



Published in final edited form as:

*Mem Cognit.* 2013 August ; 41(6): 839–849. doi:10.3758/s13421-013-0302-0.

## Monitoring the capacity of working memory: Executive control and effects of listening effort

Nicole M. Amichetti, Raymond S. Stanley, Alison G. White, and Arthur Wingfield

Volen National Center for Complex Systems, Brandeis University, MS 013, Waltham, MA 02454-9110, USA

Arthur Wingfield: wingfield@brandeis.edu

### Abstract

In two experiments, we used an interruption-and-recall (IAR) task to explore listeners' ability to monitor the capacity of working memory as new information arrived in real time. In this task, listeners heard recorded word lists with instructions to interrupt the input at the maximum point that would still allow for perfect recall. Experiment 1 demonstrated that the most commonly selected segment size closely matched participants' memory span, as measured in a baseline span test. Experiment 2 showed that reducing the sound level of presented word lists to a suprathreshold but effortful listening level disrupted the accuracy of matching selected segment sizes with participants' memory spans. The results are discussed in terms of whether online capacity monitoring may be subsumed under other, already enumerated working memory executive functions (inhibition, set shifting, and memory updating).

### Keywords

Working memory; Metamemory

---

The characterization of a central executive as a critical element in working memory has evolved over the years from that of a single resource overseeing the storage and manipulation of recent, or recently activated, memory traces, to a system that controls such additional operations as maintaining the focus of attention, set shifting, updating working memory, and inhibition (Baddeley, 1986; Cowan, 1999; Engle, 2002; Logie, 2011; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Miyake et al., 2000).

One function of the executive system not directly included in this list is an individual's ability to monitor the moment-to-moment storage capacity remaining in working memory as the memory load incrementally increases with the arrival of additional input. This ability is not tapped in a standard memory span test, in which participants hear word or digit lists of various lengths, with instructions to recall as many items as possible as soon as the list is finished (Wechsler, 1997). Such span tests emphasize the storage component of working memory, sometimes referred to as *short-term memory* (STM), with the term *working memory* being reserved for complex span tasks that also include a concurrent processing requirement (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001).

A standard feature of the STM span test is that the examiner controls the stimulus input, with the listener left to deal as well as possible with whatever list length is presented.

Whether or not with a supraspan list the listener is aware of the point at which his or her memory span is being exceeded has not been of interest. A moment's reflection, however, suggests that listeners are, at least to some degree, aware of when their immediate memory capacity is being reached. This ability can be illustrated by an *interruption- and-recall* (IAR) paradigm, in which a listener hears spoken words arriving at a steady rate for a test of recall, but with the listener having the ability to press a key to interrupt the input at the latest point at which he or she believes perfect recall will still be possible.

This task is distinguishable from traditional studies of metamemory, in which participants are asked to estimate in advance of hearing a word list how many items they believe that they will accurately be able to recall. The interest in such studies is typically the contrast between what a person believes he or she will be able to recall versus actual recall ability. As such, the focus in metamemory studies is on offline, reflective judgments, which often vary significantly in their predictive accuracy (Bjork, 1994; Connor, Dunlosky, & Hertzog, 1997; Kornell, Rhodes, Castel, & Tauber, 2011; Nelson & Narens, 1990). By contrast, the IAR paradigm tests the listener's ability to monitor the arrival of incoming information online and to continuously update the judged difference between the amount of current content versus the "space" still remaining in working memory capacity.

We report the results of two experiments, both involving recorded word lists presented for immediate recall. In the first experiment, we used the IAR procedure to obtain two measures: the size of the segments selected by participants at the point when they judged their memory capacity to have been reached, and the number of words that they were actually able to recall. The difference between the two serves as a measure of participants' accuracy for online monitoring of their memory capacity. That is, unlike traditional metamemory judgments, these judgments must be based on a continuous difference calculation occurring contemporaneously with the ever-changing input.

In this experiment, results from the IAR paradigm were compared to word lists presented using a standard memory span procedure in which the memory set size was determined by the experimenter. These latter data served as a baseline against which to compare online capacity decisions when the listener was allowed to interrupt the input for recall when he or she believed that the capacity limit had been reached.

The second experiment was based on past research showing that less favorable listening conditions, in which perceptual effort is required for successful word recognition, result in reduced recall accuracy for word lists relative to more favorable listening conditions (Piquado, Cousins, Wingfield, & Miller, 2010; Rabbitt, 1968, 1991; Surprenant, 1999). The presumption has been that successful word recognition in such cases draws resources that would otherwise be available for encoding what has been heard in memory (McCoy et al., 2005; Murphy, Craik, Li, & Schneider, 2000; Van Boxtel et al., 2000).

In the second experiment, we varied the loudness level of word-list presentations in order to ask whether the accuracy of online monitoring of working memory capacity would also suffer when lists were presented at a low, although perceptible, sound level. That is, we asked whether online capacity monitoring draws on attentional resources, as has been argued for other components of executive function (see, e.g., Barrouillet, Bernardin, & Camos, 2004). Questions of competition for limited resources are often tested with dualtask interference paradigms, such as the effect of conducting a visual tracking task while concurrently listening to speech for later recall (Anderson, Craik, & Naveh-Benjamin, 1998; Naveh-Benjamin, Guez, & Sorek, 2007; Tun, McCoy, & Wingfield, 2009). Rather than using a traditional dual-task paradigm, we simply asked whether the need for extra listening effort, which often occurs in everyday experience, would affect online capacity monitoring.

Reports of a significant increase in the incidence of hearing loss among young adults (Shargorodsky, Curhan, Curhan, & Eavey, 2010) give this question added significance.

## Experiment 1

### Method

**Participants**—The participants were 24 university undergraduates and graduate students (seven men, 17 women) ranging in age from 18 to 24 years ( $M = 20.3$  years,  $SD = 1.45$ ). As a group, the participants averaged 14.8 years of formal education at the time of testing ( $SD = 1.07$ ), with a mean Shipley vocabulary score (Zachary, 1991) of 14.3 ( $SD = 1.92$ ). All participants received audiometric screening to ensure normal hearing acuity based on a pure-tone average (PTA) of auditory thresholds across 1, 2, and 4 kHz ( $M = 4.5$  dB HL,  $SD = 3.28$ ) and speech reception thresholds (SRTs), measured as the lowest sound level at which 50 % of recorded words could be correctly identified ( $M = 9.0$  dB HL,  $SD = 2.94$ ) (Katz, 2002).

### Stimuli and procedures

**Baseline memory span:** In the baseline span condition, participants heard 24 word lists: two lists each, at list lengths of 1–12 words. The stimulus words consisted of common one- and two-syllable concrete nouns taken from the Toronto word pool (Friendly, Franklin, Hoffman, & Rubin, 1982) and recorded on computer sound files by a female speaker of American English using SoundEdit software (Macromedia, Inc., San Francisco, CA) that digitized (16-bit) at a sampling rate of 44 kHz. Recordings were equalized within and across lists for root-mean-square (RMS) intensity using MATLAB (MathWorks, Natick, MA). The various list lengths were randomized in the order of presentation, with the constraint that one example of each list length was presented first, followed by the second example of each list length, also in random order. Participants were not told in advance the length of the list that they would be hearing. Words were presented with a 1-s interstimulus interval (ISI), measured from the offset of one word to the onset of the next word.

List presentations were initiated by a participant keypress. The participants were instructed to recall aloud as many of the words from the stimulus list as possible in any order in which the words came to mind as soon as the list had finished.

**Interruption-and-recall (IAR) condition:** In this condition, participants heard 12 lists spoken by the same female speaker and presented at the same 1-s ISI as in the baseline span condition. In this case, however, all of the lists were 12 words long.

Participants were instructed to initiate presentation of a list with one keypress, and that a second keypress would terminate the word list the moment that it was pressed. Once they had interrupted the input, their task was to recall aloud as many of the selected words as possible. Participants were told that on each trial they were to attempt to select the longest list that would still allow for 100 % recall accuracy. Selecting the longest segment size that would allow for perfect recall was stressed as the primary goal of the task, and participants were instructed not to attempt putative compensatory strategies such as taking an extra item in an attempt to make up for a missed item. As in the baseline condition, the instructions were to recall aloud as many of the words as possible in any order that they came to mind. For each IAR trial, both the number of words selected and participants' recall responses were recorded.

Participants were tested individually in a sound-attenuated testing room, with word lists presented binaurally through calibrated Eartone 3A insert earphones (E-A-R Auditory Systems, Aero Company, Indianapolis, IN), via an Interacoustics AD229e Audiometer

(Interacoustics, Assens, Denmark) at a sound level of 25 dB above each individual's better-ear SRT. This sound level represents a comfortable listening level for easy audibility (Jerger & Hayes, 1977). Stimulus words were counterbalanced such that no word was heard more than once by any participant in either the baseline or the IAR conditions, but that, by the end of the experiment, each word was heard an equal number of times in the baseline and IAR conditions. The baseline span and IAR conditions were blocked in presentation, with the order of the conditions counterbalanced across participants.

## Results

Results for the baseline span condition are shown in the left panel of Fig. 1, in which we have plotted the mean number of words recalled correctly against the number of words presented in the stimulus list. The baseline span for each participant was taken as being the maximum list length that allowed for completely accurate recall on at least one of the two baseline span trials at that length. Scoring followed a free-recall protocol in which recalled words were scored as being correct if they had appeared in the most recent list, regardless of their position in the order of response (Golomb, Peelle, Addis, Kahana, & Wingfield, 2008). The broken line at 45° in the figure represents the level of perfect possible recall.

It can be seen from this panel that for list lengths of one to four, word recall is at ceiling, and at near ceiling for a list length of five, a span that compares favorably with the range typically reported for immediate word-list recall following a single presentation (see the review in Cowan, 2001). Beyond a five-item list, additional stimulus items yielded progressively smaller average recall gains that never peaked beyond a mean of 5.8 items. Also shown in the left panel of Fig. 1 is the mean number of words correctly recalled as a function of the segment sizes selected in the IAR condition. (Points are plotted only for list lengths for which there were more than ten examples on which to base a mean.) Although the accuracy curves for the baseline and IAR conditions begin to spread apart beyond a list length of five items, a comparison of mean recall accuracy collapsing across list lengths six through nine, for which exemplars for the selfselected condition were available, confirms a relatively close fit,  $t(23) = 1.65$ , n.s.

The right panel of Fig. 1 shows the distribution of the sizes of the segments selected for recall in the IAR condition. As can be seen, these self-selected segment sizes varied widely, although there were relatively few instances in which participants chose segments of four or fewer items or of nine or more items for recall. It can be seen that the modal selected segment size of six words was close to the peak mean of 5.8 words correctly recalled, as determined in the baseline condition. This close approximation between the two figures demonstrates a good ability to calibrate segment size selections with actual memory span.

## Discussion

The results of Experiment 1 demonstrate that listeners show good accuracy in judging the point at which their maximum span for accurate recall of unstructured word lists has been reached. That is, although there was trial-to-trial variability in the segment sizes selected in the IAR condition, the participants' modal selected segment size was close to the mean span for perfect list recall, as measured in the baseline span task.

It can be argued that the ability to monitor the moment-to-moment capacity remaining in working memory may involve a "feeling-of-knowing" (FOK) judgment, as described in the metamemory literature (e.g., Nelson & Narens, 1990). In this literature, however, such judgments are typically made offline without time constraints. One can thus consider capacity monitoring in terms of an FOK judgment only to the extent that such judgments can be continuously made and modified as to-be-remembered items are arriving with a 1-s ISI. It

seems unlikely that this rapid input rate would allow participants to be able to engage, and continuously update, FOK judgments, at least as they are typically characterized (Bjork, 1994; Connor et al., 1997). By contrast, the rapidity with which listeners were able to continuously calculate the difference between the current contents of their working memory and the remaining capacity may be suggestive of an automatic process. One feature of automaticity, that the process does not consume attentional resources (Posner & Boies, 1971), was addressed in Experiment 2.

## Experiment 2

A common element among most models of executive function is that of finite attentional resources that must be shared among multiple operations (Kahneman, 1973). Although attentional demands have been manipulated by such tasks as an imposed memory preload (Baddeley & Hitch, 1974) or the need to conduct a concurrent secondary task (Anderson et al., 1998), it has been argued that perceptual effort at the time of stimulus encoding will also drain attentional resources that might otherwise be available for downstream, higher-level operations. It has been shown, for example, that acoustic masking of spoken digits (Rabbitt, 1968), syllables (Surprenant, 1999), and word lists (Piquado et al., 2010) reduces their probability of recall, even when the level of acoustic masking still allows for their correct, albeit effortful, recognition. These findings are consistent with consequences that would be predicted from Barrouillet et al.'s (2004) and others' emphasis on attentional resources as a limiting factor for effective encoding and refreshing of information in working memory (Cowan, 1999; Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001).

In Experiment 2, we used the IAR paradigm for word lists presented at two sound intensity levels in order to manipulate the degree of perceptual effort required for successful word identification. To the extent that perceptual effort draws executive resources that affect accurate moment-to-moment monitoring of the usable capacity remaining in working memory, one would predict that with a reduced sound level, accuracy in online capacity monitoring would become less effective. This would be manifested as a mismatch between these selected sizes and memory span as measured in a simple baseline span test.

## Method

**Participants**—The participants were 24 university undergraduates and graduate students (12 men and 12 women) ranging in age from 18 to 26 years ( $M = 19.8$  years,  $SD = 1.97$ ) different from those who had served in Experiment 1. All participants had normal hearing acuity based on PTAs averaged over 1, 2, and 4 kHz ( $M = 5.7$  dB HL,  $SD = 3.86$ ) and on SRTs ( $M = 8.3$  dB HL,  $SD = 2.82$ ).

Participants were randomly assigned to one of two groups: a high or a low sound-level group. For the high sound-level group, lists were presented at 25 dB above the individual's SRT, the same level that had been used in Experiment 1. In the subsequent discussion, this will be referred to as *25 dB sensation level (SL)*. For the low sound-level group, lists were presented at 10 dB above the individual's SRT (i.e., *10 dB SL*). These two sensation levels were chosen to represent speech presented at an intensity level that is typically considered to be a comfortable listening level (25 dB SL) versus speech that is intelligible but that requires listening effort (10 dB SL; Jerger & Hayes, 1977). A pretest was conducted to ensure word intelligibility for stimulus words presented at the two sound levels. None of the words used for this pretest were used in the main experiment.

Care was taken to ensure that the two presentation-level groups did not differ significantly in age ( $M_{25\text{dB}} = 20.3$  years,  $SD = 2.23$ ;  $M_{10\text{dB}} = 19.3$  years,  $SD = 1.61$ ),  $t(22) = 1.26$ , n.s., years of education ( $M_{25\text{dB}} = 14.5$  years,  $SD = 1.73$ ;  $M_{10\text{dB}} = 14.0$  years,  $SD = 1.04$ ),  $t(22) =$

0.86, *n.s.*, or verbal knowledge as measured by the Shipley Vocabulary Test (Zachary, 1991;  $M_{25\text{dB}} = 14.2$ ,  $SD = 1.53$ ;  $M_{10\text{dB}} = 14.7$ ,  $SD = 1.72$ ),  $t(22) = 0.75$ , *n.s.*

### Stimuli and procedures

In the baseline span condition, participants heard 24 word lists: two lists each at list lengths of 1–12 words, presented at either 25 or 10 dB SL, depending on the participant group. As in Experiment 1, the stimulus words consisted of common concrete nouns taken from the Toronto word pool (Friendly et al., 1982) recorded by a female speaker of American English and equalized for RMS intensity within and across lists using MATLAB (MathWorks, Natick, MA). Words were again presented with a 1-s ISI, measured from the offset of one word to the onset of the next.

The instructions for the baseline span and IAR conditions were the same as those used for Experiment 1: namely, to recall aloud as many of the words from the stimulus list as possible, in any order in which the words came to mind. For both the baseline and IAR conditions, the stimulus lists were counterbalanced across participant groups, such that by the end of the experiment, each list had been heard an equal number of times by the 25-dB SL and 10-dB SL groups. The baseline span and IAR conditions were again blocked in presentation, with the order of the conditions counterbalanced across participants.

### Results

**Baseline memory span**—As might be expected from prior studies (e.g., McCoy et al., 2005; Piquado et al., 2010; Rabbitt, 1968; Surprenant, 1999), the maximum list length that allowed for completely accurate recall on at least one of the two baseline span trials at that length revealed a significant effect of stimulus intensity ( $M_{25\text{dB}} = 5.75$  words,  $SD = 0.75$ ;  $M_{10\text{dB}} = 4.33$  words,  $SD = 1.23$ ),  $t(22) = 3.40$ ,  $p < .01$ .

The left panel of Fig. 2 shows the mean number of words correctly recalled versus the number of words presented in the stimulus list for each of the two presentation intensities in the baseline span task. As in Experiment 1, scoring followed a free-recall protocol in which recalled words were scored as being correct if they had appeared in the most recent list, regardless of their position in the order of response (Golomb et al., 2008). The broken line at 45° again represents the level of perfect possible recall.

It can be seen from this panel that for list lengths of one to three words, recall is at ceiling, and at near-ceiling for a four-item list length at both intensity levels, thus confirming the intelligibility of the stimuli at both the 25-dB SL and 10-dB SL presentation levels. Two additional features can be seen in this panel. The first is that beyond a four-item list, additional stimulus items yield progressively smaller recall gains that never peak beyond means of 5.8 items for the 25-dB SL lists and 4.3 items for the 10-dB SL lists. This value of 5.8 items at 25 dB SL replicates the baseline span accuracy level obtained in Experiment 1, in which the stimuli were also presented at 25 dB SL.

The second feature seen in the left panel of Fig. 2 is that beyond a list length of four items, there is a significant decrement in recall accuracy for the 10-dB SL lists relative to the 25-dB SL lists. This was confirmed by a two-way mixed-design analysis of variance (ANOVA) conducted on recall accuracy for list lengths 5 through 12, with list length as a within-participants variable. The ANOVA revealed a significant main effect of list length,  $F(7, 154) = 3.25$ ,  $p < .01$ ,  $\eta_p^2 = .13$ , reflecting the increase in the number of items recalled as list lengths increased, albeit with a shallow slope. More important for our present interests, however, was a significant main effect of intensity level,  $F(1, 22) = 21.33$ ,  $p < .001$ ,  $\eta_p^2 = .49$ , reflecting the diminished recall for 10-dB SL lists relative to the 25-dB SL lists that held



to relatively equivalent degrees across all of the list lengths beyond the four-item lists. We found no List Length  $\times$  Presentation Level interaction,  $F(7, 22) = 0.38, p = .91, \eta_p^2 = .017$ .

**Interruption-and-recall (IAR) condition**—The middle panel of Fig. 2 shows the distribution of segment sizes that participants selected for recall for the 25-dB SL and 10-dB SL presentation levels in the IAR condition. As can be seen, the sizes of the segments selected by participants varied widely, although there were relatively few instances in which participants chose segments either of four or fewer items or of nine or more items for recall.

The most striking feature of these two distributions of selected segment sizes is the shift in the peaks of the two distributions from a modal self-selected segment length of six words for lists, at the louder 25-dB SL level, to seven words, at the reduced-intensity 10-dB SL level. This shift was verified by a chi-square frequency analysis that confirmed a significant relationship between presentation level and modal self-selected segment sizes,  $\chi^2(1) = 20.74, p < .001$ .

To place these segment size selections in context, it can be seen that for lists presented at 25 dB SL, the modal segment size of six words was close to the mean for accurate item recall of 5.8 words determined in the baseline span condition in both Experiments 1 and 2, suggesting a good ability to calibrate segment size selections with actual memory span when stimuli were heard at a comfortable listening level.

By contrast, when the need for perceptual effort was induced by lowering the presentation level to 10 dB SL, listeners appeared to lose this close calibration. That is, a reduced memory span for accurate recall of 4.3 words, as estimated in the baseline condition, was not accompanied by listeners adaptively taking shorter segment sizes for recall.

The right panel of Fig. 2 shows, for the two presentation levels, the mean numbers of words correctly recalled versus the sizes of the list lengths selected in the IAR condition. Points are plotted only for list lengths for which more than ten examples were available on which to base a mean. A comparison of mean recall accuracy collapsing across list lengths 4 through 8, for which exemplars for both intensity conditions were available, shows reduced recall accuracy for the 10-dB SL lists relative to the 25-dB SL lists,  $t(21) = 3.95, p < .001$ .

Recall in the IAR condition reflects a more complex span task than does recall in the baseline condition because listeners are required to make continuous capacity judgments while at the same time maintaining in memory the items heard at each point. As such, this dual-task characteristic of storage and executive decision-making represents a greater cognitive load than is present for the baseline span task. As would be expected if listening effort draws on already depleted resources, visual inspection of the left and right panels of Fig. 2 suggests that, while baseline and IAR spans were similar for 25-dB SL presentations, recall accuracy in the IAR condition was reduced relative to the baseline span when lists were heard at 10 dB SL.

Although baseline span data were available for list lengths up to 12 items, not all list lengths were available in the IAR condition for the 25-dB SL presentation level. (The exception was for list lengths 4 and 5, which were at or near a performance ceiling for both the baseline and IAR conditions.) Collapsing across list lengths 6–8, for which exemplars were available for both the baseline and IAR conditions at 25 dB SL did not show a significant difference,  $t(11) = 0.48, n.s.$  A similar analysis collapsing across list lengths 6–8 for the 10-dB SL condition, however, showed significantly poorer recall in the IAR condition relative to the baseline condition,  $t(11) = 3.46, p < .01$ .

An interesting question is whether participants in the IAR condition, in general or individually, maintained the same list length across trials, which might be indicative of preplanning, or whether segment sizes varied from trial to trial, which would be more indicative of online monitoring and decision-making on each trial. A participant-by-participant, trial-by-trial examination of the segment sizes selected favored the latter interpretation for the 25-dB SL presentation levels in Experiment 1 and both the 25-dB SL and 10-dB SL presentations in Experiment 2. That is, in all cases, participants' segment sizes varied from trial to trial, although usually not by more than one or two words, and they did not appear to progressively reduce this variability, or increase the accuracy of matching segment-size selections against their baseline spans over the 12 presentation trials.

This latter point was examined by calculating the absolute difference between each participant's baseline span and his or her mean selected list length for the first, second, and third portions of the 12 IAR trials. For the 25-dB SL lists, averaged over the first third of the trials, the sizes of the segments that participants selected differed from their baseline spans by an average of 1.46 words ( $\pm 0.44$ ); for the middle third of the presentation trials, they varied by 1.58 words ( $\pm 0.50$ ); and for the final third of the presentation trials, they varied by 1.65 words ( $\pm 0.43$ ). For the 10-dB SL presentations, the corresponding values were 3.35 words ( $\pm 0.58$ ), 3.83 words ( $\pm 0.75$ ), and 3.79 words ( $\pm 0.70$ ), respectively. These data were submitted to a 2 (intensity level: 25 or 10 dB SL)  $\times$  3 (trial position: first, middle, or final third) ANOVA. A Greenhouse–Geisser correction was applied to the degrees of freedom for  $F$  values to adjust for heterogeneity of variances as needed. The analysis confirmed a significant main effect of intensity level on the difference between baseline spans and segment sizes selected,  $F(1, 22) = 7.13$ ,  $p < .025$ ,  $\eta_p^2 = .25$ , but no main effect of trial position,  $F(1.48, 32.53) = 1.68$ ,  $p = .21$ ,  $\eta_p^2 = .07$ , nor an Intensity  $\times$  Trial Position interaction,  $F(1.48, 32.53) = 0.44$ ,  $p = .59$ ,  $\eta_p^2 = .02$ .

This absence of a systematic reduction in variability, or a progressive increase in the closeness of selected segment sizes to participants' baseline spans over the course of presentation trials, is more consistent with an argument for online decision-making than deliberate preplanning under strategic control.

## Discussion

The baseline span results are in agreement with other studies showing poorer recall for verbal materials when successful recognition of the to-be-recalled words comes at the cost of perceptual effort due to acoustically difficult listening conditions (McCoy et al., 2005; Murphy et al., 2000; Piquado et al., 2010; Rabbitt, 1968, 1991; Surprenant, 1999, 2007). Consistent with these findings, reducing the presentation intensity from 25 to 10 dB SL in the baseline span condition resulted in a significant reduction from a mean without-error maximum of 5.8 to 4.3 words. Importantly, this difference in baseline spans appeared even though a pretest confirmed that both presentation levels exceeded an audibility threshold, allowing for successful word recognition. Acceptable audibility was also affirmed by the performance at ceiling for both presentation levels at the short list lengths in the baseline condition.

The IAR task required participants to monitor their current memory capacity as to-be-recalled words arrived at a fairly rapid rate, and to use this information to interrupt the input when their monitoring suggested that this capacity was about to be exceeded. Our hypothesis was that, under a condition in which perceptual effort was required to identify stimulus words presented at a low sound level, cognitive resources would need to be diverted to obtain this successful front-end word recognition. Assuming an upper limit to attentional resources (Barrouillet et al., 2004; Engle, 2002; Kahneman, 1973; McCabe et al., 2010), this draw would in turn reduce the ability to contemporaneously update the judged



difference between the current memory contents and remaining memory capacity. Consistent with this expectation, we found that when speech was heard at 25 dB SL, a comfortable listening level (Jerger & Hayes, 1977), participants' ability for online monitoring was good, with the modal segment sizes selected by participants being close to their baseline spans.

One might have expected poorer recall for the baseline condition relative to the IAR condition independent of auditory presentation level, because of a potential uncertainty effect operating in the baseline condition. That is, the mixed-list design in the baseline condition would not allow a participant to predict the length of a list as it was being heard, and it has been found by Crowder (1969) and others that lists of unpredictable lengths tend to be less well recalled than lists whose lengths are known beforehand (see also Palladino & Jarrold, 2008). As can be seen with our present data, however, with a 25-dB SL presentation level in both Experiments 1 and 2, baseline and IAR spans were similar. It might of course be argued that recall might have been superior in the IAR condition relative to the baseline condition because of the absence of list-length uncertainty in the IAR task, but that this advantage was offset by the concurrent monitoring and termination-decision activity, resulting in equivalent recall levels for the IAR and baseline conditions. Although this question cannot be answered with these present data, it should be noted that both factors, list-length uncertainty in the baseline task and task complexity in the IAR task, also operated in the 10-dB SL presentation level condition. Our finding, that the baseline and IAR spans were similar for 25-dB SL presentations, but that the recall span in the IAR condition was reduced relative to the baseline span when lists were heard at the 10-dB SL condition, would thus not be attributable to this difference in task demands.

Our working hypothesis was that perceptual effort would draw from attentional resources that might otherwise have been available for effective online monitoring and rapid determination of the point at which the capacity of working memory had been reached. We made no particular prediction about whether listeners' termination decisions would tend in the direction of shorter or longer segment sizes in the resource-demanding 10-dB SL condition relative to the less perceptually demanding 25-dB SL condition. To the extent that the listeners were able to effectively monitor their running memory capacity with the more perceptually challenging 10-dB SL lists, it would be more adaptive for them to select shorter segment sizes to correspond with the consequent reduced recall spans. That this was not the case supports the view that, under conditions of perceptual effort, participants' ability for accurate capacity monitoring was compromised.

An outstanding question is why, in the perceptually more demanding 10-dB SL condition, participants tended to take longer rather than shorter segment sizes, as compared with the perceptually less demanding 25-dB SL condition. One might speculate that in the 10-dB SL condition, with its lower recall level, participants may have sensed the loss of a to-be-recalled word and attempted to compensate by taking an extra item at input before beginning their recall. If this were the case, one might ask whether the source of this effect for the 10-dB SL lists was because the attentional demands for successful processing of the early list items interfered in a cascading fashion with monitoring and encoding of the later list items.

Traditional list-learning studies with unrelated, perceptually clear verbal stimuli typically show better recall for words at the beginnings (primacy effect) and endings (recency effect) of word lists, with a relatively greater recency than primacy effect when the instructions are for free recall (Golomb et al., 2008; Hovland, 1938). A buildup of interference as a list proceeded would lead one to expect a reduction or elimination of the traditional free-recall recency advantage. However, with these data, both the baseline and IAR recalls at both

sound levels in Experiment 2 showed the typical free-recall curve, with the recency effect being greater than the primacy effect.

Although the perceptual effort attendant to a 10-dB SL presentation level led to reduced accuracy in capacity monitoring, we cannot say with certainty why this reduced accuracy took the form of selecting longer rather than shorter segments relative to the individuals' selections at 25 dB SL, which more closely matched their baseline spans. It can be said that taking additional items in the IAR condition in an attempt to compensate for reduced recall with a 10-dB SL presentation level would not be an adaptive strategy for maximizing the stated task goal of accurate recall. Indeed, as noted, participants were specifically instructed not to attempt to extend a list to hear more items to make up for a missed item. We cannot, however, rule out this possibility.

## General discussion

It can be argued that rapid output monitoring is an inherent component of successful memory processes, whether the monitoring is done to prevent repetitions of an item already given in list recall (Kahana, Dolan, Sauder, & Wingfield, 2005) or in generating category exemplars (Rosen & Engle, 1997; Wingfield & Kahana, 2002). Such output monitoring differs in character from monitoring the moment-to-moment capacity of working memory to detect a point of impending overload of storage capacity. Both abilities, however, may be challenged when executive resources are strained by a concurrent cognitive task (e.g., Rosen & Engle, 1997) or, in the present case, when resources must be allocated to effortful perception in the face of above-threshold intensity differences.

Models of working memory have long postulated a trade-off between processing and storage, whether conceived in terms of a shared general resource (Carpenter, Miyake, & Just, 1994), a limited-capacity central executive (Baddeley & Hitch, 1974; Logie, 2011), or a time-based model in which switching attention from processing to storage, or to updating and refreshing the memory trace, is constrained by the time parameters of these processes (Barrouillet et al., 2004; Barrouillet, De Paepe, & Langerock, 2012; Towse & Hitch, 1995). The concepts of "working memory" and "executive functioning" have both been used to describe this limited-capacity system, often interchangeably. Although the former has focused on the ability to store and manipulate information, and the latter on goal-directed behavior, both abilities are associated with activity in prefrontal cortex, and each contains elements of the other (McCabe et al., 2010).

Proposed components of executive function include both task-specific functions (inhibition, set shifting, and updating) and more general functions (flexibility of thinking, concept formation, and abstract thinking; Friedman et al., 2006; Friedman et al., 2008; McCabe et al., 2010). One is thus bound to ask whether capacity monitoring, as tested in the IAR paradigm, represents an additional task-specific executive function, or whether it can be considered a special case of a previously described function. This determination is made difficult for two reasons. One of these is that these functions are often defined by the paradigms used in their test (e.g., Friedman et al., 2006; Friedman et al., 2008; McCabe et al., 2010; Miyaki et al., 2000), and second, that none of these tests of specific executive functions can necessarily be considered to be "process pure" (Jacoby, 1999; Salthouse, Atkinson, & Berish, 2003).

With this caveat, one can suggest that capacity monitoring contains elements of the three most often cited task-specific executive functions (inhibition, set shifting, and updating). It can be argued, for example, that successful execution of all executive functions must rely on inhibitory control to keep tasks on target (Hasher, Lustig, & Zacks, 2007). Effective capacity

monitoring would be no exception, as it too would require inhibition of potential off-task interference for success.

A similar case can be made for the near-ubiquitous need for set shifting in any complex span task, when shifting is broadly defined as an individual's ability to shift attention between tasks or elements within a task (e.g., Fisk & Sharp, 2003). In the IAR paradigm, the listener is required to make continuous judgments of available capacity, while at the same time maintaining in memory the items heard to that point. Effective capacity monitoring would thus require a dynamic time-based cycling of attentional resources (Barrouillet et al., 2012)—in this case, cycling between maintenance and monitoring.

Following this resource argument might lead to an a priori expectation of better recall in the baseline than in the IAR condition, because the baseline condition did not require moment-to-moment termination-point decisionmaking. As such, it could be argued that resources would be free for maintenance of the words in memory, whether through subvocal articulatory rehearsal or attentional refreshing (Camos, Mora, & Oberauer, 2011). We cannot deny this possibility because, as we have indicated, this potential advantage may have been offset by the use of a random-order design for list-length presentations in the baseline span task, resulting in unpredictable list lengths from trial to trial. This could have engaged the previously cited uncertainty effect, in which unpredictable list lengths are generally not recalled as well as lists with predictable lengths (Crowder, 1969). This degree of unpredictability was not present in the IAR condition, in which the participant maintained control of the list length. This argument also implies that capacity monitoring and decision-making draw on attentional resources.

Memory updating, as a commonly cited executive function, is typically defined as the process of discarding older information as new information is available, frequently tested by running memory (Morris & Jones, 1990; Palladino & Jarrold, 2008) or *N*-back tasks (Basak & Verhaeghen, 2011). It remains to be seen whether online monitoring of available memory capacity, as measured in the IAR paradigm, and typically defined memory updating should be considered as being separate executive functions, or whether the definition of memory updating should be broadened.

Although this definitional question remains open, our data offer support for the position that the perceptual effort needed for successful identification of degraded or lowintensity stimuli (Murphy et al., 2000; Rabbitt, 1968, 1991; Stewart & Wingfield, 2009; Surprenant, 1999, 2007; Wingfield, McCoy, Peelle, Tun, & Cox, 2006; Wingfield, Tun, & McCoy, 2005; see also Cronin-Golomb, Gilmore, Nearing, Morrison, & Laudate, 2007; Dickinson & Rabbitt, 1991) can produce interference effects ordinarily associated with dual-task and memory preload paradigms in tests of resource limits in executive functions (see the review in McCabe et al., 2010).

Consistent with this position, Rönnerberg and colleagues (Rönnerberg, Rudner, Foo, & Lunner, 2008; Rönnerberg, Rudner, & Lunner, 2011; Rönnerberg, Rudner, Lunner, & Zekveld, 2010) have argued that while a clear speech signal allows for an easy match between the phonological input and potential targets in the mental lexicon, a degraded signal leads to a shift to controlled processing, in which the phonological, lexical, and semantic representations retrieved from long-term memory interact in working memory with the information extracted from the sensory input. The negative effects of stimulus clarity and the associated perceptual effort may be attributable at least in part to a slowing in stimulus encoding, such that there may be an overlap in time in which the cognitive system is concurrently conducting perceptual and encoding operations on one stimulus as another is arriving (Miller & Wingfield, 2010; Piquado et al., 2010).

In addition to this slowing account of the effects of perceptual effort on recall, there has been a suggestion that auditory stimuli heard at a reduced intensity level, even when suprathreshold, may truncate the duration of an already rapidly fading echoic trace (Baldwin, 2007; Baldwin & Ash, 2011). If correct, this would allow less time for each word to be perceptually encoded before its echoic trace had faded, resulting in potentially fewer items being encoded in memory in the 10-dB SL condition, and hence fewer items being accurately recalled. Although a limited echoic trace may have contributed to reduced recall for these suprathreshold stimuli, the performance effects of reduced stimulus intensity on recall can be reliably demonstrated.

At the practical level, the negative consequences of perceptual effort on working memory and executive function, even when the materials pass a screen for audibility, may have special ecological significance in view of the previously cited claims of an increasing incidence of slight, mild, or more severe hearing loss among university-age young adults (Shargorodsky et al., 2010). It is especially important in this context to note that college students who have a hearing loss are often unaware of this fact (Le Prell, Hensley, Campbell, Hall, & Guire, 2011; Widen, Holmes, Johnson, Bohlin, & Erlandsson, 2009).

## Acknowledgments

This research was supported by NIH Grant No. AG019714. We also gratefully acknowledge support from the W. M. Keck Foundation and NIH Training Grant No. T32 NS007292 (to R.S.S.). We thank Nelson Cowan, Carryl Baldwin, and two anonymous reviewers for valuable input on an earlier version of this article.

## References

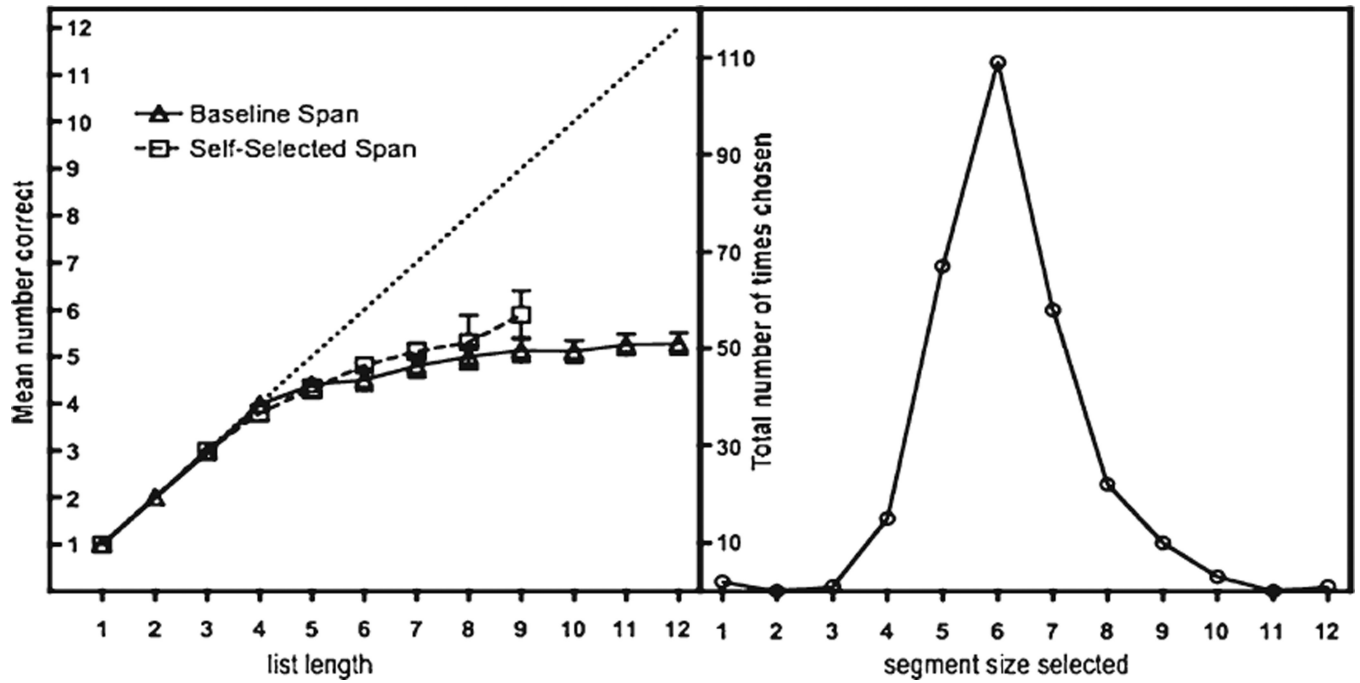
- Anderson ND, Craik FIM, Naveh-Benjamin M. The attentional demands of encoding and retrieval in younger and older adults: Evidence from divided attention costs. *Psychology and Aging*. 1998; 13:405–423. [PubMed: 9793117]
- Baddeley, A. Working memory. Oxford, UK: Oxford University Press, Clarendon Press; 1986.
- Baddeley, AD.; Hitch, GJ. Working memory. In: Bower, GH., editor. *The psychology of learning and motivation: Advances in research and theory*. Vol. Vol 8. New York, NY: Academic Press; 1974. p. 47-89.
- Baldwin CL. Cognitive implications of facilitating echoic persistence. *Memory & Cognition*. 2007; 35:774–780. [PubMed: 17848034]
- Baldwin CL, Ash IK. Impact of sensory acuity on auditory working memory span in young and older adults. *Psychology and Aging*. 2011; 26:85–91. [PubMed: 20718539]
- Barrouillet P, Bernardin S, Camos V. Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology. General*. 2004; 133:83–100. [PubMed: 14979753]
- Barrouillet P, De Paepe A, Langerock N. Time causes forgetting from working memory. *Psychonomic Bulletin & Review*. 2012; 19:87–92. [PubMed: 22184034]
- Basak C, Verhaeghen P. Aging and switching the focus of attention in working memory: Age differences in item availability but not in item accessibility. *Journals of Gerontology*. 2011; 66B: 519–526. [PubMed: 21571704]
- Bjork, RA. Memory and metamemory considerations in the training of human beings. In: Metcalf, J.; Shimamura, AP., editors. *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press; 1994. p. 185-205.
- Camos V, Mora G, Oberauer K. Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*. 2011; 39:231–244. [PubMed: 21264630]
- Carpenter, PA.; Miyake, A.; Just, MA. Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In: Gernsbacher, M., editor. *Handbook of psycholinguistics*. San Diego, CA: Academic Press; 1994. p. 1075-1122.

- Connor LT, Dunlosky J, Hertzog C. Age-related differences in absolute but not relativemetamemory accuracy. *Psychology and Aging*. 1997; 12:50–71. [PubMed: 9100268]
- Cowan, N. An embedded-processes model of working memory. In: Miyake, A.; Shaw, P., editors. *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, UK: Cambridge University Press; 1999. p. 62-101.
- Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*. 2001; 24:87–114. disc. 115-185. [PubMed: 11515286]
- Cronin-Golomb A, Gilmore GC, Nearing S, Morrison SR, Laudate TM. Enhanced stimulus strength improves visual cognition in aging and Alzheimer's disease. *Cortex*. 2007; 43:952–966. [PubMed: 17941352]
- Crowder RG. Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior*. 1969; 8:524–528.
- Dickinson CVM, Rabbitt PMA. Simulated visual impairment: Effects on text comprehension and reading speed. *Clinical Vision Sciences*. 1991; 6:301–308.
- Engle R. Working memory as executive attention. *Current Directions in Psychological Science*. 2002; 11:19–23.
- Fisk JE, Sharp CA. The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition*. 2003; 52:364–381. [PubMed: 12907181]
- Friedman NP, Miyake A, Corley RP, Young SE, DeFries JC, Hewitt JK. Not all executive functions are related to intelligence. *Psychological Science*. 2006; 17:172–179. [PubMed: 16466426]
- Friedman NP, Miyake A, Young SE, DeFries JC, Corley RP, Hewitt JK. Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology. General*. 2008; 137:201–2254. [PubMed: 18473654]
- Friendly M, Franklin PE, Hoffman D, Rubin DC. The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*. 1982; 14:375–399.
- Golomb JD, Peelle JE, Addis KM, Kahana MJ, Wingfield A. Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition*. 2008; 36:947–956. [PubMed: 18630201]
- Hasher, L.; Lustig, C.; Zacks, RT. Inhibitory mechanisms and the control of attention. In: Conway, ARA.; Jarrold, C.; Kane, MJ.; Miyake, A.; Towse, JN., editors. *Variation in working memory*. New York, NY: Oxford University Press; 2007. p. 227-249.
- Hovland CI. Experimental studies in rote learning theory: II. Reminiscence with varying speeds of syllable presentation. *Journal of Experimental Psychology*. 1938; 22:338–353.
- Jacoby LL. Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1999; 25:3–22.
- Jerger J, Hayes D. Diagnostic speech audiometry. *Archives of Otolaryngology*. 1977; 103:216–222. [PubMed: 849199]
- Kahana MJ, Dolan ED, Sauder CL, Wingfield A. Intrusions in episodic recall: Age differences in editing of overt responses. *Journals of Gerontology*. 2005; 60B:P92–P97. [PubMed: 15746023]
- Kahneman, D. *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall; 1973.
- Kane MJ, Bleckley MK, Conway ARA, Engle RW. A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology. General*. 2001; 130:169–183. [PubMed: 11409097]
- Katz, J. *Handbook of clinical audiology*. 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2002.
- Kornell N, Rhodes MG, Castel AD, Tauber SK. The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*. 2011; 22:787–794. [PubMed: 21551341]
- Le Prell CG, Hensley BN, Campbell KCM, Hall JW, Guire K. Evidence of hearing loss in a “normally-hearing” college-student population. *International Journal of Audiology*. 2011; 50:S21–S31. [PubMed: 21288064]
- Logie RH. The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*. 2011; 20:240–245.

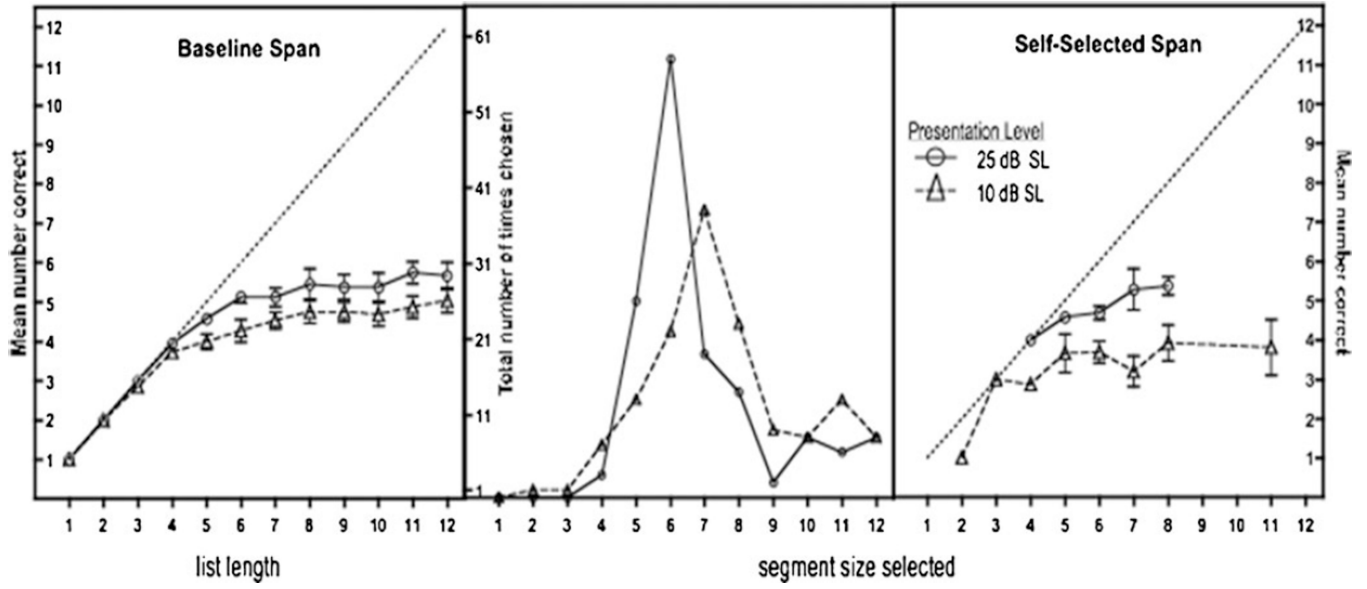


- McCabe DP, Roediger HL III, McDaniel MA, Balota DA, Hambrick DZ. The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*. 2010; 24:222–243. [PubMed: 20230116]
- McCoy SL, Tun PA, Cox LC, Colangelo M, Stewart RA, Wingfield A. Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology*. 2005; 58A:22–33. [PubMed: 15881289]
- Miller P, Wingfield A. Distinct effects of perceptual quality on auditory word recognition, memory formation and recall in a neural model of sequential memory. *Frontiers in Systems Neuroscience*. 2010; 4:14. [PubMed: 20631822]
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wagner TD. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*. 2000; 41:49–100. [PubMed: 10945922]
- Miyake A, Friedman NP, Rettinger DA, Shah P, Hegarty M. How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology. General*. 2001; 130:621–640. [PubMed: 11757872]
- Morris N, Jones DM. Memory updating in working memory: The role of the central executive. *British Journal of Psychology*. 1990; 81:111–121.
- Murphy DR, Craik FIM, Li KZH, Schneider BA. Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging*. 2000; 15:323–334. [PubMed: 10879586]
- Naveh-Benjamin M, Guez J, Sorek S. The effects of divided attention on encoding processes in memory: Mapping the locus of interference. *Canadian Journal of Psychology*. 2007; 61:1–12.
- Nelson, TO.; Narens, L. Metamemory: A theoretical framework and new findings. In: Bower, GH., editor. *The psychology of learning and motivation*. Vol. Vol. 26. New York, NY: Academic Press; 1990. p. 125-173.
- Palladino P, Jarrold C. Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology*. 2008; 61:392–399.
- Piquado T, Cousins KAQ, Wingfield A, Miller P. Effects of degraded sensory input on memory for speech: Behavioral data and a test of biologically constrained computational models. *Brain Research*. 2010; 1365:48–65. [PubMed: 20875801]
- Posner MI, Boies SJ. Components of attention. *Psychological Review*. 1971; 78:391–408.
- Rabbitt PMA. Channel capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*. 1968; 20:241–248. [PubMed: 5683763]
- Rabbitt PMA. Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngologica*. 1991; 476(Suppl):167–176.
- Rönnerberg J, Rudner M, Foo C, Lunner T. Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*. 2008; 42:S99–S105. [PubMed: 19012117]
- Rönnerberg J, Rudner M, Lunner T. Cognitive hearing science: The legacy of Stuart Gatehouse. *Trends in Amplification*. 2011; 20:1–9.
- Rönnerberg J, Rudner M, Lunner T, Zekveld AA. When cognition kicks in: Working memory and speech understanding in noise. *Noise & Health*. 2010; 12:263–269. [PubMed: 20871181]
- Rosen V, Engle RW. The role of working memory capacity in retrieval. *Journal of Experimental Psychology. General*. 1997; 126:211–227. [PubMed: 9281831]
- Salthouse TA, Atkinson TM, Berish DE. Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology. General*. 2003; 132:566–594. [PubMed: 14640849]
- Shargorodsky J, Curhan SG, Curhan GC, Eavey R. Change in prevalence of hearing loss in US adolescents. *Journal of the American Medical Association*. 2010; 304:772–778. [PubMed: 20716740]
- Stewart R, Wingfield A. Hearing loss and cognitive effort in older adults' report accuracy for verbal materials. *Journal of the American Academy of Audiology*. 2009; 20:147–154. [PubMed: 19927677]

- Surprenant AM. The effect of noise on memory for spoken syllables. *International Journal of Psychology*. 1999; 34:328–333.
- Surprenant AM. Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging, Neuropsychology, and Cognition*. 2007; 14:126–143.
- Towse JN, Hitch GJ. Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology*. 1995; 48A:108–124. [PubMed: 7754077]
- Tun PA, McCoy S, Wingfield A. Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*. 2009; 24:761–766. [PubMed: 19739934]
- Van Boxtel MPJ, van Beijsterveldt CEM, Houx PJ, Anteunis LJC, Metsemakers JFM, Jolles J. Mild hearing impairment can reduce verbal memory performance in a healthy adult population. *Journal of Clinical and Experimental Neuropsychology*. 2000; 22:147–154. [PubMed: 10649553]
- Wechsler, D. Wechsler adult intelligence scale—WAIS III. 3rd ed. San Antonio, TX: Harcourt Assessment; 1997.
- Widen SE, Holmes AE, Johnson T, Bohlin M, Erlandsson SI. Hearing, use of hearing protection, and attitudes towards noise among young American adults. *International Journal of Audiology*. 2009; 48:537–545. [PubMed: 19842807]
- Wingfield A, Kahana MJ. The dynamics of memory retrieval in older adulthood. *Canadian Journal of Experimental Psychology*. 2002; 56:187–199. [PubMed: 12271749]
- Wingfield A, McCoy SL, Peelle JE, Tun PA, Cox LC. Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology*. 2006; 17:487–497. [PubMed: 16927513]
- Wingfield A, Tun PA, McCoy SL. Hearing loss in adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*. 2005; 14:144–148.
- Zachary, RA. Shipley institute of living scale: Revised manual. Los Angeles, CA: Western Psychological Services; 1991.



**Fig. 1.** Mean numbers of words correctly reported as a function of the number of words presented (baseline span) and the number of words selected in the interruption-and-recall (IAR) condition (left panel), and the distribution of segment sizes selected in the IAR condition (right panel)



**Fig. 2.** Mean numbers of words correctly recalled as a function of the numbers of words presented (baseline span) for words presented at either 25 or 10 dB SL (left panel); the distribution of segment sizes selected in the interruption-and-recall (IAR) condition at the two sound levels (middle panel); and the numbers of words recalled as a function of the number of words selected in the IAR condition at the two presentation levels (right panel)