Identifying Links Between Latent Memory and Speech Recognition Factors

Adam K. Bosen[1], **Gianna M. Doria**[1]

[1]Boys Town National Research Hospital, Omaha, NE

Abstract

Objectives: The link between memory ability and speech recognition accuracy is often examined by correlating summary measures of performance across various tasks, but interpretation of such correlations critically depends on assumptions about how these measures map onto underlying factors of interest. The current work presents an alternative approach, wherein latent factor models are fit to trial-level data from multiple tasks to directly test hypotheses about the underlying structure of memory and the extent to which latent memory factors are associated with individual differences in speech recognition accuracy. Latent factor models with different numbers of factors were fit to the data and compared to one another to select the structures which best explained vocoded sentence recognition in a two-talker masker across a range of target-to-masker ratios, performance on three memory tasks, and the link between sentence recognition and memory.

Design: Young adults with normal hearing (N = 52 for the memory tasks, of which 21 participants also completed the sentence recognition task) completed three memory tasks and one sentence recognition task: reading span, auditory digit span, visual free recall of words, and recognition of 16-channel vocoded Perceptually Robust English Sentence Test Open-set (PRESTO) sentences in the presence of a two-talker masker at target-to-masker ratios between +10 and 0 dB. Correlations between summary measures of memory task performance and sentence recognition accuracy were calculated for comparison to prior work, and latent factor models were fit to trial-level data and compared against one another to identify the number of latent factors which best explains the data. Models with one or two latent factors were fit to the sentence recognition data and models with one, two, or three latent factors were fit to the memory task data. Based on findings with these models, full models that linked one speech factor to one, two, or three memory factors were fit to the full data set. Models were compared via Expected Log pointwise Predictive Density (ELPD) and post-hoc inspection of model parameters.

Results: Summary measures were positively correlated across memory tasks and sentence recognition. Latent factor models revealed that sentence recognition accuracy was best explained by a single factor that varied across participants. Memory task performance was best explained by two latent factors, of which one was generally associated with performance on all three tasks and the other was specific to digit span recall accuracy at lists of 6 digits or more. When these models were combined, the general memory factor was closely related to the sentence recognition factor, whereas the factor specific to digit span had no apparent association with sentence recognition.

Conclusions: Comparison of latent factor models enables testing hypotheses about the underlying structure linking cognition and speech recognition. This approach showed that multiple memory tasks assess a common latent factor that is related to individual differences in sentence recognition, although performance on some tasks was associated with multiple factors. Thus, while these tasks provide some convergent assessment of common latent factors, caution is needed when interpreting what they tell us about speech recognition.

INTRODUCTION

Speech recognition accuracy in degraded listening conditions partially depends on the listener's cognitive ability (Arlinger et al., 2009), such that measures of individual differences in working memory ability tend to correlate with speech recognition accuracy (Akeroyd, 2008). However, measures of various cognitive constructs tend to positively correlate with one another (Miyake et al., 2000). As a result, most measures of cognitive ability have been found to correlate weakly with speech recognition accuracy to some extent (Dryden et al., 2017). These correlations are difficult to interpret because they could arise from a weak direct link between the measured cognitive abilities and speech recognition, or alternatively they could arise indirectly from covariance between the measured cognitive abilities and other cognitive abilities that directly support speech recognition. Discovering such correlations is an important initial step in the development of theory, but studies in this field are often conducted with the goal of identifying latent factors, i.e. the underlying individual differences that affect performance on multiple tasks of interest. The work presented here demonstrates a method of moving beyond examining correlations by instead identifying and comparing latent factor models to explain the individual differences that link performance across cognitive and speech recognition tasks.

In the psychology literature, individual performance on specific tasks is typically modeled as arising from an underlying set of latent factors (Engle et al., 1999; Miyake et al., 2000). While the exact nature and relationship of those latent factors to one another is an area of active research (Friedman & Miyake, 2017; Kovacs & Conway, 2016; Rey-Mermet et al., 2018; Shipstead et al., 2016; Troche et al., 2021), the norm is to explicitly state and test assumptions about how task performance maps onto latent variables. This approach is advantageous in that it can identify latent factors which contribute to performance across multiple tasks, estimate individual differences in those factors, and identify tasks for which performance depends on multiple latent factors. In contrast, not using this approach makes the implicit assumption that the observed task performance directly reflects the latent variable of interest (Oberauer & Lewandowsky, 2019). Such assumptions can be problematic, as it is not always evident at face value whether two tasks measure the same

latent factor. For example, although both the Stroop and flanker tasks have long been regarded as tests of inhibition, recent work indicates that they are unrelated measures which do not measure a common latent variable (Rey-Mermet et al., 2018; Rouder & Haaf, 2019). Similarly, the antisaccade task, a measure of attentional control that requires looking away from a cued location, and complex span working memory tasks do not obviously relate to one another, yet individual differences in performance on complex span tasks predict speed and accuracy on the antisaccade task (Kane et al., 2001; Unsworth et al., 2004). Thus, there is a need to test hypotheses about what latent factors any given task measures to advance our understanding of the cognitive processes that affect speech recognition.

An influential model of the link between working memory and speech recognition is the Ease of Language Understanding (ELU) model (Rönnberg, 2003; Rönnberg et al., 2008, 2013, 2022). The premise of the ELU model is that speech recognition occurs via two processes. The first process occurs automatically and rapidly whenever sensory inputs unambiguously match a lexical representation in long-term memory, and the second process is an explicit and slower mechanism for restoring or inferring an interpretation of ambiguous input. This slow, explicit mechanism is conceptualized as working memory.

There is a wealth of empirical findings that have motivated the development of several theories regarding the structure of working memory (Oberauer et al., 2018). The ELU model is based on one of these models of working memory, Baddeley's multicomponent model (Baddeley, 2012). In this model, working memory is defined as the ability to concurrently store and process information. Working memory is distinct from short-term memory, which is the temporary storage of information without processing. The ELU model emphasizes working memory as a critical ability for speech recognition because tasks that measure working memory tend to predict speech recognition accuracy, whereas tasks that measure short-term memory do not (Rönnberg et al., 2013). From this perspective, individuals differ from one another in terms of their short-term capacity and their ability to process information in memory, and it is specifically the processing ability that accounts for correlations between working memory tasks and speech recognition accuracy. Working memory is often experimentally tested with complex span tasks, which include a primary memory task interleaved with a secondary processing task. One of the most widely used complex span tasks is reading span (Daneman & Carpenter, 1980). In reading span tasks, participants read a sequence of sentences one at a time and make a judgment about each sentence, such as whether it is factually true or semantically sensible. At the end of the sequence, they then recall the last word of each sentence. Interleaving both a storage task (storing and recalling words in order) and a processing task (judging sentence truth) is what critically distinguishes reading span from simple span tasks that only include a primary recall task, such as serial recall of lists of digits in the order they were presented. This processing requirement is the reason complex span tasks are hypothesized to relate to speech recognition while simple span tasks are not (Rönnberg et al., 2016, 2021).

However, alternative theories about the structure of working memory have been proposed. For example, some models of working memory posit the presence of a limited capacity focus of attention that is used to maintain information in a readily accessible state. When the capacity of the focus of attention is overloaded, information is displaced into an activated

state in long-term memory and must be retrieved back into the focus of attention before it can be used (Cowan, 2001; Oberauer, 2002). In these models, individuals differ from one another in both their ability to maintain information in the focus of attention and retrieve activated information from long-term memory back into the focus of attention (Unsworth & Engle, 2007b). From this perspective, maintenance and retrieval demands vary across memory tasks, although simple and complex span tasks both measure these processes to some extent (Unsworth & Engle, 2007a). Specifically, the dual-task structure of complex span tasks requires retrieval of activated items in long-term memory regardless of list length because the secondary task displaces those items from the focus of attention. In contrast, simple span requires retrieval only at longer list lengths, in which the amount of information in the list overloads an individual's ability to maintain that information in the focus of attention. In support of this idea, previous research has reported correlations between the number of items recalled at longer list lengths (5 – 7 item lists) in serial recall tasks with recall accuracy for complex span at all list lengths (Unsworth & Engle, 2006). Similarly, free recall tasks, which present long lists of items and require participants to recall as many items as they can in any order, load onto the same latent factor as complex span tasks (Unsworth & Engle, 2007b), which supports the theory that individual differences in retrieval ability affect performance in both tasks. More generally, Wilhelm et al. (2013) demonstrated that performance across a variety of memory tasks was highly correlated and loaded onto the same underlying factor, which was in turn closely related to fluid intelligence.

The storage and processing model (which the ELU model is based on) and the maintenance and retrieval model make different predictions about which memory tasks will overlap in the latent factors they measure. The storage and processing model predicts that all tasks which depend on storing information will load onto a common storage factor, and tasks with an explicit processing component will also depend on a distinct processing factor. Thus, if we use a complex span task, which includes a processing component, in conjunction with serial or free recall tasks, which do not include a processing component, we should be able to partition individual differences in storage and processing ability from one another. Specifically, all tasks should associate with a common storage factor, while complex span tasks will also depend on individual differences in processing ability. In contrast, the maintenance and retrieval model predicts that complex span tasks at all list lengths and serial or free recall tasks with long list lengths that overload the focus of attention will depend on retrieval ability, whereas short list lengths in serial or free recall tasks should only depend on maintenance ability. Thus, including tasks for which these models make different predictions will enable us to identify the underlying latent factor structure. Identifying latent factors and estimating individual differences in ability for each factor will enable us to identify which latent memory factor(s) determine individual differences in speech recognition accuracy.

For this study, three memory tasks were selected to determine how these tasks overlap in the latent memory factor(s) they measure and determine the relationship between sentence recognition and memory factors. Reading span was included for its theoretical importance in the ELU model and because it often associates with speech recognition in older adults (O'Neill et al., 2019; O'Neill, Parke, et al., 2021). Reading span requires concurrent storage and processing of information within a trial, as described above. Forward digit span was included because performance on this task has been previously found to correlate with

vocoded sentence recognition accuracy (Bosen & Barry, 2020), which is part of the listening condition used in the current study. This correlation indicates that digit span and sentence recognition abilities are related to some extent, but do not tell us whether digit span assesses the same underlying speech-related factors as reading span. Forward digit span is a serial recall task, which requires storing a sequence of digits in order but does not require any manipulation or processing of that sequence. Digit sequence length can be manipulated to control the amount of information that needs to be retrieved from activated long-term memory during recall. Visual free recall of words was included because it has been shown to associate with complex span task performance (Unsworth & Engle, 2007b) but does not include an explicit processing component. In a typical free recall task, more items are presented than can be stored in short-term memory to place a demand on retrieval ability. To our knowledge, free recall has not been previously examined with respect to speech recognition. Together, these memory tasks enable the identification of latent memory factors that are associated with sentence recognition.

The ELU model states that working memory is needed to infer missing speech information whenever the input is ambiguous (Rönnberg et al., 2013). Specifically, when few ambiguous inputs are present, individual differences in speech recognition should be governed by the automatic, rapid process. As more ambiguous inputs are introduced, i.e. as listening conditions become more difficult, then individual differences in recognition accuracy should be dominated by the explicit, slow process, with the automatic process playing a smaller role. By extension, the ELU model predicts that individual differences in working memory factor(s) should have stronger effects on speech recognition accuracy in more difficult listening conditions. Currently, it is unknown how the transition from automatic to explicit processing occurs across changes in listening condition. If the transition is sharp, then there should be a threshold below which working memory factor(s) should consistently associate with speech recognition accuracy. Alternatively, if the transition is more gradual, then the relationship between working memory and speech recognition accuracy should change as a function of listening condition difficulty. It could be the case that a single working memory factory becomes more strongly associated with recognition accuracy as difficulty increases. Alternatively, it could be that factors associated with automatic and explicit processing trade off in their relationship with speech recognition accuracy. If so, individual differences in the factors that support automatic processing would be most strongly associated with speech recognition in easy listening conditions and factors that support explicit processing would be most associated with difficult listening conditions.

Some evidence against a change in latent factors as a function of listening condition difficulty was found by Bosen & Barry (2020), in which individual differences in sentence recognition accuracy were moderately correlated across three listening conditions that were degraded with a vocoder to produce different levels of spectral resolution. At a group level, keyword recognition accuracy declined as the number of channels in the vocoder was reduced, from 88.1% for a 16-channel vocoder to 77.1% for an 8-channel vocoder and 39.1% for a 4-channel vocoder. However, individual differences in keyword recognition accuracy across the 16 and 4 channel conditions were correlated with an r = 0.66, which is only slightly smaller than the correlations between the 16 and 8 channel (r = 0.75) and the 8 and 4 channel conditions (r = 0.73). If the latent factors which supported sentence

recognition changed as a function of accuracy within the tested range then we would expect individual differences in sentence recognition to be weakly correlated, if at all, across the 16 and 4 channel conditions. However, we noted above that interpreting correlations in performance across tasks can be challenging, so the present study was designed to identify the latent factors underlying sentence recognition across a manipulation of recognition accuracy.

For this study, we measured sentence recognition accuracy in young adults with normal hearing for sentences spoken by multiple talkers which were mixed with competing two-talker speech and then vocoded. This listening condition was selected to elicit a high cognitive demand for successful sentence recognition. Changes in talker elicit a processing cost as the listener normalizes to characteristics of the new talker (Magnuson et al., 2021; Mullennix et al., 1989; Mullennix & Pisoni, 1990). Listening to speech in the presence of competing speech requires selecting and attending to the target talker while ignoring competitors, with the greatest interference from the competing talkers occurring with a two-talker masker (Freyman et al., 2004). Vocoding the combined target and masking speech further increases the cognitive demand because it removes some of the cues that support stream segregation (Bernstein et al., 2016; Brungart, 2001; Qin & Oxenham, 2003; Shinn-Cunningham & Best, 2008). The inclusion of these elements in the listening condition used here should make it more likely that we would observe changes in the latent factors that support speech recognition across manipulations of recognition accuracy, if such changes exist. Using a degraded listening condition with a high cognitive demand should also facilitate the identification of the relationship between memory and speech recognition. As a secondary motivation, this listening condition is translationally relevant because identifying indexical properties and recognizing speech in the presence of other talkers is particularly difficult for individuals with cochlear implants (Smith et al., 2019; Stickney et al., 2004) and is a listening situation these individuals frequently encounter in daily life (O'Neill, Basile, et al., 2021). Thus, the results of this experiment could inform the design of future studies to identify the cognitive constructs that facilitate speech recognition in listeners with cochlear implants.

Sentence recognition accuracy was tested across a range of target-to-masker ratios. At favorable target-to-masker ratios, we predicted that sentence recognition accuracy would not be substantially impeded by the competing speech, so the correlation between sentence recognition accuracy and performance on the memory tasks should replicate the findings of prior work which used vocoded speech without competing speech (Bosen & Barry, 2020). At more difficult target-to-masker ratios the ELU model predicts that there is a greater requirement for slow, explicit processing, which may manifest as increased effect of specific memory factor(s) on recognition accuracy in the difficult conditions.

MATERIALS AND METHODS

Young adults with normal hearing completed three memory tasks and repeated vocoded sentences in the presence of a two-talker masker at a range of target-to-masker ratios. Performance in the sentence recognition and memory tasks was compared across individuals, first by correlating average performance in each task across participants and

second by fitting latent factor models to trial-level data. Task design documentation, data, and analysis scripts are available as supplemental material at https://osf.io/j2s45/.

Participants and Experimental Environment

A total of 52 people participated in this study, 27 in the lab and 25 via remote testing. 27 young adults (21 women, 6 men, range of 19 – 34 years of age, mean of 25.2 years) were recruited for this study in our lab. These participants were screened for typical hearing (pure-tone thresholds < 20 dB HL at octave frequencies between 0.5 and 8 kHz) and did not report any developmental, intellectual, or neurological disorders that would interfere with any of the tests used here. For these participants, the tasks described below were implemented in MATLAB (Mathworks, Natick, MA, USA) and the Psychtoolbox-3 library (Kleiner et al., 2007) was used for visual presentation in the free recall task. Tasks were completed in an echo-attenuated sound booth. Visual stimuli were presented on a computer monitor located in the booth and auditory stimuli were presented from a loudspeaker, both of which were located side-by-side approximately 1 m straight ahead of the participant. Auditory stimuli were presented at an average level of 65 dB SPL. Of these 27 participants, 21 completed all tasks described below, while 6 completed the memory tasks described here and a pilot sentence recognition task which used a different approach to vocoding and mixing the target and masker speech than the one described below. Given the small in-lab sample size of the pilot sentence recognition task, those results are not reported here.

In-person data collection was discontinued at the onset of the COVID-19 pandemic. We originally intended to collect more data in the pilot sentence recognition task by recruiting participants to complete the study in a remote testing format, although concerns about differences in audio equipment and auditory fidelity across participants prevent us from reporting their sentence recognition results until we can replicate our findings under more controlled listening conditions. However, prior results indicate that assessment of auditory digit span is robust to degradations in auditory quality (Bosen & Barry, 2020; Bosen & Luckasen, 2019) and the other memory tasks were visually presented, so data collected remotely for these tasks should be comparable to those collected in the laboratory. Thus, performance on the memory tasks for these remote participants is included to determine if in-lab and remote testing yield a similar range of group performance as a reference for future studies and to increase the sample size for analyses which compare individual differences in performance across memory tasks.

25 participants (20 women, 5 men, range of 19 – 26 years of age, mean of 22 years) completed the study via remote testing. Remote testing was conducted via a WebEx video call with an experimenter, who provided instructions and links to websites which implemented the tasks. Participants were asked to wear headphones throughout the study and were free to use any headphones they had available. At the start of the experiment, participants were sent a link to a calibration noise which matched the long-term spectral characteristics of the digit stimuli and were asked to adjust the volume of their headphones so that the calibration noise was at a "loud, but comfortable" listening level. These individuals completed the digit span and free recall tasks described here using an implementation of the tasks written using the jsPsych library (de Leeuw, 2015) and hosted

on the website https://www.cognition.run/. Reading span was administered through Inquisit 6 Online (Millisecond Software, 2020), which requires participants to download a driver program to their computer to ensure reliable timing for stimulus presentation and responses.

All participants provided informed written consent and were compensated hourly for participation. This study was approved by our Institutional Review Board.

Memory Tasks

All participants completed reading span, free recall, and digit span in that order.

Reading Span

Participants completed an automated reading span task on a computer (Unsworth et al., 2005) implemented in Inquisit Lab 5 (Millisecond Software, 2019), available at https://www.millisecond.com/download/library/rspan/ at the time of publication. The version of the reading span task used here is slightly different than the one developed by Daneman & Carpenter (1980), in that the items to be remembered are distinct letters presented after each sentence rather than the last word of the sentence. This version was used to prevent participants from using semantic information from sentences to facilitate recall (Conway et al., 2005). In this task, participants were shown alternating sequences of letters and sentences between 3 and 7 letter-sentence pairs long, with each length presented three times in a random order. Participants were asked to judge whether or not the sentences were sensical and to remember the sequence of letters in order. Nonsensical sentences were created by taking a semantically sensible sentence and changing one word (e.g. "Andy was stopped by the policeman because he crossed the yellow heaven."). After each sentence a prompt appeared stating "This sentence makes sense", to which participants clicked on true or false dialog boxes with a mouse. At the end of each sequence, a keypad appeared on screen containing all 12 possible letters and participants were asked to click on the letters in the order they were presented using a mouse. The task instructions encouraged participants to maintain at least 85% accuracy on sentence judgment and percent accuracy for sentence judgment was displayed on screen to ensure participants attended both the judgment and the recall portions of the task. Practice trials of letter recall, sentence judgment, and a mixture of both were provided prior to completing the experimental trials. The time to respond in the practice sentence judgment trials was used to calculate a timeout period, defined as the mean time to respond plus 2.5 standard deviations for each participant. In experimental trials, sentence judgment responses timed out and were marked as incorrect if a response was not provided within the timeout period. This timeout period was intended to ensure participants could not chose to pause and rehearse the letter sequence after making sentence judgments.

Free Recall

The structure of this task was previously described by (Engle et al., 1999). Participants saw 10 lists of 12 words flashed on a computer screen. Each word was presented for 750 ms with a 250 ms blank screen between words. At the end of each list, participants said aloud as many words as they could remember in any order within a 30 s time limit. Participants were told it is easier to recall words from the end of the list first before recalling any other words they could remember to be consistent with the methods described by Engle et al.

(1999), although they were free to use whatever response strategy they preferred. Verbal responses were recorded and transcribed offline by the second author. Unclear responses were reviewed by both authors and consensus was reached.

We designed a novel word list for this study. A list of potential words was taken from the consonant-nucleus-consonant word lists provided by Storkel (2013). Three-phoneme, four-letter words were then selected to form two practice lists and 10 test lists. Lexical frequency for each potential word was obtained from Brysbaert & New (2009), and high frequency words were preferred because they tend to be easier to recall in other types of memory tasks (Roodenrys & Quinlan, 2000). Words with a high (>30) or low (<10) number of phonological neighbors were excluded because lexical neighborhood density can also affect recall (Roodenrys et al., 2002). Words were compiled into lists that were approximately balanced with respect to the number of neighbors, the biphone segment sum, and lexical frequency. Meaningful phrases, rhyming words and semantically related words were avoided within lists. A spreadsheet of words used and their lexical statistics is provided in the OSF repository associated with this manuscript.

Digit Span

Participants repeated spoken lists of between 2 and 9 digits, with each list length presented 5 times for a total of 40 trials. Prerecorded digits were spoken by a single female talker at a rate of one digit per second. All participants heard the same 40 lists in the same order. List length was randomized and was not known at the onset of each list. This task was previously described by Bosen and Barry (2020), with the exception that in the prior study digits were vocoded and participants heard 10 lists of each length. We conducted a sensitivity analysis to estimate the distribution of measurement error as a function of the number of lists tested at each length and found that we only needed 5 lists per length to obtain similar levels of precision as the 10-list version used in the prior study. Task reliability for a version which uses 5 lists per length was then estimated by randomly splitting the 10-list data from the prior study into two sets of 5 lists per length, summing the number of digits correct recalled by each subject in each set, and correlating sum digits recalled across sets (Parsons et al., 2019). This process was repeated 5000 times to estimate a range of reliabilities, which yielded an average reliability for the task with 5 lists per length of $r = 0.87$, with a 95% confidence interval ranging from 0.81 to 0.93. Thus, halving the length of the digit span task relative to the prior study provides a shorter task with good reliability. The scripts used to conduct these analyses are provided in the OSF repository associated with this manuscript.

Sentence Recognition Task

Following the memory tasks, participants heard and repeated Perceptually Robust English Sentence Test Open-set (PRESTO) sentences (Gilbert et al., 2013; Tamati et al., 2013). These sentences vary in talker gender, talker dialect, syntactic structure, and semantic contents, which limits the ability to strategically use these cues to facilitate sentence recognition. These sentences were mixed with a two-talker masker across a range of target-to-masker ratios and then vocoded.

The two-talker masker was generated by concatenating PRESTO sentences that were not used as target sentences, splitting the concatenated audio into two halves, and summing those halves. Silence was automatically trimmed from the onset and offset of each sentence to ensure a maximum gap between sentence offset and onset of 100 ms.

Target sentences were taken from PRESTO lists 7, 8, 13, 15, 17, and 23, as these lists have good equivalency and consistency (Faulkner et al., 2015). Each list contained 76 keywords which were unevenly distributed across 18 sentences. For each target sentence, a random sample of the masker was selected that was 1 s longer than the target sentence and aligned so that the masker started and ended 0.5 s before and after the target sentence. This delay between masker and target onset and offset helped participants identify the target speech. The gain of the masker was adjusted such that the long-term average target-to-masker ratio across sentences was 0 (list 7), +2 (list 8), +4 (list 13), +6 (list 15), +8 (list 17), and +10 dB (list 23) to produce a broad range of keyword recognition accuracy. Based on previous work with IEEE sentences in steady-state noise (O'Neill et al., 2019), the masker was expected to have little effect on sentence recognition at +10 dB target-to-masker ratio, with a substantial reduction of accuracy to a low, but non-zero, level at 0 dB.

Stimuli were presented such that the level of the target sentences in quiet was fixed at an average of 65 dB SPL, so overall stimulus level varied with target-to-masker ratio. Target-masker pairings were the same across all participants to avoid variability in target-masker interactions (Buss et al., 2021). Sentences were also presented in a fixed order, starting with the highest target-to masker ratio (+10 dB for list 23) and stepping down to the most difficult (0 dB for list 7).

Target-masker pairs were vocoded with a 16 channel, sine-wave carrier vocoder, selected to match the 16-channel vocoder used by Bosen & Barry (2020). A 16 channel vocoder is expected to yield high keyword recognition accuracy at the higher target-to-masker ratios used here. Stimuli were passed through rectangular filters with edge frequencies that were equally spaced on the Greenwood (1990) function (filter edges of 100, 158, 230, 319, 430, 566, 735, 944, 1202, 1522, 1917, 2406, 3011, 3760, 4686, 5832, and 7250 Hz). The envelope of each filter's output was calculated with a Hilbert transform and then low-pass filtered with a 300 Hz fourth-order Butterworth filter. Low-pass envelopes were then multiplied with a sine-wave carrier with a frequency of the geometric mean of the filter edges, and the products were summed to produce the vocoded stimulus.

No training was provided for the sentence recognition task. Verbal responses were recorded to every sentence for each participant and scored offline for number of keywords correct by the second author. Scoring followed the rules provided with the PRESTO sentences (Gilbert et al., 2013), which required exact matches for morphological markers. Unclear responses were reviewed and scored by both authors and consensus was reached.

Data Analysis

The data were analyzed in two ways. First, individual performance on each task was estimated by calculating task-level summary statistics. We report descriptive statistics for the range and mean performance on each task to facilitate comparison of participants in the

current study to prior work. Pairwise linear correlations were calculated between each task-level statistic and tested for significance to demonstrate that our results replicate previously discovered positive correlations between memory task performance and sentence recognition accuracy. Second, latent factor models were fit to trial-level task performance and compared against one another to identify the number of latent factors which best explains the data.

Task-Level Summary Statistics

For reading span and digit span, responses were scored by the number of edits required to transform the given response into the target list (Gonthier, 2022), such that each edit (an insertion, deletion, or swap of a letter or digit) reduced the score for that trial by one, down to a minimum of zero. This scoring method differs from the common approach of scoring based on the total number of letters recalled in the correct position across all lists (Conway et al., 2005; Friedman & Miyake, 2005). The advantage of scoring by number of edits is that it yields more intuitive scores that are not sensitive to error position. For example, if the target sequence is '12345', then responses of '2345' and '1234' would both receive a score of 4 of 5 items correct when scoring by number of edits. Scoring by correct in position would instead assign scores of 0 of 5 and 4 of 5 items correct, respectively, despite the fact that the only error in either response is a single deletion. Preliminary analysis with the latent factor models described below suffered from the inclusion of trials which were assigned a score of 0 by the correct in position scoring method, as these trials tended to have very high leverage over the model fit. Scoring by number of edits yielded similar task-level individual difference scores relative to the group average but made the distribution of trial-level scores more consistent within participants. We also scored the data by correct in position for comparison to prior work, but used edit distance scoring for all other analyses. For both scoring methods, reading span had a maximum possible cumulative score of 75 letters, and digit span had a maximum possible cumulative score of 220 digits.

For free recall, any responses that matched a word shown in the list were scored as correct, and the average number of words recalled across all 10 lists was calculated for each participant. Repeated responses were not counted more than once.

For sentence recognition, the proportion of keywords correctly identified by each participant for each target-to-masker ratio was calculated to quantify sentence recognition accuracy. Pairwise correlations of task performance were calculated across all target-to-masker ratios and the memory tasks.

Latent Model Description

Key aspects of the latent model approach are described here, with additional technical details provided in supplemental digital content 1. Instead of aggregating data across trials in each task, trial level data were coded as the number of correct responses out of the total number of items in that trial for every participant and task. These data can be modeled as samples from a binomial distribution, such that for a given trial i, task t, and participant p, the number of items correct, X_{itp}, is binomially distributed as a function of the total number of items in the trial, n_i, and the probability of a correct response for that participant and task, a_{tp} (expressed in log odds). The total number of items in the trial is dependent on the task

(either number of keywords or number of items to be remembered) and can vary across trials within a task. Incorporating the variable number of keywords in each PRESTO sentence into the model reflects the variable amount of information obtained in each trial. Each list length in the reading span and digit span tasks is modeled as a separate task to allow the probability of a correct response to vary across list lengths.

$$X_{itp} \sim Binomial(n_i, \alpha_{tp})$$

The probability of a correct response is the sum of the group-level average probability of a correct response for each task, μ_t, and individual differences in latent factor(s), η_{fp}, scaled by the amount each factor contribution to performance on each task, λ_{tf}.

$$\alpha_{tp} = \mu_t + \sum_{\forall f} \lambda_{tf} \cdot \eta_{fp}$$

This model allows each latent factor to affect performance on each task and enables us to estimate the likelihood that the contribution of each factor to each task, λ_{tf}, is non-zero, as described below. In models with more than one factor, the factors were constrained to be orthogonal to one another. The orthogonality constraint was imposed to prevent the model from "collapsing", wherein two different latent factors might take the same values during model estimation and thereby reduce the effective number of latent factors in the model. Despite this constraint, it is still possible for models with an excess of latent factors to collapse, as was observed for a few models described below. An alternative model formulation is to allow each latent factor to affect performance on a single task, but allow latent factors to be correlated with one another (see Friedman et al., 2008 for an example of both model structures applied to the same data set). Here we opted to allow factors to affect multiple tasks to examine the convergent and divergent loading of tasks onto distinct latent factors. To ensure the model had a fixed scale and was therefore identifiable, each latent factor η was restricted such that the mean across participants was zero and the standard deviation was 1 in each sample drawn by the fitting algorithm described below. To ensure that factors could not change their ordering (i.e. higher values lead to worse task performance), all λ parameters were constrained to be positive.

To examine relationships between speech and memory latent factors, individual differences in speech recognition factors, η_{speech}, were modeled as a linear regression against memory factors with regression coefficients β_f and normally distributed error terms, σ_{speech}, which were constrained to be positive.

$$\eta_{speech,\, p} \sim Normal\left(\sum_{\forall f} \beta_f \eta_{fp},\, \sigma_{speech}\right)$$

Latent Factor Model Comparison

Models that varied in the number of latent factors, η, were fit to three portions of the data to determine the number of latent factors which best accounted for task performance. For each portion of the data, we started with a model with one latent factor, then iteratively increased the number of latent factors in the model until the model with the highest number of factors demonstrated poor convergence and high correlations between latent factors. For the sentence recognition data alone, this iterative approach yielded a comparison between models with one or two latent factors. For the memory task data alone, models with one, two, and three latent factors were compared. For the entire dataset, models with one, two, or three latent memory factors and one sentence recognition factor were compared, based on the results obtained for the two separate portions of the data.

Models were fit using the Stan programming language (v2.26.1, Carpenter et al., 2017) using the RStan interface (v2.26.11, Stan Development Team 2020) in R (v4.2.0, R Development Team, 2022). Stan code describing each model and R code used to fit and analyze models are provided in the OSF repository associated with this manuscript. Stan estimates the posterior distribution of model parameters via Markov Chain Monte Carlo sampling. This approach to modeling allows us to examine the estimated posterior probability distribution of each model parameter given the data and thereby use Bayesian inference to test hypotheses. An overview of the workflow involved in Bayesian inference is described by Gelman et al., (2020). We can use the probability distributions generated by Stan to estimate values of interest, such as the Maximum A Posteriori value of a parameter (MAP, i.e. the most likely value or mode of the parameter), and to calculate the likelihood that the value of a parameter lies outside of a null hypothesis, which in our case is the likelihood that a model parameter lies outside of a negligible region around zero.

The posterior distributions generated by model fitting can be used to estimate the likelihood that the experimental data could have arisen from the given model, which can in turn be used to select the model which best explains the data from among a candidate set. Pareto smoothed importance-sampling leave-one-out cross-validation (Vehtari et al., 2017) was used to estimate the goodness of fit of each model, quantified as the expected log pointwise predictive density (ELPD). ELPD includes a penalty term for the number of effective parameters to penalize more complex models which do not have a better explanatory power than simpler models. When comparing models, both the ELPD difference between models and the standard error of that difference are calculated. If the ELPD difference between models is greater than 4 and the standard error is smaller than the difference then we select the model with the better fit as a substantially better explanation for the data (Sivula et al., 2020).

Model Parameter Significance

In addition to examining the likelihood of whole models, we can also examine the likelihood that parameters within those models are non-negligible. There are several approaches to testing parameter values in a Bayesian context, depending on the hypothesis being tested (see Makowski et al., 2019 for an overview). Here, we are interested in testing whether the contribution of each latent factor to performance on each task, λ_{tf}, and whether the

regression coefficients of the latent speech factor onto each latent memory factor, β_f, were likely to be greater than a negligible region around zero (Morey & Wagenmakers, 2014). For this test, we computed Bayes Factors as the ratio of probability density inside or outside a Region of Practical Equivalence (Kruschke, 2018; Morey & Rouder, 2011). We tested whether the effect of factors on task performance, λ, were likely to have values above a region of 0 to 0.02. This small negligible region was selected to account for the fact that effect size likely depends on overall task performance, and tasks with low overall performance, such as Free Recall in the manner the data were coded, could have small but meaningful effect sizes. We also tested whether regression coefficients, β, were likely to have values above 0.3. This negligible region was selected based on the meta-analysis by Dryden et al., (2017), which found that most cognitive measures correlated with speech recognition accuracy at around $r = 0.3$, so normalized regression coefficients less than this value are unlikely to reflect a specific relationship between the construct being measured and sentence recognition. We adopt the convention of labeling Bayes Factors greater than 3 in favor of the parameter value lying outside of the negligible region as "substantial" evidence in favor of that parameter being meaningful in the model (Wetzels et al., 2011).

RESULTS

Task Performance

The distribution of performance on each task is depicted in Figure 1. Comparison of the range of performance on each of these tasks to prior work and estimates of task reliability are provided in supplemental digital content 2. To summarize, individual differences in performance on all tasks were reliable and followed trends that have been previously reported in the literature, and the in-lab and remote testing participant groups did not differ from one another on memory task performance.

Task-level summary statistic correlations between sentence recognition and memory tasks

Individual performance relative to the group was consistent across target-to-masker ratios in the sentence recognition task and across memory tasks, as shown by the correlations in Table 1. Significant positive correlations across target-to-masker ratios indicates that individuals with relatively high/low keyword recognition accuracy had high/low accuracy across target-to-masker ratios. Memory task performance was positively correlated with sentence recognition at all target-to-masker ratios, although many of these correlations did not reach significance ($p < 0.05$).

To visualize these correlations, Figure 2 shows the association between PRESTO sentence recognition accuracy averaged across target-to-masker ratios and performance on the memory tasks. Significant correlations ($p < 0.05$) were observed for each memory task but were marginal for reading and digit span ($p = 0.02$ and $p = 0.05$, respectively).

Latent factor models of sentence recognition and memory task accuracy

Models with one or two latent individual difference factors were fit to keyword recognition accuracy for PRESTO sentences across target-to-masker ratios to test which model provided

the best explanation for these speech data. A comparison of expected log pointwise predictive density across models indicated that a model with a single latent factor was a better explanation for the data than a model with two latent factors, as summarized in Table 2. Post-hoc inspection of the two-factor model fit indicated that even though the model was constrained so that latent factors were orthogonal within each sample, the posterior modes of both factors were highly correlated (r > 0.99). Exploratory attempts to constrain the model to reduce this correlation were unable to remove this correlation or improve the model fit. Thus, the model fit comparison and correlation between factors in the two-factor model both indicate that one latent factor is the best explanation for individual differences in sentence recognition accuracy across target-to-masker ratios in this study. This finding is consistent with the strong correlations across target-to-masker ratios in Table 1.

The one-factor model fit is depicted in Figure 3A. Individual differences in the latent factor, η, significantly contributed to keyword recognition accuracy in every target-to-masker ratio (bayes factors greater than 100 for all λ parameters), and group-level keyword recognition accuracy, μ, declined from +10 dB down to 0 dB, consistent with the trend depicted in Figure 1. Note that the estimated values of contribution of the latent factor to performance are not directly comparable across target-to-masker ratios because they are on a log-odds scale and the effect of a change in log-odds on percent correct is relative to where that change occurs on the scale. For example, in the +10 dB and +4 dB target-to-masker ratio conditions, the group-level accuracy is 0.78 and 0.27, respectively, which corresponds to keyword recognition probabilities of 68.6% and 56.7%. Going up by one standard deviation in the latent factor (0.78 + 0.38 and 0.27 + 0.31) yields an increase to 76.1% and 64.1%, respectively, or a difference of 7.6% and 7.4% relative to the mean. Thus, although the contribution of the latent factor to each condition differs across target-to-masker ratios (0.38 vs 0.31), they both yield a similar change in the probability of correctly recognizing keywords because they are relative to different group-level intercepts.

The estimated posterior densities for the single latent sentence recognition factor are depicted in Figure 3B. These densities show the range of values that the latent factor could likely have for each participant and can be used in two ways. First, individual differences in sentence recognition ability can be estimated as a point value by finding the most likely value for each participant (i.e. the maximum a posteriori value or mode of the distribution), which is analogous to a random intercept in a mixed effects model. Second, density can be used to estimate a range of likely values for the parameter for each participant, which can be quantified in a variety of ways (e.g. Kruschke, 2018; Morey et al., 2016). Estimates of 95% highest density interval and visual inspection of Figure 3B both indicate that individuals are heterogeneous in their latent sentence recognition ability, which demonstrates that substantial individual differences exist in sentence recognition accuracy in this population for these stimuli and that sufficient data was collected per participant to estimate those differences across participants.

For the three memory tasks, models with one, two, or three latent factors were fit to data from all three tasks, again to select the latent structure which best explained task performance. A comparison of model fits is provided in Table 3. As in the speech model, post-hoc inspection of the three-factor model fit found that the posterior modes of all three

factors were highly correlated (r > 0.8), indicating that the three-factor model had an excess of latent factors. Further inspection of posterior densities for model parameters found that they tended to be multi-modal, indicating that there were multiple model fits that could explain the data. Multiple fits can arise from swapping parameters throughout model fitting, such that there is not clear mapping of each latent factor onto performance in each task. As in the speech model, exploratory attempts to constrain the model did not eliminate these correlations or improve the fit of the three-factor model. This evidence favors the two-factor model as the best explanation for the memory data, which is a finding that is not evident from correlations of task-level performance reported in Table 1.

The two-factor memory model is depicted in Figure 4A. Group-level accuracy, μ, declines with increasing list length for digit and reading span, consistent with the trends depicted in Figure 1. Performance on each task was determined by individual differences in two latent factors, η_1 and η_2. Scale parameters λ_1 and λ_2 are comparable within a task, so the relative contribution of each latent factor to performance on each task is a meaningful comparison. The first level of comparison of scale parameters is to examine whether Bayes Factors support each parameter lying outside the region of practical equivalence, which is depicted as coefficients in bold in Figure 4. As shown, η_1 predominantly determined performance on the free recall task and reading span at all list lengths and had some contribution to digit span at short list lengths. η_2 predominantly determined performance in the digit span task at longer list lengths and had a secondary effect on performance in the reading span task. The second level of comparison is to examine the relative magnitude of the coefficients that are likely to be nonzero. η_1 has a larger contribution to reading span performance than η_2 even when λ_2 is nonzero. Conversely, η_2 has a larger contribution to performance on the digit span task for lists of length 5 and greater, which is where most of the variance in performance on this task arises. Because performance on the digit span task for lengths of 2 and 3 was almost completely at ceiling, meaningful individual differences in performance on these list lengths were not observed, which is reflected in the fact that the scale parameters for these list lengths are unlikely to differ from zero.

Estimated posterior densities for η_1 and η_2 are shown in Figure 4B. Because these data are two-dimensional, posterior density is depicted as a shade gradient around the most likely values of each of these parameters for each participant. As shown, the most likely parameter values are uncorrelated across participants, indicating that the model was able to recover two independent factors per participant that accounted for their performance across memory tasks. Visual inspection of the gradient regions for each participant indicate that they were centered around the most likely value and took on an approximately multivariate Gaussian shape, indicating that latent factors estimates were independent of one another. As in Figure 3, the separation of posterior density between participants indicates that meaningful differences in each latent factor exist across participants in this population and the data collected per participant were sufficient to precisely estimate those differences.

Finally, models with a varying number of memory factors and a single speech factor were fit to the entire data set to select the latent structure which best explained the link between memory and sentence recognition. The decision to compare these specific models was made post-hoc after seeing the results of the models depicted in Figures 3 and 4. Only

a single speech factor was included here because the two-factor speech model completely collapsed into a single factor. The memory model with three factors appeared to have an excess of interchangeable factors that was less severe than the two-factor speech model, so we believed that it was possible that adding their relationship to the speech factor could constrain the model in such a way that allows for all three memory factors to have a unique relationship with the speech factors and memory tasks. For these models, data from digit span at list lengths 2 and 3 were excluded because they did not contribute to estimation of individual differences in latent memory factors. Preliminary model fits with one latent memory factor and the speech factor converged to a single solution, with a regression coefficient of $\beta = 0.62$ and $\sigma_{speech} = 0.78$ (Bayes Factor of 26.7 that β lay outside the region of practical equivalence). However, the model we initially used failed to converge when the number of latent memory factors was greater than one. Post-hoc inspection of posterior densities indicated that model fitting failed because the model trended toward values of β_1 = 1 and $\sigma_{speech} = 0$ (i.e. $\eta_{speech} = \eta_1$), which were parameter values at the boundaries of the initial model. Based on this observation, a second model was created that fixed η_{speech} = η_1 and was fit with two and three latent memory parameters. Comparison of models with different free or fixed model parameters is valid because they are both used to explain the same data set and thereby generate comparable estimates of likelihood of the data arising from each model. A comparison of model fits is provided in Table 4. As was expected based on the memory models reported in Table 3 and Figure 4, the model with two or three latent memory factors had similar model fits to the data, but inspection of the three-factor model found that latent factors 2 and 3 were highly correlated, which supports the selection of the two-factor model as the best explanation for the data.

The full two-factor model is depicted in Figure 5A. The effect of latent factors on their respective memory and speech task performance is similar to the component models depicted in Figures 3 and 4. The latent sentence recognition factor affected recognition accuracy across all target-to-masker ratios and was the same factor that predominantly affected performance in the free recall and reading span tasks. Because this model fixed β_1 = 1 estimating Bayes Factor is not possible for this model parameter. However, the fact that this model is a better explanation for the data than the one-factor model indicates that η_1 is closely related to η_{speech}, even if their true relationship is not unitary. Posterior densities for η_1 and η_2 are shown in Figure 5B in the same format as in Figure 4B, and similarly indicate that these latent factors are independent of one another and meaningfully vary across individuals. A post-hoc power analysis, described in supplemental digital content 3, demonstrated that the amount of data collected per participant was sufficient to select from among models with the number of latent factors tested here and to estimate individual differences in latent factors with high reliability.

DISCUSSION

In this study, we examined the relationship among three memory tasks, digit span, reading span, and free recall, and recognition of vocoded sentences in the presence of a two-talker masker across a range of target-to-masker ratios. Performance across memory tasks was positively correlated. Memory tasks had positive correlations with recognition accuracy for vocoded speech-in-speech, although only some correlations were significant. These results

indicate that all three memory tasks measure underlying abilities which are associated with sentence recognition. Subsequent latent factor modeling of accuracy in each trial of each task indicated the presence of two distinct memory factors, one of which was closely related to sentence recognition and one of which was not.

Individual differences in sentence recognition across target-to-masker ratios

As shown in Figure 3, individual differences in PRESTO sentence recognition across target-to-masker ratios were best explained by a single latent factor that differed across participants. The posterior densities of these individual differences, depicted in Figures 3B and 5B, indicates that these differences are likely to be meaningful variability within the population of young adults with normal hearing, rather than measurement error. This finding is consistent with the results from Bosen and Barry (2020), which found that individual differences in sentence recognition accuracy were correlated across spectral resolutions. Together, these results indicate that the extent to which latent factors are responsible for individual differences in sentence recognition accuracy does not depend on recognition accuracy in the task. Thus, our results indicate that the explicit, slow processing component of the ELU model consistently affects speech recognition accuracy in young adults with normal hearing within the range of performance we have observed (group averages of 88% to 23% across this study and Bosen and Barry, 2020). It is possible that the factors that affect speech recognition change outside of this range, although in such cases accuracy would be close to floor or ceiling and individual differences in speech recognition accuracy would be challenging to precisely assess. Thus, if there is a threshold for which cognition "kicks in" (Rönnberg et al., 2010), it is likely that this threshold is reached whenever recognition accuracy is not perfect, or alternatively whenever inference is needed to identify unclear words regardless of accuracy (Winn & Teece, 2022). If the transition from automatic to explicit processing occurs when recognition accuracy is nearly perfect, then alternative methodology is likely needed to identify this transition, such as examining the time course of lexical activation (Farris-Trimble et al., 2014) or listening effort (Winn & Teece, 2022).

An important design detail in the present study is that the target-masker pairs were fixed across all participants, which allows for fair comparison of individual differences in sentence recognition accuracy without the confounding effects of variability in the vocoder's effect on target intelligibility (DiNino et al., 2016), in target audibility relative to a fluctuating speech masker (Buss et al., 2021), and in cross-talker differences in intelligibility (Markham & Hazan, 2004). This finding is in agreement with prior work by Carbonell (2017), who found consistent individual differences in monosyllabic word recognition when words were vocoded, time-compressed, or presented in the presence of four competing talkers. It seems plausible that these individual differences in sentence recognition accuracy are driven by individual differences in cognitive ability, but a study that uses multiple types of speech materials and adverse listening conditions within participants is needed to determine whether multiple latent factors are needed to account for individual differences across stimuli and degraded listening conditions.

There appears to be a discontinuity in recognition accuracy in Figure 1, where accuracy drops between the +4 and +2 dB target-to-masker ratio conditions. Such a discontinuity

could indicate that the attentional mechanisms for picking out the target relative to the masker were impaired when the target-to-masker ratio was close to zero (Ihlefeld & Shinn-Cunningham, 2008). Vocoding reduces access to auditory cues that facilitate segregation of concurrent auditory streams (Qin & Oxenham, 2003), so the discontinuity could reflect loss of access to an obvious level cue to distinguish the target speech from the maskers. If a distinct attentional mechanism is responsible for stream segregation we might expect to see recognition accuracy at +2 dB and 0 dB target-to-masker ratio associate more with a second latent sentence recognition factor, but that was not the case here. It is possible that individual differences in segregation ability would manifest as a second latent factor when stream segregation cues are available, so it would be informative to repeat this study without vocoding the target and masker speech to determine if multiple latent factors are evident in sentence recognition when stream segregation cues are available.

Multiple memory factors explain cross-task performance

Examination of the most likely model to explain memory (Tables 3 and 4) supports the existence of two independent memory factors. The presence of multiple latent factors provides evidence against a single unifying explanation for individual differences in task performance, such as motivation or general cognitive ability. However, these factors did not seem to align with our expectations based on the storage and processing or maintenance and retrieval models as described by prior literature, and instead supported the existence of one factor that predominantly affected performance on free recall and reading span and a second factor that dominated digit span performance at long list lengths, as shown in Figures 4 and 5.

A post-hoc explanation for the observed results is that serial recall of digits, and to a lesser extent serial recall of letters in the reading span task, depends on domain-specific knowledge of the stimuli used for the task (Botvinick, 2005; Waris et al., 2017). Recall accuracy for digit sequences has been shown to depend on how often transitions between successive digits occur in natural language samples (Jones & Macken, 2015), which supports the idea that individual differences in digit span task performance might depend on experience specifically with storing and retrieving digit sequences, in addition to general memory ability. Previous work with older adults who hear with cochlear implants found a correlation between digit serial recall accuracy and self-reported vocabulary (Bosen et al., 2021), which further suggests that the memory factor which primarily determined digit span task performance could be reflective of verbal crystallized intelligence or experience. This factor made no contribution to sentence recognition outcomes in the best fitting model tested here. In contrast, the memory factor that accounted for performance in the free recall task and was the primary determinant of performance in the reading span task was equivalent to the latent sentence recognition factor in the best-fitting model. While it seems unlikely that these two latent factors are truly unitary, the evidence is in favor of a strong relationship between these latent factors.

Our results suggest a refinement to the Ease of Language Understanding model. There is ample evidence that performance on complex span tasks such as reading span, which include interleaved storage and processing components, correlates with speech recognition accuracy

in a variety of conditions. However, our results demonstrate that complex span tasks are not unique predictors of speech recognition accuracy. Performance on the free recall task also correlated with speech recognition accuracy and loaded onto a common latent factor with reading span. The common feature linking these tasks is the ability to retrieve recently presented verbal information (Unsworth & Engle, 2007b). Thus, we posit that the slow, explicit process of the Ease of Language Understanding model is dependent on individual differences in activation and subsequent retrieval of information in long-term memory, which is line with the role of long-term memory in the Ease of Language Understanding Model (Rönnberg et al., 2021).

While performance on all three tasks was affected by the factor related to sentence recognition ability to some extent, digit span task performance seems to be primarily determined by a second factor that is not related to sentence recognition ability. It appears that prior work (Bosen & Barry, 2020) found correlations between digit span performance and sentence recognition despite the choice of task, not because of it. Re-examination of the correlations shown in Figure 2 are consistent with this interpretation, with the weakest correlation for digit span and the strongest correlation for free recall. Our findings indicate that caution is needed when interpreting the presence or absence of correlations between various cognitive tasks and speech recognition, as it cannot be assumed that any given task is a pure measure of the cognitive construct it is purported to measure.

When does memory affect speech recognition in young adults with normal hearing?

As discussed in the introduction, performance on many cognitive tasks weakly correlates with speech recognition accuracy (Dryden et al., 2017), so it is generally unsurprising to find positive correlations such as the ones shown in Table 1 and Figure 2 so long as the tasks used have sufficient reliability (Heinrich & Knight, 2020; Parsons et al., 2019). Latent factor modeling demonstrated that all three tasks used here measure the latent memory factor which was related to sentence recognition to some extent, so it appears that observing associations between memory and sentence recognition does not depend on the use of a specific memory task. Comparison of the current results with prior literature indicates that observing a correlation between memory task performance and speech recognition accuracy in young adults with normal hearing seems to require specific conditions in the speech recognition task that are present in the PRESTO sentences, as described below.

One potential condition is adaptation to a novel listening condition. When participants who are unfamiliar with vocoded speech hear it for the first time they undergo a period of learning which improves speech recognition accuracy, which occurs within about 10 – 12 sentences (Davis et al., 2005). If the rate at which participants could learn to interpret vocoded speech were a major factor determining the presence or absence of a correlation, then we would expect accuracy for the first target-to-masker ratio tested (+10 dB) to have the strongest correlation with individual differences in memory ability because that is when learning would occur. The magnitude of correlations between memory task performance and sentence recognition accuracy in Table 1 is indeed larger for the +10 dB condition, but significant correlations are present at other target-to-masker ratios. Thus, it is possible that individual differences in learning the novel listening condition were partially responsible for

the observed link between memory and sentence recognition, although the consistency of individual differences in recognition accuracy across target-to-masker ratios suggests that a transient learning mechanism would only account for a small portion of this link.

Another potential condition is the presence of the two-talker masker. Bosen & Barry (2020) observed correlations between serial recall accuracy and vocoded PRESTO sentence recognition when vocoder spectral resolution was manipulated, rather than target-to-masker ratio as in the current study. The presence of correlations in that study without including the two-talker masker used here indicates that the inclusion of the masker does not determine the presence or absence of a correlation between memory and sentence recognition accuracy. However, a within-participant comparison of recognition accuracy with and without the two-talker masker would be needed to determine whether the magnitude of the correlation changes across these listening conditions.

The use of vocoded speech is also unlikely to determine the presence or absence of correlations, as O'Neill et al., (2019) used vocoded IEEE sentences and found no correlation between sentence recognition accuracy and performance on the reading span task in young adults with normal hearing. A similar lack of correlation was also reported by Shader et al., (2020) between vocoded IEEE sentence recognition and performance on the List Sorting working memory test from the NIH Toolbox. Contrasting their findings with the ones presented in the current manuscript indicates that the strength of the correlations that we observed were likely driven by the use of PRESTO sentences.

PRESTO sentences are somewhat unusual relative to many other speech stimuli, in that the PRESTO sentences were designed to limit talker-contingent learning by changing talkers and indexical properties of those talkers across sentences (Gilbert et al., 2013). Word recognition is slower and less accurate following changes in target talker, which indicates the presence of an obligatory cognitive cost that is incurred when switching between talkers (Magnuson et al., 2021; Martin et al., 1989; Mullennix et al., 1989; Mullennix & Pisoni, 1990). Talkers change across successive PRESTO sentences, so this switch cost would be present at the onset of each sentence. This switch cost also increases under conditions of cognitive load (Nusbaum & Morin, 1992), indicating that the switching process requires cognitive resources. More recent work has shown that the cost of switching talkers occurs even if the switch is predictable and to a previously heard talker (Kapadia & Perrachione, 2020), that the magnitude of the cost depends on the similarity of the prior talker and the new talker (Stilp & Theodore, 2020), that brief auditory precursors are sufficient to eliminate the switch cost (Morton et al., 2015), and that the switching cost is transient, dissipating within about 600 ms (Choi et al., 2021). Taken together, these findings indicate the presence of an automatic attentional mechanism for normalizing to a new talker. In addition, Choi et al. (2021) also found that expectation of a multi-talker listening condition incurred a sustained cognitive cost, demonstrated through increased reaction times, that lasted for at least 1.5 s after a talker switch. It is possible that individual differences in the cognitive abilities that underlie switching between talkers are assessed by the memory tasks we used here, which would account for the correlations we found between PRESTO sentence recognition and performance on our memory tasks. If this were the case, we would expect such correlations to arise whenever PRESTO sentences are presented in an

adverse listening condition, regardless of the nature of the listening condition. Evidence in favor of this possibility is provided by Tamati et al. (2013), who found that a group of young adults with normal hearing who had high recognition accuracy for PRESTO sentence recognition in a six-talker masker (without vocoding, as was used in the present study) also had better serial recall ability than a group of young adults with low recognition accuracy. Associations between speech recognition in noise and performance on the reading span task are generally weaker in young adults with normal hearing than in older adults and individuals with hearing loss across a variety of other types of speech materials (Füllgrabe & Rosen, 2016; Gordon-Salant & Cole, 2016), which further supports the idea that recognizing PRESTO sentences depends on individual differences in cognitive ability in young adults with normal hearing, while other materials do not. While these possibilities are intriguing, additional within-participant comparisons of recognition accuracy for PRESTO sentences and other speech materials across various listening conditions are needed to test whether relationships between speech recognition and memory vary across speech materials and/or listening conditions.

Steps toward identifying the latent structure linking cognition and speech recognition

The apparent importance of variable indexical properties in the PRESTO sentences for eliciting a correlation between speech recognition accuracy and performance on memory tasks suggests that working memory could be required to identify and extract speech cues (Heald & Nusbaum, 2014) in dynamic listening conditions. An important caveat of this interpretation of our data is that, as noted in the introduction, distinct cognitive abilities tend to be correlated with one another. Thus, it is possible that even though our memory tasks provided a convergent estimate of individual ability, we only measured a cognitive construct that is correlated with the constructs which mechanistically support speech recognition, rather than the supporting constructs themselves. Given that identifying the overall structure linking various cognitive abilities is a vast and active field of research we will not attempt to review all possible mechanisms which could support speech recognition, but rather provide a brief overview of plausible mechanisms underlying the link between memory task performance and speech recognition accuracy that could be readily tested.

Working memory and attentional control are closely related cognitive abilities (Shipstead et al., 2015; Unsworth & Robison, 2015). Working memory guides attentional control by maintaining perceptual templates and actions in a readily available state, and attentional control determines what information is held in working memory (Oberauer, 2019). Adapting to a change in talkers seems to be an automatic attentional mechanism as described above, so working memory maintenance could support adaptation by holding templates for targets of perception (i.e. indexical properties of the target talker) and by controlling action (the mapping of acoustic to linguistic information for that talker). In visual working memory experiments, such as remembering a specific color hue, evidence indicates that a single item to be remembered can be held in a continuous representation, while other items are stored in a categorical representation (Hardman et al., 2017). If a similar limitation was present in verbal working memory then the continuous properties of the target speech would be limited to a singular store as well, which would be consistent with the finding that switch costs are incurred every time the target talker changes (Choi et al., 2021). In addition to the transient

automatic switching cost, sustained costs for multi-talker listening condition could reflect maintenance of multiple speech maps in memory at once. Correlations between working memory ability and motor control have been observed in experiments where individuals must switch between response actions across successive experimental trials, such as being cued to looking toward or look away from a visual target (Unsworth et al., 2004). Similar correlations have also been observed between working memory and attentional filtering, so long as the filtering demands of the task change on a trial-by-trial basis (Robison et al., 2018). If attentional control and working memory maintenance are mechanisms underlying speech recognition, then alternating between adverse listening conditions or talkers should yield declines in speech recognition accuracy that are contingent on working memory ability, with individuals who have low working memory ability showing greater declines because they have more frequent lapses in attentional maintenance (McVay & Kane, 2012; Unsworth et al., 2012; Unsworth & McMillan, 2014) of indexical properties or novel mappings and/or have difficulty maintaining multiple speech maps in a readily accessible state. Working memory ability would also predict the speed with which individuals can adapt to a change in talker.

Recognizing speech in a novel adverse condition likely also depends on learning and retrieval abilities. Familiarity with a talker improves speech recognition (Nygaard et al., 1994; Souza et al., 2013). Such improvement can occur rapidly, as demonstrated by increases in recognition accuracy for vocoded speech within about a dozen exposures to vocoded speech (Davis et al., 2005). Improvements in recognition accuracy persist across multiple training sessions (Nogaki et al., 2007) for at least one week after training (McGettigan et al., 2014). From a working memory perspective, individual differences in working memory ability predict the ability to use previously presented information to solve novel problems (Harrison et al., 2015). If working memory ability supports learning to recognize speech in novel conditions, then working memory ability should be correlated with individual differences in the rate at which individuals learn to extract speech cues. Training participants to recognize synthetic speech improves speech recognition accuracy while maintaining information in memory, which suggests that training diminishes working memory demands of speech recognition (Francis & Nusbaum, 2009). This suggests that working memory ability should be correlated with speech recognition prior to learning the novel condition, but such correlations should diminish or be absent after learning. Individuals who have a high ability to encode into and retrieve relevant information to and from long-term memory, such as recalling speech mappings for particular accents or listening conditions, would be predicted to be better at adapting to changes in talker or listening condition for conditions they have prior experience with.

We posit that these working memory mechanisms may be active when listening in any condition where the mapping of speech cues changes over time. For example, signal processing algorithms in hearing aids can distort the speech signal in a manner that changes over time, and evidence suggests that individuals with low working memory ability are adversely affected by this distortion to a greater extent than individuals with high working memory ability (Souza et al., 2015). Correlations between working memory ability and speech recognition are also most evident shortly after device activation or alteration (Foo et al., 2007; Lunner, 2003; Ng et al., 2014; Rudner et al., 2009), suggesting the presence

of a learning phase in which device users must learn to interpret novel distorted speech signals. The effect of experience listening with a hearing aid on the correlation between working memory ability and speech recognition also depends on listening condition, with the relationship persisting with experience in the presence of difficult multi-talker maskers but not in steady-state noise (Ng & Rönnberg, 2020). In cochlear implants, the mapping of acoustic cues to speech percepts can vary dramatically across individuals (Harnsberger et al., 2001) and changes in one acoustic dimension, such as stimulus level, can alter perception of other stimulus dimensions, such as pitch, in a manner that is idiosyncratic (Carlyon et al., 2010). In both types of devices, it is plausible that these features of listening with an assistive device create a condition wherein the listener is frequently adapting to changes in the mapping of speech cues, which could be why associations between working memory ability and speech recognition accuracy are most often found in these populations (O'Neill et al., 2019; O'Neill, Parke, et al., 2021; Rönnberg et al., 2016; Tamati et al., 2020). Souza et al. (2015) suggest that individuals with low working memory ability may benefit more from devices with low alteration in how they process acoustic input. If this is true, it would indicate that assessing individual differences in working memory ability could play a role in clinical decisions about which signal processing algorithms are appropriate for a patient.

The age of the majority of individuals with assistive devices could also play a role. Older adults are susceptible to declines in working memory ability regardless of hearing status (Luo et al., 2021) and working memory is more closely associated with speech recognition accuracy in older adults than in younger adults (Gordon-Salant & Cole, 2016; O'Neill et al., 2019). The ability to automatize learned behaviors and thereby bypass the cognitive processing bottleneck declines with advancing age (Maquestiaux, 2016), which could indicate that older adults are less able to automatically retrieve lexical representations from long-term memory when listening to speech and instead tend to engage in slow, explicit processing. Overall, it is likely that speech material, listening condition, hearing status, experience, age, and cognitive ability all interact to determine individual differences in speech recognition accuracy.

Limitations

This relatively small study focused on a small portion of the broad set of cognitive abilities and adverse listening conditions that are relevant to speech perception. Thus, while we demonstrated convergent validity of the memory tasks used here, we did not demonstrate their divergence from related concepts, such as fluid intelligence or attention. A next step in this line of research would be to replicate this study using a broader set of cognitive tasks, adverse listening conditions, and/or participant populations to identify the boundaries at which the results presented here do not generalize.

As described above, the link between PRESTO sentence recognition and working memory may arise from the unique properties of the PRESTO sentence task. Many of the sentences in this task are long and complex, which could place a particularly heavy demand on working memory that may not generalize to shorter sentences or individual words (Heinrich & Knight, 2016). This feature of the PRESTO sentences is not inherently undesirable because speech recognition outcomes depend on both cognitive and auditory factors

(Akeroyd, 2008; Lunner, 2003), but the interpretation of sentence recognition accuracy and its relationship to those underlying sources of variability needs to be carefully considered when designing assessments and interventions. More generally, the extent to which any type of listen-and-repeat task generalizes to meaningful assessment of a listener's receptive communication ability needs to be considered when attempting to identify the latent factors that impede real-world communication (Beechey, 2022).

The model of task performance used here is also relatively simplistic compared to what is known in the field, as we modeled performance in each task as a set of independent samples from a binomial distribution. We also opted to only calculate a single metric from each task, although it can be informative to measure several trial-level measures of performance (Unsworth et al., 2009) or model trial-level error patterns in addition to modeling the probability of a correct response (Oberauer & Lewandowsky, 2019). The use of a relatively simple model here ensured it was tractable and provides a baseline to determine how much adding complexity to capture known features of working memory and speech recognition processes alters the observed relationships between latent factors. Thus, while the models proposed here likely require further refinement, our goal was to provide a foundation from which to develop our understanding of the latent factors that link speech recognition to cognition. The OSF repository associated with this manuscript provides examples to facilitate adoption of this approach and direct testing of alternative hypotheses against those tested here.

The approach used here also requires more work to develop and validate the proposed models. The ability to articulate hypotheses as a set of statistical formulae allows for flexibility that is not available with more rigid model fitting frameworks, although that flexibility can easily lead to models that are intractable or fail to converge, as we found for our initial attempt to fit the model depicted in Figure 5. As an alternative, structural equation modeling is an option to test competing hypotheses about the latent structure which produced a dataset (e.g. Rosseel, 2012), but is more limited in the types of statistical distributions it can model. Packages are available to conduct Bayesian generalized linear multi-level modeling (Bürkner, 2017), which enables the use of Bayesian statistical analysis without requiring the user to implement their hypothesis in a statistical programming language so long as the hypothesis can be articulated as a multi-level model. These alternatives should be considered for research questions which conform to their constraints, but these constraints do not need to limit the types of questions that can be answered, as shown here.

Conclusions

Our results indicate that individual differences in recognition accuracy for vocoded PRESTO sentences in the presence of competing speech generalize across target-to-masker ratios and strongly associate with individual differences in performance across multiple memory tasks. In studies which examine associations between specific cognitive mechanisms and speech recognition, multiple tasks should be used to provide convergent measures of the constructs of interest and dissociate them from related constructs. These measures should be used to test hypotheses about the underlying structure which connects cognition to speech

recognition. Identifying the latent factors that link performance across tasks will enable us to identify how and under what conditions such cognitive factors enable speech recognition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Financial Disclosures/conflicts of interest:

REFERENCES

Akeroyd MA (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(SUPPL. 2). 10.1080/14992020802301142

Arlinger S, Lunner T, Lyxell B, & Kathleen Pichora-Fuller M (2009). The emergence of cognitive hearing science. Scandinavian Journal of Psychology, 50(5), 371–384. 10.1111/j.1467-9450.2009.00753.x [PubMed: 19778385]

Baddeley A (2012). Working Memory: Theories, Models, and Controversies. Annual Review of Psychology, 63(1), 1–29. 10.1146/annurev-psych-120710-100422

Beechey T (2022). Is speech intelligibility what speech intelligibility tests test? The Journal of the Acoustical Society of America, 152(3), 1573–1585. 10.1121/10.0013896 [PubMed: 36182275]

Bernstein JGW, Goupell MJ, Schuchman GI, Rivera AL, & Brungart DS (2016). Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-sided deaf cochlear implantees. Ear and Hearing, 37(3), 289–302. 10.1097/AUD.0000000000000284 [PubMed: 26886027]

Boothroyd A, & Nittrouer S (1988). Mathematical treatment of context effects in phoneme and word recognition. The Journal of the Acoustical Society of America, 84(1), 101–114. 10.1121/1.396976 [PubMed: 3411038]

Bosen AK, & Barry MF (2020). Serial recall predicts vocoded sentence recognition across spectral resolutions. Journal of Speech, Language, and Hearing Research, 63(4), 1282–1298. 10.1044/2020_JSLHR-19-00319

Bosen AK, & Luckasen MC (2019). Interactions Between Item Set and Vocoding in Serial Recall. Ear and Hearing, 40(6), 1404–1417. 10.1097/aud.0000000000000718 [PubMed: 31033634]

Bosen AK, Sevich VA, & Cannon SA (2021). Forward Digit Span and Word Familiarity Do Not Correlate With Differences in Speech Recognition in Individuals With Cochlear Implants After Accounting for Auditory Resolution. Journal of Speech, Language, and Hearing Research, 64(8), 3330–3342. 10.1044/2021_jslhr-20-00574

Botvinick MM (2005). Effects of domain-specific knowledge on memory for serial order. Cognition, 97(2), 135–151. 10.1016/j.cognition.2004.09.007 [PubMed: 16226560]

Brungart DS (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. The Journal of the Acoustical Society of America, 109(3), 1101–1109. 10.1121/1.1345696 [PubMed: 11303924]

Brysbaert M, & New B (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 977–990. 10.3758/BRM.41.4.977 [PubMed: 19897807]

Burkholder R, & Pisoni D (2004). Digit span recall error analysis in pediatric cochlear implant users. International Congress Series, 1273, 312–315. [PubMed: 21423864]

Bürkner PC (2017). brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software, 80(1). 10.18637/jss.v080.i01

Buss E, Calandruccio L, Oleson J, & Leibold LJ (2021). Contribution of Stimulus Variability to Word Recognition in Noise Versus Two-Talker Speech for School-Age Children and Adults. Ear and Hearing, 1–10. [PubMed: 32639255]

Carbonell KM (2017). Reliability of individual differences in degraded speech perception. The Journal of the Acoustical Society of America, 142(5), EL461–EL466. 10.1121/1.5010148 [PubMed: 29195433]

Carlyon RP, Lynch C, & Deeks JM (2010). Effect of stimulus level and place of stimulation on temporal pitch perception by cochlear implant users. The Journal of the Acoustical Society of America, 127(5), 2997–3008. 10.1121/1.3372711 [PubMed: 21117749]

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, & Riddell A (2017). Stan: A probabilistic programming language. Journal of Statistical Software, 76(1). 10.18637/jss.v076.i01

Choi JY, Kou RSN, & Perrachione TK (2021). Distinct mechanisms for talker adaptation operate in parallel on different timescales. Psychonomic Bulletin & Review, in press.

Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, & Engle RW (2005). Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin & Review, 12(5), 769–786. 10.3758/BF03196772 [PubMed: 16523997]

Cowan N (2001). The magical number 4 in short term memory: A reconsideration of storage capacity. Behavioral and Brain Sciences, 24(4), 87–186. 10.1017/S0140525X01003922 [PubMed: 11515286]

Daneman M, & Carpenter PA (1980). Individual Differences in Working Memory and Reading. Journal of Verbal Learning, 19, 450–466.

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, & McGettigan C (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. Journal of Experimental Psychology: General, 134(2), 222–241. 10.1037/0096-3445.134.2.222 [PubMed: 15869347]

Depaoli S, Winter SD, & Visser M (2020). The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. Frontiers in Psychology, 11(November), 1–18. 10.3389/fpsyg.2020.608045 [PubMed: 32038435]

de Leeuw JR (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior Research Methods, 47(1), 1–12. 10.3758/s13428-014-0458-y [PubMed: 24683129]

de Winter JCF, Dodou D, & Wieringa PA (2009). Exploratory factor analysis with small sample sizes. Multivariate Behavioral Research, 44(2), 147–181. 10.1080/00273170902794206 [PubMed: 26754265]

DiNino M, Wright RA, Winn MB, & Bierer JA (2016). Vowel and consonant confusions from spectrally manipulated stimuli designed to simulate poor cochlear implant electrode-neuron interfaces. The Journal of the Acoustical Society of America, 140(6), 4404–4418. 10.1121/1.4971420 [PubMed: 28039993]

Dryden A, Allen HA, Henshaw H, & Heinrich A (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. Trends in Hearing, 21, 1–21. 10.1177/2331216517744675

Engle RW, Laughlin JE, Tuholski SW, & Conway ARA (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. Journal of Experimental Psychology: General, 128(3), 309–331. 10.1037/0096-3445.128.3.309 [PubMed: 10513398]

Fandakova Y, Sander MC, Werkle-Bergner M, & Shing YL (2014). Age differences in short-term memory binding are related to working memory performance across the lifespan. Psychology and Aging, 29(1), 140–149. 10.1037/a0035347 [PubMed: 24660802]

Farris-Trimble A, McMurray B, Cigrand N, & Tomblin JB (2014). The Process of spoken word recognition in the face of signal degradation. J Exp Psychol Hum Percept Perform, 40(1), 308–327. 10.1037/a0034353 [PubMed: 24041330]

Faulkner KF, Tamati TN, Gilbert JL, & Pisoni DB (2015). List Equivalency of PRESTO for the Evaluation of Speech Recognition. Journal of the American Academy of Audiology, 26(6), 582–594. 10.3766/jaaa.14082 [PubMed: 26134725]

Foo C, Rudner M, Rönnberg J, & Lunner T (2007). Recognition of speech in noise with new hearing instrument compression release settings requeres explicit cognitive storage and processing capacity. Journal of the American Academy of Audiology, 18(7), 618–631. 10.3766/jaaa.18.7.8 [PubMed: 18236648]

Francis AL, & Nusbaum HC (2009). Effects of intelligibility on working memory demand for speech perception. Attention, Perception & Psychophysics, 71(6), 1360–1374.

Freyman RL, Balakrishnan U, & Helfer KS (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. The Journal of the Acoustical Society of America, 115(5), 2246–2256. 10.1121/1.1689343 [PubMed: 15139635]

Friedman NP, & Miyake A (2005). Comparison of four scoring methods for the reading span test. Behavior Research Methods, 37(4), 581–590. 10.3758/BF03192728 [PubMed: 16629290]

Friedman NP, & Miyake A (2017). Unity and Diversity of Executive Functions: Individual Differences as a Window on Cognitive Structure. Cortex, 86, 186–204. [PubMed: 27251123]

Friedman NP, Miyake A, Young SE, Defries JC, Corley RP, & Hewitt JK (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. Journal of Experimental Psychology, 137(2), 201–225. 10.1037/0096-3445.137.2.201.Individual [PubMed: 18473654]

Füllgrabe C, & Rosen S (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. Frontiers in Psychology, 7(AUG), 1–8. 10.3389/fpsyg.2016.01268 [PubMed: 26858668]

Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P-C, & Modrák M (2020). Bayesian Workflow. http://arxiv.org/abs/2011.01808

Gilbert JL, Tamati TN, & Pisoni DB (2013). Development, Reliability, and Validity of PRESTO: A new High-Variability Sentence Recognition Test. J Am Acad Audiol, 24(1), 26–36. [PubMed: 23231814]

Gonthier C (2022). An easy way to improve scoring of memory span tasks: The edit distance, beyond "correct recall in the correct serial position." Behavior Research Methods. 10.3758/s13428-022-01908-2

Gordon-Salant S, & Cole SS (2016). Effects of Age and Working Memory Capacity on Speech Recognition Performance in Noise Among Listeners With Normal Hearing. Ear and Hearing, 37(5), 593–602. 10.1097/AUD.0000000000000316 [PubMed: 27232071]

Greenwood DD (1990). A cochlear frequency-position function for several species – 29 years later. The Journal of the Acoustical Society of America, 87(6), 2592–2605. 10.1121/1.399052 [PubMed: 2373794]

Hardman KO, Vergauwe E, & Ricker TJ (2017). Categorical Working Memory Representations are used in Delayed Estimation of Continuous Colors. J Exp Psychol Hum Percept Perform, 43(1), 30–54. [PubMed: 27797548]

Harnsberger JD, Svirsky MA, Kaiser AR, Pisoni DB, Wright R, & Meyer TA (2001). Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. The Journal of the Acoustical Society of America, 109(5), 2135–2145. 10.1121/1.1350403 [PubMed: 11386565]

Harrison TL, Shipstead Z, & Engle RW (2015). Why is working memory capacity related to matrix reasoning tasks? Memory and Cognition, 43(3), 389–396. 10.3758/s13421-014-0473-3 [PubMed: 25331277]

Heald SLM, & Nusbaum HC (2014). Speech perception as an active cognitive process. Frontiers in Systems Neuroscience, 8(March), 35. 10.3389/fnsys.2014.00035 [PubMed: 24672438]

Heinrich A, & Knight S (2020). Reproducibility in Cognitive Hearing Research: Theoretical Considerations and Their Practical Application in Multi-Lab Studies. Frontiers in Psychology, 11. 10.3389/fpsyg.2020.01590

Heinrich A, & Knight S (2016). The Contribution of Auditory and Cognitive Factors to Intelligibility of Words and Sentences in Noise. Adv Exp Med Biol, 894, 37–45. 10.1007/978-3-319-25474-6 [PubMed: 27080644]

Ihlefeld A, & Shinn-Cunningham B (2008). Spatial release from energetic and informational masking in a selective speech identification task. The Journal of the Acoustical Society of America, 123(6), 4369–4379. 10.1121/1.2904826 [PubMed: 18537388]

Jones G, & Macken B (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. Cognition, 144, 1–13. 10.1016/j.cognition.2015.07.009 [PubMed: 26209910]

Kane MJ, A Conway AR, Kathryn Bleckley M, & Engle RW(2001). A Controlled-Attention View of Working-Memory Capacity. J Exp Psychol Gen, 130(2), 169–183. [PubMed: 11409097]

Kane MJ, & Engle RW (2000). Working-Memory Capacity, Proactive Interference, and Divided Attention: Limits on Long-Term Memory Retrieval. J Exp Psychol, 26, 336–358.

Kapadia AM, & Perrachione TK (2020). Selecting among competing models of talker adaptation: Attention, cognition, and memory in speech processing efficiency. Cognition, 204(February), 104393. 10.1016/j.cognition.2020.104393 [PubMed: 32688132]

Kleiner M, Brainard D, & Pelli D (2007). What's new in Psychtoolbox-3? In Perception 36 ECVP Abstract Supplement (Vol. 36, Issue 0, p. 14). 10.1068/v070821

Kovacs K, & Conway ARA (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. Psychological Inquiry, 27(3), 151–177. 10.1080/1047840X.2016.1153946

Kruschke JK (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. Advances in Methods and Practices in Psychological Science, 1(2), 270–280. 10.1177/2515245918771304

Legendre P (2013). Model II regression user's guide, R edition. R Vignette, 4, 1–14. http://ftp-nyc.osuosl.org/pub/cran/web/packages/lmodel2/vignettes/mod2user.pdf

Lunner T (2003). Cognitive function in relation to hearing aid use. International Journal of Audiology, 42(sup1), 49–58. 10.3109/14992020309074624

Luo DX, Azuma T, Kolberg C, & Pulling KR (2021). The Effects of Stimulus Modality, Task Complexity, and Cuing on Working Memory and the Relationship with Speech Recognition in Older Cochlear Implant Users. Journal of Communication Disorders, 95(November 2021), 106170. 10.1016/j.jcomdis.2021.106170 [PubMed: 34839068]

Lustig C, May CP, & Hasher L (2001). Working memory span and the role of proactive interference. Journal of Experimental Psychology: General, 130(2), 199–207. 10.1037/0096-3445.130.2.199 [PubMed: 11409099]

Magnuson JS, Nusbaum HC, Akahane-Yamada R, & Saltzman D (2021). Talker familiarity and the accommodation of talker variability. Attention, Perception, and Psychophysics, 83(4), 1842–1860. 10.3758/s13414-020-02203-y

Makowski D, Ben-Shachar MS, Chen SHA, & Lüdecke D (2019). Indices of Effect Existence and Significance in the Bayesian Framework. Frontiers in Psychology, 10(December), 1–14. 10.3389/fpsyg.2019.02767 [PubMed: 30713512]

Maquestiaux F (2016). Qualitative attentional changes with age in doing two tasks at once. Psychonomic Bulletin and Review, 23(1), 54–61. 10.3758/s13423-015-0881-9 [PubMed: 26106060]

Markham D, & Hazan V (2004). The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception test. Journal of Speech, Language, and Hearing Research, 47(4), 725–737. 10.1044/1092-4388(2004/055)

Martin CS, Mullennix JW, Pisoni DB, & Summers WV (1989). Effects of Talker Variability on Recall of Spoken Word Lists. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(4), 676–684. [PubMed: 2526857]

May CP, Hasher L, & Kane MJ (1999). The role of interference in memory span. Memory and Cognition, 27(5), 759–767. 10.3758/BF03198529 [PubMed: 10540805]

McGettigan C, Rosen S, & Scott SK (2014). Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: Implications for cochlear implantation. Frontiers in Systems Neuroscience, 8(FEB). 10.3389/fnsys.2014.00018

McVay JC, & Kane MJ (2012). Drifting from Slow to "D'oh!" Working Memory Capacity and Mind Wandering Predict Extreme Reaction Times and Executive- Control Errors. J Exp Psychol Learn Mem Cogn, 38(3), 525–549. 10.1037/a0025896.Drifting [PubMed: 22004270]

Miyake A, Friedman NP, Emerson MJ, Witzki a H., Howerter A, & Wager TD (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cognitive Psychology, 41(1), 49–100. 10.1006/cogp.1999.0734 [PubMed: 10945922]

Morey RD, Hoekstra R, Rouder JN, Lee MD, & Wagenmakers EJ (2016). The fallacy of placing confidence in confidence intervals. Psychonomic Bulletin and Review, 23(1), 103–123. 10.3758/s13423-015-0947-8 [PubMed: 26450628]

Morey RD, & Rouder JN (2011). Bayes Factor Approaches for Testing Interval Null Hypotheses. Psychological Methods, 16(4), 406–419. 10.1037/a0024377 [PubMed: 21787084]

Morey RD, & Wagenmakers EJ (2014). Simple relation between Bayesian order-restricted and point-null hypothesis tests. Statistics and Probability Letters, 92, 121–124. 10.1016/j.spl.2014.05.010

Morton JR, Sommers MS, & Lulich SM (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. 10.1121/1.4913456 [PubMed: 25786955]

Mullennix JW, & Pisoni DB (1990). Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics, 47(4), 379–390. 10.3758/BF03210878 [PubMed: 2345691]

Mullennix JW, Pisoni DB, & Martin C (1989). Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America, 85(1), 365–378. 10.1121/1.397688 [PubMed: 2921419]

Ng EHN, Classon E, Larsby B, Arlinger S, Lunner T, Rudner M, & Rönnberg J (2014). Dynamic relation between working memory capacity and speech recognition in noise during the first 6 months of hearing aid use. Trends in Hearing, 18, 1–10. 10.1177/2331216514558688

Ng EHN, & Rönnberg J (2020). Hearing aid experience and background noise affect the robust relationship between working memaory and speech recognition in noise. International Journal of Audiology, 59(3), 208–218. 10.1080/14992027.2019.1677951 [PubMed: 31809220]

Nogaki G, Fu QJ, & Galvin JJ (2007). Effect of training rate on recognition of spectrally shifted speech. Ear and Hearing, 28(2), 132–140. 10.1097/AUD.0b013e3180312669 [PubMed: 17496666]

Nusbaum HC, & Morin TM (1992). Paying attention to differences among talkers. In Speech perception, speech production, and linguistic structure (pp. 113–134).

Nygaard LC, Sommers MS, & Pisoni DB (1994). Speech perception as a talker-contingent process. Psychological Science, 5(1), 42–46. 10.1111/j.1467-9280.1994.tb00612.x [PubMed: 21526138]

O'Neill ER, Basile JD, & Nelson P (2021). Individual Hearing Outcomes in Cochlear Implant Users Influence Social Engagement and Listening Behavior in Everyday Life. Journal of Speech, Language & Hearing Research, 1–18.

O'Neill ER, Kreft HA, & Oxenham AJ (2019). Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions. The Journal of the Acoustical Society of America, 146(1), 195–210. 10.1121/1.5116009 [PubMed: 31370651]

O'Neill ER, Parke MN, Kreft HA, & Oxenham AJ (2021). Role of semantic context and talker variability in speech perception of cochlear-implant users and normal-hearing listeners. The Journal of the Acoustical Society of America, 149(2), 1224–1239. 10.1121/10.0003532 [PubMed: 33639827]

Oberauer K (2002). Access to information in working memory: Exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(3), 411–421. 10.1037/0278-7393.28.3.411 [PubMed: 12018494]

Oberauer K, Lewandowsky S, Awh E, Brown G, Conway A, Cowan N, Donkin C, Farrell S, Hitch GJ, Hurlstone MJ, Ma WJ, Morey CC, Nee DE, Sweppe J, Vergauwe E, & Ward G (2018). Benchmarks for models of short-term and working memory. Psychological Bulletin, 144(9), 885–958. 10.1037/bul0000153 [PubMed: 30148379]

Oberauer K (2019). Working memory and attention - A conceptual analysis and review. Journal of Cognition, 2(1), 1–23. 10.5334/joc.58 [PubMed: 31517219]

Oberauer K, & Lewandowsky S (2019). Simple measurement models for complex working-memory tasks. Psychol Rev, 126(6), 880–932. [PubMed: 31524425]

Parsons S, Kruijt A-W, & Fox E (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. Advances in Methods and Practices in Psychological Science, 2(4), 378–395. 10.1177/2515245919879695

Qin MK, & Oxenham AJ (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. The Journal of the Acoustical Society of America, 114(1), 446–454. 10.1121/1.1579009 [PubMed: 12880055]

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Redick TS, Broadway JM, Meier ME, Kuriakose PS, Unsworth N, Kane MJ, & Engle RW (2012). Measuring working memory capacity with automated complex span tasks. European Journal of Psychological Assessment, 28(3), 164–171. 10.1027/1015-5759/a000123

Rey-Mermet A, Gade M, & Oberauer K (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. Journal of Experimental Psychology: Learning Memory and Cognition, 44(4), 501–526. 10.1037/xlm0000450 [PubMed: 28956944]

Robison MK, Miller AL, & Unsworth N (2018). Individual Differences in Working Memory Capacity and Filtering. J Exp Psychol Hum Percept Perform, 44(7), 1038–1053. [PubMed: 29683715]

Rönnberg J, Holmer E, & Rudner M (2021). Cognitive hearing science: three memory systems, two approaches, and the ease of language understanding model. Journal of Speech, Language, and Hearing Research, 64(2), 359–370. 10.1044/2020_JSLHR-20-00007

Rönnberg J, Lunner T, Ng EHN, Lidestam B, Zekveld AA, Sörqvist P, Lyxell B, Träff U, Yumba W, Classon E, Hällgren M, Larsby B, Signoret C, Pichora-Fuller MK, Rudner M, Danielsson H, & Stenfelt S (2016). Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. International Journal of Audiology, 55(11), 623–642. 10.1080/14992027.2016.1219775 [PubMed: 27589015]

Rönnberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, Dahlström O, Signoret C, Stenfelt S, Pichora-Fuller MK, & Rudner M (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Frontiers in Systems Neuroscience, 7(July), 31. 10.3389/fnsys.2013.00031 [PubMed: 23874273]

Rönnberg J, Rudner M, Lunner T, & Zekveld A (2010). When cognition kicks in: Working memory and speech understanding in noise. Noise and Health, 12(49), 263. 10.4103/1463-1741.70505 [PubMed: 20871181]

Rönnberg J, Rudner M, Foo C, & Lunner T (2008). Cognition counts: A working memory system for ease of language understanding (ELU). International Journal of Audiology, 47(SUPPL. 2). 10.1080/14992020802301167

Rönnberg J (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. International Journal of Audiology, 42(SUPPL. 1). 10.3109/14992020309074626

Roodenrys S, Hulme C, Lethbridge A, Hinton M, & Nimmo LM (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(6), 1019–1034. 10.1037//0278-7393.28.6.1019 [PubMed: 12450329]

Roodenrys S, & Quinlan PT (2000). The effects of stimulus set size and word frequency on verbal serial recall. Memory, 8(2), 71–78. [PubMed: 10829124]

Rosseel Y (2012). lavaan: An R Package for Structural Equation Modeling. 2Journal of Statistical Software, 48(2).

Rouder JN, & Haaf JM (2019). A psychometrics of individual differences in experimental tasks. Psychonomic Bulletin & Review, 26(2), 452–467. 10.3758/s13423-018-1558-y

Rudner M, Foo C, RÖnnberg J, & Lunner T (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. Scandinavian Journal of Psychology, 50(5), 405–418. 10.1111/j.1467-9450.2009.00745.x [PubMed: 19778388]

Schönbrodt FD, & Perugini M (2013). At what sample size do correlations stabilize? Journal of Research in Personality, 47(5), 609–612. 10.1016/j.jrp.2013.05.009

Shader MJ, Yancey CM, Gordon-Salant S, & Goupell MJ (2020). Spectral-Temporal Trade-Off in Vocoded Sentence Recognition. Ear and Hearing, Ci, 1. 10.1097/aud.0000000000000840

Shinn-Cunningham BG, & Best V (2008). Selective Attention in Normal and Impaired Hearing. Trends in Amplification, 12(4), 283–299. 10.1177/1084713808325306 [PubMed: 18974202]

Shipstead Z, Harrison TL, & Engle RW (2015). Working memory capacity and the scope and control of attention. Attention, Perception, and Psychophysics, 77(6), 1863–1880. 10.3758/s13414-015-0899-0

Shipstead Z, Harrison TL, & Engle RW (2016). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. Perspectives on Psychological Science, 11(6), 771–799. 10.1177/1745691616650647 [PubMed: 27899724]

Sivula T, Magnusson M, Matamoros AA, & Vehtari A (2020). Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison. March. http://arxiv.org/abs/2008.10296

Smith GNL, Pisoni DB, & Kronenberger WG (2019). High-Variability Sentence Recognition in Long-Term Cochlear Implant Users: Associations With Rapid Phonological Coding and Executive Functioning. Ear and Hearing, 40(5), 1149–1161. 10.1097/AUD.0000000000000691 [PubMed: 30601227]

Souza PE, Arehart K, & Neher T (2015). Working memory and hearing aid processing: Literature findings, future directions, and clinical applications. Frontiers in Psychology, 6(DEC), 1–12. 10.3389/fpsyg.2015.01894 [PubMed: 25688217]

Souza PE, Gehani N, Wright R, & McCloy D (2013). The advantage of knowing the talker. Journal of the American Academy of Audiology, 24(8), 689–700. 10.3766/jaaa.24.8.6 [PubMed: 24131605]

Stan Development Team. (2020). RStan: the R interface to Stan. http://mc-stan.org/

Stickney GS, Zeng F-G, Litovsky R, & Assmann P (2004). Cochlear implant speech recognition with speech maskers. The Journal of the Acoustical Society of America, 116(2), 1081–1091. 10.1121/1.1772399 [PubMed: 15376674]

Stilp CE, & Theodore RM (2020). Talker normalization is mediated by structured indexical information. Attention, Perception, and Psychophysics, 82(5), 2237–2243. 10.3758/s13414-020-01971-x

Storkel HL (2013). A corpus of consonant-vowel-consonant (CVC) real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age-of-acquisition. Behav Res Methods, 45(4), 1159–1167. [PubMed: 23307574]

Tamati TN, Gilbert JL, & Pisoni DB (2013). Some Factors Underlying Individual Differences in Speech Recognition on PRESTO: A First Report. Journal of the American Academy of Audiology, 24(7), 616–634. 10.3766/jaaa.24.7.10 [PubMed: 24047949]

Tamati TN, Ray C, Vasil KJ, Pisoni DB, & Moberly AC (2020). High- and Low-Performing Adult Cochlear Implant Users on High-Variability Sentence Recognition: Differences in Auditory Spectral Resolution and Neurocognitive Functioning. Journal of the American Academy of Audiology, 13(275), 1–13. 10.3766/jaaa18106

Thalmann M, Souza AS, & Oberauer K (2019). How does chunking help working memory? Journal of Experimental Psychology: Learning Memory and Cognition, 45(1), 37–55. 10.1037/xlm0000578 [PubMed: 29698045]

Troche SJ, von Gugelberg HM, Pahud O, & Rammsayer TH (2021). Do executive attentional processes uniquely or commonly explain psychometric g and correlations in the positive manifold? A structural equation modeling and network-analysis approach to investigate the process overlap theory. Journal of Intelligence, 9(3). 10.3390/jintelligence9030037

Unsworth N, Brewer GA, & Spillers GJ (2009). There's more to the working memory capacity-fluid intelligence relationship than just secondary memory. Psychonomic Bulletin and Review, 16(5), 931–937. 10.3758/PBR.16.5.931 [PubMed: 19815801]

Unsworth N, & Engle RW (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. Journal of Memory and Language, 54(1), 68–80. 10.1016/j.jml.2005.06.003

Unsworth N, & Engle RW (2007a). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order

abilities. Psychological Bulletin, 133(6), 1038–1066. 10.1037/0033-2909.133.6.1038 [PubMed: 17967093]

Unsworth N, & Engle RW (2007b). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. Psychological Review, 114(1), 104–132. 10.1037/0033-295X.114.1.104 [PubMed: 17227183]

Unsworth N, Heitz RP, & Engle RW (2005). An automated version of the operation span task. Behavior Research Methods, 37(3), 498–505. [PubMed: 16405146]

Unsworth N, & McMillan BD (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. Acta Psychologica, 150, 14–25. 10.1016/j.actpsy.2014.04.001 [PubMed: 24793128]

Unsworth N, Redick TS, Spillers GJ, Gene A, & Brewer G. a. (2012). Variation in working memory capacity and cognitive control : Goal maintenance and microadjustments of control. Q J Exp Psychol, 65(2), 326–355.

Unsworth N, & Robison MK (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. Psychonomic Bulletin and Review, 22(3), 757–765. 10.3758/s13423-014-0747-6 [PubMed: 25324180]

Unsworth N, Schrock JC, & Engle RW (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. Journal of Experimental Psychology: Learning Memory and Cognition, 30(6), 1302–1321. 10.1037/0278-7393.30.6.1302 [PubMed: 15521806]

Unsworth N, Spillers GJ, & Brewer GA (2010). The Contributions of Primary and Secondary Memory to Working Memory Capacity: An Individual Differences Analysis of Immediate Free Recall. Journal of Experimental Psychology: Learning Memory and Cognition, 36(1), 240–247. 10.1037/a0017739 [PubMed: 20053060]

Vehtari A, Gelman A, & Gabry J (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing, 27(5), 1413–1432. 10.1007/s11222-016-9696-4

Waris O, Soveri A, Ahti M, Hoffing RC, Ventus D, Jaeggi SM, Seitz AR, & Laine M (2017). A latent factor analysis of working memory measures using large-scale data. Frontiers in Psychology, 8(JUN), 1–14. 10.3389/fpsyg.2017.01062 [PubMed: 28197108]

Watanabe S (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research, 11, 3571–3594.

Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, & Wagenmakers EJ (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. Perspectives on Psychological Science, 6(3), 291–298. 10.1177/1745691611406923 [PubMed: 26168519]

Wilhelm O, Hildebrandt A, & Oberauer K (2013). What is working memory capacity, and how can we measure it? Frontiers in Psychology, 4(JUL), 1–22. 10.3389/fpsyg.2013.00433 [PubMed: 23382719]

Winn MB, & Teece KH (2022). Effortful Listening Despite Correct Responses: The Cost of Mental Repair in Sentence Recognition by Listeners With Cochlear Implants. Journal of Speech, Language, and Hearing Research, 65(10), 3966–3980. 10.1044/2022_JSLHR-21-00631

Wolf EJ, Harrington KM, Clark SL, & Miller MW (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. Educ Psychol Meas, 76(6), 913–934. 10.1177/0013164413495237.Sample [PubMed: 25705052]
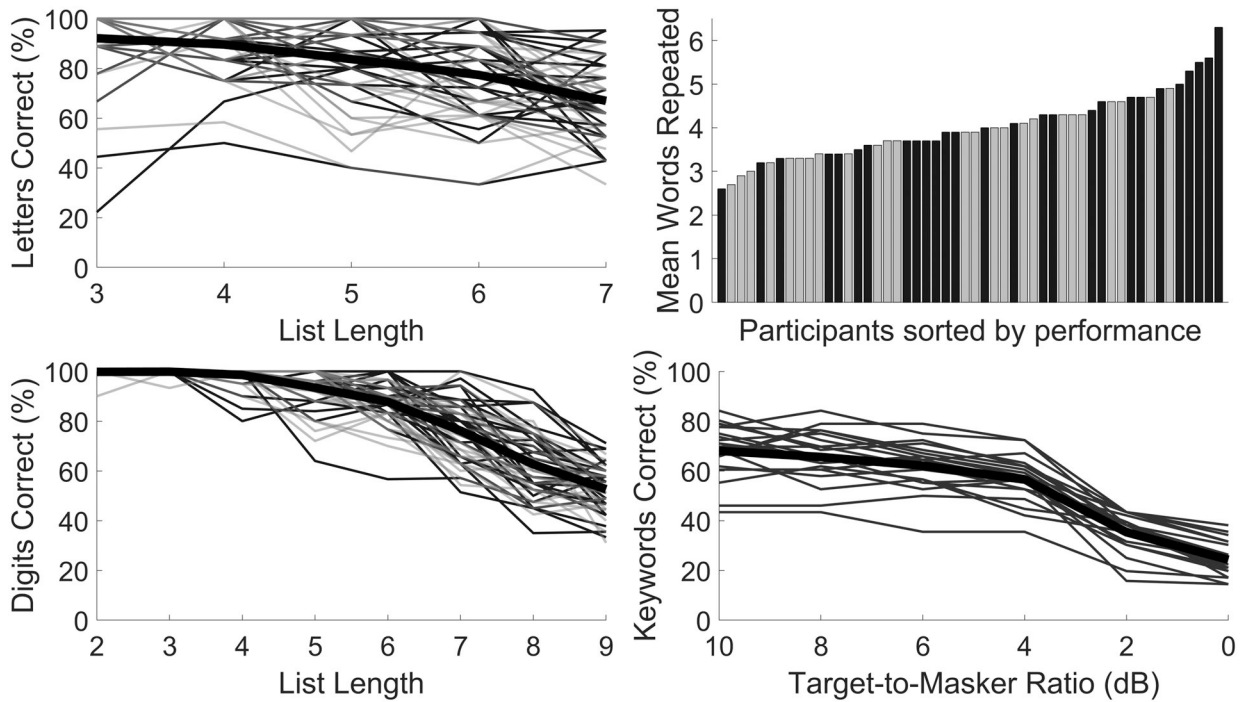
Figure 1.

Performance on working memory and speech tasks. For reading span (top left) and digit span (bottom left), edit distance score as a percentage of maximum possible score is shown as a function of list length. Thin lines show recall accuracy for each participant, and the thick line shows the group average. In-lab participants are depicted with black lines and remote participants are depicted with gray lines. For free recall (top right), the number of items in each list was fixed, so the average number of items recalled across all lists is shown as a bar for each participant. In-lab participants are depicted with black bars and remote participants are depicted with gray bars. For speech recognition (bottom right), keyword recognition accuracy is shown a as a function of target-to-masker ratio. Thin lines show keyword recognition accuracy for each participant, and the thick line shows the group average.
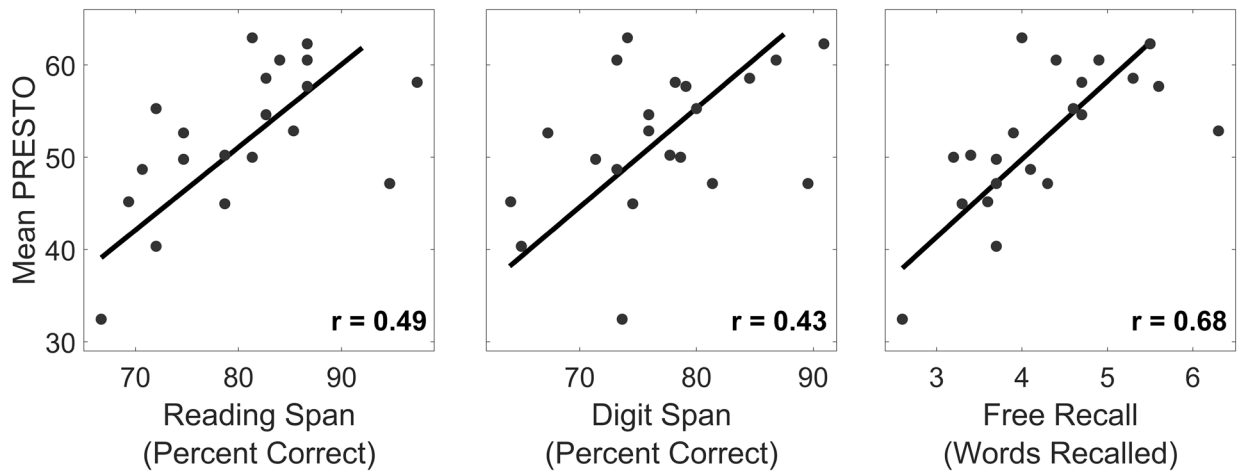
Figure 2.
Mean PRESTO sentence recognition accuracy across target-to-masker ratios as a function of performance in each memory task. Each dot shows data from one individual, and lines show standard major axis regression fits (Legendre, 2013) across individuals. All correlations were significant ($p < 0.05$).
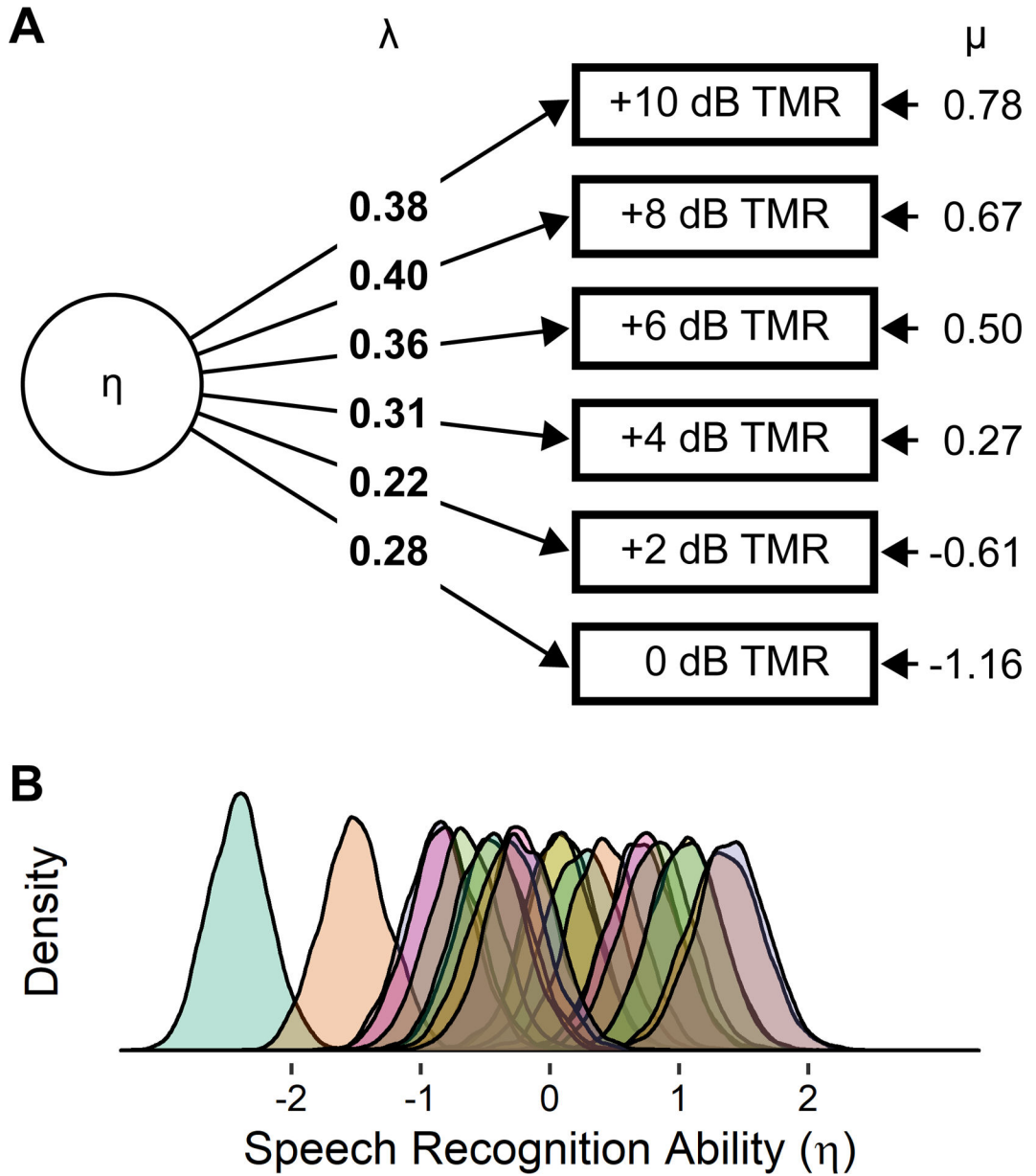
Figure 3.
The best-fitting latent factor model for speech recognition across target-to-masker ratios. (A) Group-level keyword recognition accuracy for each target-to-masker ratio is given as μ (reported in log odds), and individual performance varied relative to the group based on the product of their individual speech recognition ability η and the amount that individual ability contributed to accuracy for each target-to-masker ratio, λ. (B) Each shaded region depicts the posterior probability density of speech recognition ability η for one participant.
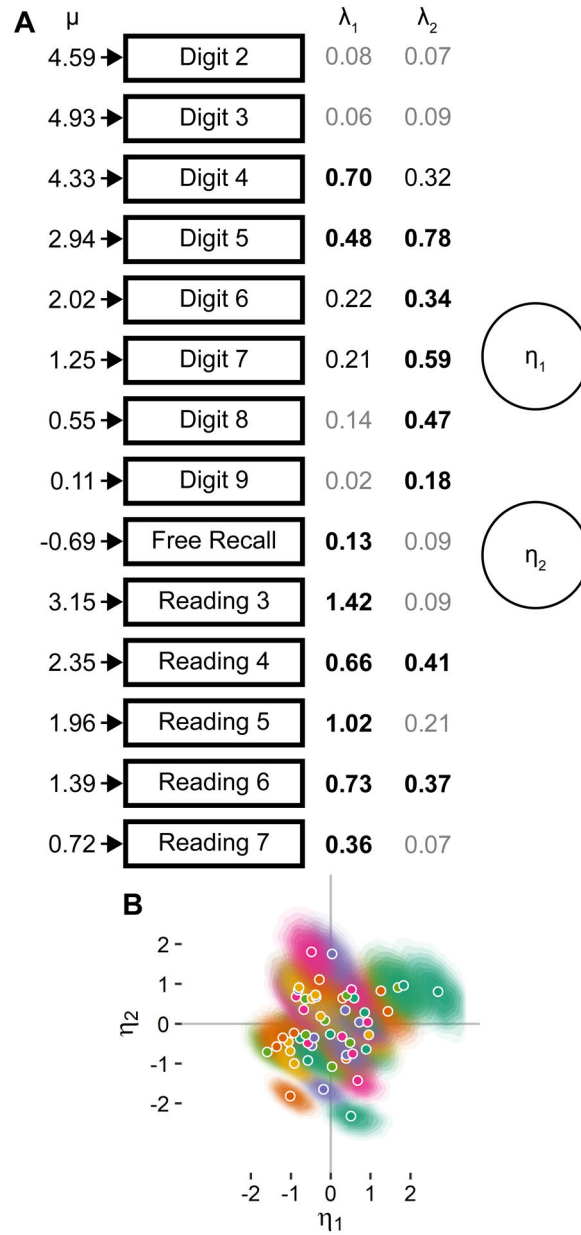
Figure 4.
The best-fitting latent factor model for working memory tasks and across list lengths within digit and reading span tasks. (A) Parameters are depicted as in Figure 3, although arrows depicting the contribution of each latent factor, η, to accuracy on each task, λ, are omitted for clarity. Model parameters λ that had a Bayes Factor greater than 3:1 in favor of lying outside of the region of practical equivalence (see text for details) are in bold, parameters with Bayes Factors between 3:1 and 1:1 are shown in black, and parameters with Bayes Factors less than 1:1 are shown in gray. (B) The two latent working memory factors differ across individuals and are not correlated across the group. Most likely values of each latent factor for each participant are depicted as circles, and shaded gradients around each most likely value depict the corresponding probability density for that participant.
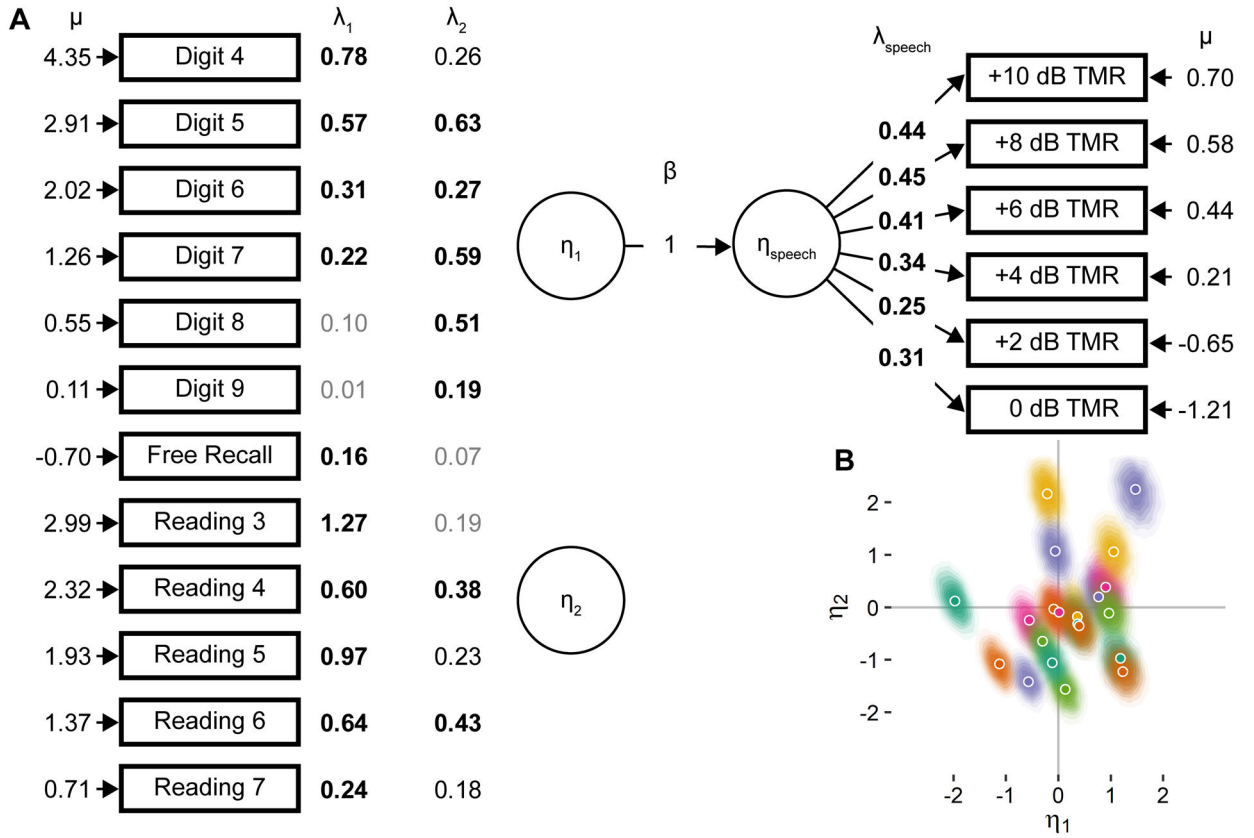
Figure 5.
The best-fitting latent factor model linking individual differences in latent working memory abilities, η_1 and η_2, to individual differences in speech recognition ability, η_{speech}. (A) The parameters linking individual differences in latent factors to performance in each task and task condition are depicted as in Figures 3 and 4. In this model, the regression coefficient β_1 was fixed to 1 because preliminary models failed to converge when attempting to approach this boundary value. (B) The two latent individual difference factors obtained by this model fit are depicted as in Figure 4B for participants who completed both the memory and speech tasks.

Table 1.

Correlation of speech recognition accuracy at each target-to-masker ratio and performance for each memory task.

Correlations	10 dB	8 dB	6 dB	4 dB	2 dB	0 dB	Reading Span	Digit Span	Free Recall
10 dB	-								
8 dB	**0.74**	-							
6 dB	**0.70**	**0.84**	-						
4 dB	**0.71**	**0.84**	**0.87**	-					
2 dB	**0.44**	**0.53**	**0.70**	**0.47**	-				
0 dB	**0.48**	**0.68**	**0.65**	**0.56**	**0.75**	-			
Reading Span	**0.69**	**0.44**	0.37	0.37	0.22	0.35	-		
Digit Span	**0.64**	0.41	0.31	0.40	0.15	0.20	**0.50**	-	
Free Recall	**0.68**	**0.54**	**0.66**	**0.55**	**0.58**	0.42	**0.44**	**0.40**	-

Pearson's linear correlation coefficients (r) are reported for each pair of values across participants. Correlation coefficients that have p-values less than 0.05 are in bold.

Table 2.

Comparison of model fits to PRESTO sentence recognition accuracy.

Model	$ELPD_{loo}$	$ELPD_{loo}$ SE	p_{loo}	P_{loo} SE	ELPD	ELPD SE
One Factor	−3891.3	48.8	57.1	1.6	-	-
Two Factors	−3900.5	49.2	87.3	2.4	−9.3	3.3

Expected Log pointwise Predictive Density ($ELPD_{loo}$), number of effective model parameters (p_{loo}), and the standard errors for both estimates (SE) were estimated for both model fits using the LOO package in R (Vehtari et al., 2017). Model fits were compared to obtain an estimated difference in expected log posterior density relative to the best-fitting model (ELPD) and standard error for the difference. Comparison of model fits indicates that the one-factor model has a substantially better fit to the observed data.

Table 3.

Comparison of model fits to reading span, digit span, and free recall task performance.

Model	ELPD$_{loo}$	ELPD$_{loo}$ SE	p$_{loo}$	P$_{loo}$ SE	ELPD	ELPD SE
Two Factors	−3843.7	62.9	163.4	7.3	-	-
Three factors	−3846.8	62.8	219.1	9.4	−3.1	4.1
One Factor	−3866.1	63.6	99.1	4.6	−22.4	10.4

Model metrics are reported as in Table 2. Comparison of model fits indicates that the two-factor model has a substantially better fit to the observed data than the one-factor model and had a marginally better fit than the three-factor model.

Table 4.

Comparison of model fits to performance on all memory and speech tasks.

Model	ELPD$_{loo}$	ELPD$_{loo}$ SE	p$_{loo}$	P$_{loo}$ SE	ELPD	ELPD SE
Three factors	−7704.7	75.7	254.9	8.9	-	-
Two Factors	−7706.4	76.0	195.0	6.3	−1.7	6.4
One Factor	−7737.1	76.3	153.6	4.8	−32.4	11.8

Model metrics are reported as in Tables 2 and 3. Comparison of model fits indicates that the two-factor and three-factor models have equivalent fit to the data and have a substantially better fit to the data than the one-factor model.