Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 5, 1106–1118

Copyright 2004 by the American Psychological Association
0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.5.1106

The Irrelevant Sound Phenomenon Revisited: What Role for Working Memory Capacity?

C. Philip Beaman
University of Reading

High-span individuals (as measured by the operation span [OSPAN] technique) are less likely than low-span individuals to notice their own names in an unattended auditory stream (A. R. A. Conway, N. Cowan, & M. F. Bunting, 2001). The possibility that OSPAN accounts for individual differences in auditory distraction on an immediate recall test was examined. There was no evidence that high-OSPAN participants were more resistant to the disruption caused by irrelevant speech in serial or in free recall. Low-OSPAN participants did, however, make more semantically related intrusion errors from the irrelevant sound stream in a free recall test (Experiment 4). Results suggest that OSPAN mediates semantic components of auditory distraction dissociable from other aspects of the irrelevant sound effect.

The effects of auditory distraction, variously named the *irrelevant sound*, *irrelevant speech*, or *unattended speech* effects (Beaman & Jones, 1997; Jones, Miles & Page, 1990; Salamé & Baddeley, 1982) on an immediate recall task are well documented. In brief, a series of changing utterances or pure tones presented during, or immediately subsequent to, the presentation of a to-be-recalled list produce a decrement in recall performance usually estimated at approximately 30% to 50% (Ellermeier & Zimmer, 1997; Jones, Beaman, & Macken, 1996), even though participants are asked to ignore any noises they may hear. The main aim of this line of research is to understand the extent of the analysis of an unattended message and how attentional selection interacts with immediate memory (for an in-depth review of current knowledge, see Banbury, Tremblay, Macken, & Jones, 2001).

Findings from research into the effects of irrelevant sound have informed a number of contemporary models of memory function, including the working memory model of Baddeley (1986), the object-oriented episodic record model of Jones (1993), and the feature model of Neath and Nairne (Nairne, 1990; Neath & Nairne, 1995; Neath, 2000). All of these models have enjoyed some degree of support from research into the irrelevant sound effect, but despite the ongoing high level of research into the phenomenon, there are a number of questions yet to be answered. Possibly the most salient of these questions is concerned with individual differences in susceptibility to auditory distraction or irrelevant sound interference and how, if there are such differences, they relate to other measures or mechanisms of attentional control.

A study by Ellermeier and Zimmer (1997) established that there are substantial and reliable individual differences in the susceptibility of participants to irrelevant sound interference. Ellermeier and Zimmer documented an average increase in errors of approximately 50% when memory was tested in the presence of auditory distracters, with performance ranging from a decrease in recall errors of 15% to an increase of 329%. Crucially, Ellermeier and Zimmer confirmed that these differences in performance remained stable over time, although they were unable to determine what the basis of these individual differences might be. They were, however, able to rule out a number of possibilities: There was no difference between the genders in the extent to which they were affected by the irrelevant sound distracters (although men were more likely to deny any effect than women), and there was no correlation between short-term memory span and susceptibility to irrelevant sound ($r = .01$, *ns*; a result recently replicated by Neath, Farley, & Surprenant, 2003).

This latter result is intriguing and also somewhat surprising. Increasing the number of to-be-recalled items in any given list has a catastrophic effect on overall levels of recall (Guildford & Dallenbach, 1925) and if, as a number of models suggest, irrelevant speech takes up space within a *phonological store* (Baddeley, 1986) or on an *episodic surface* (Jones et al., 1996), one might anticipate that those individuals with restricted memory capacity would be particularly vulnerable to the effects of concurrent irrelevant speech. Consistent with this idea, there is evidence that increasing the number of items within the irrelevant speech stream also increases the size of the irrelevant speech effect (Bridges & Jones, 1996; Campbell, Beaman, & Berry, 2002). However, it is also possible that, in looking for a relationship between memory span and the irrelevant sound effect, the choice of span measure to be used is crucial. Although Ellermeier and Zimmer (1997) failed to find a correlation between short-term memory span and susceptibility to auditory distraction, a different working-memory span measure has been shown to differentiate between participants who are likely to report hearing their own name on an "unattended" channel while simultaneously shadowing an "attended" message and those who are less likely to recognize their own name in the unattended stream (Conway, Cowan, & Bunting, 2001). In this "cocktail party phenomenon" (Cherry, 1953), high-span participants were less likely than low-span participants to recognize their own name in the unattended stream. Presumably, high-span participants were capable of tighter attentional control, filtering out or

---

I thank Al Harvey for critical readings of earlier versions of this article, Natalie Lynch for running Experiment 3, and Sophie Scott for useful discussions. Nelson Cowan and Emily Elliott suggested Experiment 3.

Correspondence concerning this article should be addressed to C. Philip Beaman, School of Psychology, University of Reading, Earley Gate, Whiteknights, Reading RG6 6AL, United Kingdom. E-mail: c.p.beaman@reading.ac.uk

attenuating the processing of information in the irrelevant auditory stream more effectively than low-span participants. From this result, one might conclude that working memory-span measures are a useful means of distinguishing individuals capable of focusing attention and ignoring irrelevant information (e.g., Engle, 1999).

Elsewhere in the literature, working memory span has proved to be more informative than the short-term memory span used by Ellermeier and Zimmer (1997; the number of digits correctly recalled in a control condition) when examining individual differences in cognitive tasks (e.g., Daneman & Carpenter, 1980). Research that has used the working memory span technique has, for example, established a relationship between working memory span and a number of measures of inhibitory capacity (see Daneman & Merikle, 1996, and Engle, Kane, & Tuholski, 1999, for reviews). Characterizations of the role of working memory span have suggested that individual differences in span are important in situations where “there is value in maintaining some task information *in the face of distraction and interference*” [italics added] or when “there is value in *suppressing or inhibiting information irrelevant to the task*” [italics added] (Engle, Kane, et al., 1999, p. 104).

Although simple short-term memory span and working memory span measures are related, they are not identical and can be distinguished (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Kail & Hall, 2001). It is possible, therefore, that a working memory span measure might correlate with susceptibility to irrelevant sound, whereas short-term memory span does not. In the study by Conway et al. (2001), the version of working memory span known as the *operation span* (OSPAN) task (Turner & Engle, 1989) was used. In the OSPAN task, a series of displays are presented. Each display contains a simple mathematical problem and an unrelated word. The OSPAN score is based on the length of the series that can be processed with the subject still able to recall all of the words. Individuals who score highly on this sort of task have been shown to be more efficient at blocking out or inhibiting distracting information in a number of domains (e.g., Conway & Engle, 1994; Conway, Tuholski, Shisler, & Engle, 1999; Gernsbacher, 1993; Hasher & Zacks, 1988; Rosen & Engle, 1998); thus, a measure of this type seems more likely to yield the expected correlate with susceptibility to auditory distraction than the simpler digit recall score reported by Ellermeier and Zimmer (1997). However, although the OSPAN task has proved extremely useful at predicting performance in a number of cognitive domains, it is a complex construct that is less well understood than the simpler short-term memory span task.

In Experiment 1, I examined the OSPAN task used by Conway et al. (2001). In the study by Conway et al., OSPAN predicted the likelihood of processing one’s own name when presented on a supposedly unattended auditory channel, suggesting that it is a good measure of ability to control involuntary attention shifts. The OSPAN measure should, therefore, act as a predictor of susceptibility to irrelevant sound interference if attention is needed to block or inhibit the irrelevant sound from entering the processing system or provoking an orienting response. Conway et al. replicated the procedure of Turner and Engle (1989) in administering the working memory test. In Turner and Engle’s procedure, the participant’s task, when presented with a display of a mathematical operation and an unrelated word (e.g., IS $(6 + 4)/2 = 5$? DOG) is to say the equation aloud, answer “yes” or “no” as to whether the equation is true, and then say the word. After a certain number of these displays, the participant is prompted to recall the words in correct serial order. The key feature in this procedure, for current purposes, is that the display is read aloud. If one imports the results of the literature on serial and free recall to the operation span procedure, it is clear that the read-aloud instructions could have one of two possible consequences: Either the words could be converted into an auditory memory trace, producing the well-known *modality effect,* which enhances recall performance for the final few words (Bjork & Whitten, 1974; Conrad & Hull, 1968), leading to an overall superiority in performance on read-aloud span measures compared with silent span measures, or an auditory feedback loop could be created, which in effect provides a self-initiated irrelevant sound condition leading to a performance decrement.

The former possibility is unlikely to entail too many consequences for the interpretation of working memory span results, although observing a modality effect would strengthen claims for different processing streams for auditory-verbal and visual-verbal memory (Penney, 1989). The latter possibility, however, implies that reading aloud the equation and its solution will create an irrelevant sound stream concurrent with the memory task. By definition, therefore, participants capable of scoring highly in this memory span task will be those capable of ignoring the auditory feedback created by their own concurrent articulations. Thus, the predictive value of OSPAN, in these circumstances, may be derived from the way in which the task is administered. Because the OSPAN score is based on the number of wholly correct lists, this problem of auditory feedback in the read-aloud span task may be particularly acute. The smallest amount of auditory distraction could have potentially devastating results on the OSPAN score. It is important, therefore, to determine whether the read-aloud presentation of the OSPAN task does lead to poorer span performance than a read-quietly version of the same task.

Experiment 1

Method

Participants. Thirty-eight students of the University of Reading participated in return for course credit. All reported normal hearing and normal or corrected-to-normal vision.

Materials and design. Participants were tested on two span measures, adapted from the OSPAN procedure employed by Conway et al. (2001). All materials were presented using a Macintosh Performa 5400/160 Power personal computer running HyperCard software. The stimuli were presented, center-justified, on the Macintosh screen in 28-point Geneva font. Two examples were presented followed by 15 experimental trials for each span measure (aloud and quiet). A randomly selected half of the participants were given the “aloud” span test first, and the remaining half were presented with the “quiet” span first. The stimuli for the OSPAN procedure consisted of a number of visual displays containing a mathematical operation and an unrelated word. The participant was required to respond “yes” or “no” to the mathematical operation by clicking on the appropriate button at the foot of the computer screen and, after a series of such visual displays, to recall the words presented when cued by visual presentation of the word *RECALL* presented in the same font and location as the previous visual stimuli. The number of visual displays presented (*list-length*) varied between two and six items before recall was cued. There were three trials at each list-length, and order of presentation of the list lengths was randomly determined but was the same random order for each participant.

Procedure. Participants were told that they were to be tested on a number of working-memory span tests. The procedure was to answer questions about a series of mathematical operations and then recall the words presented next to the mathematical operations. Participants were given two examples of this procedure before beginning the experimental trials. They were told that they should determine whether the operation was correct and respond by clicking on the appropriate button. Half the participants were given the aloud procedure instructions first, the remainder received the quiet procedure instructions first. In the aloud procedure, participants were asked to read the equation aloud before deciding whether it was true, then they read aloud the word. In the quiet procedure, participants were asked to read both the equation and the word silently. Participants proceeded to the next display by clicking on the *next* button at the foot of the computer screen. Participants were asked to proceed through the series of displays as quickly and accurately as possible. At the end of each series of displays, the recall cue, the word *RECALL* visually presented in the same manner as the OSPAN stimuli, appeared. Participants were asked, when cued to recall, to write on the response sheet provided all the words they could remember from each series in the order in which they had been presented. Participants in the read-aloud condition were reassured that they were not required to read aloud the recall cue, merely to begin their recall attempt. After the participants had completed 15 trials under the first presentation condition (aloud or quiet), they were asked to complete the final 15 trials using the remaining presentation mode. The scoring procedure for the OSPAN tasks was identical to that of Conway et al. (2001). The OSPAN score was the cumulative number of words recalled in those series that were perfectly recalled in correct serial order, with no points awarded for imperfect recall of a series. Serial position curves for the average number of correct responses in each list length (2–6) and each condition (read aloud or read quietly) were also produced.

Results

Both methods of presenting the span task resulted in a surprisingly wide range of working memory span scores, from between 8 (minimum) and 60 (maximum) in the quiet presentation condition and between 2 and 60 in the read-aloud presentation condition. A paired sample *t* test on the span scores showed that there was a significant difference between the quiet and aloud versions of the working memory operation span measure, $t(37) = -4.93$, $p < .001$. Performance was superior in the quiet presentation condition, as would be expected if reading aloud created an irrelevant sound–articulation feedback loop (mean OSPAN in quiet feedback equaled 30; mean OSPAN in auditory feedback equaled 21.55). There was also, however, a strong correlation between the two measures ($r = .74$, $N = 38$, $p < .001$).

Although not the main focus of the current investigation, it is informative to examine the pattern of errors distributed across the lists as revealed by the serial position data (see Figure 1). Data are seldom, if ever, reported in this much detail in a span scoring procedure, as typically the purpose of the span procedure is to identify a level of performance at which individuals are unlikely to make errors. Serial position data, in contrast, identify the list positions in which errors, averaged over participants, are most likely to occur. The serial position data presented here are, however, of considerable interest because they show a different pattern of errors for the read-quietly and read-aloud conditions. Primacy effects appear for both conditions but, consistent with Bjork and Whitten's (1974) continuous, or intermittent, distracter technique, a recency effect for the last 1 or 2 items appears, regardless of list length, only in the read-aloud condition (although Bjork and Whitten's procedure was applied to free rather than serial recall). In the read-quietly condition, all semblance of recency vanishes beyond lists longer than three to-be-recalled items. Read-aloud recency is sufficient to bring performance up to the same level of performance for the final item, but unlike previous modality-effect investigations, the level of performance is no higher than in the read-quietly condition. Because the recall cue is visual and recall itself is written, this failure to observe an auditory advantage at the final serial position cannot be the accidental by-product of an auditory suffix and is, presumably, a function of the overall low level of performance in the read-aloud condition.

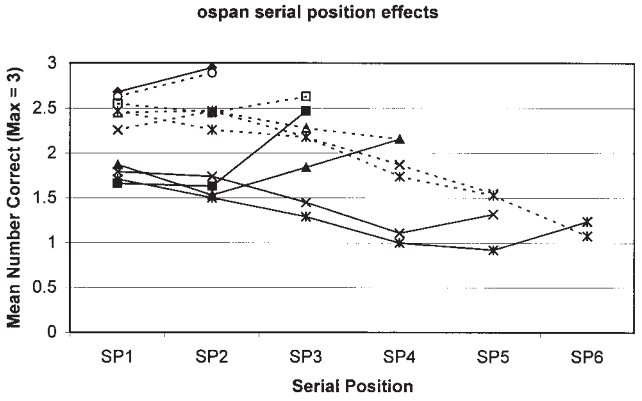

Figure 1. Experiment 1: The effects of mode of presentation on the number of correct recalls in working memory operation span task by serial position. Dashed lines (with open symbols) indicate read-quietly conditions; solid lines (with filled symbols) indicate read-aloud conditions for each of the span lengths presented (2–6). In both conditions, diamonds indicate lists of length 2, squares indicate length 3, triangles indicate length 4, Xs indicate length 5, and crossed Xs indicate lists of length 6. Max = maximum; ospan = operation span; SP = serial position.

Discussion

The results of Experiment 1 confirm the suspicion that reading-aloud the stimuli as they appear in an OSPAN task results in a substantial underestimate of a participant's working memory capacity. From the point of view of an individual differences investigation, this underestimate may, however, be unimportant. Provided that the two OSPAN tests discriminate between those with low and high spans on the same basis, there are no consequences for using either testing procedure when examining the effect of working memory span on any given cognitive task. However, the finding of modality-based differences in span performance points the way to further investigation of the separate processing components of the OSPAN task. The read-aloud version of the task evidently benefits from the extra auditory information underlying the auditory-recency evident in Figure 1, although this extra information is insufficient to overcome the disadvantages of reading aloud the mathematical operations irrelevant to the memory task. Once again, there are two obvious explanations for this disruption to memory performance.

The first possibility is that the act of articulating irrelevant material was disruptive because of non-specific attentional demands related to setting up and running the appropriate articulatory motor program (Meiser & Klauer, 1999). The other possibility is that, as previously indicated, articulating the mathematical operations creates an irrelevant auditory-feedback loop that must be successfully ignored in order to perform well at the OSPAN task. If the latter possibility is correct, then it may be that studies such as that of Conway et al. (2001), showing a difference in sensitivity to irrelevant or to-be-ignored auditory information between low- and high-OSPAN participants, reflect the de facto incorporation of a measure of ability to ignore auditory distracters when administering a read-aloud version of the working memory span task that would not be present in read-quietly versions of the same task.

Experiment 2

The aim of Experiment 2 was to determine whether OSPAN predicts susceptibility to the disruptive effects of irrelevant sound, a finding that would provide a locus for the individual differences in irrelevant sound effects currently lacking from models of immediate memory. The experiment also addressed the issue of whether an unintended element of the read-aloud procedure of administering the OSPAN task, the necessity to ignore one's own articulations, might have contributed to Conway et al's finding that high-span participants were less likely to report hearing their own name on an irrelevant auditory channel than were low-span participants. Participants were administered a standard serial recall test with irrelevant speech and performance on both read-aloud and read-quietly OSPAN measures will then be correlated with the extent of the irrelevant sound interference to determine whether either or both measures predict individual differences in disruption to cognitive (memory) operations by irrelevant sound distracters. A measure of proactive interference, the extent of positional intrusions from previous lists, was also compared with the span measures to confirm whether either version of the OSPAN task is predictive of resistance to types of interference other than irrelevant sound.

Method

Participants. The participants who had previously taken part in Experiment 1 were contacted and recruited to participate in Experiment 2. Approximately equal time had elapsed for each of these participants before they took part in Experiment 2. Because the participants were retested in approximately the same order as they had participated in Experiment 1, this time lag was no longer than 1 month for any individual.

Materials and design. All materials were presented using a Macintosh Performa 5400/160 Power personal computer running HyperCard software. Auditory stimuli consisted of a series of five single syllable consonant-vowel-consonant nonwords recorded in a male voice at 8-bit resolution and a sampling rate of 22 KHz using SoundEdit software. Each nonword was digitally edited to a length of 500 ms with 100 ms interstimulus interval. The auditory stimuli were presented using Sony MDR-CD470 stereo headphones.

The to-be-recalled stimuli comprised 60 lists of digits, 20 five-digit lists, 20 seven-digit lists, and 20 nine-digit lists. List lengths were varied in order to produce a measure of irrelevant sound interference that was not contaminated by possible floor or ceiling effects for particular individuals or particular memory loads. All participants saw the same lists in the same order. The five-digit lists were presented first, followed by the seven- and then the nine-digit lists. Lists were made up of random selections, without replacement, of the digits 1–9. The digits were presented on the computer screen at a rate of one per second (500 ms on, 500 ms off). Digits were presented center-justified in 48-point Geneva font. Following each list there was a 9-s retention interval before the visual presentation of the cue *Recall* in the same manner as the digits. The recall cue remained on for 2 s. Half of the lists of each length were randomly selected and for those lists, irrelevant sound was played during the retention interval. The irrelevant sound consisted of the list of nonwords *chun, drike, foon, gluck, jerf.* This list was spoken in a male voice and played twice.

Procedure. Participants were told that they would see lists of digits appearing one at a time on the computer screen, and after a short pause a recall cue would appear. When they saw the recall cue, they were required to write down on the response sheet provided the digits they had seen in the order in which they appeared. Participants were also informed that some of the time they would hear speech over the headphones. They were asked to ignore the speech and were reassured that they would not be tested on it in any way.

Results

The size of the irrelevant sound effect was operationalized, as in Ellermeier and Zimmer (1997), by subtracting the total number of errors in the control conditions from the total number of errors in the irrelevant sound conditions to produce an irrelevant sound susceptibility (ISS) score for each participant. The results of this procedure are shown in Figure 2, broken down by length of to-be-recalled list. Summing over the different list lengths, this resulted in a range of ISS scores from −3 to 83, with mean ISS of 27.79 ($SD = 18.53$). Analysis showed that neither the ISS nor the OSPAN scores differed significantly from normality (Kolmogorov-Smirnov test, Lilliefors correction, $p > .05$ in both cases). A regression analysis indicated that OSPAN (quiet) failed to account for a significant amount of the variance in irrelevant sound disruption, $R^2 = .002$, $F(1, 36) = .26$, $p = .61$. A second analysis found that the amount of irrelevant sound disruption accounted for by OSPAN (aloud), controlling for the earlier span measure, was slight and also nonsignificant, R^2 change $= .03$, $F(1, 35) = .93$, $p = .34$. The results of these analyses thus fail to provide any evidence that OSPAN underlies individual differences in the irrelevant sound effect.

A further correlation between the OSPAN scores and a measure of positional intrusions from past lists (e.g., Conrad, 1959, 1960; Estes, 1991; Nairne, 1991) was also carried out to determine whether the span measure predicted any capability to inhibit irrel-

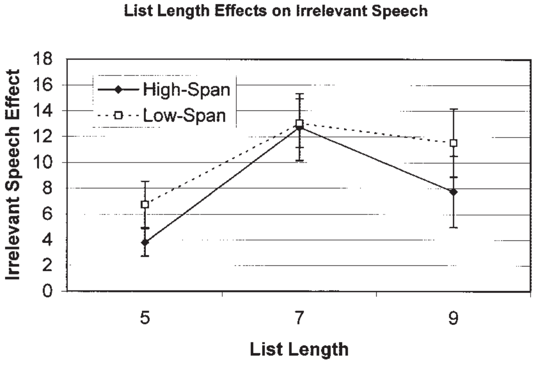

Figure 2. Experiment 2: Irrelevant sound effects on a serial memory task with memory load varying between five, seven, and nine to-be-recalled digits. Effect size is calculated as the number of errors in irrelevant sound conditions minus the number of errors in quiet conditions. Error bars are standard errors.

evant information in a serial recall task. Previous investigations of working memory span as a measure of controlled attention have established that high-span participants are less likely to suffer from proactive interference in prose recall (Rosen & Engle, 1998), that span itself is influenced by interference, and that differences in span may be due to differences in the ability to overcome interference (Lustig, May, & Hasher, 2001). On this basis, a negative correlation between OSPAN and positional intrusions from preceding lists is expected in serial recall. Two positional intrusion measures were used. The total positional intrusions measure was a simple count of intrusions from immediately preceding lists. An error was counted as a positional intrusion if the erroneous item was recalled at the same serial position on the immediately preceding list. High-OSPAN participants were expected to make fewer overall errors; thus, to compensate for differing total number of errors between participants and between experimental conditions, the total number of positional intrusions was also converted into a proportion of the total number of errors. Significant negative correlations were found between the total number of errors classified as positional intrusions and OSPAN[1] ($r = -.54$, $N = 38$, $p = .002$), which confirms that low-span participants in both measures were more likely to show failures of inhibition than were high-span participants. The correlation between the OSPAN score and the proportion of total errors classified as positional was also significant ($r = -.28$, $N = 38$, $p < .05$).

Because of the strong rank correlation between the aloud and quiet working memory spans ($\rho = .74$, $p < .001$) categorical classification of the individuals as high or low span for the purposes of analysis was carried out on the basis of the aloud OSPAN score. This and later analyses will continue to use the aloud OSPAN score for purposes of maintaining consistency with the existing literature. Classification was by median split, with the top 50% of participants, those scoring 20 or more, assigned to the high-span group. This group had an OSPAN range of 20–60 ($M = 33.37$, $SD = 12.31$). The low-span group had an OSPAN range of 2–19 ($M = 9.74$, $SD = 4.92$), giving a mean difference in OSPAN scores between the two groups of 23.63. When this was entered into a mixed analysis of variance (ANOVA), with all power calculations computed using $\alpha = .05$, a significant effect of irrelevant sound on the total number of errors made in the serial recall task was observed, $F(1, 36) = 86.41$, $MSE = 56.60$, $p < .001$, effect size (partial η^2) = .706, observed power = .999, that interacted with list length, $F(2, 72) = 6.75$, $MSE = 41.28$, $p = .004$, partial $\eta^2 = .158$, observed power = .907 (with Huynh-Feldt correction for nonsphericity) but not with OSPAN, $F(1, 36) = 1.39$, $MSE = 56.60$, $p = .25$, partial $\eta^2 = .037$, observed power = .209.

The main effect of list length was significant, $F(2, 72) = 176.30$, $MSE = 111.94$, $p < .001$, partial $\eta^2 = .83$, observed power = .999, and interacted significantly with OSPAN, $F(2, 72) = 4.86$, $MSE = 111.94$, $p = .01$, partial $\eta^2 = .119$, observed power = .786. The between-participants factor of OSPAN was significant, $F(1, 36) = 161.63$, $MSE = 776.69$, $p = .048$, partial $\eta^2 = .104$, observed power = .512, but the interaction between span, list length and speech was not significant, $F < 1$.

An alternative classification was also attempted using the top and bottom 25% of the OSPAN scores to identify high- and low-span groups. This classification produced two groups ($N = 10$ in each group) with mean OSPAN scores of 42.30 ($SD = 10.47$, range: 28–60) and 6.20 ($SD = 2.44$, range: 2–10), giving a mean difference in OSPAN scores between the two groups of 36.10. Thus, a stronger manipulation of OSPAN was obtained but at the expense of smaller group size. When this reclassification was attempted, however, none of the factors involving span and list length proved to be significant. The effects of irrelevant sound remained statistically reliable, $F(1, 18) = 40.26$, $MSE = 64.91$, $p < .001$, partial $\eta^2 = .691$, observed power = .999, but did not interact significantly with OSPAN, $F < 1$, partial $\eta^2 = .002$, observed power = .054.

Discussion

This study provided several novel observations from a single experimental design; thus, it would be helpful to summarize each before attempting an interpretation of these results. To begin with, both OSPAN measures correlated negatively with the number of positional intrusions from previous lists in the serial recall task. This result extends past research into the relationship between working memory span and proactive interference. Previous research has shown that working memory span predicts susceptibility to proactive interference in paired associate (Rosen & Engle, 1998) and prose recall (Lustig, May, & Hasher, 2001) tasks. The data reported here add to this literature by demonstrating that proactive interference within short-term memory paradigms, such as serial recall, can also be predicted on the basis of working memory (OSPAN) results. It is interesting to note that the nature of this interference differs from that reported earlier, in being interference from past positional (order) information rather than simply item information. These results support the idea that OSPAN is a useful measure of capability to withstand proactive interference from many different types of information. It is important that both the modalities of OSPAN measurement (read-quietly and read-aloud) correlated with the incidence of positional intrusions, confirming that the OSPAN task, however administered, predicts resistance to some forms of interference.

However, OSPAN (quiet) did not correlate with ISS, and OSPAN (aloud) added no further predictive power when the contribution of OSPAN (quiet) was controlled for. It would appear, therefore, that contrary to expectations, although there is no confound in Conway et al's (2001) read-aloud OSPAN measure, OSPAN is not a good measure of susceptibility to auditory distraction even if an auditory distraction component (reading aloud the displays) is built into the OSPAN measure. This was further confirmed by analysis of variance, which although displaying significant effects of irrelevant sound, list length and OSPAN failed to show any significant interaction between OSPAN and irrelevant sound. Two means of classifying participants on the basis of their OSPAN were attempted. The first method classified half the participants as low-span and half as high-span, but this produced a nonsignificant result and only a very small effect size (partial $\eta^2 < .04$). The second method attempted to boost the strength of the OSPAN manipulation by comparing only the upper and lower quartiles of the OSPAN scores, thus reducing the

---

[1] Correlations with the aloud span are reported here because this is consistent with the existing literature; however, similar correlations carried out using the quiet span gave identical results.

possible overlap between the two groups (a method used by Engle and colleagues, e.g., Engle, 1999; Engle, Kane et al., 1999). The resulting loss in sample size, however, had the effect of reducing the effect size to .002, resulting in a severe drop in observed power. At least as far as the irrelevant sound effect is concerned, it appears that strong manipulations of OSPAN are an unsatisfactory means of investigating the effect without simultaneously maintaining a large sample size. This issue will be considered further in later experiments.

The failure to find any connection between span and susceptibility to irrelevant sound interference raises questions concerning how these data are to be reconciled with those of Conway et al. (2001). In the study by Conway et al., OSPAN differentiated between participants processing their own name when presented on an unattended channel (low-span participants) and those who showed no evidence of such processing (high-span participants). A simple explanation of the Conway et al. data is that low-span participants were those who were unable to focus their attention fully on the primary task and allowed their attention to wander to the supposedly unattended channel. If this is correct, however, then the greater degree of attentional drift or orienting to the irrelevant sound exhibited by low-span participants clearly did not produce any further detriment to their serial recall scores relative to that of their high-span colleagues. This calls into question the extent to which attentional drift can account for the appearance of the irrelevant sound effect. Attentional drift (or lack of attentional control) does, however, provide an elegant explanation for Elliott's (2002) finding that young children are more susceptible to irrelevant speech effects than are older children and adults. Thus, alternative explanations for the failure to find the anticipated relationship between the irrelevant sound effect must be considered before any strong conclusions are justified.

Experiment 3

The results of the study so far can be summarized as follows: Working memory span, as a measure, is susceptible to variations in the means by which the measure is administered (Experiment 1), but these fluctuations are unlikely to account for any differences between the sensitivity to the cocktail party effect in the study by Conway et al. (2001). Contrary to predictions, there is no sign that ability to resist the disruptive effects of irrelevant speech (Experiment 2) is in any way influenced or mediated by working memory span as measured by the OSPAN technique. There are a number of differences between the previous experiment and that of Conway et al., however, which are worthy of consideration when trying to reconcile the inconsistent effects of OSPAN between the irrelevant speech and cocktail party paradigms. These concern the way in which the analysis was conducted and the timing of presentation of the supposedly unattended stimuli.

To take the methodological issue first, Conway et al. (2001) presented the unattended auditory stream concurrently with the presentation of the attended (shadowed) information, as required in a dichotic listening task. In the experiments presented here, however, the irrelevant sound was presented subsequent to the presentation of the to-be-recalled material. Arguably, if OSPAN is to be identified as a primary determinant in susceptibility to the irrelevant sound effect, this difference is unimportant because a number of studies have investigated the issue of timing of irrelevant sound and have concluded that irrelevant sound presented subsequent to the to-be-recalled list is equivalent in its effects to irrelevant sound presented concurrently with the to-be-recalled list (Baddeley & Salamé, 1986; Beaman & Jones, 1998; Macken, Mosdell, & Jones, 1999; Miles, Jones, & Madden, 1991). Thus, any influence of OSPAN on the irrelevant sound effect at encoding that is not also present when the irrelevant sound effect is presented during a retention interval would imply that OSPAN as a measure of auditory distractibility does not generalize particularly well beyond a particular, circumscribed situation. However, an effect of OSPAN if the irrelevant speech was limited to the encoding period (i.e., simultaneous presentation of the two [attended and unattended] sets of stimuli) would provide a link to the previous results of Conway et al. and suggest how the current, apparently incompatible, series of data could be reconciled with the earlier findings. One possibility is that although OSPAN is not—as originally hypothesized—a good determinant of resistance to irrelevant sound generally, working memory, as measured by OSPAN, is specifically required to filter out irrelevant (auditory) stimulation during relevant stimulation. Thus, a further experiment examining the influence of OSPAN when the irrelevant speech is played concurrently with the to-be-attended stimuli could help to resolve the issue.

In terms of the analyses, it is also the case that the division of participants into high- and low-span groups in Experiment 2 was not carried out exactly in the manner of Conway et al. (2001), and this may have reduced the sensitivity of the subsequent analyses. In these two experiments, classification of the two groups was by median split, whereas in the earlier study, although half the sample were classified as high-span and half as low-span, the basis of this classification was OSPAN scores that fell into the upper and lower quartiles of a larger sample of participants who carried out the OSPAN task. There is the potential, therefore, that the similarity between high- and low-span participants may have been greater in Experiment 2 than in Conway et al., reducing the size of any effect of OSPAN and thus also reducing power. Because the effect size observed in Experiment 2 was small (partial $\eta^2 = .037$ for span by irrelevant sound interaction), this may be sufficient to account for the apparent discrepancy in the results. A further experiment separating the high- and low-span groups by means of comparing only the top- and bottom-scoring quartiles will act as a check on this possibility. The earlier attempt to examine this possibility in Experiment 2 failed because the increase in difference between the high- and low-span groups was insufficient to overcome the loss of statistical power associated with the reduction in sample size. In accordance, Experiment 3 used a sample size sufficiently large that the number of participants in each of the groups remains high even when the participants in the middle two quartiles are discarded.

Method

Participants. Eighty undergraduate and postgraduate students from the University of Reading participated in return for either course credit or a small honorarium. All reported normal hearing and normal or corrected-to-normal vision.

Materials and design. The materials for the OSPAN task were identical to those of Experiment 1; however, only the stimuli for the read-aloud condition of that experiment were used. The to-be-recalled stimuli and procedures from the seven-digit lists of Experiment 2 were also used. The to-be-ignored auditory stimuli were identical to those of Experiment 2, but unlike the previous experiment, the auditory stimuli were played to the participants at the same time as they observed the to-be-attended stimuli.

Procedure. Participants were asked to carry out the serial recall task as in Experiment 2. The one alteration to this procedure was that the participants were not required to wait before recalling the to-be-attended stimuli; that is, instead of a 9-s retention interval following presentation of the to-be-attended list, there was only a 500-ms interstimulus interval between the presentation of the final list item and the recall cue. Following completion of this task, participants were then asked to carry out the read-aloud working memory-span test of Experiment 1. The experimental session lasted approximately 40 min in total.

Results

Analysis was carried out on the number of operation errors within the OSPAN task to confirm that participants were engaging in this part of the task and not treating the session as a memory test only. The incidence of operation errors ranged from 2% to 35%, with a mean rate of operation errors of 8.08% ($SD = 7.67\%$; chance performance = 50%). Therefore, there is no evidence that participants were neglecting this part of the task.[2]

The top 25% of the working memory scores were identified as high-span participants, and the bottom 25% were identified as low-span participants. OSPAN ranged from 2 to 54. In the low-span group (lower quartile), mean span was 4.9 (range: 2–7, $SD = 1.8$) and in the high-span group (upper quartile) mean span was 32.2 (range: 20–54, $SD = 10.3$). Repeated measures ANOVA was carried out using these two groups; all power calculations were computed using $\alpha = .05$. ANOVA revealed that there was a main effect of irrelevant speech, $F(1, 38) = 86.72$, $MSE = 32.91$, $p < .001$, observed effect size was .695 (partial η^2) and observed power was .999. There was also a main effect of OSPAN, $F(1, 38) = 18.79$, $MSE = 195.4$, $p < .001$, partial $\eta^2 = .331$, observed power = .988. However, irrelevant speech once again failed to interact significantly with OSPAN ($F < 1$, partial $\eta^2 = .019$, observed power = .133). Results are shown in Figure 3. As shown in the figure, the size of the irrelevant speech effect was numerically greater in the high-span condition (mean ISS = 13.05) than in the low-span condition (mean ISS = 10.85), although, of course, this difference was not significant.

As in the previous experiment, the effect of OSPAN on the positional intrusion error ratio was analyzed. Because only seven-item lists were used in this experiment, the smaller number of overall errors means that the proportion of positional intrusions is an inherently more variable, and hence less reliable, measure, particularly among participants whose overall error rate was low. For example, a single error could—in theory—constitute up to 100% of the total errors. This highlights the problems associated with any simple comparison of OSPAN and the proportion of positional errors when, in some cases, total number of errors is low. To overcome these difficulties, only those participants whose total error rate reached 10% or more were included in an analysis using a new measure, the ratio of observed to expected number of positional intrusions. This measure involved calculating, for each participant, the ratio of the total number of positional intrusions observed to the number of errors of that type expected by chance (in practice, this expected value was estimated as the total number of errors for each participant multiplied by an assumed 1 in 9 probability of a positional error appearing by chance; Ng & Maybery, 2002). When calculated in this manner, t tests showed a marginally significant effect of OSPAN on the conservative ratio measure of Ng and Maybery (2002), $t(35)= 1.35$, $p < .10$, and a significant effect of OSPAN on the total number of positional intrusions, $t(38) = 3.05$, $p < .008$. Note that because span and number of errors in serial recall are inversely related, the exclusion criterion (< 10% error rate) means that only members of the high-span group were removed from the ratio measure analysis.

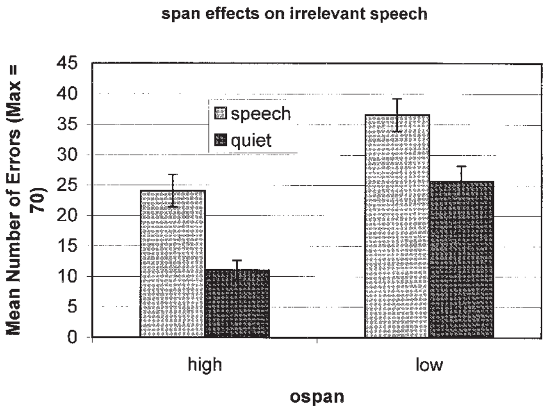

Figure 3. Experiment 3: The disruptive effect of irrelevant speech on the serial recall of high- and low-span participants. Error bars are standard errors. Max = maximum; ospan = operation span.

Discussion

This experiment showed that simultaneously increasing the number of participants and the strength of the OSPAN manipulation while also replicating the Conway et al. (2001) procedure as closely as possible had no effect on the data and produced once again a nonsignificant result and one, furthermore, in which (contrary to predictions) irrelevant sound disruption was numerically greater among high-span participants. On this basis, it is possible to state with some confidence that susceptibility to irrelevant speech distraction is not mediated by the working memory mechanism measured by OSPAN. However, neither the reason for the inconsistency in results obtained here, between ignoring irrelevant speech in a serial recall task and ignoring irrelevant speech in a dichotic listening task obtained by Conway et al., nor the reason behind the effects of OSPAN on positional intrusion errors is quite so clear cut.

The examination of positional intrusion errors is a side issue (albeit an interesting one) and is dealt with only briefly, because it is clear that considerable further experimentation is necessary to provide a comprehensive account of the situation. As previously mentioned, the total number of positional intrusions was significantly affected by OSPAN, and equally, there was also a tendency for OSPAN to influence the more conservative ratio measure, although it is disappointing to note that differences were only marginally significant using this measure. This is clearly an issue that requires further experimentation and, if possible, the applica-

---

[2] Participants scoring greater than 20% operation errors were sometimes excluded from subsequent analyses. Here, only 3 participants fell into this category and excluding their data did not affect the results.

tion of a well-formed model of interference effects (e.g., the SIMPLE model of Brown, Neath, & Chater, 2004) to examine the possible mechanisms involved.

Turning to the question of the discrepancy between the non-effects of OSPAN on irrelevant sound disruption and the previously documented effects of OSPAN on dichotic listening, two points are worthy of note. The first point is that there remains a discrepancy between the irrelevant sound and dichotic listening literature in that irrelevant sound is ordinarily presented alongside a serial recall task where the to-be-recalled material is visually presented. Dichotic listening, however, by definition, requires the presentation of two auditory streams. It is possible that OSPAN affects only the ability to select between two streams presented in the same modality. If this is the case, then OSPAN cannot serve as the basis for individual differences in the irrelevant sound effect for two reasons. First, on the occasion when auditory and visual presentation of to-be-recalled material has been compared in irrelevant sound effects, disruption has been equivalent regardless of the modality of the to-be-recalled material (e.g., Campbell, Beaman, & Berry, 2002; Hanley & Broadbent, 1987; Jones, Macken, & Nicholls, 2004). It therefore seems unlikely that individual variations in susceptibility to irrelevant sound would differ according to the modality of the to-be-recalled material. Second, and more importantly, the individual differences established by Ellermeier and Zimmer (1997) were established on the standard basis of visual presentation of the to-be-recalled material. Any effect of OSPAN that is concerned only with auditory–auditory selection cannot therefore apply to these data; thus, by extension, it cannot apply to individual variation within the majority of irrelevant sound effect studies.

The further point worthy of note is that, in addition to the to-be-ignored and to-be-attended material, both being auditory in dichotic listening, in the cocktail party version of dichotic listening used by Conway et al. (2001), processing of the supposedly unattended message was measured by the awareness participants displayed of a highly meaningful stimulus, their own name, being presented on the unattended channel. This is a different measure to the effects of irrelevant sound on serial recall that could arguably be observed without the participant displaying any great awareness of the content of the to-be-ignored sound (see Cowan & Wood, 1997, for a discussion of related issues). Most crucially, the irrelevant sound effect generally does not show any effect of semantics (Buchner, Irmen, & Erdfelder, 1996) whereas one's own name—the unattended stimulus in the study by Conway et al.—is salient precisely because it is intrinsically meaningful to the person involved. It may be, therefore, that OSPAN is not a good predictor of auditory distraction per se but of meaningful auditory distracters specifically. If this is the case, it is necessary to compare the disruptive effects of meaningful and nonmeaningful stimuli by using a primary task other than immediate recall of digits.

A study by Neely and LeCompte (1999) demonstrated that a semantic irrelevant speech effect can be observed if free, rather than serial, recall is required and if the to-be-ignored stimuli are semantically related to the to-be-attended stimuli. Controlled attention in this situation might serve to act on material at the semantic level rather than the phonological or acoustic levels considered in the previous experiments. It is interesting that recent imaging data provided by Scott and colleagues (Scott, Rosen, Wickham, & Wise, 2004; Scott, Davis, Rosen, Beaman, & Wise, 2004) support the idea that unattended or masking speech in a dichotic listening paradigm selectively activates different brain regions depending on whether the masking speech is intelligible. Both intelligible and acoustically matched unintelligible (spectrally rotated) speech activate right hemisphere superior temporal gyrus, but only the intelligible speech also activated left hemisphere temporal lobe regions. As previously established (e.g., Jones & Macken, 1993), acoustic complexity, specifically dynamic pitch variations, is sufficient for an irrelevant sound effect, but an effect of semantics in the irrelevant sound stream occurs only rarely and under specific circumstances. Involvement of OSPAN in dampening down activation due to processing of intelligible speech would explain not only Conway et al.'s (2001) results, which are dependent on processing of an inherently meaningful stimulus, but also why no connection was found between working memory span and auditory distraction in the current study, in which no semantic effects of irrelevant sound were anticipated. If OSPAN reflects the ability to reject stimuli at a different stage in processing than that represented by the irrelevant sound distracters, possibly during semantic analysis, high-span participants could exclude the highly meaningful stimulus in Conway et al.'s cocktail party paradigm quite efficiently but this capability would be of no use in attempting to ignore the rather less meaningful stimuli (nonwords and tones) of Experiments 2 to 3. In accordance, Experiment 4 used a modification of Neely and LeCompte's (1999) free recall procedure to determine whether OSPAN could account for the effects of irrelevant speech at the semantic level.

Experiment 4

Method

Participants. Thirty-seven undergraduate students of the University of Reading participated in return for course credit.

Materials and design. To-be-recalled and to-be-ignored stimuli were taken from Battig and Montague's (1969) category norms. Thirty lists of 16 words each were constructed as the to-be-recalled stimuli. Twenty lists of 8 words each were constructed as the to-be-ignored auditory stimuli. The to-be-ignored words consisted of the 8 most frequently produced exemplars of each category, and the to-be-recalled words consisted of the 16 next most frequently produced responses (exemplars 9–24). The order of presentation of the words in each to-be-ignored and to-be-recalled list was randomly determined but was the same random order for each participant. The to-be-ignored stimuli were recorded in a male voice at a sampling rate of 22 KHz and eight-bit resolution using SoundEdit software, and were digitally edited to a length of 500 ms. Presentation of the auditory (to-be-ignored) stimuli was via Sony MDR-CD470 stereo headphones, and presentation of the to-be-recalled (visual) stimuli was on the centre of the screen of a Macintosh 5300cs Powerbook running Hypercard software.

For the first phase of the experiment, a randomly determined half of the to-be-ignored word lists were paired with to-be-recalled words from the same category, and the remainder were paired with to-be-recalled lists from different categories. For the final phase of the experiment, these lists were swapped over; thus, those to-be-recalled lists that had earlier been paired with related to-be-ignored words (exemplars of the same category) were now paired with unrelated word lists (exemplars from a different category). To-be-recalled words were presented center-justified in 48-point Geneva font at a rate of two words per second. Following the end of the list was an 8-s retention interval during which, in the distraction conditions, the to-be-ignored stimuli were played twice. Order of presentation of the three auditory conditions (related distracters, unrelated distracters, quiet) was randomly determined but remained in the same random order for each participant. The end of the retention interval was signaled by the visual presentation of a *RECALL NOW* cue, which stayed on the screen for 2 s.

Procedure. Participants were told that they would be presented with lists of words, presented one word at a time, on the computer screen. They were asked to try and remember these words and, when cued by the appearance of the cue *RECALL NOW,* to write down, in any order, all the words they could remember on the paper provided. Participants were given 30 s to recall the words from the list. The start of the next list was signaled by a warning tone at the end of the 30 s. Participants were also informed that for many of the lists they would hear other words over the headphones. They were asked to try and ignore these words and were reassured that they would not be tested on them in any way. Following this initial phase of 30 lists for free recall, participants were tested for their OSPAN in an identical manner to Experiment 3. Finally, participants were once again presented with the free-recall task. They were explicitly informed that the words they would be trying to remember were identical to those they had seen in the initial phase of the experiment. No such information was given regarding the to-be-ignored stimuli, and no participant enquired. The experimental sessions lasted, in total, approximately 1 hr 15 min.

Results

Half of the sample, designated as high-span participants ($N = 18$), had an OSPAN range of 16–55 ($M = 25.39$, $SD = 9.6$). The remaining half, designated as low-span participants ($N = 19$) had an OSPAN range of 2–13 ($M = 7$, $SD = 3.54$). One-way repeated measures ANOVAs were carried out on the number of items correctly recalled in each of the three speech conditions (quiet, unrelated speech, related speech) and the number of intrusion errors in each of the conditions. For all power calculations, alpha was set to .05. Data are shown in Figure 4. There was a significant effect of speech condition on the number of items correctly recalled, $F(2, 70) = 64.41$, $MSE = 51.04$, $p < .001$, effect size (partial η^2) = .648, observed power = .999, but there was no effect of OSPAN, $F(1, 35) = 1.56$, $MSE = 1101.03$, $p = .22$, partial $\eta^2 = .043$, observed power = .229, and no interaction effect, $F(2, 70) < 1$, partial $\eta^2 = .021$, observed power = .17. Paired sample *t* tests also revealed that the means for the quiet and unrelated speech conditions were significantly different, $t(36) = 7.84$, $p < .002$, as were the means for the related and unrelated speech conditions, $t(36) = 3.41$, $p < .004$ (Bonferroni corrected), confirming that a semantic irrelevant speech effect was present, although the size of the effect was not modified by OSPAN.

The number of related-item intrusions—erroneous intrusions into the recall protocol of any of the eight most popular category exemplars—was also collated and analyzed. For the quiet auditory condition, these items, although related to the to-be-remembered items, were not presented to the participants at any stage during the experiment, and therefore their appearance in recall protocols presumably reflects spreading semantic activation and possible guesswork rather than failure to inhibit auditory distracters. For both the related and unrelated speech condition, these items may have been encountered as to-be-ignored stimuli on previous lists, but crucially, for the related speech conditions, these items were also present in the current set of to-be-ignored auditory distracters. When the number of related-item intrusions was analyzed, a significant effect of speech condition on the number of intrusions observed was apparent, $F(2, 70) = 34.97$, $MSE = 6.61$, $p < .001$, partial $\eta^2 = .5$, observed power = .999, but there was no main effect of OSPAN, $F(1, 35) = 1.67$, $MSE = 42.41$, $p = .21$, partial $\eta^2 = .046$, observed power = .242. It is important to note, however, that there was a significant interaction between speech condition and OSPAN, $F(2, 70) = 4.48$, $MSE = 6.61$, $p < .02$, partial $\eta^2 = .113$, observed power = .747. Independent samples *t* tests reveal that low-span participants produced significantly more related-item intrusions in the related speech condition than did high-span participants, $t(35) = 1.97$, $p < .028$.

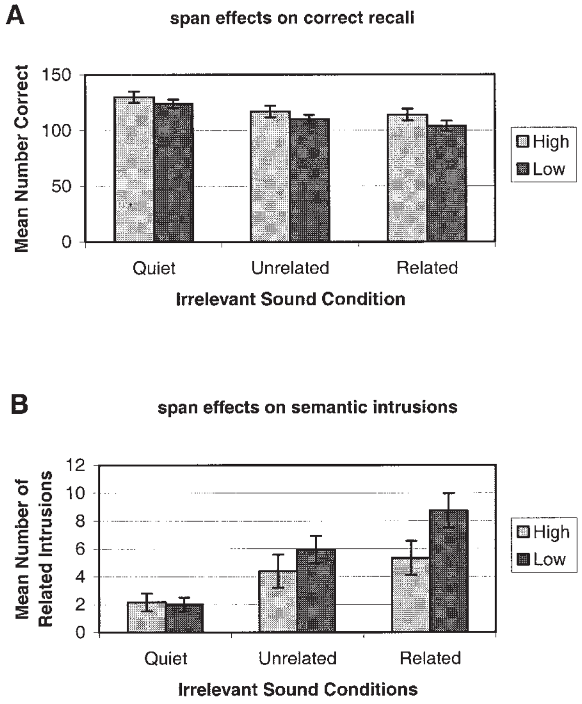

Figure 4. Experiment 4: The number of correct responses (Panel A) and related intrusion errors (Panel B) in free recall. Panel B breaks down the number of related intrusion errors by working memory span. Error bars are standard errors.

Discussion

The results of this experiment are illuminating for a number of reasons. First, they demonstrate that with different types of recall test (free recall) and auditory distracter (semantically related), the OSPAN measure still fails to distinguish between individuals who suffer from high or low levels of irrelevant sound interference when the dependent measure is the number of correct recalls. Second, the results are the first replication of Neely and LeCompte's (1999) demonstration of semantically mediated irrelevant speech effects in free recall, and they extend the findings of those authors in also showing a reliable effect of intrusion by irrelevant speech items into the free recall protocols in related-speech conditions. Third, and most importantly, the results demonstrate that intrusions from semantically related irrelevant speech occur more frequently among low- than high-OSPAN participants, providing a conceptual replication previously lacking for the results of Conway et al. (2001) and also suggesting a means by which the results of this experimental series can be reconciled with the data from Conway et al.'s dichotic listening task. In both Conway et al.'s task and the present experiment, high-span participants were less aware of the content of the auditory distracters when given the opportunity to recall auditory items (in Conway et al's study, such recall of one's own name is marked as correct; in the current study, the recall would appear as an intrusion error), but this experimental series also demonstrates that awareness of the contents of the irrelevant speech stream (as measured by the number of irrelevant speech intrusion errors) is, to a large extent, independent of the decrease in the number of correctly recalled items occasioned by the presence of such distracters.

One could argue that, in the current experiment, low-OSPAN participants were simply more likely to guess than high-OSPAN participants, resulting—under the fairly lenient scoring criterion of free recall—in a failure to find any significant differences between the two groups in terms of total number recalled and also producing the significant differences in intrusion errors between the two groups. Such an argument is only partially sustainable, however. Without a correction for guessing the possibility that the recall performance (total number correctly recalled) of low-OSPAN participants was inflated by an enhanced propensity to guess cannot be ruled out. However, a control for guessing is available when examining the number of intrusion errors.

To be classified as an intrusion error, the error must be a member of the eight most popular category items from the relevant Battig and Montague (1969) category that constitute the to-be-ignored stimuli. An estimate of guessing when the participant has not previously heard the intrusion item therefore corresponds to the two left-hand bars of the bottom panel of Figure 4 (quiet condition). An estimate of guessing when the participant has previously heard the intrusion error paired with another to-be-recalled list corresponds to the two middle bars of Figure 4 (unrelated condition). In neither of these conditions are there discernible effects of OSPAN. Only in the right-hand pair of bars, which document "guessing" (when the intrusion item was part of the current to-be-ignored set related condition) do low- and high-span groups finally separate, resulting in the speech by OSPAN interaction effect on intrusion errors documented earlier but having no main effect of OSPAN. Thus, if low-span participants are more prone to guessing than are high-span participants, then this only manifests itself when the guesses correspond to items from the current to-be-ignored set.

The conclusion drawn is that differences in guessing (intrusion errors) between these groups only occur when plausible candidates (to-be-ignored stimuli) are more readily available to the low-span participants than the high-span participants because, as in Conway et al. (2001), the to-be-ignored information has not been inhibited as efficiently by the low-span participants. When plausible candidates have not been recently presented, there are no differences between the two groups. The relative lack of intrusion errors shown by high-span participants when possible intrusions are presented as part of the to-be-ignored stream must therefore be either a consequence of superior suppression of the to-be-ignored stimuli or of superior source monitoring. Because the former possibility is one that is consistent with the dichotic listening data presented by Conway et al. (which cannot be explained by differences in source monitoring), this is the simplest and most elegant explanation of these results.

General Discussion

The size of the effect of OSPAN on ISS was fairly consistent across the studies, ranging from .037 (Experiment 2) down to a minimum of .019 (Experiment 3). In Experiment 3, this effect is actually in the opposite direction to that hypothesized; thus, given these data, one can be fairly confident that if working memory span does influence the irrelevant sound distraction effect, the influence is numerically very slight. Such an influence would be extremely difficult to detect given the large number of participants that analysis would require and arguably would be of little or no theoretical value. It would certainly have minimal effect on the large and statistically reliable individual variations in susceptibility to irrelevant sound documented by Ellermeier and Zimmer (1997), and similarly, is unlikely to be the same effect as that observed relatively easily by Conway et al. (2001) when comparing the awareness of high- and low-span participants of their own name on an unattended speech channel.

This leads to the first conclusion from this series of experiments: It is disappointing to conclude that a locus for individual differences in ISS cannot be found in working memory capacity as measured by OSPAN. Whatever is being measured by OSPAN is not related to any capacity to inhibit or ignore auditory distracters of the kind represented by the irrelevant sound effect. The second conclusion is, however, less negative: Working memory span picks up some (hitherto unreported) capacity to inhibit positional intrusions from previously important lists (Experiments 2 and 3) and semantically related items from to-be-ignored auditory distracters (Experiment 4). These conclusions, and their implications, are examined in turn.

Implications for the Irrelevant Sound Effect

The first set of implications of this set of studies for the irrelevant sound effect is predominantly negative. No evidence has been found that pinpoints the locus of individual differences in susceptibility unambiguously. This is disappointing because any factor influencing the extent to which an individual suffers from the irrelevant sound effect must, of necessity, be a mechanism governing the very existence of the effect, an issue that is still a matter of some controversy (Baddeley, 1986; Cowan, 1995; Jones et al., 1996; Neath, 2000). The current set of data can, however, be put alongside those of Ellermeier and Zimmer (1997) and Neath et al. (2003) in ruling out one possible explanation of the effect. Working memory capacity, at least insofar as it is measured by OSPAN, has no effect on the disruption caused in the irrelevant sound paradigm.

More positively, the current data provide further evidence that semantically related irrelevant speech causes more disruption than unrelated speech, but only if the primary task also involves a relatively high degree of semantic processing. The data also suggest that OSPAN affects the processing of the meaning of unattended speech. No effect of this was found in the total-number-recalled measure of Experiment 4, but evidence was found in support of this hypothesis in the number of related intrusions observed in the same experiment. High-OSPAN participants produced significantly fewer of these errors when the related intrusions came from a current set of to-be-ignored stimuli but were equivalent in performance to low-OSPAN participants when the related intrusions were not part of a current to-be-ignored list. Because the irrelevant sound effect is, under the circumstances in which it is usually tested, unaffected by the meaning of the irrelevant speech (Buchner et al., 1996; Salamé & Baddeley, 1982) but is affected by acoustic factors concerned with dynamic pitch variation (Jones & Macken, 1993; Jones, Macken, & Murray, 1993), this suggests that, whatever else working memory span might influence, it mediates semantic factors in the irrelevant sound effect that are dependent on processing intelligible speech rather than lower level acoustic processing (Scott et al., 2004).

Implications for Working Memory Span

The current set of experiments also provides an interesting new perspective on operation span measure of working memory. Experiment 1 showed quite clearly that OSPAN itself is subject to the effects of presentation modality, for reasons that are yet unclear. The notion of an auditory feedback loop is insufficient to explain the differences as otherwise the effects of such auditory feedback would correlate highly with the effects of irrelevant speech in a serial recall task, which they do not (Experiment 2). Close examination of Figure 1 reveals that, in any case, the presentation modalities also differ qualitatively, with read-aloud presentation providing a characteristic auditory-recency effect for the final item at any given list length, a recency effect noticeably absent from the read-quietly presentation. This is obviously an area for future study, but a number of conclusions can be drawn from the data obtained thus far.

The first, and arguably most crucial, point is that the difference in presentation modality does not affect the relative performance of those undergoing the span tests: Participants who scored high in one procedure also scored high in the other. Therefore, there is no reason to suppose that altering the absolute levels of performance within the span test using this sort of manipulation will in any way affect the correlations between span and other measures of cognitive performance. The second observation is that the span task differs fundamentally from other tasks in which modality of presentation is manipulated and that also involve the interleaving of processing tasks with the presentation of to-be-recalled material, most notably Bjork and Whitten's (1974) continuous (or more accurately, intermittent) distracter technique. In both instances, a modality effect is observed toward the end of the list, but the intermittent distracter technique does not usually result in such a strong visual advantage elsewhere. Again, the reasons for this difference in the results are unclear, but recent evidence has been presented that suggests recall requirements may, in certain circumstances, favor visual-only presentation (Beaman, 2002; Beaman & Morton, 2000; Metcalfe & Sharpe, 1985), and it may prove to be important that (as previously noted) Bjork and Whitten's technique asked for free recall, whereas OSPAN requires serial recall.

A further notable set of findings for the working memory span literature is the relationship between span and positional intrusions (Experiments 2 and 3) and between span and irrelevant speech intrusions (Experiment 4). In regard to positional intrusions, the current set of studies provide the first evidence that working memory capacity reflects the capacity to inhibit order as well as item information from previous, no longer relevant lists. These data are provisional and obviously require further supporting research (beyond the scope of this article) to provide a comprehensive account of the findings. The observed relationship between OSPAN and irrelevant speech intrusions, however, confirms and reinforces the report by Conway et al. (2001) that OSPAN contributes to the cocktail party effect. Both studies provide evidence that high-span participants are less likely than low-span participants to report the contents of speech designated as irrelevant to the primary task. To this extent, Experiment 4 provides a bridge between the dichotic listening procedure employed by Conway et al. and the irrelevant sound–immediate recall tasks presented here. Speculatively, it is suggested that the link is via the intelligibility and meaningfulness of the to-be-ignored sound. Sound does not need to be meaningful to provide an irrelevant sound effect, and in many circumstances, meaningful speech is no more disruptive than meaningless speech (Buchner et al., 1996; Salamé & Baddeley, 1982). In circumstances in which meaning does exert an influence, as in the free recall procedure of Neely and LeCompte (1999) adopted for Experiment 4, OSPAN distinguishes between participants' erroneous recall of to-be-ignored material in the same manner as, in Conway et al.'s (2001) studies, it distinguishes between recall of hearing the participants' own name on the unattended channel. The crucial common feature, it is suggested, is that high-span participants are less aware in both situations of the content of the irrelevant or unattended stream, presumably because, in both cases, they have more effectively inhibited the irrelevant information (as would be predicted on the basis of Engle's theory, Engle, 1999; Engle, Kane, et al., 1999, and on the earlier theory of Hasher & Zacks, 1988). Thus, the results of this series of experiments imply that the capacity to inhibit irrelevant semantic information reflected in the OSPAN working memory task (Experiment 4) may be distinct from the capacity to resist nonsemantic interference from irrelevant sound in serial recall (Experiments 2 and 3), which is influenced only in a very minor way, if at all, by the mechanisms underlying OSPAN.

References

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Clarendon Press.

Baddeley, A. D., & Salamé, P. (1986). The unattended speech effect: Perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 525–529.

Banbury, S., Tremblay, S., Macken, W. J., & Jones, D. M. (2001). Auditory distraction: Phenomena, models and practical implications. *Human Factors, 43,* 12–29.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46.

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology, 55A,* 371–389.

Beaman, C. P., & Jones, D. M. (1997). The role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 459–471.

Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *Quarterly Journal of Experimental Psychology, 51A,* 615–636.

Beaman, C. P., & Morton, J. (2000). The separate but related origins of the modality effect and the recency effect in free recall. *Cognition, 77,* B59–B65.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189.

Bridges, A. M., & Jones, D. M. (1996). Word-dose in the disruption of serial recall by irrelevant speech. *Quarterly Journal of Experimental Psychology, 49A,* 919–939.

Brown, G. D. A., Neath, I., & Chater, N. (2004). *A ratio model of scale-invariant memory and identification.* Manuscript submitted for publication.

Buchner, A., Irmen, L., & Erdfelder, E. (1996). On the irrelevance of semantic information for the "irrelevant speech" effect. *Quarterly Journal of Experimental Psychology, 49A,* 765–779.

Campbell, T., Beaman, C. P., & Berry, D. C. (2002). Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory, 10,* 199–214.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975–979.

Conrad, R. (1959). Errors of immediate memory. *British Journal of Psychology, 50,* 349–359.

Conrad, R. (1960). Serial order intrusions in immediate memory. *British Journal of Psychology, 51,* 45–48.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10,* 135–136.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8,* 331–335.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354–373.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory and Cognition, 27,* 1042–1050.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford, England: Clarendon Press.

Cowan, N., & Wood, N. L. (1997). Constraints on awareness, attention, processing and memory: Some experiments with ignored speech. *Consciousness & Cognition, 6,* 182–203.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422–433.

Ellermeier, W., & Zimmer, K. (1997) Individual differences in susceptibility to the "irrelevant speech effect." *Journal of the Acoustical Society of America, 102,* 2191–2199.

Elliott, E. M. (2002). The irrelevant speech effect and children: theoretical implications of developmental change. *Memory & Cognition, 30,* 478–487.

Engle, R. W. (1999). What is working memory capacity? In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297–314). Washington, DC: American Psychological Association.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, England: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Estes, W. K. (1991). On types of item coding and sources of recall in short-term memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays in honor of Bennet B. Murdock* (pp. 155–174). Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4,* 294–298.

Guildford, J. P., & Dallenbach, K. M. (1925). The determination of memory span by the method of constant stimuli. *American Journal of Psychology, 36,* 621–628.

Hanley, J. R., & Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology, 78,* 287–297.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of language and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. In: A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control. A tribute to Donald Broadbent* (pp. 87–104). Clarendon, England: Oxford University Press.

Jones, D. M., Beaman, C. P., & Macken, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 209–238). Hove, England: Psychology Press.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 369–381.

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition, 21,* 318–328.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 656–674.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proof-reading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology, 4,* 89–108.

Kail, R., & Hall, L. K. (2001) Distinguishing short-term memory from working memory. *Memory & Cognition, 29,* 1–9.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 130,* 199–207.

Macken,W. J., Mosdell, N., & Jones, D. M. (1999) Explaining the irrelevant sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 810–814.

Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1272–1299.

Metcalfe, J., & Sharpe, D. (1985). Ordering and reordering in the auditory and visual modalities. *Memory & Cognition, 13,* 435–441.

Miles, C., Jones, D. M., & Madden, C. A. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 578–584.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition, 19,* 332–340.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7,* 403–423.

Neath, I., Farley, L. A., & Surprenant, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology, 56A,* 1269–1278.

Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review, 2,* 429–441.

Neely, C. B., & LeCompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & Cognition, 27,* 37–44.

Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology, 55A,* 391–424.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398–422.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39,* 418–436.

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 150–164.

Scott, S. K., Davis, J., Rosen, S., Beaman, C. P., & Wise, R. (2004). *A positron emission tomography (PET) study of informational masking.* Unpublished manuscript.

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America, 115,* 813–821.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task-dependent? *Journal of Memory and Language, 28,* 127–154.

Received June 2, 2003
Revision received February 5, 2004
Accepted February 10, 2004 ■

Call for Nominations

The Publications and Communications (P&C) Board has opened nominations for the editorships of ***Clinician's Research Digest, Emotion, JEP: Learning, Memory, and Cognition, Professional Psychology: Research and Practice,*** and ***Psychology, Public Policy, and Law*** for the years 2007–2012. Elizabeth M. Altmaier, PhD; Richard J. Davidson, PhD, and Klaus R. Scherer, PhD; Thomas O. Nelson, PhD; Mary Beth Kenkel, PhD; and Jane Goodman-Delahunty, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2006 to prepare for issues published in 2007. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations also are encouraged.

Search chairs have been appointed as follows:

- ***Clinician's Research Digest:*** William C. Howell, PhD
- ***Emotion:*** David C. Funder, PhD
- ***JEP: Learning, Memory, and Cognition:*** Linda P. Spear, PhD, and Peter Ornstein, PhD
- ***Professional Psychology:*** Susan H. McDaniel, PhD, and J. Gilbert Benedict, PhD
- ***Psychology, Public Policy, and Law:*** Mark Appelbaum, PhD, and Gary R. VandenBos, PhD

Candidates should be nominated by accessing APA's EditorQuest site on the Web. Using your Web browser, go to http://editorquest.apa.org. On the Home menu on the left, find Guests. Next, click on the link "Submit a Nomination," enter your nominee's information, and click "Submit."

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Karen Sellman, P&C Board Search Liaison, at ksellman@apa.org.

The deadline for accepting nominations is **December 10, 2004,** when reviews will begin.