International Journal of Audiology

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

The role of working memory in speech recognition by hearing-impaired older listeners: does the task matter?

Dorina Strori & Pamela E. Souza

To cite this article: Dorina Strori & Pamela E. Souza (2022): The role of working memory in speech recognition by hearing-impaired older listeners: does the task matter?, International Journal of Audiology, DOI: 10.1080/14992027.2022.2128445

To link to this article: https://doi.org/10.1080/14992027.2022.2128445

Published online: 26 Oct 2022.

INTERNATIONAL JOURNAL OF AUDIOLOGY
https://doi.org/10.1080/14992027.2022.2128445

ORIGINAL ARTICLE

The role of working memory in speech recognition by hearing-impaired older listeners: does the task matter?

Dorina Strori[a] and Pamela E. Souza[a,b]

[a]Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA; [b]Knowles Hearing Center, Northwestern University, Evanston, IL, USA

ABSTRACT

Objective: Working memory refers to a cognitive system that holds a limited amount of information in a temporarily heightened state of availability, for use in ongoing cognitive tasks. Research suggests a link between working memory and speech recognition. In this study, we investigated this relationship using two working memory tests that differed in regard to the operationalisation of the link between working memory and attention: the *auditory visual divided attention test* (AVDAT) and the widely used reading span test.
Design: The relationship between speech-in-noise recognition and working memory was examined for two different working memory tests that varied in methodological and theoretical aspects, using a within-subject design.
Study sample: Nineteen hearing-impaired older listeners participated.
Results: We found a strong link between the reading span test and speech-in-noise recognition and a less robust link between the AVDAT and speech-in-noise recognition. There was evidence for the role of selective attention in speech-in-noise recognition, shown via the new AVDAT measure.
Conclusion: Our findings suggest that the strength of the relationship between speech-in-noise recognition and working memory may be influenced by the match between the demands and the stimuli of the speech-in-noise task and those of the working memory test.

ARTICLE HISTORY
Received 2 November 2021
Revised 29 June 2022
Accepted 15 September 2022

KEYWORDS
Working memory; selective and divided attention; speech-in-noise recognition; working memory measures; behavioral measures; speech perception

Introduction

Working memory and speech recognition

Working memory (WM), one of the most studied cognitive constructs across various disciplines, has been defined and conceptualised in several ways (see Baddeley 2012 and Cowan 2017 for reviews). A general definition of WM that has been deemed applicable across different theories of WM and a wide range of implementations of the concept refers to WM as a system of components that holds a limited amount of information in a temporarily heightened state of availability, for use in ongoing information processing tasks (Cowan 2017).

WM capacity has been linked to speech recognition, particularly in adverse listening conditions, such as in the presence of noise and/or hearing loss (e.g. Souza and Arehart 2015; Strori, Bradlow, and Souza 2021; Zekveld, Rudner, Johnsrude, & Rönnberg (2013); for reviews, see Akeroyd 2008; Besser et al. 2013; Souza, Arehart, and Neher 2015). The Ease of Language Understanding (ELU) model developed by Rönnberg et al. (2013) offers a comprehensive description of the relationship between WM capacity and speech recognition. To provide a brief overview, in the ELU model, lexical retrieval is facilitated by an unambiguous match between language input and the respective phonological representation stored in long-term memory, with retrieval occurring in an automatic and relatively effortless manner. When the incoming language input is degraded (e.g. by background noise and/or hearing loss), lexical retrieval is impaired by the difficulty in matching the new information to the corresponding phonological representation(s). Consequently, WM is explicitly engaged to facilitate the match. This view has been supported by several studies that found a more robust relationship between WM capacity and speech recognition in noise than in quiet and for older adults with hearing impairment (see Akeroyd 2008; Besser et al. 2013, for reviews) compared to young adults without hearing impairment (Füllgrabe and Rosen 2016).

Working memory and attention

Recognising speech in realistic situations, such as in the presence of noise, requires the listener to process a rapidly incoming auditory stream, attend to the relevant part of this stream (speech) while ignoring the irrelevant background noise, concurrently extract information and store that information for integration with subsequent input and later retrieval. It is thus reasonable to expect that speech recognition draws upon both WM and attention resources. More specifically, selective attention (the ability to direct attention to the relevant information and ignore co-occurring irrelevant information in the background), divided attention (the ability to attend to two or more streams of information), and the ability to temporarily manipulate and store task-relevant information in WM will all impact how speech is processed and recognised. WM and attention are considered to be closely linked by a broad consensus in the literature surrounding these multi-faceted constructs (e.g. Cowan 1998; Kane et al. 2001). One predominant view of attention is that of a

CONTACT Dorina Strori dorina.strori@gmail.com Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

limited resource for information processing (Wickens 1980). According to theories that link WM to attention, the limited capacity of WM reflects a limited cognitive resource, which also serves functions typically attributed to attention. The link between WM and attention can be conceptualised in several ways that differ in terms of the functions that draw on the limited attentional resource (see Oberauer 2019 for a detailed treatment of this topic). Here we focus on two conceptualisations that are relevant for the purposes of the WM measures used in the present study: (1) attention as a limited resource for *storage and processing of information* (e.g. Daneman and Carpenter 1980) and (2) attention as a limited resource for *controlling purposes* (Schneider and Shiffrin 1977).

In the first conceptualisation, attention is shared between "storage" and "processing" task demands. That is, the same attentional resource is required to keep representations available in WM and to carry out other cognitive processes, such as judging the plausibility of a sentence or selecting a response to a stimulus. A central assumption of this view is that attention-demanding cognitive processes/tasks compete with concurrent storage demands. The second conceptualisation is referred to as *controlled attention*, where a central assumption is that the process of *controlling* the allocation of attention consumes the limited resource, rather than the process of attending to an object/task per se (Shiffrin and Schneider 1977). Contrary to the first view, *controlled attention* assumes that the limited attentional resource is needed for the control of what we attend to, not for keeping representations of objects and events in WM.

These conceptualizations have different implications in regard to WM. Specifically, in a situation in which WM receives both relevant and irrelevant information, according to the *storage and processing* view, attention limits the amount of information that can be retained in WM, not the extent to which the irrelevant information is kept out of WM (i.e. the filtering efficiency or the ratio of relevant-to-irrelevant stimuli in WM). Consequently, individuals with lower WM capacity retain a smaller amount of both relevant and irrelevant information, but the filtering efficiency is independent of WM capacity. In contrast, the *controlledattention* view assumes that the limited attentional resource determines the filtering efficiency. Therefore, individuals with lower WM capacity retain the same amount of information as those with higher capacity, but different WM capacities reflect differences in the filtering efficiency.

Measures of working memory and attention in speech recognition

In speech recognition and hearing science research, WM has been predominantly measured by complex span tasks that require the participant to simultaneously process/manipulate and store information for later recall. One of the most widely used complex span tests is the *reading span test* (RST) developed by Rönnberg and colleagues (adapted from Daneman and Carpenter 1980). The participant reads sentences on a computer monitor, presented in lists of varying size, one at a time. After reading each sentence, the participant makes a semantic judgement on the sentence while concurrently trying to retain its first and/or last words. At the end of a list of sentences, the participant is prompted to recall the test items (the first and/or last words in the sentences). The load of the task is controlled by gradually increasing the number of sentences in a recall list. The score of the test (an estimation of WM capacity) is the percentage of correctly recalled target words. In terms of the theoretical framework concerning the WM and attention link, the RST (and its variants) incorporate the *storage and processing* view of this link.

The popularity of the RST in the speech and hearing literature may be attributed to the number of supporting studies that found a relationship between individual scores on this test and speech recognition performance in older listeners with hearing loss. However, there is also evidence indicating the absence of such a link (e.g. Desjardins and Doherty 2014; see Souza, Arehart, and Neher 2015 for a comprehensive review). In addition to the mixed results in the literature, broadly speaking, the dominant use of a single test of WM, such as the RST, may be constraining. From a methodological perspective, no single task can be deemed a perfect or pure measure of a cognitive construct (Conway et al. 2005). From a theoretical perspective, it means the examination of only one of the conceptualizations of the cognitive construct in question. Only a limited number of studies of older hearing-impaired listeners have used multiple tests to either derive a composite/weighted score of capacity from several similar complex span tests (Ng and Rönnberg 2020; Nagaraj 2017), or to compare the efficacy of individual complex span tests that are largely similar (e.g. different versions of the RST in Souza and Arehart 2015).

Additionally, unlike the *storage and processing* view represented by the widely used complex span tests, the *controlled attention* view has received limited attention in the speech recognition and hearing literature. A study by Meister et al. (2013) found reduced performance in older compared to young adults for speech recognition tasks in a multi-talker setting that required divided attention, and a strong relationship between performance in speech tasks requiring selective attention and working memory capacity. More recently, Gallun and colleagues developed a new measure of WM, *the Auditory Visual Divided Attention Test* (AVDAT) (Gallun and Jakien 2019). This new test was adapted from measures originally developed by Cowan and colleagues (Cowan et al. 2006), where WM is operationalised as depending on the *selective attention* system (*controlled attention*), as proposed by Cowan (1998)[1]. That is, the test combines the *storage* aspect of WM with the *control* of attention. The AVDAT involves several separate components that are categorised as single or dual modality. The two single modality components involve either auditory (lists of digits) or visual stimuli (list of letters), and can be categorised as simple span tasks of WM. The four dual modality components involve the concurrent presentation of both auditory and visual stimuli (lists of digits and letters). In all the components, the task is to store and recall a list of stimuli (auditory or visual) and the task load is controlled by gradually increasing the size of the recall list. The two types of the dual modality components differ in terms of whether the response list is *cued* (the participant knows in advance which modality list will be reported and can *selectively* attend to it while ignoring the other), or *uncued* (the participant does not know in advance which modality will be reported and has to *divide* attention between the two modalities). Gallun and Jakien (2019) examined the relationship between performance on the AVDAT and speech-on-speech recognition in a complex auditory environment that involved competing talkers in either the same (co-located) or different (separated) locations as the target speech. They found that performance on the AVDAT was correlated with speech performance, albeit, different components of the test predicted performance in different speech task conditions. Specifically, the dual modality component with the cued visual response (a *selective* attention component) was a significantpredictor of speech performance (represented as target-to-masker ratio) in the separated speech condition and of spatial release (the difference between performances in the separated and co-located conditions). The dual modality component with the uncued visual response (a *divided* attention component) was a significant predictor of speech performance in the co-located speech condition.

Given the importance of WM and attention in speech recognition, a WM test that incorporates separate measures of selective and divided attention can be a useful tool for tackling the role of individual abilities related to these constructs in speech recognition in adverse listening situations. AVDAT is a promising tool in this respect. However, further study is needed to better understand and consolidate it. In addition, the considerations above display the need for studies that implement in tandem different tests of WM that incorporate different theoretical conceptualizations of the link between WM and attention.

The current study

The present study examined the relationship between speech-in-noise recognition and WM capacity in older adults with hearing loss using a new WM test (AVDAT) and the widely used reading span test (RST). The most crucial methodological contrast regards the secondary "processing" task that requires processing of the incoming stimuli beyond merely attending to them. Namely, the RST, a complex WM span test, involves such a secondary task (semantic judgement of sentences), whereas the AVDAT does not. In the dual modality components of the AVDAT, the only task is to solely attend to either one (cued tasks) or both (uncued tasks) of the two concurrently presented stimuli lists, for recall at the end of the presentation. Table 1 displays the characteristics of the two WM tests.

Our goal was to examine the relationship between speech recognition and different WM tests that include methodological and theoretical contrasts concerning the link between WM and attention. We anticipated better speech-in-noise recognition performance to be related to higher scores on the RST, in line with existing literature (e.g. Souza and Arehart 2015). Regarding the novel AVDAT measure of working memory, given that understanding speech in noise can be assumed to engage both selective and divided attention, we hypothesised that there would be a link between the cued (selective attention) and uncued (divided attention) dual-modality components of the AVDAT and speech recognition, regardless of the test modality. More specifically, we reasoned that the ability to selectively attend to and process target speech while ignoring concurrent background noise would likely be related to the ability to selectively direct attention to the relevant information (the cued recall list) and store it for later retrieval, while ignoring the irrelevant, competing information (performance on the cued dual modality AVDAT tasks). Similarly, the ability to recognise/process and retain each incoming word in a sentence in order to retrieve it at the end of the sentence (speech recognition performance) may be related to the ability to divide attention between two concurrent sources of information and store this information for later retrieval (performance in the uncued dual modality AVDAT tasks). In regard to the relationship between the two WM measures, we may anticipate a correlation between the RST and the dual modality components of the AVDAT if both tests tap into WM to a similar extent. However, given the methodological and theoretical differences that the RST and the dual modality components of the AVDAT incorporate, they may tap into different WM and related cognitive mechanisms (including attention), in which case a weaker correlation between them could be expected. To assess other factors which might affect the relationship between WM and speech-in-noise recognition in our participant sample, we also included the measure of peripheral hearing loss.

Materials and methods

Participants

Participants included 19 adults (11 female) aged 63–89 years (mean age = 73.4 years) with symmetrical sensorineural loss (Figure 1). Nine participants wore hearing aids bilaterally. The mean pure tone average measured at 0.5, 1 and 2 kHz was 35.53 dB (range: 11.67–66.67 dB) in the right ear and 34.12 dB (range: 15–65 dB) in the left ear. The mean word recognition in quiet scores were 96% (range 66–100%) in the right ear and 95% (range 70–100%) in the left ear. All listeners passed the MoCA cognitive screening test (Nasreddine et al. 2005), scoring at least 23 out of 30 points with a group mean score of 26.6 (range 23–30). The inclusion of participants scoring 23 or higher is below the originally proposed passing score of 26, but aligns with work demonstrating good sensitivity and specificity for patients with broad sociodemographic backgrounds and/or having a hearing loss (e.g. Luis, Keegan, and Mullan 2009; Shen, Sherman, and Souza 2020). All listeners were native speakers of American English and reported normal or corrected-to-normal vision. All listeners were compensated at an hourly rate for their time.

Tests and procedure

The present data consisted of two different WM tests along with audiometric results and one speech-in-noise test. Each measure is described in detail below.

Working memory

Reading span test (RST)

The abbreviated English-language version of the Reading Span Test, developed by Rönnberg and colleagues (Ng et al. 2013) was delivered to the participants. This test involves information processing (semantic judgement) and information storage (recall). The stimuli of the test consist of short sentences that are all grammatically correct, but can be semantically plausible or implausible (e.g. "The captain saw his boat" [plausible], "The train sang a song" [implausible]). Participants were asked to read sentences on a computer screen, which appeared one word or word pair at a time. Words or word pairs were presented at a rate of 0.8 seconds per word or word pair, with an interstimulus interval of 75 ms. At the conclusion of each sentence, participants judged whether the sentence made semantic sense by replying "yes" for plausible and "no" for implausible sentences. Two lists each of 2, 3, 4 and 5 sentences were presented in ascending order of length. At the end of each list, participants were randomly queried to recall either the first or the last words from the list of sentences, in any order. The assignment of "first versus last" word recall was randomised across participants. Participants completed a practice list of 2 sentences before moving to the experimental lists. The experimenter recorded the

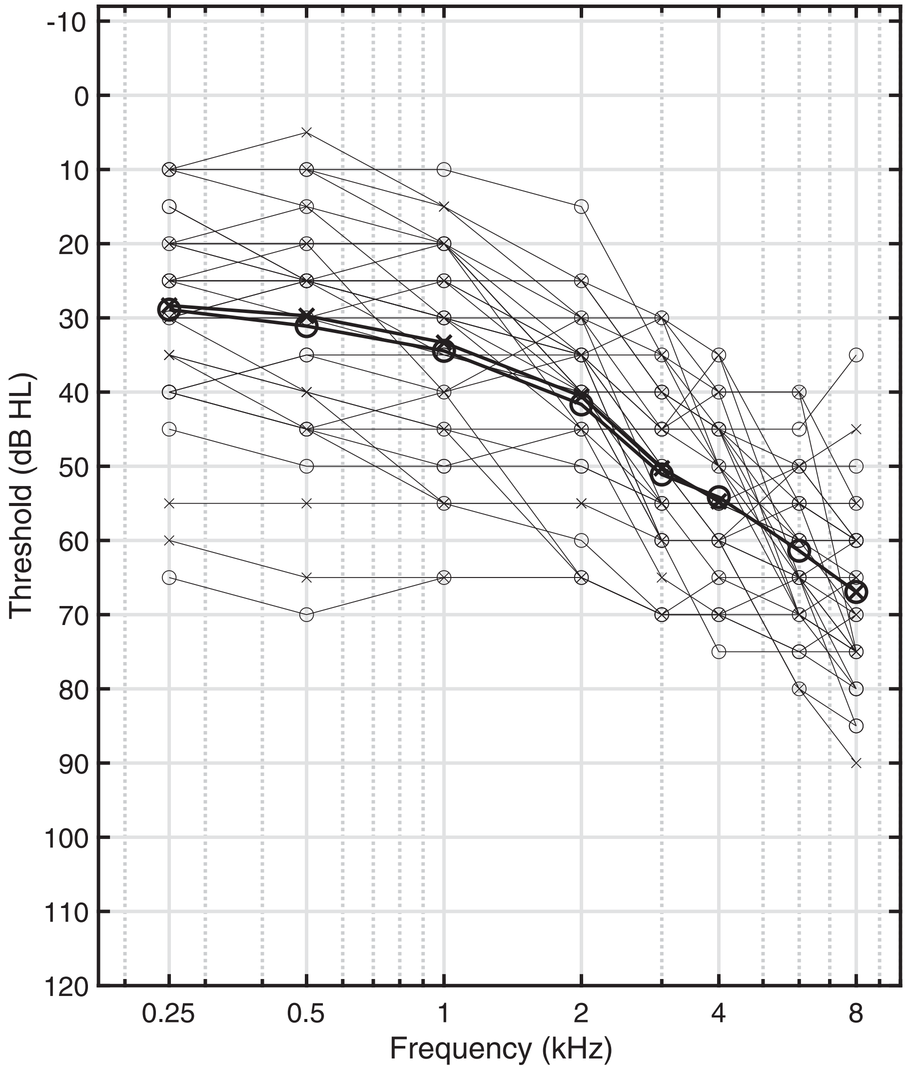

Figure 1. Audiograms of both ears for the participants ($N = 19$). Individual thresholds: thin lines, and group average: thick bold lines. Right ear is indicated by the 'o' marker and left ear by the 'x' marker.

correctly repeated words on a printed form of the visually presented test. The percentage of correctly recalled words (out of the total number of target words in the 28 total sentences) was taken as the measure of WM capacity.

Auditory visual divided attention test (AVDAT)

The materials for this test consisted of auditory (digits) and visual stimuli (letters). Participants completed single- and dual-modality span tasks where they were asked to report in the order presented a list of digits presented aurally via insert earphones and/or letters presented visually on a computer located in front of them, as described in detail below. The tasks were completed in the following order: visual letter span, followed by the auditory digit span task, followed by the dual-modality task. All tasks were implemented on a Windows computer using MATLAB R2018b software (MATLAB n.d.).

Visual letter span task. The stimuli consisted of the letters A, C, E, F, H, I, L, O, and R that were presented in 90-point font at a distance of approximately 48 cm from the participant and at a rate of one letter per second. The visual stimuli appeared in white in the centre of a black background, following an orienting stimulus (fixation cross) that appeared for 3 seconds before the trial began. Every participant received the same sequence of letter lists, beginning with three three-item practice lists and proceeding to the test that included two lists per length, starting with three-item lists and increasing by one item at a time to a maximum of nine items. The test ended if both lists of a particular length were recalled incorrectly. For each list, the stimuli that made up the standard sequence had been drawn randomly, without replacement. After the presentation of each list, the participant was asked to recall the list of letters in the order presented and enter their responses via a graphical user interface (GUI) that displayed keypads showing letters and numbers (all presented in 90-point font). Each item selected as a response appeared at the top of the screen. Participants used a mouse to select their responses. The score represented the average sequence length correctly recalled and was calculated by averaging the total number of correctly recalled items across the total number of lists presented.

Auditory digit span task. The auditory stimuli involved digitised recordings of the digits 1–9 spoken by a male talker, which were time-compressed or expanded in order to have a duration of exactly 500 ms and presented at a rate of 1 sec per item. All auditory measures were delivered without hearing aids, via ER-2 insert earphones at a root-mean-square (RMS) level of 40 dB above the Speech Reception Threshold (SRT) measured for spondaic words presented in quiet. This presentation level never exceeded 80 dB SPL. The auditory digit span task was administered and scored in the same fashion as the visual letter span task.

Dual-modality memory task. This task involved the synchronous presentation of auditory and visual lists. There were three different types of trials that were distinguished by the stimulus arrangement and task instructions. In the first type of trial (attend auditory), a cue appeared (a picture of an ear) which indicated to participants that they had to listen carefully to a digit list and ignore a synchronous list of letters that were visually presented. In the second type of trial (attend visual), a cue appeared (a picture of an eye) which indicated to participants that they had to attend to a list of letters that were visually presented and ignore a list of spoken digits presented synchronously. In the third type of trial (attend unknown) a cue appeared (a question mark) which indicated to participants that they were to attend to both the visual and auditory lists. After presentation of the lists, participants were prompted to recall one of the lists, in the order presented. Dual-modality conditions were thus divided into those in which participants knew in advance the modality that they would report and as such, could *selectively* attend to that modality and those in which participants were not informed in advance which modality they would be asked to report and as such, had to attend to and *divide* their attention between both modalities concurrently. The four conditions were presented in a randomly interleaved fashion, the trial types were randomly determined and an equal number of each cue type was presented for each task at each list length. Each list length was presented twice, starting with the 3-item lists and increasing by one item at a time to a maximum of 7-item lists. Similar to the single-modality tasks, the score was calculated by averaging the total number of correctly recalled items across the total number of lists presented.

Speech-in-noise recognition

Speech-in-noise recognition was measured using the QuickSIN (Killion et al. 2004) administered binaurally via insert (ER-3A) earphones. The test required the participant to repeat back sentences spoken by a female talker and played in four-talker babble (three males, one female) background noise. The sentences are low-context and each one includes five key words. The sentences were presented in lists of six, each one at a signal-to-noise ratio ranging from +25 dB (first sentence) to 0 dB (last sentence) in 5 dB-steps. Three lists were administered to each participant, one practice and two test lists. The first list presented served as a practice list to familiarise the participant with the task and allow for speech level adjustment. The test was recorded on a compact disc and routed through an Interacoustics AC 40 audiometer (all the stimuli were preloaded onto the audiometer). Speech presentation levels were fixed and were adjusted based on the participant's hearing loss. In line with the protocol for the QuickSIN, speech levels were set to 70 dB HL for most listeners (with a pure-tone average of 45 dB or lower[2]), and to a "loud but ok" level for listeners with a pure-tone average greater than 45 dB in either ear. The level was decreased to 65 dB HL for five participants and to 60 dB HL for one participant. The test score represents the signal-to-noise ratio required for the listener to repeat 50% of the words correctly. The final score consisted of the average of the two test lists' scores.

Results

The data were analysed in the R environment (R Core Team 2021, version 4.0.4). Participants' mean speech in noise score measured via the QuickSIN was 3.87 dB SNR loss (Standard Deviation (SD) = 1.80; range 1.5–8 dB). The mean amount of hearing loss across both ears measured by the pure tone average at 0.5, 1, and 2 kHz was 34.74 dB (SD: 12.34 dB; range: 11.67–65 dB). Table 2 displays the mean scores and corresponding standard deviations for each WM test/test component.

The mean, standard deviation and range of WM scores measured via the reading span test (RST) were consistent with previous studies of similar-aged groups (e.g. Souza and Arehart 2015). For the AVDAT scores, the mean values, standard deviation and

Table 1. Overview of the characteristics of the two cognitive tests used in the present study.

Test	Modality	Stimuli	Recall Item	Processing/Secondary Task	Attention-WM link
RST	Visual	Short, grammatical sentences varying in plausibility and presented in blocks of increasing size	First or last words (in each sentence block)	Semantic judgement on the sentence	Storage and processing
AVDAT	Auditory and visual	Digits presented aurally and letters presented visually as lists of increasing length	Sequences of digits (audio) or letters (visual) in the correct order	No processing task	Storage and control of attention

Table 2. Mean, standard deviation (SD) and the range of scores across participants ($N = 19$) in each WM test/test component.

		Single Modality		Dual Modality			
	RST (% of correctly recalled words)	Auditory (SM-A) (average digit sequence length)	Visual (SM-V) (average letter sequence length)	Auditory Cued (DM-AC) (average digit sequence length)	Visual Cued (DM-VC) (average letter sequence length)	Auditory Uncued (DM-AU) (average digit sequence length)	Visual Uncued (DM-VU) (average letter sequence length)
Mean	41.67	4.10	3.66	3.69	2.94	2.85	1.57
SD	12.04	0.43	0.43	0.68	0.94	0.65	0.70
Range	21.43–64.29	3.38–4.75	2.88–4.5	2.5–4.8	0.1–4.3	1.6–3.9	0.4–2.6

ranges for the single- and dual-modality components were relatively similar to the corresponding values in Gallun and Jakien (2019), with some slight differences that may be attributed to the differences in the age range and population samples in the two studies (only older listeners in our study compared to both young and older adults in the reference study).

Correlations

A correlation analysis was performed to assess the relationship between WM capacity measured by the reading span test (RST) and each of the components of the auditory visual divided attention test (AVDAT), speech-in-noise scores, hearing and age (results are displayed in Table 3). Normality checks, conducted on each variable via Shapiro-Wilks tests, revealed that all but one of the variables (namely, DM-VC) were normally distributed. Pearson correlation was implemented for the normally distributed variables and Spearman correlation for tests involving the non-normally distributed variable. As observed in Table 3, participants' QuickSIN scores and their performance on the RST were significantly correlated ($r = -.54$, $p = .02$), whilst no significant correlations were found between QuickSIN and performance on any of the task components of the AVDAT. Further, no significant correlations were found between the RST score and any of the scores of the *selective*, or *divided* attention components of the AVDAT, revealing a weak relationship between performances in these two WM tests.

Linear regression analysis

Speech-in-noise scores were analysed in relation to the predictors of interest which included: WM scores measured by the two tests; the amount of hearing loss represented by the average of pure-tone audiometric thresholds across 0.5, 1 and 2 kHz, averaged over both ears; and age (Age); by means of linear regression models. In the case of the AVDAT, only the four dual-modality components were included in regression analyses: the cued auditory *selective* attention task (DM-AC), the cued visual *selective* attention task (DM-VC), the auditory *divided* attention task (DM-AU), and the visual *divided* attention task (DM-VU). This was motivated by the fact that these components were the ones that tapped into divided and selective attention, while the other two components were simple digit/letter spans. Each of these four components was considered a separate predictor and entered separately in a regression model. The primary aims of the regression analyses were to assess: 1) the contribution of WM capacity in explaining additional variance in speech-in-noise recognition scores after the contribution of hearing loss (as measured by the PTA) was accounted for and 2) the individual contribution of WM capacity (as measured by the RST and the AVDAT components) in explaining variance in speech-in-noise recognition (QuickSIN scores) regardless of hearing loss (i.e. as a stand-alone cognitive factor). Accordingly, multiple regression models were implemented in an incremental fashion for the first aim and simple regression models for the second. The multiple regression models included two predictors, wherein PTA was the first predictor, followed by one of the WM scores (RST or one of the AVDAT's dual-modality components) or Age (the first column in Table 4 depicts the equations of the models that explained variance in QuickSIN scores). The WM scores from each test (RST and AVDAT) were entered in separate regression models (i.e. were never included in the same model) and in the case of the AVDAT, each component was entered in a separate model (i.e. no two or more AVDAT components were entered in the same model). All the numerical predictors were centred around their mean value before being included in the corresponding regression models. Each model was assessed for outliers after being fit with *ordinary least squares linear regression (OLS)*. In the case of influential outliers (with a high Cook's Distance) or with a large residual, *robust regression (RR)* was used for the model(s) in question to avoid any potential issues of problematic or over-estimation by the OLS model(s) in the presence of outliers (Cook, Hawkins, and Weisberg 1992; Huber and Ronchetti 2009). Residual and quantile plots of the linear models indicated that the assumptions of normality and linearity were satisfied (Hair et al. 2010).

The added effect of a predictor (improvement in the model fit) was assessed in an incremental fashion by performing likelihood ratio tests between the models with and without the predictor of interest. A predictor was included in a multiple regression model only if it contributed to a significant improvement in the fit of the model (explained additional residual variance).

As displayed in the correlation analysis in Table 3, RST and PTA were significantly correlated ($r = -0.50$, $p = .03$), however their variance inflation factors (VIFs) were < 2, indicating no serious concerns of multicollinearity that would impede their inclusion in the same linear model(s) (Hair et al. 2010). No significant correlations between PTA and any of the AVDAT

Table 3. Correlations between the variables of interest: speech-in-noise recognition (QuickSIN scores), the two different WM measures (RST and the individual components of the AVDAT), hearing (PTA), and age.

	RST	DM-AU	DM-VU	DM-AC	DM-VC	SM-A	SM-V	PTA	Age
QuickSIN	**−.54***	−.23	−.27	−.40	−.24	−.13	−.42	**.59****	.31
RST		.19	.13	.17	−.02	.33	−.13	**−.50***	**−.55***
DM-AU			.06	.33	.07	−.09	.31	−.13	−.23
DM-VU				**.47***	**.46***	.32	**.66****	−.37	−.26
DM-AC					**.47***	**.52***	**.53***	−.03	−.14
DM-VC						.20	**.53***	−.32	.12
SM-A							.32	−.31	−.08
SM-V								−.31	−.02
PTA									.27

RST: Reading span test; DM-AU: Dual modality task with the auditory stimuli as recall target, not cued; DM-VU: Dual modality task with the visual stimuli as recall target, not cued; DM-AC: Dual modality task with the auditory stimuli as recall target, cued; DM-VC: Dual modality task with the visual stimuli as recall target, cued; SM-A: Single modality task, auditory digit stimuli (average digit span); SM-V: Single modality task, visual letter stimuli (average letter span); PTA: Across-ears pure tone average measured at 0.5, 1 and 2 kHz. Spearman correlation used for tests including the non-normally distributed variable DM-VC. *$p < .05$; **$p < .01$, *p*-values not corrected for multiple comparisons.

Table 4. Model-comparison statistics for the effects of WM capacity (measured by the RST and the AVDAT) and hearing loss (represented by the PTA) on QuickSIN scores.

		Speech-in-Noise Recognition		
PREDICTORS	*Model*	χ^2/*F*	*df*	*Pr(>χ2)/Pr(>F)*
RST	M_1 (OLS)	7.01	1	.02 (*)
PTA	M_2 (RR)	9.46	1	.002 (**)
PTA + RST	M_3 (RR)	5.61	1	.02 (*)
PTA + DM-AC	M_4 (RR)	6.12	1	.01 (*)

The type of linear regression for each model is indicated in parentheses; OLS: Ordinary least squares regression; RR: Robust regression. For OLS models, the output of the model comparisons includes F, df and Pr(>F), and for the RR models the output consists of $\chi 2$, df and Pr(>$\chi 2$).
Amount of hearing loss (PTA) is the across-ears average of hearing thresholds measured at 0.5, 1 and 2 kHz. RST: Reading Span Test; DM-AC: Dual Modality Auditory Cued AVDAT component.

Table 5. Summaries of the models wherein the addition of a predictor of interest, either alone (M_1 and M_2), or after the effect of hearing loss (PTA) had been controlled for (M_3 and M_4), explained significant variance in QuickSIN scores.

	Speech-in-Noise Recognition				
Predictors					
M_1 (RST only)	*ß*	*SE*	*t*	*Pr(>|t|)*	*R^2/Adjusted R^2*
Intercept	3.87	.04	10.81	< .001 (***)	
RST	−0.08	.03	−2.65	.02 (*)	.29 / .25
M_2 (PTA only)	*ß*	*SE*	*t*	*Pr(>|t|)*	*R^2/Adjusted R^2*
Intercept	3.86	.35	11.06	< .001 (***)	
PTA	.08	.03	3.08	.007 (**)	.32 / .28
M_3 (PTA + RST)	*ß*	*SE*	*t*	*Pr(>|t|)*	*R^2/Adjusted R^2*
Intercept	3.80	.37	10.14	< .001 (***)	
PTA	.06	.05	1.22	.24	
RST	−0.05	.02	−2.37	.03 (*)	.43 / .36
M_4 (PTA + DM-AC)	*ß*	*SE*	*t*	*Pr(>|t|)*	*R^2/Adjusted R^2*
Intercept	3.86	.31	12.52	< .001 (***)	
PTA	.08	.03	3.04	.008 (**)	
DM-AC	−1.02	.41	−2.47	.02 (*)	.46 / .39

Amount of hearing loss (PTA) is the across-ears average of hearing thresholds measured at 0.5, 1 and 2 kHz. RST: Reading Span Test; DM-AC: Dual Modality Auditory Cued AVDAT component.

components were found (Table 3). Table 4 displays the output of the model comparisons and Table 5 provides the summaries of the models with the most predictive power/best fit (simple and multiple regression models).

As shown in Tables 4 and 5, participants' performance on the reading span test (M_1) and their amount of hearing loss (M_2) were significant predictors of speech-in-noise scores (QuickSIN) when entered individually in the respective simple linear regression models. In line with our expectation, there was a main effect of verbal WM capacity measured by the reading span test (RST) on listeners' sentence-in-noise scores, which is also consistent with existing literature on speech recognition in noise. As displayed by the model summary in Table 5, listeners with higher WM capacities (higher RST scores) displayed better speech-in-noise scores. Importantly, as shown by the R^2 value (M_1), RST alone accounted for 29% of the variance in speech-in-noise performance (25% adjusted variance), comparable to the amount of variance explained by PTA alone (M_2), 32% (28% adjusted variance). None of the AVDAT components were significant predictors of speech-in-noise scores when included individually in the simple linear regression models.

RST remained a significant predictor of speech-in-noise scores after the effects of hearing loss were taken into account (M_3). The addition of RST as a predictor in M_3 resulted in an additional 8% of explained variance compared to the case where only the effect of hearing loss (PTA) was included in M_2 (adjusted R^2 difference between M_3 and M_2).

In regard to the AVDAT, only one of its components - the auditory cued dual modality task (DM-AC) – was a significant predictor after the effects of hearing loss had been controlled (M_4). The inclusion of the dual-modality component with the cued auditory response (DM-AC) in M_4 led to an additional 11% of variance explained. There was no effect of age on QuickSIN scores after the effects of hearing loss were taken into account.

Discussion

Working memory, attention and speech recognition

The results of the present study revealed that performance on the RST was correlated to and predicted speech-in-noise recognition scores both before and after the effect of hearing loss was accounted for. This result was in line with our prediction and agrees with prior work that examined the relationship between WM capacity measured by the RST and speech-in-noise recognition (e.g. Souza and Arehart 2015). The linear regression analysis revealed that on its own, the RST accounted for a comparable amount of explained variance in speech-in-noise scores to that accounted for by hearing loss alone. In addition, the RST remained a significant predictor of QuickSIN scores after the effect of hearing loss was accounted for in the statistical models.

In regard to the AVDAT, we found that one of its components which taps into selective attention, the cued auditory dual-modality task, was a significant predictor of speech-in-noise scores, in combination with hearing loss. This finding was in line with our prediction regarding selective attention and speech-in-noise recognition, and suggests that the task of recognising speech in the presence of background noise relies on the ability to select in advance one source of information - the relevant/cued one – in the presence of irrelevant information competing for attention. This result extends those of Gallun and Jakien (2019), who found that the dual modality components of the AVDAT with the visual modality as the response (cued and uncued) were predictors of speech-on-speech performance. Contrary to our prediction regarding the relationship between divided attention and speech-in-noise recognition, none of the uncued dual modality components of the AVDAT that tap into divided attention were related to or predicted speech-in-noise scores. A possible explanation may rely on the different implementation of the divided/shared attention phenomenon in the two WM tests and on the extent to which the demands of the tests match with those of the speech task. In the case of the RST, the limited attention resource needs to be shared between the processing (semantic judgement of sentences) and storage (remembering the first and last words of the sentences) demands of the task. In the case of the ADVAT, there is no secondary task that requires processing of stimuli beyond merely attending to them and attention is shared between the two concurrent streams of non-word stimuli (digits and letters) that the participant needs to store for later retrieval. The speech-in-noise task requires the participant to recognise (i.e. process) the incoming words in the sentence stimuli and store them for integration with subsequent words and retrieval at the end of the sentence. As such, the sharing of attention between processing and storage of incoming speech input in the QuickSIN may display more overlap with the sharing of attention between processing and storage involved in the RST. Nevertheless, it should also be noted that while both the QuickSIN and the RST involve processing of words and sentence stimuli, the depth of processing may be different in each of them. That is, while the task demands of the RST require the participant to recognise and comprehend words in sentences in order to judge the sentence's plausibility, the task of repeating back the words in a sentence in the QuickSIN may only require the recognition of the words, without evoking a deeper level of processing to comprehend them.

Overall, our results are in line with previous work demonstrating that WM capacity is related to recognition of speech in noise by hearing-impaired older listeners (e.g. Besser et al. 2013). In regard to the link between WM and attention, our results provide support for the *storage and processing* conceptualisation of this link (shown via the RST) and more limited support for the *controlled attention* view (shown via one of the selective attention tasks of the AVDAT). While selective attention (shown via the cued auditory response dual modality component of the

AVDAT) seemed to play a role in recognising words in sentences degraded by the combination of background noise and hearing impairment, this role was limited to only one cued response modality (auditory). More research is needed to determine the role (if any) of the selective attention task of the AVDAT with the cued visual response modality in speech-in-noise recognition.

In contrast to Gallun and Jakien (2019), we found more limited evidence regarding the AVDAT and its relationship to speech recognition performance, with only one of its dual modality components being a significant predictor of speech-in-noise scores, compared to two dual modality components found to predict speech-on-speech performance in all the speech tasks of Gallun and Jakien (2019). Further, the components that were significant predictors of speech performance were different between the two studies: the cued auditory dual modality component in the present study and the cued and uncued visual dual modality components in Gallun and Jakien (2019). Methodological differences between our study and that of Gallun and Jakien (2019) may have been a factor in these differences. Specifically, our sentence stimuli were open-set, provide some semantic cues to the listener and display a certain degree of variability in their linguistic structure. The Gallun and Jakien (2019) sentence stimuli were closed-set, with an identical structure across them and with little-to-no linguistic and semantic information, which may have matched well with the cued and uncued visual dualmodality tasks of the AVDAT. Further, Gallun and Jakien (2019) measured speech-on-speech recognition in a multi-talker environment, which may evoke different phenomena from speech-in-noise recognition, such as informational masking, and consequently, employ different processing mechanisms.

To summarise, besides demonstrating the role of WM in speech recognition, our findings also indicate that a stronger match between the demands of the speech task and those of the WM test may capture the link between speech recognition and WM capacity more robustly compared to when the match between the demands of the speech task and the WM test is weaker. In comparison to the AVDAT, the combination of the cognitive demands and the type of stimuli of the RST seem to have been a better match for the cognitive demands and the type of stimuli of the speech-in-noise recognition task in the present study. Similarly, in Gallun and Jakien (2019), the demands and the type of stimuli of the speech-on-speech recognition tasks may have overlapped with the cognitive demands and the stimuli of the selective and divided attention tasks of the AVDAT to a larger extent compared to our speech recognition task and stimuli.

Which working memory test?

In the present study, the weak correlation between the two WM tests may indicate that they are not assessing the same construct to the same degree or manner. The more general and complex question of how to assess WM accurately and consistently across studies remains an open one. Our results suggest that a combination of factors that include methodological characteristics of the tests, the theoretical framework behind them and importantly, the match between the cognitive demands of the speech task and the cognitive abilities tapped by the WM tests, may govern the emergence of a relationship between WM and speech recognition. A relevant consideration for researchers studying the link between WM and speech recognition when choosing a WM test could be to decide what aspect(s) of WM and/or its relation to other cognitive abilities (such as attention) would be of interest for the speech recognition task in question. Lastly, although our sample size met the accepted range of suggested number of observations per predictor included in a regression model (Harrell 2001), a larger sample would allow for replication and expansion to a wider range of hearing loss and/or participant age.

Notes

1. It should be noted that the original measures developed by Cowan et al. (2006) include only tasks that tap into selective attention, whereas the AVDAT involves both selective and divided attention tasks.
2. Unless the tester had audibility concerns due to a sloping hearing loss.

Acknowledgements

The authors thank Patrick Zacher and Kendra Marks for their assistance during data collection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval

This study was reviewed by the Northwestern University Institutional Review Board (IRB number STU00203677).

Informed consent from participants

Each participant received verbal and written description of the study and written informed consent was obtained from participants prior to the commencement of the study.

Funding

This research was supported by NIH R01 DC006014 and NIH R01 DC012289 and by the Knowles Hearing Center.

Data availability statement

The data that support the findings of this study are available from the corresponding author, [DS], upon reasonable request.

References

Akeroyd, M. A. 2008. "Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults." *International Journal of Audiology* 47 (Supp 2), SS53–SS71. doi:10.1080/14992020802301142.

Baddeley, A. 2012. "Working memory: theories, models, and controversies." *Annual Review of Psychology* 63:1–29. doi:10.1146/annurev-psych-120710-100422.

Besser, J., T. Koelewijn, A. A. Zekveld, S. E. Kramer, and J. M. Festen. 2013. "How linguistic closure and verbal working memory relate to speech recognition in noise - A review." *Trends in Amplification* 17 (2):75–93. doi:10.1177/1084713813495459.

Conway, A. R., M. J. Kane, M. F. Bunting, D. Z. Hambrick, O. Wilhelm, and R. W. Engle. 2005. "Working memory span tasks: A methodological review and user's guide." *Psychonomic Bulletin & Review* 12 (5):769–786. doi:10.3758/BF03196772.

Cook, R. D., D. M. Hawkins, and S. Weisberg. 1992. "Comparison of model misspecification diagnostics using residuals from least mean of squares and least median of squares fits." *Journal of the American Statistical Association* 87 (418):419–424. doi:10.1080/01621459.1992.10475221.

Cowan, N. 1998. “Attention and memory: An integrated framework.” *Oxford Psychology Series, No. 26*. New York, NY: Oxford University Press.

Cowan, N. 2017. “The many faces of working memory and short-term storage.” *Psychonomic Bulletin & Review* 24 (4):1158–1170. doi:10.3758/s13423-016-1191-6.

Cowan, N., N. M. Fristoe, E. M. Elliott, R. P. Brunner, and J. S. Saults. 2006. “Scope of attention, control of attention, and intelligence in children and adults.” *Memory & Cognition* 34 (8):1754–1768. doi:10.3758/BF03195936.

Daneman, M, and P. A. 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.

Desjardins, J. L, and K. A. Doherty. 2014. “The effect of hearing aid noise reduction on listening effort in hearing-impaired adults.” *Ear & Hearing* 35 (6):600–610. doi:10.1097/AUD.0000000000000028.

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* 7:1268. doi:10.3389/fpsyg.2016.01268.

Gallun, F. J, and K. M. Jakien. 2019. “The ability to allocate attentional resources to a memory task predicts speech-on-speech masking for older listeners.” In *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany.

Hair, J. F. J., W. C. Black, B. J. Babin, and R. E. Anderson. 2010. *Multivariate Data Analysis*. Upper Saddle River: Prentice Hall.

Harrell, F. E. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Huber, P, and E. M. Ronchetti. 2009. *Robust Statistics*. 2nd ed. Hoboken, NJ: Wiley,

Kane, M. J., M. K. Bleckley, A. R. A. Conway, and R. W. Engle. 2001. “A controlled-attention view of working-memory capacity.” *Journal of Experimental Psychology: General* 130 (2):169–183. doi:10.1037/0096-3445.130.2.169.

Killion, M. C., P. A. Niquette, G. I. Gudmundsen, L. J. Revit, and S. Banerjee. 2004. “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners.” *The Journal of the Acoustical Society of America* 116 (4):2395–2405. doi:10.1121/1.1784440.

Luis, C. A., A. P. Keegan, and M. Mullan. 2009. “Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US.” *International Journal of Geriatric Psychiatry* 24 (2):197–201. doi:10.1002/gps.2101.

MATLAB n.d. *Statistics toolbox release 2018b*. Natick, MA: The MathWorks, Inc.

Meister, H., S. Schreitmüller, L. Grugel, M. Ortmann, D. Beutner, M. Walger, and I. G. Meister. 2013. “Cognitive resources related to speech recognition with a competing talker in young and older listeners.” *Neuroscience* 232:74–82. doi:10.1016/j.neuroscience.2012.12.006.

Nagaraj, N. K. 2017. “Working Memory and Speech Comprehension in Older Adults With Hearing Impairment.” *Journal of Speech, Language, and Hearing Research : JSLHR* 60 (10):2949–2964. doi:10.1044/2017_JSLHR-H-17-0022.

Nasreddine, Z. S., N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow. 2005. “The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment.” *Journal of the American Geriatrics Society* 53 (4):695–699. doi:10.1111/j.1532-5415.2005.53221.x.

Ng, E. H. N, and J. Rönnberg. 2020. “Hearing aid experience and background noise affect the robust relationship between working memory and speech recognition in noise.” *International Journal of Audiology* 59 (3):208–218. doi:10.1080/14992027.2019.1677951.

Ng, E. H., M. Rudner, T. Lunner, M. S. Pedersen, and J. Rönnberg. 2013. “Effects of noise and working memory capacity on memory processing of speech for hearing-aid users.” *International Journal of Audiology* 52 (7):433–441. doi:10.3109/14992027.2013.776181

Oberauer, K. 2019. “Working memory and attention - A conceptual analysis and review.” *Journal of Cognition* 2 (1):36. doi:10.5334/joc.58.

R Core Team 2021. *R: A language and environment for statistical computing*,” Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Rönnberg, J., T. Lunner, A. Zekveld, P. Sörqvist, H. Danielsson, B. Lyxell, Ö. Dahlström, C. Signoret, S. Stenfelt, M. K. Pichora-Fuller, et al. 2013. “The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances.” *Frontiers in Systems Neuroscience* 7:31. doi:10.3389/fnsys.2013.00031.

Schneider, W, and R. M. Shiffrin. 1977. “Controlled and automatic human information processing: I. Detection, search, and attention.” *Psychological Review* 84 (1):1–66. doi:10.1037/0033-295X.84.1.1.

Shen, J., M. Sherman, and P. E. Souza. 2020. “Test administration methods and cognitive test scores in older adults with hearing loss.” *Gerontology* 66 (1):24–32. doi:10.1159/000500777.

Shiffrin, R. M, and W. Schneider. 1977. “Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory.” *Psychological Review* 84 (2):127–190. doi:10.1037/0033-295X.84.2.127.

Souza, P, and K. H. Arehart. 2015. “Robust relationship between reading span and speech recognition in noise.” *International Journal of Audiology* 54 (10):705–713. doi:10.3109/14992027.2015.1043062

Souza, P., K. H. Arehart, and T. Neher. 2015. “Working memory and hearing aid processing: Literature findings, future directions, and clinical applications.” *Frontiers in Psychology* 6:1894. doi:10.3389/fpsyg.2015.01894.

Strori, D., A. R. Bradlow, and P. E. Souza. 2021. “Recognizing foreign-accented speech of varying intelligibility and linguistic complexity: Insights from older listeners with or without hearing loss.” *International Journal of Audiology* 60 (2):140–150. doi:10.1080/14992027.2020.1814431

Wickens, C. D. 1980. “*The structure of attentional resources*.” In *Attention & Performance*, Vol. VIII, edited by R. S. Nickerson, 239–257. Hillsdale, NJ: Erlbaum.

Zekveld, A. A., M. Rudner, I. S. Johnsrude, and J. Rönnberg. 2013. “The effects of working memory capacity and semantic cues on the intelligibility of speech in noise.” *The Journal of the Acoustical Society of America* 134 (3):2225–2234. doi:10.1121/1.4817926.