance and covariance parameters. Statistical modelling indicated no significant effects or interactions for positive responses to verbatim, paraphrase, or inference probes, and WMC was also not significant in any of these models. Statistical analysis of positive responses ("yes") to wrong probes did reveal a significant effect of listening condition ($F(3, 156) = 6.50$, $p < 0.001$), with increased positive responses occurring in conditions with acoustic background competition compared to quiet. No significant effects of talker condition ($F(1, 52) = 0.72$, $p = 0.40$), or the interaction between talker condition and listening condition ($F(3, 156) = 0.48$, $p = 0.69$), were detected. WMC was not a significant predictor of positive responses to wrong probes ($F(1, 52) = 0.54$, $p = 0.46$). An ICC calculation indicated clustering of observations at the level of participant (ICC = 0.51), as well. Tukey-Kramer HSD all pairwise comparisons revealed an increase in positive responses to wrong probes between the speech babble ($LSM = 0.23$, $SE = 0.024$) and quiet ($LSM = 0.14$, $SE = 0.024$) conditions ($t(53) = 3.68$, $p = 0.002$), the one-talker speech ($LSM = 0.22$,

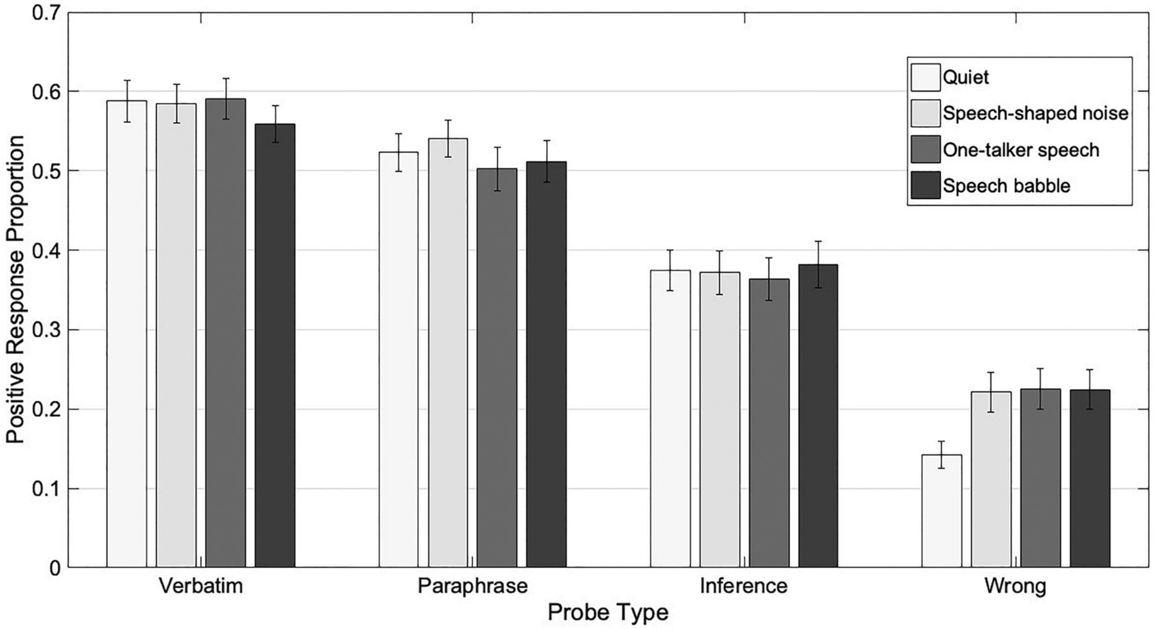

Figure 2. Proportion of positive responses to each recognition memory probe type for the narrative comprehension experiment (data are collapsed across talker condition). Mean positive responses and standard errors are displayed via grouped bar graphs for verbatim, paraphrase, inference, and wrong probe types across each listening condition (*quiet, one-talker speech, speech babble, speech-shaped noise*).

$SE = 0.023$) and quiet ($LSM = 0.14$, $SE = 0.024$) conditions ($t(53) = 3.55$, $p = 0.003$), and the speech-shaped noise ($LSM = 0.22$, $SE = 0.023$) and quiet ($LSM = 0.14$, $SE = 0.024$) conditions ($t(53) = 3.59$, $p = 0.003$).

The increase in the rate of positive responses is in line with the idea that people are struggling to create an adequate event model of the described events of the target narratives when there is background competition present. That is, they are only recording the general thematic nature of the references, and not building an adequate understanding of the described situations. This is happening in the case of there being no impact of background noise on the processing and memory of the language signal itself.

Discussion

The aim of these experiments was to examine the influence of acoustic challenge on spoken narrative comprehension and memory in young adults at an SNR that ensured accurate target speech recognition. An intelligibility assessment conducted prior to the main experiment indicated near-perfect performance in recognising random, context-less segments of the target speech in the conditions of acoustic challenge at 5 dB SNR for young adults with normal hearing. Thus, the three conditions with background competition present (*one-talker speech, speech babble, speech-shaped noise*) allowed for the manipulation of acoustic challenge across listening conditions without negatively influencing initial target speech recognition. While near-ceiling level speech recognition was ensured across listening conditions as a result of the advantageous SNR that we used, it is likely that greater degrees of listening effort and/or cognitive resource engagement may have occurred in conditions with acoustic background competition present.

In the narrative comprehension experiment, recognition accuracy in two separate target talker condition groups (*fixed* and *varied target talker*) was measured in four listening conditions (*quiet, one-talker speech, speech babble, speech-shaped noise*) across three levels of memory representation (*surface form, propositional, event model*). Results indicated that young adults had poorer memory recognition accuracy at the event model level when listening to narratives in the conditions with acoustic background competition present (*one-talker speech*, *speech babble*, and *speech-shaped noise*), compared to quiet. In a subsequent analysis of positive responses to each probe type (*verbatim, paraphrase, inference, wrong*) across the four listening conditions, it was found that increased proportions of positive responses to wrong probes in the acoustic challenge conditions was driving this reduction in event model level recognition accuracy. In other words, when narratives had been presented in the conditions of acoustic challenge, people had trouble distinguishing between ideas that were and were not consistent with the described events, but only shared some vague, general thematic consistency. Event model representations characterise a comprehensive understanding of what is being said across the words and sentences of the narrative (van Dijk and Kintsch 1983). Such representations are essential when it comes to indexing the comprehension of connected speech. The results of this experiment indicate that it is the robustness of this level of narrative representation that may be negatively influenced by acoustic background competition.

No significant differences were detected as a result of talker condition. That is, participants that heard all four target narratives produced by the same talker did not perform differently than those that heard the narratives spoken by four unique talkers. While previous work has demonstrated negative effects of talker variability on subsequent memory for speech stimuli (Bradlow, Nygaard, and Pisoni 1999), it is possible that the infrequent and predictable voice changes in this paradigm allowed listeners to adjust to each talker with relative ease (Best et al. 2016). This lack of effect also falls in line with previous work showing that talker adaptation in varied target talker conditions seems to occur rapidly (Morton, Sommers, and Lulich 2015). This also demonstrates that the unique target talkers used in the *varied target talker* condition likely did not require differing degrees of listening effort and/or cognitive resource allocation in initial speech processing to the extent that narrative comprehension was influenced in a significant manner compared to the *fixed target talker* condition.

Our results indicate a negative effect of acoustic challenge on event model memory, and, more specifically, on the likelihood of incorrectly identifying a thematically consistent, but situationally incorrect sentence as having been heard during the narrative. This effect was detected even though target narratives were presented at a positive SNR allowing near-perfect speech recognition. Therefore, deleterious effects of acoustic background competition on memory encoding and comprehension may be occurring at central levels of speech processing, potentially as a result of increased information processing resource allocation to initial speech decoding and lexical access (Gordon, Daneman, and Schneider 2009; Sörqvist and Rönnberg 2012; Ward et al. 2016). Although these results appear to be in line with predictions of the current frameworks of speech processing (such as ELU, Rönnberg et al. 2013), WMC was not a significant predictor of performance in any condition, although it was marginally so at the propositional level.

Inconsistencies related to the importance of WMC in accounting for variance in target speech processing in conditions of acoustic challenge exist across the literature (Füllgrabe, Moore, and Stone 2015; Füllgrabe and Rosen 2016). This may be related to the level of memory representation targeted by the task demands and/or the nature of the stimuli used in the experimental paradigm. The reading span task measures verbatim memory for single words in combination with a sentence processing task. Scores on this task correlate well with other tasks that are similar in nature (i.e. processing and remembering isolated words or sentences with little or no surrounding context). Thus, this is near transfer. The processing of narrative materials goes well beyond this, and is likely not captured by traditional WMC measures (cf., Radvansky and Copeland 2004). Therefore, it makes sense that the cognitive resources used to build memory representations for such speech materials differ from those measured by traditional working memory tests. The exact nature of these resources remains elusive and needs to be explored in future work.

In conditions with acoustic background competition, listeners must attend to the target speech of interest and segregate it from interfering environmental sounds. The competing signals often overlap and combine in both time and frequency with the target, forcing listeners to process a spectrally and temporally sparse signal and subsequently allocate top-down resources to focus upon and disambiguate sounds belonging to the target stream while simultaneously inhibiting sounds of the masker stream (Edwards 2016; Shinn-Cunningham and Best 2015). It is likely that the process of auditory stream segregation may drag on the limited pool of information processing resources and negatively influence subsequent speech comprehension processes when listening to speech in noisy environments (Bregman 1990; Heinrich, Schneider, and Craik 2008; Wingfield 2016), with the inhibitory effects possibly being greater for speech maskers than for comparable non-speech maskers (Sörqvist and Rönnberg 2012; although this effect was not indicated in our results).

Our results diverged from our initial hypothesis, in that conditions with competing speech backgrounds (*one-talker speech* and *speech babble*) did not negatively influence narrative memory to a greater extent than the steady-state noise background (*speech-shaped noise*). While this effect has been shown in previous work (Sörqvist and Rönnberg 2012), equivalence of initial speech recognition between listening conditions has rarely been controlled for. This lack of difference in target narrative comprehension between masker conditions that were hypothesised to be more informational in nature (i.e. *one-talker speech* and *speech babble*) versus mainly energetic in nature (*speech-shaped noise*) was surprising. However, previous research has indicated that informational masking declines dramatically when the target sentence is louder than the masker sentence (i.e. a positive SNR; Arbogast, Mason, and Kidd 2005; Freyman, Balakrishnan, and Helfer 2008). It is possible that the advantageous SNR that we utilised (in which a strong level cue to designate the target speech exists) allowed for relatively easy segregation of the target and masker stimuli even when the masker was competing speech. In listening situations with competing speech present, stream segregation builds up over the course of a sentence (Ezzatian et al. 2012). Due to our advantageous (positive) SNR, the benefits of rich semantic context in our narrative materials, and the overall length of our narratives (on the order of connected sentences and paragraphs per trial, rather than isolated words or sentences per trial), it is very possible that target/masker stream segregation was quite easy for our normally hearing listeners, rendering speech processing differences in the presence of speech maskers versus the noise masker non-significant.

The idea that information processing resources are allocated towards early stage language processes and away from conceptual integration, inference, and language comprehension in situations with reduced target signal fidelity (such as sensory impairment or competing background noise) has been demonstrated in the reading literature. For example, Gao et al. (2011, 2012) tested the hypothesis that visual noise added to a reading task (i.e. signal degradation of the text) increases difficulty in initial text decoding and lexical access, thereby reducing the resources available for higher-level cognitive functions such as semantic integration and comprehension. They used an online resource allocation paradigm, as well as an offline memory recall task, to examine the effect of varying degrees of visual noise on sentence processing during reading in young adults. Specifically, using the resource allocation approach, the authors decomposed reading times via regression into time/resource measurements for word-level processing (similar to initial speech recognition in listening) and textbase-level (or propositional) processing during a moving window reading paradigm (Stine-Morrow, Miller, and Hertzog 2006). The researchers also measured memory for the text material via a recall task after reading.

In both Gao et al. (2011), and Gao, Levinthal, and Stine-Morrow (2012) (Experiment 2), results indicated that visual noise drew participants' attentional resources away from textbase-level processing towards word-level processing, with greater degrees of disruption in resource allocation occurring in response to greater degrees of visual noise. This online result was mirrored in the recall task, in which the quality of participants' text recall was reduced when reading in greater degrees of visual noise. They took this evidence as an indication that extra time and information processing resources were spent on orthographic decoding and lexical access in conditions with visual noise, thereby reducing resources that could have been used for higher-level conceptual integration and language comprehension. Interestingly, recall performance indicated that people were less likely to recall core, "main" ideas from the text presented in the noisier conditions. When discussing these results, it was concluded that visual noise caused participants to produce "fuzzier", or less robust, conceptual representations of the central text ideas, disrupting effective concept integration and semantic analysis in language comprehension (Gao et al. 2011, Gao, Levinthal, and Stine-Morrow 2012).

These results bear resemblance to our own, in which, even with accurate initial speech recognition (which was also confirmed in each level of visual noise in the Gao et al., research via a lexical decision task), downstream comprehension was disrupted by target signal degradations. While our narrative comprehension measure was offline, it is possible that our listeners may have allocated greater processing resources towards lower levels of speech processing (i.e. speech recognition, lexical access, construction of surface form representations), thereby leaving fewer resources for the inference generation and conceptual integration of event model construction in conditions of acoustic challenge.

Data from Ward et al. (2016) (discussed in the introduction), provide further evidence that target speech signal degradations (in their case, via noise vocoding) can result in people recalling fewer "main" idea units from spoken narratives (Ward et al. 2016). Converging evidence from both visual and auditory research suggests that challenges in initial recognition of linguistic stimuli using distinct target signal manipulations (visual noise: Gao et al. 2011; Gao, Levinthal, and Stine-Morrow 2012; noise vocoding; Ward et al. 2016; acoustic background competition: present work), may involve a re-allocation of some domain-general information processing resources to earlier stages of language processing and away from higher-level comprehension processes. It is possible that the narrative comprehension deficits displayed in conditions of acoustic background competition in the present work may be related to potentially greater degrees of listening effort that participants had to put forth in the noisy conditions compared to quiet. A growing body of literature is lending credence to the idea that various acoustic challenges require listeners to engage cognitive processing resources to recognise and comprehend target speech, and, consequently, make listening more effortful (Peelle 2018). Future work on the effects of acoustic challenge on speech comprehension may consider using online measures (such as eye-tracking and/or pupillometry) in combination with behavioural tasks (such as recognition memory and/or memory recall) to examine changes in resource allocation.

One striking outcome of the current study is how the comprehension and memory of different levels of speech are influenced by background noise as a function of the context in which they are embedded. When people are presented with isolated words or sentences, memory for these materials is disrupted by background noise (Koeritzer et al. 2018; Murphy et al. 2000). Moreover, processing at these levels is often related to WMC scores (Ljung, Israelsson, and Hygge 2013). However, what the current study reveals is that this performance deficit is largely absent when words and sentences are embedded within a larger narrative context. Memory for the language, per se, is the same in noise and quiet, but memory for what the language is about, the event model, is compromised. Moreover, this processing deficit may be unrelated to traditional WMC measures. This suggests that work that focuses on lists of isolated words and sentences, outside of any meaningful context, falls short of capturing the speech processing struggles people encounter in the world from environmental challenges (i.e. background noise), hearing deficits, or both. The nature of the cognitive resources and processes that are compromised by acoustic challenges in speech processing are unknown at this time, and may not be well-captured by traditional conceptions of working memory capacity.

In conclusion, the current data demonstrate that acoustic challenges can negatively influence the comprehension of spoken narratives, even in conditions with otherwise near-perfect initial speech recognition. This seems to be driven by listeners producing less robust event model memory representations, indicated by the greater likelihood of identifying wrong (or inaccurate) probe sentences as having been a part of the narratives heard in conditions with acoustic background competition. Information processing resources may be shunted to lower levels of speech processing in challenging acoustic conditions, such as those used here, thereby leaving fewer resources for memory encoding and comprehension. Distinct processes involved in the construction of accurate memory representations for the described events of a narrative may also be particularly influenced by challenges in initial speech processing. Frequent reports from patients in audiology clinics indicating that understanding speech in the real world is challenging beyond simply hearing the sounds are potentially related to these phenomena.

Notes

1. For written text, this is often referred to as the textbase level.
2. Often, speech babble maskers are produced with multiple unique voices. Here, we used one consistent voice to create this masker (see also Spahr et al. 2012 and Iyer et al. 2010). Perceptually, this still resulted in a babble noise, which can be accessed via our Open Science Framework page (see link in Data Availability Statement).
3. Also referred to as a norming study in the memory literature.
4. Planned recruitment for the *varied target talker* condition was 30 participants. However, data collection was stopped as a result of the coronavirus pandemic.
5. Although the Schmalhofer and Glavanov procedure uses signal detection theory to parse the levels of memory representation, it is not conventional to use responses to probes that were not actually presented to participants as the signal in this way (i.e., for propositional and event model recognition accuracy). Therefore, we also examine positive response proportions (i.e., the mean rates of responding "yes" to the four different probe types) in a separate analysis of these data.

Acknowledgements

We are thankful for our undergraduate and graduate research assistants. Their help with voice recordings and data collection during this project was invaluable. We are also grateful to Jessica Salley, Nardine Taleb, and Jessica Kong for their assistance during this project.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this paper.

Data availability statement

Stimulus materials and memory data are available from https://osf.io/pwamf/.

References

Arbogast, T. L., C. R. Mason, and G. Kidd. Jr. 2005. "The Effect of Spatial Separation on Informational Masking of Speech in Normal-Hearing and Hearing-Impaired Listeners." *The Journal of the Acoustical Society of America* 117 (4 Pt 1): 2169–2180. doi:10.1121/1.1861598.

Best, V., G. Keidser, K. Freeston, and J. M. Buchholz. 2016. "A Dynamic Speech Comprehension Test for Assessing Real-World Listening Ability." *Journal of the American Academy of Audiology* 27 (7): 515–526. doi:10.3766/jaaa.15089.

Boersma, P., and D. Weenink. 2017. *Praat: Doing Phonetics by Computer* [computer program]. http://praat.org/

Bohay, M., D. P. Blakely, A. K. Tamplin, and G. A. Radvansky. 2011. "Note-Taking, Review, Memory and Comprehension." *American Journal of Psychology* 124: 63–73. doi:10.5406/amerjpsyc.124.1.0063.

Bradlow, A. R., G. M. Torretta, and D. B. Pisoni. 1996. "Intelligibility of Normal Speech I: Global and Fine-Grained Acoustic-Phonetic Talker Characteristics." *Speech Communication* 20 (3): 255–272. doi:10.1016/S0167-6393(96)00063-5.

Bradlow, A. R., L. C. Nygaard, and D. B. Pisoni. 1999. "Effect of Talker, Rate, and Amplitude Variation on Recognition Memory for Spoken Words." *Perception & Psychophysics* 61 (2): 206–219. doi:10.3758/BF03206883.

Bregman, A. S. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Carhart, R., T. W. Tillman, and E. S. Greetis. 1969. "Release from Multiple Maskers: Effects of Interaural Time Disparities." *The Journal of the Acoustical Society of America* 45 (2): 411–418. doi:10.1121/1.1911389.

Daneman, M., and P. Carpenter. 1980. "Individual Differences in Working Memory and Reading." *Journal of Verbal Learning and Verbal Behavior* 19 (4): 450–466. doi:10.1016/S0022-5371(80)90312-6.

Dubno, J. R., J. B. Ahlstrom, and A. R. Horwitz. 2000. "Use of Context by Young and Aged Adults with Normal Hearing." *The Journal of the Acoustical Society of America* 107 (1): 538–546. doi:10.1121/1.428322.

Durlach, N. I., C. R. Mason, G. Kidd, Jr., T. Arbogast, H. S. Colburn, and B. G. Shinn-Cunningham. 2003. "Note on Informational Masking (L)." *The Journal of the Acoustical Society of America* 113 (6): 2984–2987. doi:10.1121/1.1570435.

Edwards, B. 2016. "A Model of Auditory-Cognitive Processing and Relevance to Clinical Applicability." *Ear and Hearing* 37 (S1): 85S–91S. doi:10.1097/AUD.0000000000000308.

Ezzatian, P., L. Li, M. K. Pichora-Fuller, and B. A. Schneider. 2012. "The Effect of Energetic and Informational Masking on the Time-Course of Stream Segregation: Evidence That Streaming Depends on Vocal Fine Structure Cues." *Language and Cognitive Processes* 27 (7-8): 1056–1088. doi:10.1080/01690965.2011.591934.

Fisher, J. S., and G. A. Radvansky. 2018. "Patterns of Forgetting." *Journal of Memory and Language* 102: 130–141. doi:10.1016/j.jml.2018.05.008.

Fletcher, C. R., and S. T. Chrysler. 1990. "Surface Forms, Textbases, and Situation Models: Recognition Memory for Three Types of Textual Information." *Discourse Processes* 13 (2): 175–190. doi:10.1080/01638539009544752.

Freyman, R. L., U. Balakrishnan, and K. S. Helfer. 2008. "Spatial Release from Masking with Noise-Vocoded Speech." *The Journal of the Acoustical Society of America* 124 (3): 1627–1637. doi:10.1121/1.2951964.

Füllgrabe, C., and S. Rosen. 2016. "On the (un)Importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds." *Frontiers in Psychology* 07: 1–8. doi:10.3389/fpsyg.2016.01268.

Füllgrabe, C., B. C. J. Moore, and M. A. Stone. 2015. "Age-Group Differences in Speech Identification Despite Matched Audiometrically Normal Hearing: Contributions from Auditory Temporal Processing and Cognition." *Frontiers in Aging Neuroscience* 6: 1–25.

Gao, X., B. R. Levinthal, and E. A. Stine-Morrow. 2012. "The Effects of Ageing and Visual Noise on Conceptual Integration During Sentence Reading." *Quarterly Journal of Experimental Psychology (2006)* 65 (9): 1833–1847. doi:10.1080/17470218.2012.674146.

Gao, X., E. A. Stine-Morrow, S. R. Noh, and R. T. Eskew, Jr. 2011. "Visual Noise Disrupts Conceptual Integration in Reading." *Psychonomic Bulletin & Review* 18 (1): 83–88. doi:10.3758/s13423-010-0014-4.

Gordon, M. S., M. Daneman, and B. A. Schneider. 2009. "Comprehension of Speeded Discourse by Younger and Older Listeners." *Experimental Aging Research* 35 (3): 277–296. doi:10.1080/03610730902769262.

Griffin, A. M., S. F. Poissant, and R. L. Freyman. 2020. "Auditory Comprehension in School-Aged Children with Normal Hearing and with Unilateral Hearing Loss." *Language, Speech, and Hearing Services in Schools* 51 (1): 29–41. doi:10.1044/2019_LSHSS-OCHL-19-0020.

Hautus, M. J. 1995. "Corrections for Extreme Proportions and Their Biasing Effects on Estimated Values of *d*." *Behavioral Research Methods* 27 (1): 46–51. doi:10.3758/BF03203619.

Heinrich, A., B. A. Schneider, and F. Craik. 2008. "Investigating the Influence of Continuous Babble on Auditory Short-Term Memory Performance." *Quarterly Journal of Experimental Psychology* 61 (5): 735–751. doi:10.1080/17470210701402372.

Heldner, M., and J. Edlund. 2010. "Pauses, Gaps and Overlaps in Conversations." *Journal of Phonetics* 38 (4): 555–568. doi:10.1016/j.wocn.2010.08.002.

Iyer, N., D. S. Brungart, and B. D. Simpson. 2010. "Effects of Target-Masker Contextual Similarity on the Multimasker Penalty in a Three-Talker Diotic Listening Task." *The Journal of the Acoustical Society of America* 128 (5): 2998–3010. doi:10.1121/1.3479547.

Kalikow, D. N., K. N. Stevens, and L. L. Elliott. 1977. "Development of a Test of Speech Intelligibility in Noise Using Sentence Material with Controlled Word Predictability." *The Journal of the Acoustical Society of America* 61 (5): 1337–1351. doi:10.1121/1.381436.

Kidd, G., Jr., and H. S. Colburn. 2017. "Informational Masking in Speech Recognition." In *The Auditory System at the Cocktail Party*, edited by J. Middlebrooks, J. Simon, A. Popper, R. Fay, 75–109. Cham: Springer.

Kintsch, W. 1988. "The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model." *Psychological Review* 95 (2): 163–182. doi:10.1037/0033-295x.95.2.163.

Kintsch, W., D. Welsch, F. Schmalhofer, and S. Zimny. 1990. "Sentence Memory: A Theoretical Analysis." *Journal of Memory and Language* 29 (2): 133–159. doi:10.1016/0749-596X(90)90069-C.

Kjellberg, A., R. Ljung, and D. Hallman. 2008. "Recall of Words Heard in Noise." *Applied Cognitive Psychology* 22 (8): 1088–1098. doi:10.1002/acp.1422.

Klaus, J., and H. Schriefers. 2016. *Measuring Working Memory Capacity: A Reading Span Task for Laboratory and Web-Based Use*. Unpublished manuscript submitted for publication. https://osf.io/nj48x

Koeritzer, M. A., C. S. Rogers, K. J. Van Engen, and J. E. Peelle. 2018. "The Impact of Age, Background Noise, Semantic Ambiguity, and Hearing Loss on Recognition Memory for Spoken Sentences." *Journal of Speech, Language, and Hearing Research : JSLHR* 61 (3): 740–751. doi:10.1044/2017_JSLHR-H-17-0077.

Kryter, K. D. 1962. "Methods of Calculation and Use of the Articulation Index." *The Journal of the Acoustical Society of America* 34 (11): 1689–1697. doi:10.1121/1.1909094.

Ljung, R., K. Israelsson, and S. Hygge. 2013. "Speech Intelligibility and Recall of Spoken Material Heard at Different Signal-to-Noise Ratios and the Role Played by Working Memory Capacity." *Applied Cognitive Psychology* 27 (2): 198–203. doi:10.1002/acp.2896.

Martin, C. S., J. W. Mullennix, D. B. Pisoni, and W. V. Summers. 1989. "Effects of Talker Variability on Recall of Spoken Word Lists." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 15 (4): 676–684. doi:10.1037//0278-7393.15.4.676.

Mattys, S. L., M. H. Davis, A. R. Bradlow, and S. K. Scott. 2012. "Speech Recognition in Adverse Conditions: A Review." *Language and Cognitive Processes* 27 (7-8): 953–978. doi:10.1080/01690965.2012.705006.

McCabe, D. P., H. L. Roediger, M. A. McDaniel, D. A. Balota, and D. Z. Hambrick. 2010. "The Relationship Between Working Memory Capacity and Executive Functioning: Evidence for a Common Executive Attention Construct." *Neuropsychology* 24 (2): 222–243. doi:10.1037/a0017619.

McCoy, S. L., P. A. Tun, L. C. Cox, M. Colangelo, R. A. Stewart, and A. Wingfield. 2005. "Hearing Loss and Perceptual Effort: Downstream Effects on Older Adults' Memory for Speech." *The Quarterly Journal of Experimental Psychology Section A* 58 (1): 22–33. doi:10.1080/02724980443000151.

Moore, B. 2013. *An Introduction to the Psychology of Hearing*. Boston, MA: Koninklijke Brill.

Morton, J. R., M. S. Sommers, and S. M. Lulich. 2015. "The Effect of Exposure to a Single Vowel on Talker Normalization for Vowels." *The Journal of the Acoustical Society of America* 137 (3): 1443–1451. doi:10.1121/1.4913456.

Murphy, D., F. Craik, Z. H. Karen, and B. Schneider. 2000. "Comparing the Effects of Aging and Background Noise on Short-Term Memory Performance." *Psychology and Aging* 15 (2): 323–334. doi:10.1037/0882-7974.15.2.323.

Narvaez, D., G. A. Radvansky, N. A. Lynchard, and D. E. Copeland. 2011. "Are Older Adults More Attuned to Morally-Charged Information?" *Experimental Aging Research* 37 (4): 398–434. doi:10.1080/0361073X.2011.590756.

Peelle, J. E. 2018. "Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior." *Ear and Hearing* 39 (2): 204–214. doi:10.1097/AUD.0000000000000494.

Pichora-Fuller, M. K. 2008. "Use of Supportive Context by Younger and Older Adult Listeners: Balancing Bottom-up and Top-Down Information Processing." *International Journal of Audiology* 47 (S2): S72–S82. doi:10.1080/14992020802307404

Pichora-Fuller, M. K., B. Schneider, and M. Daneman. 1995. "How Young and Old Adults Listen to and Remember Speech in Noise." *The Journal of the Acoustical Society of America* 97: 593–608. doi:10.1121/1.412282

Pichora-Fuller, M. K., S. E. Kramer, M. A. Eckert, B. Edwards, B. Hornsby, L. E. Humes, U. Lemke, et al. 2016. "Hearing Impairment and Cognitive

Energy: The Framework for Understanding Effortful Listening (FUEL).” *Ear and Hearing* 37 (S1): 5S–27S. doi:10.1097/AUD.0000000000000312.

Piquado, T., J. I. Benichov, H. Brownell, and A. Wingfield. 2012. “The Hidden Effect of Hearing Acuity on Speech Recall, and Compensatory Effects of Self-Paced Listening.” *International Journal of Audiology* 51 (8): 576–583. doi:10.3109/14992027.2012.684403.

Rabbit, P. M. 1968. “Channel-Capacity, Intelligibility and Immediate Memory.” *Quarterly Journal of Experimental Psychology* 20: 241–248. doi:10.1080/14640746808400158

Rabbit, P. M. 1991. “Mild Hearing Loss Can Cause Apparent Memory Failures Which Increase with Age and Reduce with IQ.” *Acta Otolaryngolica* 476: 167–176.

Radvansky, G. A., and D. E. Copeland. 2004. “Working Memory Span and Situation Model Processing.” *The American Journal of Psychology* 117 (2): 191–213. doi:10.2307/4149022.

Radvansky, G. A., and J. M. Zacks. 2014. *Event Cognition*. Oxford, UK: Oxford University Press.

Radvansky, G. A., D. E. Copeland, and R. A. Zwaan. 2003. “Aging and Functional Spatial Relations in Comprehension and Memory.” *Psychology and Aging* 18 (1): 161–165. doi:10.1037/0882-7974.18.1.161.

Radvansky, G. A., D. E. Copeland, and W. von Hippel. 2010. “Stereotype Activation, Inhibition, and Aging.” *Journal of Experimental Social Psychology* 46 (1): 51–60. doi:10.1016/j.jesp.2009.09.010.

Radvansky, G. A., R. A. Zwaan, J. M. Curiel, and D. E. Copeland. 2001. “Situation Models and Aging.” *Psychology and Aging* 16 (1): 145–160. doi:10.1037/0882-7974.16.1.145.

Rönnberg, J., T. Lunner, A. Zekveld, P. Sörqvist, H. Danielsson, B. Lyxell, O. Dahlström, et al. 2013. “The Ease of Language Understanding (ELU) Model: Theoretical, Empirical, and Clinical Advances.” *Frontiers in Systems Neuroscience* 7: 31.

Sachs, J. S. 1967. “Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse.” *Perception & Psychophysics* 2 (9): 437–442. doi:10.3758/BF03208784.

Schmalhofer, F., and D. Glavanov. 1986. “Three Components of Understanding a Programmer’s Manual: Verbatim, Propositional, and Situational Representations.” *Journal of Memory and Language* 25 (3): 279–294. doi:10.1016/0749-596X(86)90002-1.

Schneider, B. A., M. Daneman, D. R. Murphy, and S. K. See. 2000. “Listening to Discourse in Distracting Settings: The Effects of Aging.” *Psychology and Aging* 15 (1): 110–125. doi:10.1037//0882-7974.15.1.110.

Shannon, R. V., F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. 1995. “Speech Recognition with Primarily Temporal Cues.” *Science (New York, N.Y.)* 270 (5234): 303–304. doi:10.1126/science.270.5234.303.

Shinn-Cunningham, B., and V. Best. 2015. “Auditory Selective Attention.” In *The Handbook of Attention*, edited by J. M. Fawcett, E. F. Risko, & A. Kingstone, 99–117. Cambridge, MA: The MIT Press.

Smeds, K., F. Wolters, and M. Rung. 2015. “Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios.” *Journal of the American Academy of Audiology* 26 (2): 183–196. doi:10.3766/jaaa.26.2.7.

Sommers, M. S., and S. M. Danielson. 1999. “Inhibitory Processes and Spoken Word Recognition in Young and Older Adults: The Interaction of Lexical Competition and Semantic Context.” *Psychology and Aging* 14 (3): 458–472. doi:10.1037/0882-7974.14.3.458.

Sörqvist, P., and J. Rönnberg. 2012. “Episodic Long-Term Memory of Spoken Discourse Masked by Speech: What Is the Role for Working Memory Capacity?” *Journal of Speech, Language, and Hearing Research : JSLHR* 55 (1): 210–218. doi:10.1044/1092-4388(2011/10-0353).

Spahr, A. J., M. F. Dorman, L. M. Litvak, S. Van Wie, R. H. Gifford, P. C. Loizou, L. M. Loiselle, T. Oakes, and S. Cook. 2012. “Development and Validation of the AzBio Sentence Lists.” *Ear and Hearing* 33 (1): 112–117. doi:10.1097/AUD.0b013e31822c2549.

Stanislaw, H., and N. Todorov. 1999. “Calculation of Signal Detection Theory Measures.” *Behavioral Research Methods* 31 (1): 137–149. doi:10.3758/BF03207704.

Stine-Morrow, E. A. L., L. M. S. Miller, and C. Hertzog. 2006. “Aging and Self-Regulated Language Processing.” *Psychological Bulletin* 132 (4): 582–606. doi:10.1037/0033-2909.132.4.582.

Stone, M. A., C. Fullgrabe, and B. C. J. Moore. 2012. “Notionally Steady Background Noise Acts Primarily as a Modulation Masker of Speech.” *The Journal of the Acoustical Society of America* 132 (1): 317–326. doi:10.1121/1.4725766.

Surprenant, A. M. 1999. “The Effect of Noise on Memory for Spoken Syllables.” *International Journal of Psychology* 34 (5-6): 328–333. doi:10.1080/002075999399648.

Tulving, Endel, and Donald M. Thomson. 1973. “Encoding Specificity and Retrieval Processes in Episodic Memory.” *Psychological Review* 80 (5): 352–373. doi:10.1037/h0020071.

Tye-Murray, N., M. S. Sommers, B. Spehar, J. Myerson, S. Hale, and N. Rose. 2008. “Auditory-Visual Discourse Comprehension by Older and Younger Adults in Favorable and Unfavorable Conditions.” *International Journal of Audiology* 47 (sup2): S31–S37. doi:10.1080/14992020802301662.

van Dijk, T. A., and W. Kintsch. 1983. *Strategies in Discourse Comprehension*. New York: Academic Press.

Ward, C. M., C. S. Rogers, K. J. Van Engen, and J. E. Peelle. 2016. “Effects of Age, Acoustic Challenge, and Verbal Working Memory on Recall of Narrative Speech.” *Experimental Aging Research* 42 (1): 97–144. doi:10.1080/0361073X.2016.1108785.

Watson, C. S. 2005. “Some Comments on Informational Masking.” *Acta Acustica United with Acustica* 91: 502–512.

Weisser, A., and J. M. Buchholz. 2019. “Conversational Speech Levels and Signal-to-Noise Ratios in Realistic Acoustic Conditions.” *The Journal of the Acoustical Society of America* 145 (1): 349–360. doi:10.1121/1.5087567.

Wingfield, A. 2016. “Evolution of Models of Working Memory and Cognitive Resources.” *Ear and Hearing* 37 (S1): 35S–43S. doi:10.1097/AUD.0000000000000310.

Wingfield, A., P. A. Tun, C. K. Koh, and M. J. Rosen. 1999. “Regaining Lost Time: Adult Aging and the Effect of Time Restoration on Recall of Time-Compressed Speech.” *Psychology and Aging* 14 (3): 380–389. doi:10.1037//0882-7974.14.3.380.

Winn, M. B. 2016. “Rapid Release from Listening Effort Resulting from Semantic Context, and the Effects of Spectral Degradation and Cochlear Implants.” *Trends in Hearing* 20: 233121651666972–233121651666917. doi:10.1177/2331216516669723.

Yost, W. A. 2013. *Fundamentals of Hearing, an Introduction*. 5th ed. Lieden, The Netherlands: Brill Academic Publishers.

Zwaan, R. A. 1994. “Effect of Genre Expectations on Text Comprehension.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20 (4): 920–933. doi:10.1037/0278-7393.20.4.920.

Zwaan, R. A. 1999. “Situation Models: The Mental Leap into Imagined Worlds.” *Current Directions in Psychological Science* 8 (1): 15–18. doi:10.1111/1467-8721.00004.

Zwaan, R. A., and G. A. Radvansky. 1998. “Situation Models in Language Comprehension and Memory.” *Psychological Bulletin* 123 (2): 162–185. doi:10.1037/0033-2909.123.2.162.