

In press 2007, *Journal of Experimental Psychology: General*

A Central Capacity Limit to the Simultaneous Storage
of Visual and Auditory Arrays in Working Memory

J. Scott Saults and Nelson Cowan
University of Missouri – Columbia

Address correspondence to: J. Scott Saults
Department of Psychological Sciences
University of Missouri
207 McAlester Hall
Columbia, MO 65211

E-mail: SaultsJ@missouri.edu.

Telephone: 573-882-9634

Fax: 573-882-7710

Original submission: August 6, 2006; First revision submitted: December 31, 2006;

Second revision submitted: May 12, 2007, accepted: May 16, 2007; This version: June 4, 2007.

Running head: Simultaneous Storage of Visual and Auditory Arrays

Abstract

If working memory is limited by central capacity (e.g., the focus of attention; Cowan, 2001) then storage limits for information in a single modality should also apply to the simultaneous storage of information from different modalities. We investigated this by combining a visual-array comparison task with a novel auditory-array comparison task in five experiments. Participants were to remember only the visual or only the auditory arrays (unimodal memory conditions) or both arrays (bimodal memory conditions). Experiments 1-2 showed significant dual-task tradeoffs for visual but not auditory capacity. In Experiments 3-5, modality-specific memory was eliminated using post-perceptual masks. Dual-task costs occurred for both modalities and the number of auditory and visual items remembered together was no more than the higher of the unimodal capacities (visual, 3-4 items). The findings suggest a central capacity supplemented by modality- or code-specific storage and point to avenues for further research on the role of processing in central storage.

Keywords:

working memory

central capacity limits

auditory and visual memory

change detection

attention allocation

A Central Capacity Limit on the Simultaneous Storage of Visual and Auditory Arrays in Working Memory

Since its beginning, cognitive psychology has tried to understand and measure the limits of human information processing. This quest was launched, in part, by George Miller's seminal article on "The magical number seven, plus or minus two" (Miller, 1956). This engaging and provocative article was such an inspiration to pioneers of cognitive psychology that we often forget its admonitions. Observation and experience provide such compelling evidence of our mental limitations that we were easily persuaded by Miller's examples that seven could approximate some general limit to human cognition. However, Miller warned how the persistence of seven might be more coincidence than magic. He also demonstrated that items are often recoded to form larger, meaningful chunks in memory that allow people to circumvent the limit to about seven items in immediate recall. For example, whereas a series of nine random letters may exceed one's immediate memory capacity, the sequence *IRS-FBI-CIA* includes three familiar chunks (if one knows the acronyms for three U.S. government agencies, the Internal Revenue Service, the Federal Bureau of Investigation, and the Central Intelligence Agency), well within the adult human capacity. The implication is that it is impossible to measure capacity in chunks without somehow assessing the number of items per chunk. Probably for this reason, 50 years after Miller's magical article, there is still no agreement as to what the immediate memory limit is and, indeed, whether such a limit exists (for a range of opinions see Cowan, 2001, 2005a).

Following Miller (1956), one of the most influential attempts to answer this question of the nature of immediate memory limits was Baddeley and Hitch's (1974) theory of working memory

(WM), a set of mechanisms for retaining and manipulating limited amounts of information to use in ongoing cognitive tasks. Theories of WM try to explain various mental abilities, like calculation and language comprehension, in terms of underlying limitations in how much information can be held in an accessible form (Baddeley, 1986). An important departure from Miller was that the WM of Baddeley and Hitch had no single, central capacity store or storage limitation. Instead, a central executive relied on specialized subsystems, the phonological loop and visuospatial sketch pad, each with its own mechanisms for storing and maintaining certain kinds of information. These dedicated storage buffers were supposed to be independent of one another and to lose information across several seconds, with no stated limit in how many items could be retained at once.

More recently, Cowan (2001) revived Miller's idea with "The magical number four in short-term memory". This article re-evaluated much of the evidence that Miller had considered in the light of subsequent research. Cowan agreed that evidence pointed to a general, central capacity limitation, but thought Miller had overestimated that limit, mostly because he had underestimated the role of chunking. Cowan (1999, 2001) proposed that the limit was closer to four than seven and that it indicated the number of chunks of information that could be held at one time in the focus of attention. The evidence comprised studies using stimulus sets for which chunking and rehearsal were presumably not used, such as briefly-presented visual arrays or spoken lists in the presence of articulatory suppression. He did not dispute the notion of a WM composed of several specialized mechanisms and stores, but did argue for the need to incorporate some kind of centralized, general purpose, capacity limited storage system also. Baddeley (2000) also has seen the need for the temporary storage of various types of abstract information outside of the specialized buffers and added another mechanism, the episodic buffer,

which could have the same type of limit as Cowan's (1999, 2001) focus of attention (Baddeley, 2001).

To learn if there is a common capacity limit shared between modalities (such as vision and hearing) or codes (such as verbal and spatial materials), what is needed first is a procedure capable of measuring capacity. Recently, considerable research to quantify memory capacity has made use of procedures in which an array of objects is presented and another array is presented, identical to the first or differing in a feature of one object (Luck & Vogel, 1997). This type of procedure can be analyzed according to simple models based on the assumption that k items from the first array are recorded in WM and the answer is based on WM if possible, whereas otherwise it is based on guessing (Pashler, 1988). According to one version of that model (Cowan, 2001; Cowan et al., 2005, Appendix A), WM is used both to detect a change if one has occurred and to ascertain the absence of a change if none has occurred. This leads to a formula in which $k = I * [p(\text{hits}) - p(\text{false alarms})]$, where I is the number of array items, $p(\text{hits})$ is the proportion of change trials in which the change is detected, and $p(\text{false alarms})$ is the proportion of no-change trials in which the participant indicated that a change did occur. This formula has been shown to result in k values that increase to an asymptotic level with between 3 and 4 objects (Cowan et al., 2005), in good agreement with many other indices of capacity (Cowan, 2001). It also increases in a linear fashion with the time available for consolidation of information into WM before a mask arrives (Vogel, Woodman, & Luck, 2006) and provides a close match to physiological responses of a capacity-limited mechanism in the brain (Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006). The array-comparison procedure seems to index WM for objects, inasmuch as the performance for combinations of different features, such as color and shape, seems roughly equivalent to the performance for the least

salient feature in the combination (Allen, Baddeley, & Hitch, 2006; Luck & Vogel, 1997; Wheeler & Treisman, 2002). We use this capacity measure for visual arrays of small, colored squares, and also extend it to auditory arrays of concurrently-spoken digits. We do so to determine whether the sum of visual and auditory capacities in a bimodal task can be made equal to the capacities in either modality within unimodal tasks and, if so, under what conditions.

If there is some kind of central, limited-capacity cognitive mechanism or storage system, one should be able to observe interference between tasks presumed to operate on very different kinds of information, like visuospatial and auditory/verbal WM tasks. However, several studies have found no interference between visuospatial and verbal tasks and, therefore, argue against any necessary reliance on a common resource (Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Luck & Vogel, 1997).

Recent studies by Morey and Cowan (2004, 2005) suggested that conflict between visuospatial and verbal WM tasks could be observed. They combined a visual-array comparison task (Luck & Vogel, 1997) with various verbal memory load conditions. Morey and Cowan (2004) found that performance on the array comparison task suffered when paired with a load of seven random digits, but not with a known seven-digit number. Morey and Cowan (2005) clarified the nature of this cross-domain interference by showing that primarily verbal retrieval, rather than verbal maintenance of the memory, interfered with visual memory maintenance. This cross-domain interference was taken as evidence that both tasks relied, in part, on a central cognitive resource. They suggested that this central resource might be the focus of attention (Cowan, 1995; 2001) or the episodic buffer of Baddeley (2000, 2001).

Nevertheless, the procedure of Morey and Cowan (2004, 2005) is not ideal for quantifying cross-domain conflicts in WM. The observed conflict was no larger than an item or two and the

total information remembered, verbal and visual items together, exceeded the total of 3 to 5 items hypothesized by Cowan (1999, 2001). We believe that the amount of inter-domain conflict observed in these experiments was limited because verbal storage benefits from resources other than the focus of attention. Indeed, the model of Cowan (1988, 1995, 1999) includes not only that central capacity, but also passive sources of temporary memory in the form of temporarily-activated elements from long-term memory, which exist outside of the focus of attention. This activation was said to include not only categorical information but also sensory memory over a period of several seconds (e.g., Cowan, 1988).

Although activated memory can be used for stimuli in all modalities, it may be especially important for materials that can be verbalized as an ordered sequence, because of the benefits of covert verbal rehearsal (Baddeley, 1986) and sequential grouping (Miller, 1956). To see how rehearsal can overcome the central storage limit, consider an often-invoked "juggler simile" for rehearsal, in which juggling increases the number of balls one can keep off the ground compared to just holding them in one's hands concurrently. One's hands represent the focus of attention in the model of Cowan (1988, 1999). The analogy for rehearsal is throwing the balls into the air (the simile for temporarily reactivating their long-term memory representations). The act of rehearsal increases the number of units that can remain active at once without being in the focus of attention. Where the simile may break down is that it implies that attention is integrated into phonological rehearsal loop, whereas attention and rehearsal may actually become separate resources after rehearsal of a sequence becomes automated (Baddeley, 1986; Guttentag, 1984). Rehearsal also might be used mentally to separate a series into distinct subgroups, a mental process that most adults have said they used in a serial recall task (Cowan et al., 2006). Given the availability of rehearsal and sequential grouping, adults may need to attend only the currently-

recited item (cf. McElree & Doshier, 2001), reserving most of the available attentional capacity for use on the spatial arrays.

Studies of immediate memory for both auditory and visual stimuli support the notion of separable central and non-central stores. The latter presumably includes sensory memory and other activated elements of long-term memory. Passive storage of such features is suggested by residual temporary memory for unattended speech (e.g., Cowan, Lichy, & Grove, 1990; Darwin, Turvey, & Crowder, 1972; Norman, 1969) and temporary mnemonic effects of unattended visual stimuli (e.g., Fox, 1995; Sperling, 1960; Yi, Woodman, Widders, Marois, & Chun, 2004). For auditory stimuli, work on the stimulus suffix effect (Crowder & Morton, 1969) reveals separable effects related to automatic versus attention-demanding storage. A final item that is not to be recalled (a suffix item) but is acoustically similar to the list items impedes performance in the latter part of the list. Greenberg & Engle (1983) found that such a suffix impeded performance of the last one or two items whether or not the suffix had to be attended, although attention modulated the suffix effect earlier in the list (see also Manning, 1987; Routh & Davidson, 1978). For recall of visual stimuli, irrelevant-speech effects (Salamé & Baddeley, 1982) demonstrate the passive, automatic nature of sensory storage. Speech to be ignored concurrent with a visually-presented verbal list impedes recall of that list. Hughes, Vachon, and Jones (2005) found that the recall of visually presented digits could be disrupted by unexpected, distracting changes in irrelevant speech, but that the factors underlying this distraction effect differed from those underlying the predominant type of irrelevant-speech interference, which depends primarily on changing state in the irrelevant speech stream rather than surprise. Thus, a wide variety of studies show that temporary memory includes components affected by attention and qualitatively different, passive components unaffected by attention.

Assuming that the passive memory tends to be available in parallel for different modalities and codes, as a considerable amount of evidence suggests (Baddeley, 1986; Brooks, 1968; Cowan, 1995; Nairne, 1990), presenting stimuli in two modalities or codes concurrently should make available not only the central, capacity-limited store, but also auxiliary sources of information from passive storage in each modality or code (e.g., central memory, visuospatial storage, and phonological storage). If all such sources of memory make a contribution to performance, then the measure of capacity for stimuli presented in two modalities or requiring two different codes should surpass memory for unimodal, unimodal presentations. Cowan (2001) assumed that passive sources of memory do not contribute to results in the two-array comparison procedure because the second array overwrites the features from the first array before a decision can be made. However, that sort of overwriting may not be complete, so passive sources of memory may contribute to performance after all.

In the experiments reported here, we have tried to devise a more direct evaluation of Cowan's (1995, 2001) theory of central capacity. In four of five experiments we presented both visual and auditory arrays at the same time, or nearly so, to minimize the use of rehearsal and grouping strategies. In unimodal conditions, participants were to remember only the visual or only the auditory stimuli and were subsequently tested for their memory of that modality. In a bimodal condition, participants were to remember both auditory and visual stimuli and could be tested for their memory of either modality. The visual task was like that used by Morey and Cowan (2004, 2005), except that we did not cue a particular stimulus in the second array for comparison with the first array. We dispensed with the cue because of the difficulty of devising an equivalent cue for the auditory modality, and because this cue seems to have very little effect on the results of visual-array experiments (Luck & Vogel, 1997). The auditory arrays comprised

four digits presented concurrently, each from an audio speaker in a different location. Like the visual memory task, the auditory memory task required participants to compare an initial array with a subsequent probe array and decide whether any one stimulus in the second array differed from the corresponding stimulus in the same location in the first array. To assist perception further, each audio speaker was assigned to a specific voice (male child, female adult, male adult, and female child) but digits presented in each audio speaker differed from trial to trial.

Although the visual array task of Luck and Vogel (1997) has been used in several recent experiments, our auditory array task, intended as an analogue of the visual array task, is more novel. Although many pioneering studies of attention and memory (e.g., Broadbent 1954; Cherry, 1953; Moray, 1959) used dichotic presentations of two concurrent sounds, relatively few presented more than two concurrent, spatially separated sounds, and those that did mostly examined selective attention rather than memory (e.g., see Broadbent, 1958 and Moray, 1969). There have been at least four previous studies of short-term memory involving spatial arrays of three or more sounds. Darwin et al. (1972) and Moray, Bates, and Barnett (1965) simulated three and four locations, respectively, over headphones, and examined the recall of concurrent lists from the different locations. However, their procedures had two limitations as measures of memory capacity. First, they confounded spatial and sequential aspects of memory. Second, they used recall measures, which produces output delay and interference that could reduce capacity estimates. This latter problem was avoided by Treisman and Rostron (1972) and Rostron (1974), who used loudspeaker arrays to present simultaneously two, three, or four tones, each from a different location. In their item recognition procedures, a sequence of two successive arrays was followed by a probe tone from one speaker. Participants were to indicate whether the probe tone had occurred in the first array and whether it had occurred in the second array, but memory for

spatial location was not tested. We presented just one memory array of concurrent digits from multiple loudspeakers, and our recognition task required knowledge of whether the identity of the digit assigned to any loudspeaker had changed.

We had no direct evidence that simultaneous arrays of spoken digits could not be converted into a verbal sequence and rehearsed. However, this seems much less natural and thus more difficult and effortful than the relatively automatic rehearsal of a verbal sequence. It would have to include the spatial information and, in our last experiment, it would have to include associations between each digit and the voice in which it was spoken (because it moved to a different, unpredictable location in the probe stimulus) and that seems especially unlikely. Therefore, we were reasonably confident that, if a central store existed, it would be needed for our auditory arrays as well as visual arrays and would show trading relations between visual and auditory information in storage.

In previous studies to combine multi-item memory tasks in two different modalities (Cocchini et al., 2002; Morey & Cowan, 2004, 2005), the stimuli in the two modalities were presented one after another. This has the consequence that the encoding of stimuli in the second modality occurs during maintenance of memory for the first-presented modality (Cowan & Morey, in press). We circumvented this situation in four of five experiments by making the encoding of the first arrays in the visual and auditory modalities concurrent. Also, previous studies examined recall in one modality and then the other. This results in retrieval in one modality concurrent with further maintenance in the other modality. To avoid that, we tested on only one modality per trial.

In brief, if remembering visual and auditory arrays occurs based only on a central capacity such as the focus of attention, then the total number of visual and auditory items that can be

remembered, together, from simultaneous visual and auditory arrays, should be no more the number of items that can be remembered from either alone, and that should be close to four items (Cowan, 2001). In contrast, contributions of modality-specific forms of memory should yield an increase in the total number of items remembered from bimodal as compared to unimodal arrays.

EXPERIMENT 1

Method

Participants

Twenty-eight undergraduate students, who reported having normal or corrected-to-normal vision and hearing and English as their first language, participated for course credit in an introductory psychology at the University of Missouri in Columbia. Of these, data from three participants were incomplete due to equipment problems. Another excluded participant completed the experiment after the orders of experimental blocks were counterbalanced. Twenty-four participants, 13 male and 11 female, were included in the final sample.

Design

The basic procedure involved a presentation of concurrent visual and auditory arrays, followed by a second array in one modality for comparison with the first array. In bimodal trial blocks, illustrated in Figure 1, the participant did not know which modality would be tested. This experiment was a 2 X 2 X 2 X 2 factorial within-participants design, with memory load condition (unimodal or bimodal), probe modality (visual or auditory), visual array set size (4 or 8), and trial type (change or no change between the sample and probe arrays in the tested modality) as factors. Memory load condition was manipulated by instructions to remember a single modality, visual or auditory, in the unimodal trial blocks, or both modalities in the bimodal trial block. There were a total of 160 experimental trials consisting of 80 unimodal trials

and 80 bimodal trials. The probe modality refers to which modality was presented for comparison with the previous array. Finally, the set size was manipulated for the visual stimuli by presenting 4 visual stimuli in some blocks and 8 visual stimuli in other blocks, always combined with 4 auditory stimuli. This was done because the visual stimulus yielding the appropriate range of difficulty could not be known in advance, whereas pilot data clearly indicated that 4 spoken stimuli were adequate to avoid ceiling effects.

Apparatus and Stimuli

Testing was conducted in 45-50 minute sessions individually with each participant, in a quiet room on a Pentium 4 computer with a 17 inch monitor. Auditory output was generated by an audio card with eight discrete audio channels.

Visual stimuli consisted of arrays of 4 or 8 colored squares, 6.0 mm X 6.0 mm (subtending a visual angle of about 0.7 degrees at a viewing distance of 50 cm), arranged in random locations within a rectangular display area, 74 mm wide by 56 mm high (8.5 degrees X 6.4 degrees), centered on the screen. The locations of squares were restricted so that all squares were separated, center to center, by at least 17.5 mm (subtending a visual angle of 2.0 degrees at viewing distance of 50 cm) and were at least 17.5 mm from the center of the display area. The color of each square was randomly selected, with replacement, from seven easily discriminable colors (red, blue, violet, green, yellow, black, and white). Each square had a dark gray shadow extending about 1 mm below and to its right. This shadow was intended to enhance contrast with the medium gray background, thereby accentuating a square's edges and location. The homogeneous gray background encompassed the screen's entire viewing area.

Auditory stimuli consisted of arrays of four spoken words, one from each of the four audio speakers. These auditory stimuli consisted of the digits 1-9, digitally recorded with 16-bit

resolution at a sampling rate of 22.05 kHz. To help distinguish the four channels and assist localization, each audio speaker played digits recorded from a different person, including a male adult, a female adult, a male child, and a female child. We also expected acoustic differences between the voices (e.g., different fundamental and formant frequencies) to improve their intelligibility by minimizing mutual interference and simultaneous masking (Brungart, Simpson, Ericson, & Scott, 2001; Assmann, 1995). Each audio speaker always played digits spoken by the same voice throughout the experiment. The actual durations of these spoken digits varied naturally within a range of about 260 to 600 ms. Auditory stimuli were reproduced by four small loudspeakers arranged in a semicircle. When participants faced the computer screen at a viewing distance of 50 cm, the speakers were 30 and 90 degrees to the left and to the right of their line-of-sight, at a distance of 1 meter. Each speaker was mounted on a stand so that the center of the speaker was 121 cm above the floor.

Each auditory array consisted of four spoken digits, randomly sampled with replacements from the digits 1-9, played at the same time, one from each audio speaker. The intensity of the digits from each audio speaker varied, because pilot testing indicated that setting their intensities to the same decibel level did not equate them for salience or intelligibility when digits from the four speakers were played simultaneously. Therefore, the intensity of each audio voice was repeatedly adjusted by 2 different persons during pilot testing and averages of these intensities were used to set the loudness levels of the four voices so that no single voice consistently stood out from the others. The intensities of the individual voices were about 65 to 75 dB and their combined intensity, played together, was about 71-78 dB(A).

Procedure

At the beginning of the experiment, the participant was given the following general

instructions: “In each trial of this experiment, you will compare two groups of spoken words (digits) and/or colored squares and try to detect any difference in the colors of corresponding squares or the spoken words. You are to press the “/” if you DO detect a difference, and the “z” if you do NOT notice any difference.” The experimenter read these instructions aloud while the participant followed the text printed on the screen. Thereafter, each block was preceded by more specific instructions displayed on the screen, as appropriate to that block. For example, the unimodal auditory condition with 4 visual stimuli had the following instructions: “In each of the following trials you will compare two groups of 4 spoken digits. These 4 spoken words will occur all at once, but each word will come from a different speaker. At the same time you hear the words, you will also see 4 colored squares, but you don't need to remember the colors in these trials.” Each set of instructions was followed by four practice trials and then 20 (in each of four unimodal blocks) or 40 (in each of two bimodal blocks) experimental trials. Immediately after the participant’s response on every trial, feedback was displayed which indicated “RIGHT” or “WRONG” and “There was a change” or “There was no change.” Thus, all trials in all blocks initially presented both auditory and visual stimuli. However, in the unimodal conditions, participants were instructed to attend to and remember either the visual or the auditory stimuli and then received a probe with stimuli only in that modality. In the bimodal conditions, participants were instructed to attend to and remember both the visual and auditory stimuli and then received a probe with stimuli in either one of the two modalities. They were told that any change which occurred would be in the modality of the probe stimuli.

Bimodal trials are illustrated in Figure 1. Each trial was initiated by the participant pressing the “Enter” key. A second later the words “Get ready” (on two lines in letters about 5 mm high) appeared in the center of the screen for 2 seconds, and then a 6 mm x 6 mm fixation cross

appeared in the center of the screen for 2 seconds. Next, the first visual array appeared for 600 ms then disappeared. On trials with a visual probe, 2 second after the onset of the first array, another array appeared for 600 ms, and then disappeared. This second array was replaced by a question mark (“?”) in the center of the screen, signaling that the participant should respond by pressing the “/” key if they noticed a change or the “Z” key if they did not notice any change. On half of the trials with a visual probe, one square in this second visual array was a different color than the corresponding square in the first array. The response keys were also marked on the keyboard with “≠” and “=” symbols to remind the participant of these instructions. On each trial, the presentation of the first auditory array of four spoken digits began at the same time as the onset of the visual array. On trials with an auditory probe, two seconds after the onset of the first auditory array, a second array of digits was presented. On half of the trials with an auditory probe, one digit in this second auditory array was different than the corresponding digit from the same speaker in the first array. After the second array, a question mark (“?”) appeared, signaling that the participant should respond by pressing the “/” key if they noticed a change or the “Z” key if they did not notice any change. The same sequence occurred in unimodal conditions but the probe was restricted to a single modality throughout a trial block.

The two unimodal conditions for each modality were run in adjacent blocks of 20 trials each, one with four auditory and four visual stimuli and the other with four auditory and eight visual stimuli. Each block of 20 trials included 10 change and 10 no-change trials, randomly ordered with a restriction of no more than 4 consecutive trials with the same answer. The bimodal trials were run in adjacent blocks of 40 trials each, one with 4 auditory and 4 visual stimuli and the other with 4 auditory and 8 visual stimuli. Each was composed of 20 visual probe and 20 auditory probe trials randomly intermixed with a restriction of no more than 4 consecutive trials

with the same probe modality. Each type of probe included 10 change and 10 no-change trials, randomly ordered with a restriction of no more than 4 consecutive trials with the same answer. New trials were generated for each participant.

The orders for the three conditions were counterbalanced between participants using all six possible permutations, combined with each order of the visual set sizes, blocks with 8 visual stimuli first or blocks with 4 visual stimuli first, within each adjacent pair of blocks comprising these conditions. For a particular participant, the order of adjacent blocks of visual set sizes was the same across all three conditions. This arrangement yields 12 different possible orders of trial blocks. Participants were randomly assigned to the 12 orders so that each order was run twice among the 24 participants.

Results

Although we think that capacity estimates (Cowan, 2001; Cowan et al., 2005) are theoretically more meaningful and easier to interpret and understand, we also analyzed the accuracy data to show that this traditional measure yielded comparable results. Table 1 shows the average change-detection accuracies for change and no-change trials for all conditions in Experiment 1. These data were analyzed in a 4-way repeated measure ANOVA with accuracy as the dependent variable and with memory load condition (unimodal or bimodal), probe modality (visual or auditory), visual array set size (4 or 8 squares), and trial type (change or no change) as within-participant factors. Most importantly, memory load condition exhibited a significant main effect, $F(1,23)=7.391$, $MSE=0.02$, $\eta_p^2=0.24$, $p<.05$, but did not interact with any other factor. Accuracy was better when only one modality was to be remembered, in the unimodal condition, $M=0.83$, $SEM=0.01$, than when both modalities were to be remembered, in the bimodal condition, $M=0.79$, $SEM=0.01$. There was a significant three-way interaction of probe modality

X visual array size X trial type, $F(1,23)=13.87$, $MSE=0.02$, $\eta_p^2=0.38$, because the set size of the visual arrays made more difference for the visual change trials than for the visual no-change trials or either type of auditory probe trial. This trend also produced a significant interaction of visual set size and trial type, $F(1,23)=9.27$, $MSE=0.02$, $\eta_p^2=0.29$. Visual set size had a significant main effect, $F(1,23)=56.20$, $MSE=0.01$, $\eta_p^2=0.71$, and interacted with probe modality, $F(1,23)=23.06$, $MSE=0.02$, $\eta_p^2=0.50$, as expected, making more difference for visual probe trials than for auditory probe trials. There were also significant main effects of probe modality, $F(1,23)=22.90$, $MSE=0.03$, $\eta_p^2=0.50$, and trial type, $F(1,23)=50.78$, $MSE=0.05$, $\eta_p^2=0.69$, of little theoretical consequence.

An analysis of capacity estimates showed that a dual-modality memory load reduced capacity at least for visual stimuli, but not as much as one would expect if the limit on the total number of items in memory were the same in the unimodal and bimodal conditions. Capacity estimates were calculated from the accuracy data based on the formula from Cowan et al. (2005): $k = I[p(\text{hits}) - p(\text{false alarms})]$, where I is the number of array items as explained in the introduction. The mean visual and auditory capacity estimates for all conditions and set sizes are shown in Table 2. A 3-way ANOVA with capacity as the dependent variable and with memory load condition (unimodal or bimodal), probe modality (visual or auditory), and visual array size (4 or 8 squares) as within-participant factors revealed a significant main effect of memory load condition, $F(1,23)=8.20$, $MSE=1.11$, $\eta_p^2=0.26$. The average capacity was 3.25 (SEM=0.33) for the unimodal memory condition and 2.81 (SEM=0.32) for the bimodal memory condition, a difference of 0.44 items due to the memory load in the non-tested modality. Also significant was the main effect for probe modality, $F(1,23)=73.90$, $MSE=2.03$, $\eta_p^2=0.76$, because visual capacity was higher than auditory capacity overall. The interaction of probe modality with memory load

condition was not significant. The main effect of visual set size was significant, $F(1,23)=5.66$, $MSE=1.25$, $\eta_p^2=0.20$, as was the interaction of visual set size with probe modality, $F(1,23)=9.14$, $MSE=1.64$, $\eta_p^2=.28$, and the reason for these effects was clarified by separate ANOVAS indicating that visual set size had a significant effect on visual performance, but not on auditory performance. Clearly, with 4 visual items the visual capacity measure approaches ceiling level, whereas somewhat higher capacity emerges with 8 visual items (Table 2).

The estimated visual capacity, averaged across visual set sizes, was greater for the unimodal condition ($M=4.26$, $SEM=0.25$) than for the bimodal condition, ($M=3.57$, $SEM=0.25$), representing a tradeoff of about 0.69 items, producing an effect of memory load condition in a separate ANOVA for this modality, $F(1,23)=9.93$, $MSE=1.24$, $\eta_p^2=.28$, whereas there was no such effect in an ANOVA of auditory capacity, $F(1,23)=0.95$, $MSE=0.76$, $\eta_p^2=.04$ (unimodal $M=2.23$, $SEM=0.15$; bimodal $M=2.06$, $SEM=0.13$).

If the visual array task estimates a central capacity and bimodal memory is limited to this central capacity, then the total number of items, auditory and visual, remembered on an average trial in the bimodal condition should be no more than that remembered in the unimodal visual condition. To examine this, visual and auditory capacities were calculated for each participant in the bimodal trial block, based on visual probe and auditory probe trials, respectively, and these were added to estimate the total bimodal capacity. These bimodal total capacities were compared to the unimodal visual capacities in a 2 X 2 repeated measure ANOVA, with memory load condition and visual set size as the two factors. The effect of memory load condition was significant, $F(1,23)=25.54$, $MSE=1.76$, $\eta_p^2=.53$, indicating that the average total capacity in the bimodal blocks, $M=5.63$, $SEM=0.32$, was more than the average capacity in the unimodal visual blocks, $M=4.26$, $SEM=0.25$. This difference can be seen in the first panel of Figure 2 by

comparing the solid gray bar above the unimodal label to the striped bar above the bimodal label. Participants remembered nearly 1.4 more visual and auditory items, together, in the bimodal condition than visual items, alone, in the unimodal condition.

Discussion

As a prequel to discussing capacity we note an asymmetry across stimulus modalities. Despite our use of voice cues in the auditory array task, it was more difficult than the visual array task. Recognition of spoken words has been found to suffer in the presence of simultaneous speech from spatially distributed sources, especially as the number of distracters increase from one to three (Drullman & Bronkhorst, 2000; Lee, 2001). The relatively poor spatial resolution of the auditory system and its susceptibility to masking probably contribute to this difficulty. Memory for auditory information in multiple channels is better when sequences of stimuli are presented in each spatial channel (Darwin et al., 1972; Moray, Bates, & Barnett, 1965) but we wanted to avoid possibilities for sequential chunking and attention switching. Consequently, unlike the visual array task, performance in the auditory array task in the unimodal condition probably is constrained more by perceptual than by mnemonic limitations, which is why we considered only unimodal visual capacity to be a possible estimate of a central memory capacity.

This experiment does suggest that auditory and visual arrays compete for some limited cognitive resource. The additional memory load provided by the auditory arrays decreased the amount of information remembered from the visual arrays by about 0.69 items, on average. This dual-memory cost is within the range of dual-task costs reported by Morey and Cowan (2005) when they had participants recite a 6-digit load during the maintenance period of a visual array comparison task. These results show that significant cross-domain interference can occur with a smaller memory load and without overt articulation and retrieval of the verbal load when all to-

be-remembered items in both domains are presented at the same time. Thus, our results provide additional evidence for the use of some kind of central storage shared by the auditory and visual tasks.

On the other hand, we failed to find strong evidence that memory capacity for visual and auditory arrays together was primarily limited by the same central resource as memory for only visual arrays. Participants remembered about 1.4 more auditory and visual items together than predicted by a central capacity based on memory for only visual arrays. Consequently, we still have not proven that this central resource is like a focus of attention holding three or four items, although that still could be the case if those extra items can somehow escape the constraints of that central capacity. The next experiment examines the effect of increasing the retrieval demands in the bimodal task to see if that prevents non-central memory sources from contributing to performance.

EXPERIMENT 2

Morey and Cowan (2005) found that more interference between tasks occurred if there was explicit retrieval of the verbal load during maintenance of the visual array. They enforced explicit retrieval by requiring overt articulation of the memory load. In our first experiment, the probe array in the bimodal condition was presented in only one modality or the other. Therefore, the participant only had to retrieve that modality. In the next experiment, the bimodal probe always includes both an auditory and a visual array. Half of the trials are change trials and when a change occurs, only one stimulus in one modality is different. Thus, participants have to maintain information in both modalities until they can look for a possible change in either modality. If retrieval processes necessarily use central resources, then the additional retrieval required in the bimodal condition should increase the amount of cross-domain interference.

Method

Participants

Twenty-four undergraduate students, 12 male and 12 female, who reported having normal or corrected to normal vision and hearing and English as their first language, participated for course credit in an introductory psychology at the University of Missouri in Columbia. Data for one male participant were excluded from the analyses because average recognition accuracy was no better than chance.

Apparatus and Stimuli

The apparatus and stimuli were the same as the previous experiment with one exception. That difference was that the bimodal conditions always presented both visual and auditory stimuli at the time of testing (see Figure 1). As before, both the auditory and visual probe stimuli occurred two seconds after the onset of the memory stimuli in both the unimodal and bimodal conditions.

Design and Procedure

Participants' instructions were the same as the previous experiment, except that participants were told to expect both visual and auditory stimuli at the time of testing in the bimodal conditions. They were also told that any change would occur in only one modality, auditory or visual, and in only one stimulus in that modality. Instructions in the unimodal, Remember Visual and Remember Auditory conditions were the same as in the previous experiment, as were all other aspects of the procedure.

Results

Both accuracies and capacities in the bimodal conditions had to be calculated differently for this and subsequent experiments than they were for the first experiment. In this experiment, both

modalities were presented as probes in bimodal memory trials. Consequently, there was no distinction between visual and auditory no-change trials. Thus, accuracies for the bimodal visual and auditory conditions were computed as the accuracy for hits on the visual and auditory change trials, respectively, averaged with the pooled accuracy for correct rejections on all no-change trials in the same block. This is why Table 1 has only one entry for visual no-change and auditory no-change trials for each block of the bimodal memory condition. Estimates of bimodal capacities also were based on these accuracy calculations incorporating the pooled rate of correct rejections in each bimodal block. Therefore, the analyses for this and subsequent experiments refer to ‘change modality’ rather than ‘probe modality’ to distinguish between memory trials relevant to auditory versus visual capacities, even though probe and change modality are the same for unimodal memory trials.

As can be seen in Table 1, the accuracy data for this experiment were a little lower, overall, but otherwise similar to the first experiment. The results of an ANOVA of these data were nearly identical to those of the previous experiment and will not be reported here because they were consistent with the more informative capacity analysis that follows.

Average visual and auditory capacity estimates for all conditions and set sizes in Experiment 2 are shown in Table 2. Individual capacity estimates were analyzed in 3-way ANOVA with memory load condition (unimodal or bimodal), change modality (visual or auditory), and visual array size (4 or 8 squares) as within-participant factors. This ANOVA revealed significant main effects of memory load condition, $F(1,22)=5.86$, $MSE=1.47$, $\eta_p^2=.21$, and change modality, $F(1,22)=124.97$, $MSE=1.04$, $\eta_p^2=.85$. As in Experiment 1, the average capacity was higher for the unimodal memory condition, $M=3.03$, $SEM=0.14$, than for the bimodal memory condition, $M=2.60$, $SEM=0.14$, and the difference of 0.43 items was quite similar to Experiment 1. Other

effects also were similar to Experiment 1. The interaction of memory load with change modality was again not significant. There was a significant main effect of visual set size, $F(1,22)=5.32$, $MSE=1.11$, $\eta_p^2=.19$, and a significant interaction of visual set size with change modality, $F(1,22)=9.56$, $MSE=0.85$, $\eta_p^2=.30$, that occurred because the visual capacity estimate was higher with 8 items in the visual array than with 4 items, which allowed near-ceiling-level performance.

Separate 2-way ANOVAs for each change modality showed a significant main effect of memory load condition for the visual modality, $F(1,22)=5.52$, $MSE=1.89$, $\eta_p^2=.20$. The estimated visual capacity, averaged across visual set sizes, was greater for the unimodal condition ($M=3.99$, $SEM=0.19$) than for the bimodal condition, ($M=3.32$, $SEM=0.24$), representing an average dual-task cost of about 0.67 items, about the same as the previous experiment. Once more there was no comparable effect for auditory change trials, $F(1,22)=2.01$, $MSE=0.42$, $\eta_p^2=.08$, $p=.17$ for memory load condition (unimodal $M=2.07$, $SEM=0.13$; bimodal $M=1.88$, $SEM=0.11$). Thus, like the previous experiment, the capacity tradeoff in the bimodal condition, compared to the unimodal condition, was significant only for visual capacity.

As in Experiment 1, to examine whether total bimodal capacity exceeded a hypothetical central capacity, based on unimodal visual capacity, visual and auditory capacities were calculated for each participant and bimodal trial block and then added to yield a total bimodal capacity. These bimodal total capacities were compared to the unimodal visual capacities in a 2 X 2 repeated measure ANOVA, with memory load condition and visual set size as the two factors. The effect of memory load condition was significant, $F(1,22)=14.40$, $MSE=2.32$, $\eta_p^2=.40$, indicating that the average total capacity in the bimodal blocks, $M=5.20$, $SEM=0.29$, was greater than the average capacity in the unimodal visual blocks, $M=3.99$, $SEM=0.19$. This difference is illustrated in the second panel of Figure 2 by the different heights of the solid gray

bar above the unimodal label and the striped bar above the bimodal label. Participants remembered about 1.2 more visual and auditory items, together, in the bimodal condition than visual items, alone, in the unimodal condition.

Discussion

The additional retrieval required by including both modalities in the bimodal probe made remarkably little difference. Although capacity estimates were generally lower than those found in Experiment 1, the pattern of effects and dual-task costs were nearly identical. For visual capacities, this cost was 0.67 items, compared to 0.69 in Experiment 1. The amount of extra capacity in the bimodal condition, based on the sum of the visual and auditory capacities compared to a central capacity estimated by the visual array task, amounted to about 1.2 items in this experiment, compared to 1.4 items in the previous experiment. It might have been expected that the two modalities in the probe would be checked sequentially, in which case there should be an attention cost for whichever modality was checked second.

At this point it is important to consider the contribution of mnemonic resources outside of attention. By using spatial arrays in both modalities, we presumably have eliminated or greatly diminished the contribution of one such resource, verbal rehearsal. Nevertheless, there are other forms of attention-free memory to be considered. As discussed above, a major source is sensory memory, explored on the basis of memory for unattended stimuli since the beginning of cognitive psychology (e.g., Broadbent, 1958) and incorporated into the model of Cowan (1988) as one family of activated features from long-term memory. Although sensory memory was somewhat ignored in the model of Baddeley (1986), one thing that these models have in common is the prediction that there exist attention-free types of memory that should be vulnerable to similar interfering stimuli. Visual or spatial forms of memory should be vulnerable to visual

interference, whereas acoustic forms of memory should be vulnerable to acoustic interference. In the approach of Baddeley (1986) this occurs because visuospatial and phonological forms of memory are separate modules, whereas in the approach of Cowan (1988, 1995, 1999), it occurs because the amount of interference depends on the amount of similarity in the activated features of the remembered and interfering stimuli, whether sensory or categorical (cf. Nairne, 1990).

In the following three experiments, we allow enough time for encoding of the arrays into WM and then present masking stimuli in each modality to eliminate any sensory and modality-specific forms of memory, hopefully providing a purer measure of central memory. Underlying this approach is the assumption that the masks will not interfere with information in the focus of attention, inasmuch as they will be designed to contain no new information and to be predictable and therefore not distracting (Cowan, 1995). Halford, Maybery, and Bain (1988) and Cowan, Johnson, and Saults (2005) similarly suggested that one special characteristic of items in the focus of attention is that they do not need to be retrieved (inasmuch as they are already retrieved) and therefore should be immune to the interference that occurs when items from the present list are similar to items from a previous list. The basic finding from these studies was that there was little proactive interference in a probe recognition task provided that the list contained 4 or fewer items, presumably few enough to be held in the focus of attention. We expect that the focus of attention in our procedure should be similarly spared from retroactive interference from the masks. Although we do not test that assumption directly, an important expected consequence is that including post-perceptual masks should leave information only in the focus of attention, which should result in a response pattern in which the sum of auditory and visual items retained in a bimodal task should be no greater than the items retained in the unimodal visual task.

EXPERIMENT 3

In the present experiment, we interposed auditory and visual masks between the initial and probe arrays. The masks occurred in the same locations as the items in the arrays and had similar sensory features. These masks were presented 1 s after the target arrays, late enough to allow consolidation of the array from sensory memory into WM (Vogel et al., 2006; Woodman & Vogel, 2005) but capable of overwriting sensory memory after that. If auditory and/or visual sensory memories are useful in these comparison tasks, subtracting their contributions might limit all storage to a central capacity.

We assumed this interval would ensure that the masks only interfered with the maintenance of modality-specific features but not with encoding or consolidation of information into WM. Using a similar visual array comparison task, Vogel et al. (2006) found that consolidation required only about 50 ms per item before a mask terminated the consolidation process. We are not aware of similar research on consolidation of auditory arrays but an upper bound can be estimated on the basis of backward recognition masking of a single item. A same-modality mask becomes ineffective when it is presented about 250 ms after the target to be identified, in both the auditory modality (Massaro, 1972) and the visual modality (Turvey, 1973). To find these masking functions, one must use a target set that is difficult enough to avoid ceiling effects well before 250 ms (Massaro, 1975). Thus it can be surmised that, to identify four spoken digits, even if they were maximally difficult and were identified one at a time, would take no more than 1 s of consolidation time.

Method

Participants

Twenty-four undergraduate students, 10 male and 14 female, who reported having normal or

corrected to normal vision and hearing and English as their first language, participated for course credit in an introductory psychology at the University of Missouri in Columbia.

Apparatus and Stimuli

The apparatus and stimuli were the same as the previous experiment except that masking stimuli were presented between the initial stimuli and the probe stimuli (see Figure 1). The visual masking stimuli consisted of an array identical to the initial array except that each square had 1 mm horizontal stripes of the seven colors used for the squares, arranged in random orders. The auditory masking stimuli consisted of combinations of the recordings of all nine digits in each voice played from the same auditory channel as the corresponding voice in the memory array. Their combined intensity was about 74 dBA, close to the average combined intensity of the initial and probe auditory stimuli. The onset of masking stimuli was one second after the onset of the initial stimuli in the same modality. The same visual and auditory masking stimuli were presented after the initial visual and auditory stimuli on every trial in all conditions.

Design and Procedure

Participants' instructions were the same as the previous experiment, except that the masking stimuli were described and referred to as "mixed irrelevant stimuli" presented between the initial and probe stimuli. Participants were instructed to ignore these stimuli. They were told that both visual and auditory stimuli would be presented at the time of testing in the bimodal conditions and that any change would occur in only one modality, auditory or visual, and in only one stimulus in that modality. Instructions in the unimodal memory conditions were the same as in the previous experiment, as were other aspects of the procedure.

Results

Proportion correct measures, which appear in Table 1, again will not be reported in

inferential statistics as the results simply reinforce those for the more informative capacity estimates, which appear in Table 2. A 3-way ANOVA with capacity as the dependent variable and change modality (visual or auditory), memory load condition (unimodal or bimodal), and visual array size (4 or 8 squares) as within-participant factors revealed significant main effects of memory load condition, $F(1,23)=12.60$, $MSE=2.25$, $\eta_p^2=.35$, and change modality, $F(1,23)=113.53$, $MSE=1.68$, $\eta_p^2=.83$, as in the previous experiments. As in Experiment 2 the interaction between these factors was not significant, $F(1,23)=2.24$, $MSE=0.99$, $\eta_p^2=.09$, $p=.14$. Unlike the previous two experiments, here the main effect of visual set size was not significant, $F(1,23)=0.00$, $MSE=1.49$, $\eta_p^2=.00$; nor was the interaction of visual set size with change modality significant, $F(1,23)=0.92$, $MSE=0.96$, $\eta_p^2=.04$.

Separate 2-way ANOVAs for each change modality confirmed that the pattern of results was more similar across modalities in this experiment than in the previous experiments. The 2-way ANOVA for the visual change trials showed a significant main effect of memory load condition, $F(1,23)=9.49$, $MSE=2.44$, $\eta_p^2=.29$, and no other effects were significant. The estimated visual capacity, averaged across visual set sizes, was significantly greater for the unimodal condition ($M=3.62$, $SEM=0.22$) than for the bimodal condition, ($M=2.63$, $SEM=0.24$), an average cost of about 0.98 items. Likewise, the 2-way ANOVA for the auditory change trials showed a significant effect of memory load condition, $F(1,23)=9.27$, $MSE=0.79$, $\eta_p^2=.29$, and no other significant effects. The estimated auditory capacity, averaged across visual set sizes, was significantly greater in the unimodal condition ($M=1.41$, $SEM=0.12$) than in the bimodal condition, ($M=0.85$, $SEM=0.10$), an average cost of 0.55 items. Thus, this experiment showed significant dual-task costs for both visual and auditory capacities.

To examine whether total bimodal capacity exceeded a hypothetical central capacity, visual and auditory capacities were calculated for each participant and bimodal trial block and then added to yield a total bimodal capacity. These total capacities were compared to unimodal visual capacities in a 2 X 2 repeated measure ANOVA, with memory load condition and visual set size as factors. This ANOVA yielded no significant effects. In contrast to the previous two experiments, here the total bimodal capacity, $M=3.49$, $SEM=0.30$, was actually slightly less than the unimodal visual capacity, $M=3.62$, $SEM=0.21$, though this difference was not statistically reliable, $F(1,23)=0.12$, $MSE=3.46$, $\eta_p^2=.01$. These results are illustrated by the third panel of Figure 2, in which the striped bar representing the total bimodal capacity is similar in height to the gray bar representing the unimodal visual capacity. Requiring simultaneous auditory memory detracts from visual memory a number of items similar to the number of auditory items retained.

Discussion

The addition of the sensory masks between initial and probe stimuli increased dual-task costs and led to memory tradeoffs in both modalities. Capacity estimates for both the visual and auditory conditions were significantly lower in the bimodal than unimodal conditions, amounting to about 1 visual item and 0.5 auditory items. Moreover, the total capacity of auditory and visual items, together, in the bimodal condition was slightly less than unimodal capacity for visual items, alone. That is, memory capacity for auditory and visual items retained together was no more than that for visual items retained alone.

These results are consistent with a bimodal capacity limited by a central attention, for example what Cowan (1999) called the focus of attention, a general, pervasive limitation in the number of items that can be simultaneously maintained by conscious attention, or what Baddeley (2000, 2001) referred to as an episodic buffer in WM. The finding seems inconsistent with

theories in which all working-memory storage is attributed to modality-specific or code-specific specialized stores (Baddeley, 1986; Cocchini et al., 2002), although it is possible that a central executive mnemonic process is capacity-limited rather than a central storage component (a possibility that we address in the General Discussion section).

Although comparison between experiments using different participants requires caution, the methodologies in Experiments 2 and 3 only differed by the irrelevant masking stimuli introduced in Experiment 3. Therefore, comparing capacity estimates from these two experiments should estimate the effect of the masking stimuli and provide an approximation of contributions from modality-specific stores in Experiment 2. Based on capacity estimates in Table 2, the masking stimuli reduced bimodal auditory and visual capacities by about 1.0 and 0.7 items, respectively. Because these stores presumably hold features, not items, it is more accurate to say that specialized stores apparently contributed sufficient modality-specific features to improve recognition performance by the equivalent of 1.7 items, total, in the bimodal condition of Experiment 2. Are these plausible contributions from sensory storage? At first glance, the 0.7 item difference in visual capacity might seem implausible, because visual sensory memory is generally considered to be short-lived, although very large in capacity (e.g., Sperling, 1960). However, the duration of visual sensory memory has never been resolved and, in some procedures, appears to persist for up to 20 seconds (Cowan, 1988). Moreover, an improvement in visual performance in the bimodal task can come from the use of auditory sensory memory to free up capacity, and conventional estimates of auditory sensory memory are several seconds, long enough to persist in our procedure, as well as being large in capacity (e.g., Darwin et al., 1972). Activated features in long-term memory that are not truly sensory in nature, such as the

phonological and visuospatial stores of Baddeley (1986), also may have been sources of information in Experiment 2 that were eliminated through post-perceptual masking in Experiment 3.

According to Vogel et al (2006), there should have been more than enough time before the mask to consolidate all visual items, but we do not know the time course for consolidating the auditory arrays or how the two modalities compete during encoding and consolidation. Inasmuch as visual and auditory arrays were presented at the same time in this experiment, this bimodal capacity limit might be caused at least partly by competition or interference during encoding and consolidation (cf. Bonnel & Hafter, 1998), rather than by storage limitations. The next experiment avoids this possibility by staggering the auditory and visual arrays so they can be encoded at different times. Although consolidation and maintenance of the different modalities will still overlap, Woodman and Vogel (2005) have shown these processes are separate stages that do not compete, at least for visual stimuli.

EXPERIMENT 4

Our goal in designing this experiment was to use the same methodology as the previous experiment while eliminating any temporal overlap of auditory and visual stimuli. However, staggering the presentation of the arrays and masks meant increasing the overall presentation time or reducing their durations. We choose to keep their durations the same, which meant that we had to increase the overall retention interval. Stimulus timings differed but otherwise the procedure was the same as the previous experiment. The stimulus timings allowed ample time between each memory array and its mask, while allowing more time between the masks and the probe array so that it could be anticipated and easily identified as such. One other difference related to the staggered presentations was that we randomly intermixed 50% of the trials with auditory arrays first and 50% with visual arrays first. To compensate for decreasing the number

of trials in each cell (for each order of presentation), we ran twice as many participant as we had in the previous experiments.

Method

Participants

Forty-eight undergraduate students, 21 male and 27 female, who reported having normal or corrected to normal vision and hearing and English as their first language, participated for course credit in an introductory psychology at the University of Missouri in Columbia.

Apparatus and Stimuli

The apparatus and stimuli were the same as the previous experiment except that the presentation of auditory and visual stimuli was offset with a stimulus onset asynchrony (SOA) of 720 ms so they did not overlap and other timing parameters were adjusted to accommodate the longer duration necessary for presenting both modalities. The order of visual and auditory presentation was counterbalanced in each block of trials so that half of the trials of each kind, auditory and visual change and no-change, presented the visual stimuli first and the rest presented the auditory stimuli first. The durations of visual and auditory stimuli were the same as those used in the previous experiments. For each modality, the onset of masking stimuli occurred 1480 ms after the onset of the initial stimuli and the onset of the probe stimuli occurred 4200 ms after the onset of the initial stimuli. These timings ensured that auditory and visual stimuli never overlapped. For example, in a bimodal trial with visual stimuli first, the initial visual array would appear for 600 ms. Then, after a blank interval of 120 ms, the auditory stimuli would occur, lasting no longer than 600 ms. After a blank interval of 160 ms, the visual mask appeared for 600 ms. Then, after another blank interval of 120 ms, the auditory masking stimuli occurred. The visual probe was presented for 600 ms after a longer blank interval, 1400 ms after the offset of

the last mask. Then, the auditory probe stimuli were presented 120 ms after the offset of the visual probe stimuli and a question mark appeared 600 ms later to cue the response.

The timings of stimuli in the unimodal trials were exactly the same, except that each trial included the presentation of only visual or auditory probe stimuli, whichever was to be remembered, and probe stimuli in the other modality were omitted and replaced with a blank gray screen and silence for the same 600-ms duration.

Design and Procedure

Participants' instructions were the same as they were for the previous experiment, except that the staggered presentation was also explained. Participants were instructed to ignore the "mixed irrelevant stimuli" (masks) and were told that both visual and auditory stimuli would be presented, one after the other, at the time of testing in the bimodal conditions and any change would occur in only one modality, auditory or visual. The orders for the various conditions were counterbalanced across participants, as in the previous study.

Results

Once more the statistical analysis of proportion correct scores, shown in Table 1, will not be presented and would only support the more informative analysis of capacity scores, shown in Table 2. A 4-way ANOVA with capacity as the dependent variable and with memory load condition (unimodal or bimodal), change modality (visual or auditory), modality order (visual-first or auditory-first) and visual array size (4 or 8 squares) as within-participant factors revealed significant main effects of memory load condition, $F(1,47)=64.87$, $MSE=2.63$, $\eta_p^2=.58$, and change modality, $F(1,47)=298.43$, $MSE=2.79$, $\eta_p^2=.86$, as in previous experiments. The interaction between these factors was also significant, $F(1,47)=53.13$, $MSE=2.32$, $\eta_p^2=.33$. This interaction was due to the bimodal memory condition hurting visual capacity more than auditory

capacity. The average visual capacity fell from 3.96 ($SEM=0.14$) in the unimodal condition to 2.49 ($SEM=0.15$) in the bimodal condition, a change of about 1.47 items. In comparison, the average auditory capacity fell from 1.35 ($SEM=0.09$) in the unimodal condition to 0.94 ($SEM=0.10$) in the bimodal condition, a change of 0.42 items. The main effect of modality order was not significant, $F(1,47)=1.32$, $MSE=2.16$, $\eta_p^2=.03$, $p=.26$, nor were any interactions with modality order. The order factor might be more relevant defined relative to the change modality, so this analysis was repeated with order defined in terms of change modality first versus non-change modality first, instead of visual first or auditory first, but the main effect of order still was insignificant, $F(1,47)=0.19$, $MSE=1.32$, $\eta_p^2=.004$, $p=.66$, as were all interactions with order. The interaction of visual set size with change modality was not significant.

Separate 2-way ANOVAs for each modality showed very similar patterns of effects. The 2-way ANOVA for the visual change trials showed a significant main effects of memory load condition, $F(1,47)=53.81$, $MSE=3.85$, $\eta_p^2=.53$, but no other effects were significant. Likewise, the 2-way ANOVA for the auditory change trials showed a significant effect of memory load condition, $F(1,47)=15.18$, $MSE=1.10$, $\eta_p^2=.24$, and no other significant effects. In this experiment, like the previous one, dual-task capacity tradeoffs occurred for both the visual change trials and the auditory change trials.

To evaluate the central capacity hypothesis, total bimodal capacities were calculated for each participant and trial block and compared to unimodal visual capacities in a 2 X 2 repeated measure ANOVA, with memory load condition and visual set size as the two factors. The effect of memory load condition was significant, $F(1,47)=5.05$, $MSE=5.37$, $\eta_p^2=.10$, but in the opposite direction to that found in the first two experiments. In this case, the average total capacity for the bimodal blocks, $M=3.43$, $SEM=0.20$, was significantly *lower* than the average capacity for the

unimodal visual blocks, $M=3.96$, $SEM=0.14$. This difference can be seen in the fourth panel of Figure 2, where the height of the unimodal gray bar exceeds that of the adjacent bimodal striped bar by 0.53 items.

Discussion

The results of Experiment 4 were remarkably similar to the previous experiment. Not only was the overall pattern of effects similar, but even the individual capacity estimates were similar. This can be seen by comparing the heights of the bars for Experiment 3 and Experiment 4 in Figure 2. Obviously, encoding constraints from the simultaneous presentations of auditory and visual memory arrays in Experiment 3 cannot explain the performance tradeoffs in the bimodal conditions. The same tradeoffs occurred when auditory and visual presentations did not overlap in Experiment 4. Therefore, bimodal capacity seems to have been limited mainly by how much information could be simultaneously maintained in storage.

The only notable change from the results of Experiment 3 was that the unimodal visual and bimodal total capacities were different in Experiment 4 and reversed from Experiments 1 and 2. Unlike the previous experiment, in this experiment unimodal visual capacity, which we have suggested might correspond to a general capacity limit, was significantly greater than a total capacity estimated by the sum of bimodal visual and auditory capacities. This might occur if alternating auditory and visual presentations selectively helped unimodal visual performance and/or impaired bimodal performance. In Figure 1, evidence for the former is most apparent, in that the unimodal visual capacity seems to improve between Experiments 3 and 4 whereas other capacity values remain very similar between these experiments (although this is a speculative comparison). Perhaps in Experiment 4 the alternating, rather than simultaneous, auditory stimuli were easier to ignore during encoding in unimodal visual trials. In contrast, when both modalities

require attention on bimodal trials, alternating presentations might not help because neither modality can be ignored. In any case, the small but significant advantage of unimodal visual capacity compared to bimodal total capacity in Experiment 4 might be attributed to any inefficiency or capacity overhead associated with coordinating information alternating between two modalities (cf. Broadbent, 1958).

One limitation in the previous experiments is that it does not establish the generality of the central storage medium because both auditory and visual arrays involved association between categorizable items and their locations in space. If the auditory task could be accomplished without remembering digit locations, then interference with the visual task would implicate a more general central storage medium. Testing this possibility is one goal of the next experiment.

Another limitation of the previous experiments has to do with the timing of the masking stimuli. In Experiments 1 and 2, the total time available for processing the stimuli to be remembered, from their onset until the probe onset, was 2000 ms. That is longer than the time from the onset of the stimuli to be remembered until the onset of the mask in Experiment 3 (1000 ms) or the time before the same-modality mask in Experiment 4 (1480 ms). It is possible that the masks work as an irrelevant change in physical properties of the display that interrupts processing (e.g., Lange, 2005) or impairs short-term consolidation (Jolicoeur & Dell'Acqua, 1998), which would violate our assumption that our mask operates post-perceptually. The next experiment examines various mask delays to see whether shorter processing time might have contributed to the different pattern of results in Experiments 3 and 4 compared to Experiments 1 and 2.

EXPERIMENT 5

This experiment used the same basic methodology as Experiment 3, with simultaneous

presentation of auditory and visual stimuli, except for two key changes (Figure 1, bottom). The first was to make spatial associations irrelevant to the auditory change detection task. This was accomplished by randomly assigning the four voices to the different loudspeakers for the initial auditory presentation of each trial and then randomly reassigning them for the probe presentation, so that each voice occurred in a different loudspeaker than it had in the initial memory array. Now, the instructions were to detect whether each voice spoke the same digit in the final presentation as it had in the initial presentation, or whether one voice said a different digit than it had initially. We were pleasantly surprised to find that subjects could do this task sufficiently well.

The second major change was to introduce three different mask delays, with 600, 1000, or 2000 ms SOAs from the initial auditory stimuli. The 600-ms delay was the minimum possible given the duration of the stimuli used in the previous experiments, and here as well; the 1000-ms delay was like that used in Experiment 3; and the 2000-ms delay allowed the same uninterrupted processing time of the initial memory arrays as in Experiments 1 and 2. The different mask delays were randomly intermixed in each of three blocks of trials, consisting of unimodal auditory, unimodal visual, and bimodal auditory/visual conditions.

The experiment was structured like Experiment 3 but increasing the masking delay from 1000 to 2000 ms also necessarily increased the memory (target-to-probe) SOA from 2000 ms to 3000 ms in order to maintain the same mask-to-probe SOA (1000 ms) as before. We worried that even that increase would not be sufficient because the mask might tend to be grouped with the probe given the overall temporal arrangement of stimuli. To investigate that question, we used a target array - to - probe array SOA of 3000 ms for half of the participants and 4000 ms for the other half (Figure 1). To compensate for the added variables in this experiment, we simplified

the block design by using only a single set size of 6 visual items, rather than separate blocks of 4 and 8 visual items as in the previous experiments.

Method

Participants

Thirty-nine undergraduate students, who reported having normal or corrected to normal vision and hearing and English as their first language, participated for course credit in an introductory psychology at the University of Missouri in Columbia. Of these, data from two initial participants were incomplete due to a programming mistake. Another participant was excluded because he did not follow instructions. Thirty-six participants (18 male and 18 female) were included in the final sample.

Apparatus and Stimuli

The apparatus and stimuli were the same as Experiment 3 except for the number of visual stimuli and the variable delays before presentation of auditory and visual masks and the number of visual stimuli on each trial. Only 6 colored squares were presented on each trial. There were only three blocks of trials, in counterbalanced order: unimodal visual trials, unimodal auditory trials, and bimodal visual and auditory trials. Durations of visual and auditory stimuli were the same as those used in the previous experiments. For each modality, the onset of masking stimuli occurred 600, 1000, or 2000 ms after the onset of the initial stimuli. The onset of the probe stimuli occurred 3000 ms after the onset of the initial stimuli for one group and 4000 ms after the onset of the initial stimuli for the other group. This procedure is illustrated at the bottom of Figure 1.

The masking stimuli were constructed differently for this experiment than for the previous experiments. Rather than combining each digit spoken by the same voice initially presented in

each loudspeaker, the mask in each channel was composed of voices randomly selected so that each voice produced 2 or 3 of the 9 digits concurrently presented in each speaker. This eliminated any chance that the mask might provide a cue about which voice had just been presented in that loudspeaker.

Design and Procedure

Participants' instructions were the same as in Experiment 3, except for details regarding the auditory change they were to listen for in the detection task. In this experiment, participants were to respond 'different' whenever they noticed that a particular voice said a different digit in the last group, compared to what the same voice said in the first group. Below is a portion of the instructions for the unimodal auditory condition:

First, you will hear 4 spoken digits. These 4 spoken digits will occur all at once, but each digit will be spoken by a different voice and will come from a different loudspeaker. At the same time, you will also see 6 colored squares, but you don't need to remember the colors in these trials. After a variable delay of about a half second to 2 seconds, you will hear and see the mixed irrelevant stimuli. Just ignore these sounds and squares. A little later, you will again hear 4 spoken digits, this time without any colored squares. Each voice will occur in a different loudspeaker than it did the first time, but the location of the voice is irrelevant, because it always changes. However, sometime the digit spoken by one of the voices in the last group will be different from the corresponding digit spoken by the same voice in the first group. Other times, each digit in the last group will be exactly the same as the digit spoken by the same voice in the first group.

There were three blocks of trials, two unimodal and one bimodal. Each unimodal block consisted of 48 trials including 16 (8 change and 8 no-change trials) at each mask delay, 600, 1000, or 2000 ms. In auditory change trials, each of the 4 voices changed with equal frequency. These trials were randomly ordered with a restriction of no more than 4 consecutive trials with

the same answer. The bimodal block consisted of 96 trials, 48 trials like those in each of the unimodal blocks, except that both modalities were always presented as probes. These were randomly intermixed with restrictions of no more than 4 consecutive trials with the same change modality and no more than 4 consecutive trials with the same answer. As in previous experiments, new trials were generated for each participant.

The order of the three conditions was counterbalanced between participants using all six possible permutations. This arrangement yields 6 different possible orders of trial blocks. Participants were randomly assigned to the 6 orders so that each order was run three times among the 18 participants in each of two groups, with memory delays of 3 seconds and 4 seconds.

Results

Inferential statistics based on proportion correct were similar to the more informative capacity estimates. A 4-way mixed ANOVA with capacity as the dependent variable and change modality (visual or auditory), memory load condition (unimodal or bimodal), and mask delay (600, 1000, or 2000 ms) as within-participant factors and memory delay (3000 ms or 4000 ms) as a between-participant factor revealed significant main effects of memory load condition, $F(1,34)=51.38$, $MSE=1.92$, $\eta_p^2=.60$, and change modality, $F(1,34)=150.00$, $MSE=2.89$, $\eta_p^2=.82$, as in previous experiments. The interaction between these factors also was significant, $F(1,34)=8.56$, $MSE=2.01$, $\eta_p^2=.20$. This interaction was due to the bimodal memory condition hurting visual capacity more than auditory capacity. The average visual capacity fell from 4.02 ($SEM=0.15$) in the unimodal condition to 2.67 ($SEM=0.19$) in the bimodal condition, a change of about 1.35 items. In comparison, the average auditory capacity fell from 1.61 ($SEM=0.15$) in the unimodal condition to 1.06 ($SEM=0.16$) in the bimodal condition, a change of 0.56 items. The

main effect of mask delay was significant, $F(2,68)=4.64$, $MSE=0.99$, $\eta_p^2=.12$. Newman-Keuls post-hoc tests yielded a significant difference between average capacities with a mask delays of 600 ms ($M=2.15$, $SEM=0.13$) and 1000 ms ($M=2.15$, $SEM=0.12$), although there was only a marginally significant difference ($p=0.07$) between 600 ms and 2000 ms ($M=2.37$, $SEM=0.12$). When the 600 ms delay was excluded from this ANOVA, the main effect of mask delay was not significant, $F(1,34)=1.73$, $MSE=0.77$, $\eta_p^2=.05$, $p=.20$. The 3-way interaction involving change modality, memory load, and mask delay was significant (for means see Table 3), $F(2,68)=4.77$, $MSE=1.22$, $\eta_p^2=.12$, although the interpretation is unclear. This unexpected interaction does not affect our hypotheses.

Separate 3-way ANOVAs for each change modality showed load effects for both. The 3-way ANOVA for the visual change trials showed significant main effects of memory load condition, $F(1,34)=56.29$, $MSE=1.76$, $\eta_p^2=.62$, and mask delay $F(2,68)=4.86$, $MSE=1.01$, $\eta_p^2=.13$, but their interaction was not significant, $F(2,68)=2.10$, $MSE=1.56$, $\eta_p^2=.06$. No effects of memory delay were significant. The estimated visual capacity was significantly greater for the unimodal condition ($M=4.02$, $SEM=0.15$) than for the bimodal condition, ($M=2.67$, $SEM=0.19$), an average cost of about 1.35 items. The effect of mask delay was due to reduced capacity at the shortest delay. According to Newman-Keuls post-hoc tests, visual capacity at the 600 ms mask delay ($M=3.04$, $SEM=0.16$) was significantly different from visual capacity at both the 1000 ms mask delay ($M=3.49$, $SEM=0.17$) and the 2000 ms mask delay ($M=3.50$, $SEM=0.19$), but the latter two capacities did not differ. The 3-way ANOVA for auditory change trials showed a significant effect of memory load condition, $F(1,23)=9.27$, $MSE=0.79$, $\eta_p^2=.29$, but no other main effects or interactions were significant. The estimated auditory capacity, averaged across memory and mask delays, was significantly greater in the unimodal condition ($M=1.61$, $SEM=0.15$) than in

the bimodal condition, ($M=1.06$, $SEM=0.16$), an average cost of 0.56 items. Thus, this experiment, like Experiment 4, showed significant dual-task costs for both visual and auditory capacities.

To examine whether total bimodal capacity exceeded a hypothetical central capacity, visual and auditory capacities were calculated and summed for each participant and bimodal trial block to yield estimates of total bimodal capacities. These total capacities were compared to unimodal visual capacities in a 2 X 3 X 2 mixed ANOVA, with memory load condition (bimodal or unimodal) and mask delay (600 ms, 1000 ms, or 2000 ms) as within-subject factors and memory delay (3000 ms or 4000 ms) as a between-subject factor. This ANOVA yielded a significant main effect for mask delay, $F(2,68)=4.61$, $MSE=1.94$, $\eta_p^2=.12$, but no other main effects or interactions were significant. The effect of mask delay was parallel to that found above, with capacity increasing as the mask delay increases from 600 to 1000 ms, then leveling out. Newman-Keuls tests showed that average capacity with a mask delay of 600 ms ($M=3.47$, $SEM=0.23$) was less than capacities with a 1000 ms mask delay ($M=4.05$, $SEM=0.23$) and a 2000 ms mask delay ($M=4.10$, $SEM=0.20$), but capacities at the longer two mask delays were not different from each other.

The finding most relevant to our main hypothesis was that the total bimodal capacity, $M=3.73$, $SEM=0.28$, was slightly less than the unimodal visual capacity, $M=4.02$, $SEM=0.15$, though this was not a significant difference, $F(1,34)=1.18$, $MSE=3.89$, $\eta_p^2=.12$, $p=0.28$. The same pattern emerged when only the masking intervals with asymptotically high performance (1000 and 2000 ms) were used. These results are illustrated in the fifth panel of Figure 2, where the striped bar representing the total bimodal capacity, averaged across both memory delays (3- and 4-s) and across the mask delays with asymptotic performance, is nearly the height of the

gray bar representing the unimodal visual capacity. At each of the three mask delays, the sum of bimodal capacities was no larger than the visual unimodal capacity. The bimodal sum and unimodal visual capacities (taken from Table 3 averaged across memory delays) were, for a 600-ms masking interval, 3.33 versus 3.61; for a 1000-ms masking interval, 4.03 versus 4.08; and for a 2000-ms interval, 3.84 versus 4.38.

Discussion

Results of the present experiment were little different than those of Experiments 3 and 4. Not only was the overall pattern of effects the same, but even unimodal and bimodal capacity estimates were similar, averaged across mask and memory delays (See Figure 2). There is no evidence that making the location of auditory stimuli irrelevant reduced the amount of interference between visual and auditory memory in the bimodal condition. The average cost of the bimodal load for visual capacity was about 1.35 items, compared to 0.98 items in Experiment 3 and 1.47 items in Experiment 4. The average cost of the bimodal load for auditory capacity was about 0.57 items, compared to 0.55 items in Experiment 3 and 0.42 items in Experiment 4. This experiment, like Experiments 3 and 4, showed significant dual-task costs for both visual and auditory capacities and a total capacity of auditory and visual items, together, in the bimodal condition, slightly less than the unimodal capacity for visual items, alone, consistent with a central capacity limitation shared by auditory and visual items.

Even though memory for the location of auditory items was irrelevant, sensory and/or perceptual processing of spatial cues probably still was necessary to separate and identify auditory items. So, we have to consider the possibility that interference was caused by competition between visual and auditory spatial encoding. However, staggered presentation of auditory and visual stimuli in Experiment 4 should have reduced encoding competition, but it did

not reduce interference between the two modalities. There is no reason to think that competition at encoding for spatial processing resources should be any more important in the present experiment, when spatial memory was irrelevant to auditory change detection. Taken together, these results indicate that dual-task interference in the bimodal memory condition does not depend on encoding or remembering spatial information in both modalities and suggest, instead, that the two modalities compete for a limited central capacity more general than spatial memory.

Although there was a significant effect of mask delay, results were not consistent with the hypothesis that the difference between Experiments 1 and 2 versus 3 and 4 was that the masks disrupted target array perception in Experiments 3 and 4. There were no statistically significant differences between capacities with mask delays of 1000 ms versus 2000 ms, so we feel that the perceptual process had ended by 1000 ms. Yet, the pattern of results obtained in Experiments 3 and 4, with the sum of bimodal capacities not exceeding unimodal visual capacity, was replicated in these conditions.

The shortest mask delay, 600 ms, did yield lower capacities, leaving open the possibility that masks presented sooner than 1000 ms after onset of the memory arrays might interrupt perceptual processing of the target arrays and impair their short-term consolidation into WM (Jolicoeur & Dell'Acqua, 1998; Vogel et al., 2006). There are possible implications of this for the interpretation of Experiment 4, in which the first target array was followed by the second one, in the other modality, after only 720 ms of consolidation time. If cross-modal stimuli interrupted consolidation, and if this interruption occurred more in the bimodal condition than in the unimodal condition, then part of the interference between visual and verbal memory in Experiment 4 could have been caused by competition for entry into WM (a consolidation bottleneck). However, the present experiment shows the same constant capacity even with mask

delays allowing asymptotic performance (1000 - 2000 ms). Furthermore, a consolidation bottleneck should apply even more heavily in Experiments 3 and 5, in which presentations in the two modalities were concurrent, than in Experiment 4 with its staggered presentation. However, bimodal performance levels were similar in all three experiments.

Last, the present experiment provides additional assurance that auditory comparisons were made on the basis of categorical information in WM. Given the low performance levels compared to visual arrays, and given participants' subjective impressions of those arrays, it is clear that some perceptual limit is likely to apply to them. One might entertain the possibility that each acoustic array is perceived as a complex, holistic sensory pattern in which the digits are not even identified, and that this complex sensory pattern is simply judged to be the same as or different from the probe array. However, this experiment provides two arguments against that hypothesis. First, as in Experiments 3 and 4, the mask should have eliminated the acoustic sensory representation as in aforementioned studies of the suffix effect (e.g., Crowder and Morton, 1969). Second, in Experiment 5 only, the auditory target and probe arrays were rather different on an acoustic basis no matter whether they were to be judged same or different. Specifically, the spatial location of each voice differed between the two arrays, and the judgment pertained to whether the linkage between voices and digits also differed. This comparison could not reasonably have been made on the basis of a holistic sensory pattern.

GENERAL DISCUSSION

Summary of Findings

To our knowledge, this is the first study in which a 1-to-1 tradeoff was obtained between memory in two different codes or modalities, as in our Experiments 3 through 5, and therefore it stands as a new type of evidence in support of a central component of WM storage. There have

been previous attempts to combine two types of memory but they have not succeeded at showing a 1-to-1 tradeoff. We believe the reason is that, as in our Experiments 1 and 2, these previous studies have used methods that allow forms of memory that do not trade off with one another. This evidence serves to strengthen the notion that an adequate theory of WM should include a component limited in capacity across stimulus domains.

The basic thrust of the present experiments is quite straightforward. Recognition memory for items in visual and auditory spatial arrays was examined in both unimodal and bimodal memory load situations. The task was to determine if a change in one item (visual or auditory) occurred or if there was no change. It was found that there was an incomplete tradeoff between visual and auditory memory in Experiment 1 (in which only one modality was present in the comparison probe stimulus) and Experiment 2 (in which both modalities were present concurrently). In particular, the sum of visual and auditory items retained from a bimodal memory load was greater than the number of visual (or auditory) items retained given a unimodal memory load. Inasmuch as both modalities were present on unimodal trials, with the same stimulus arrangements in both cases, and the load was manipulated by instructions and blocking of trials, this excess in bimodal memory cannot be attributed to stimulus-dependent perceptual factors. However, it can be attributed to the contribution of modality-specific stores such as auditory and visual sensory memory or other activated features of long-term memory that carry modality- or domain-specific information (for a review see Cowan, 1995) or phonological and visuospatial stores (Baddeley, 1986). Performance in these experiments cannot be accounted for entirely on the basis of modality-specific storage either, though (unless the capacity-limited component is a process rather than a store; see below), because the sum of visual and auditory items retained from a bimodal memory load was far less than twice the unimodal average. Thus,

memory is not independent in the two modalities and, although there was some gain when a memory load was distributed across modalities, there also was evidence for a central form of memory that is shared across modalities. It could be the focus of attention (Cowan, 1995, 1999) or the episodic buffer (Baddeley, 2000, 2001).

Experiments 3 - 5 served to isolate that central form of memory. Masking stimuli in Experiments 3 and 4 were placed at least 1 s after the stimulus array to allow time for information to be encoded (i.e., consolidated) into WM and then to overwrite the sensory memory in both modalities (Cowan, 1988, 1995; Nairne, 1990). In Experiment 3 (which resembled Experiment 2 except for the insertion of the masks), the sum of auditory and visual items retained from a bimodal array did not differ significantly from the highest unimodal capacity, the visual capacity. This suggests a tradeoff of allocation of a central memory resource, across modalities or to a single modality. The lower performance in the unimodal auditory case is attributed to limitations in perception of simultaneous sounds; the key point is that the number of auditory items stored reduced concurrent visual storage by a comparable number of items. In Experiment 4 (which differed from Experiment 3 in that the visual and auditory arrays were not presented simultaneously), the findings were similar except that there actually was a slight, significant superiority of unimodal memory over the sum of auditory and visual bimodal memories. That effect, the opposite of what was observed in Experiments 1 and 2, might be attributed to the cost of switching or sharing attention between modalities (Broadbent, 1958) and is consistent with the idea that a central memory was the main or sole source of storage in this study when sensory memory or other modality-specific memory was overwritten using masking stimuli.

Experiment 5 examined two concerns that limited interpretations of the preceding experiments. The first concern was that both auditory and visual arrays involved association between categorizable, verbalizable items and their locations in space. This means that bimodal interference might have been limited to a spatial short-term memory. In Experiment 5, spatial cues were irrelevant to the auditory task, which was based, instead, on remembering item-voice associations. Nevertheless, Experiment 5 showed the same pattern of interference between modalities as in Experiments 3 and 4, matching what would be expected if the primary source of memory were a capacity-limited store.

A second concern was that the masks in Experiments 3 and 4 might have reduced processing time compared to Experiments 1 and 2. Therefore, it was not clear whether the masks have brought the total bimodal capacity in line with unimodal visual capacity by overwriting modality-specific storage, as we suggested, or instead by allowing less time to process and consolidate items under dual-task conditions that could slow processing. In Experiment 5, we compared unimodal and bimodal capacities at mask delays of 600, 1000, and 2000 ms. The 2000-ms mask delay showed no better bimodal capacity than the 1000-ms mask delay. The 600-ms mask delay did reduce capacity estimates, but this effect was not related to memory load, and therefore failed to show that interrupted processing would selectively impair bimodal capacity. At all mask delays, bimodal capacity failed to exceed either unimodal capacity, consistent with Experiments 3 and 4.

Comparing across experiments, Figure 2 suggests that the effects of the masks on unimodal performance were larger for audition than for vision. This difference is consistent with a large auditory modality advantage in serial recall (Penney, 1989) that suggests a certain greater robustness or persistence of acoustic memory compared to visual memory. However, the

auditory modality advantage in serial recall instead may be related to the advantage of the acoustic modality for sequential stimuli, as Penney suggested. If so, the larger effect of auditory masks compared to visual masks in unimodal performance in the present study using arrays may not be related to the traditional modality effect. It could have more to do with the perceptual difficulty of encoding multiple speech sounds at the same time and the importance of sensory memory when perception is difficult. Thus, the precise role of auditory and visual stores warrants further investigation. In bimodal performance, either mask could affect performance in either modality because central capacity can be reallocated across modalities as needed.

The limit for central storage in our set of experiments was about 3-4 items, as expected from a great deal of past research (for reviews see Cowan, 2001, 2005a; Cowan et al., 2005). The larger limit of about seven items (Miller, 1956) can be attributed to the use of rehearsal and grouping strategies for verbal sequences (Cowan, 2001). These strategies are presumably difficult or impossible to implement for briefly-presented simultaneous arrays like ours. Attempts to examine memory for larger multi-item chunks of verbal information (Chen & Cowan, 2005; Cowan, Chen, & Rouder, 2004; for foundations of this work see Johnson, 1978; Slak, 1970; Tulving & Patkau, 1962) produce similar capacity estimates of 3 to 4 chunks, except when the list is limited to a verbal sequence short enough to be rehearsed, in which case participants may not need to rely upon the central memory mechanism (Chen & Cowan, 2005). WM for objects in knowledge-intense situations such as chess appears similar, although there is some debate about the exact capacity given the difficulties of identifying chunks (Gobet & Clarkson, 2004).

Potential Boundary Conditions for the Findings

It is helpful to contrast our findings to other studies with results discrepant from ours to

determine the boundary conditions for observing a 1-to-1 tradeoff between WM for stimuli in two modalities. A good example of a previous study combining two types of memory is Sanders & Schroots (1969). They present two lists to be recalled, including lists of consonants paired with lists of consonants, digits, tones, or spatial locations (and also two lists of tones). As the similarity between lists increased, the amount of interference also increased. However, for these sequential stimuli there was no metric of capacity and the conditions did not allow a way to quantify the tradeoff. The similarity-dependent nature of the tradeoffs would be expected given that no masking stimuli were present to eliminate modality- and code-specific sources of memory. If two lists draw on the same type of code (e.g., both verbal) then one cannot determine how much of that tradeoff depends on a central, modality-independent form of memory.

Cocchini et al. (2002) and Morey and Cowan (2004, 2005) did directly compare unimodal and bimodal memory situations using very different stimuli (verbal lists combined with spatial arrays) and found some tradeoff between modalities, but very little except when the verbal lists were spoken aloud (cf. Baddeley, 1986). Given the dissimilarity, the main basis of tradeoff would be from sharing a central memory. However, if that central memory is the focus of attention (Cowan, 1988, 1999, 2001), it would be expected that that tradeoff should be small because covert verbal rehearsal, which takes up very little attention (Guttentag, 1984), can be used to retain verbal lists, freeing up the focus of attention for memory of the spatial arrays.

A number of recent findings have suggested that, unlike spoken lists, spatial arrays require attention, or what is typically termed central executive function, for maintenance in working memory (e.g., Fisk & Sharp, 2003; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). However, the attentional involvement for non-sequential spatial tasks appears to be smaller than in sequential tasks (Rudkin, Pearson, & Logie, 2007) so, to observe a 1-to-1 tradeoff between

visual and auditory items, our use of visual as well as acoustic masks to interfere with attention-free sources of memory in Experiments 3 - 5 may well have been critical.

Alternative Theoretical Accounts of Capacity Limits in WM

According to our theoretical framework (Cowan, 1988, 1999, 2001, 2005a; Cowan et al., 2005), temporarily activated forms of long-term memory, which can include sensory memory, are vulnerable to interference from other stimuli with similar features (cf. Nairne, 1990). The difference between visual color stimuli and spoken digit stimuli that we used should be large enough to reduce cross-modal interference in memory. In contrast, by using arrays of auditory and visual stimuli and post-perceptually masking them in Experiments 3 through 5, we prevent all but a central form of memory. According to our theoretical framework, this is the focus of attention. Similarly, Baddeley (2000, 2001) could account for the same findings designating modality-specific buffers as the ones vulnerable to masking and proposing that the episodic buffer is subject to a tradeoff between items regardless of the modality, not unreasonable assumptions given that framework.

There are important remaining questions that must be answered before the nature of the WM tradeoff we have observed can be understood fully. The theory of Cowan (1988, 2001, 2005a) includes attention as a storage device but similar results might be achieved if attention is used only as a processor, as in the central executive function of Baddeley (1986) or Baddeley and Logie (1999). In specific, its function might be to fend off interference from subsequent stimuli (such as our masks) and help to keep the information loaded in the passive, attention-free types of memory. That is quite difficult to distinguish from the notion that the presence of information in the focus of attention protects it from interference (e.g., Cowan, Johnson, & Saults, 2005; Halford et al., 1988). Confirming that storage and processing are related, Oberauer & Göthe

(2006) have provided additional evidenced that storage and processing trade off within a limited focus of attention, except that items not involved in ongoing operations can be set aside within activated memory without competing for limited attentional resources. It is not yet clear if this relation occurs because of common storage or common processing (but see below). In any case, our study of 1-to-1 tradeoffs and the boundary conditions needed to observe them helps to constrain the theory of central memory, without resolving the role of storage and processing in that memory.

A question that is related to potential processing limitations, discussed above, has to do with the role of the masking stimuli. We have argued that the masking stimuli overwrite passive, modality-specific features, eliminating contributions from sensory memory and other modality-specific types of memory. If, on the other hand, interference from the masking stimuli is not modality-specific but more general, then other interpretations are possible. The irrelevant masking stimuli might enter WM and replace representations of relevant stimuli. That is, the masks may be distractions that temporarily occupy part of a limited focus of attention that otherwise could be used to maintain information in WM. However, that interpretation provides no account of our key finding that, when the masks were added, the bimodal total capacity was roughly equivalent to the unimodal visual capacity. It also leaves unexplained the observation that the mask generally had much more effect on auditory performance than on visual performance (see Figure 2). That finding is consistent with the common observation that auditory memory tends to be used over a much longer period (e.g., Penney, 1989).

Recent evidence seems to support the existence of a capacity-limited region of storage. McElree (1998, 2001) suggested that the focus of attention stores a single chunk because the last one (or the one that the participant is asked to hold) has a faster retrieval dynamic in a probe

recognition task than the other items, which tend not to differ. Verhaeghen, Cerella, and Basak (2004) found that the retrieval dynamic demonstrates that, with task practice, the focus of attention for items in a list expands from a single item to four items (but see Oberauer, 2006). Oberauer (2002, 2005) found that reaction times were affected by the number of items in a set designated as task-relevant, but not the number of items in a set designated as task-irrelevant, which still had to be remembered for later. He suggested that all items reflect activated elements of long-term memory but that only several items in a task-relevant set could be in a capacity-limited region, and that only one of those items could be in the focus of attention and benefit from the fastest reaction time. An alternative interpretation of his findings is that the entire capacity-limited region is in the focus of attention but with higher priority for some items than for others (Cowan, 2005a).

Recent brain research also helps to resolve this issue and suggestions that there is a separate, general WM storage function in parietal regions that may integrate input across modalities. Cowan (1995, 2005a) proposed that the storage function of attention is represented largely in the parietal lobes of the brain, whereas the related processing of the central executive is represented largely in the frontal lobes. Supporting this suggestion, Postle et al. (2006) found that, although both the frontal and parietal lobes were active for both storing or manipulating information in WM, transcranial magnetic stimulation to frontal lobe locations disrupted only manipulation, not storage; comparable stimulation to parietal lobe locations disrupted both, which would be expected inasmuch as manipulation depends on storage. It is not yet clear if these parietal areas are related to the seat of attention, but Cowan (1995, 2005a) reviewed evidence that they are (e.g., the evidence on a distinction between anterior and posterior attention systems discussed by Posner & Peterson, 1990).

Cowan (1988, 1995, 1999) suggested that the focus of attention is the basis of capacity-limited central storage and that it is controlled partly by the central executive. Cowan (2001) then summarized evidence that storage is limited to about 4 chunks (or items, if they are not chunked together). However, that formulation does not explain how storage and processing trade off with one another. This question was addressed by Cowan (2005a, 2005b). In cases of difficult processing, it is presumably necessary to use the focus of attention to hold on to a goal that goes against other prepotent but inappropriate inclinations (e.g., Kane, Bleckley, Conway, & Engle, 2001). For example, it is difficult to attend to a classroom lecture when there is a shouting crowd that can be heard outside of the classroom. The need to retain the goal of listening to the lecturer in such a situation presumably requires at least one slot in the focus of attention and detracts from space that otherwise could be spent retaining data from the lecture. In contrast, in quiet listening situations, the task goal may become automatic and not require a slot from the focus of attention.

This study can be viewed as part of a larger effort to identify the structure of WM. A close reading of the seminal article by Baddeley and Hitch (1974) suggests that they entertained the possibility of a central memory that must share resources with processing, as well as separate dedicated, passive stores (the phonological and visuospatial stores) that operate without drawing on this central resource. Baddeley (1986) later simplified the model by eliminating the storage function of the central resource, making it a processing resource only (termed the central executive). However, the insufficiency of that solution was suggested by Baddeley's (2000, 2001) later addition of an episodic buffer to account for WM for types of information that did not fit the description of phonological or visuospatial information, such as semantic information and associations between different types of features. Still, Baddeley (2000, 2001) has left open the

question of whether storage (via the episodic buffer) and processing (via the central executive) share a resource or not. A large body of research that examines individual differences in WM and cognitive aptitudes has continued to assume that storage and processing must share resources (e.g., Cowan et al., 2005; Daneman & Carpenter, 1980; Kane et al., 2004).

Just as we have obtained reciprocity between visual and auditory WM storage using capacity measures, there is a study that has obtained reciprocity between two processes, pursuit tracking and auditory discrimination, using an event-related potential (ERP) measure (Sirevaag, Kramer, Coles, & Donchin, 1989). As the pursuit tracking task increased in difficulty, the amplitude of the P-300 component of the ERP increased, and the amplitude of the P-300 to the concurrent auditory discrimination task decreased in such a way that the sum of the two amplitudes remained constant. One can envision a program of research in which both behavioral and physiological measures are used to determine not only whether two types of storage or two types of processing display reciprocity, but also whether storage and processing trade off with one another in a manner consistent with the notion that both of them depend on a common resource. Evidence that visual storage and auditory retrieval trade off (Morey & Cowan, 2004, 2005) tentatively points toward that conclusion.

Another basis for suspecting that storage and processing trade off within the focus of attention comes from work on the nature of complexity in processing. Halford, Wilson, and Phillips (1998) set out a framework in which the complexity of processing depends on how many elements must be associated with one another for a correct understanding of the material. For example, in order to understand a crossover interaction between two 2-level variables, one must keep in mind that $A > B$ at Level 1 but that $A < B$ at Level 2, and this is more complex than keeping in mind two main effects (e.g., $A_1 > A_2$ and $B_1 > B_2$) (Halford, Baker, McCredden, &

Bain, 2005). Given that the associations can be kept in mind only if the elements being associated are kept in mind, it stands to reason that the complexity of processing that can be accomplished should vary with the amount that must be held in the focus of attention concurrently, although there is yet little work to address that suggestion.

Building on the Research

This series of experiments makes another point about the larger effort to identify the structure of WM. It shows the potential consequences of ignoring sensory memory in one's theoretical model. Even though one of the first models of human information processing included sensory stores (Broadbent, 1958) and there have been few challenges to the basic evidence for brief sensory storage, Cowan's embedded process theory is one of the few recent models of WM that have any provision for sensory storage. Maybe most WM theorists do not believe that sensory storage makes much difference to the work that WM does. Our present research, however, illustrates one methodological pitfall of ignoring sensory storage, because our evidence of a central capacity limit was obscured by contributions of passive storage that may well have been sensory in nature (although it alternatively could be viewed as the result of modality-specific passive buffers, as in Baddeley, 1986) and was only revealed when the stimuli to be recalled were post-perceptually masked. This suggests, furthermore, that contributions from sensory storage might be important in other WM assessment tasks and even practical tasks that are thought to depend on WM capacity.

One important question raised by our present findings, that need to be addressed by future research, is whether the same central capacity limit we have found for visual and auditory arrays can apply to combinations of auditory and visual sequences and arrays. As we noted above, previous attempts have shown only modest tradeoffs, yielding total capacities that exceed the 3

to 4 item limit we found in these studies. However, none of these previous attempts have used procedures that prevent chunking and eliminate contributions from automatic sources, like rehearsal and sensory memory. A developmental study by Cowan et. al. (2005) did use tasks that presented sequential stimuli that might help accomplish this, a running memory span task and a change-detection task using sequences of tone. The running memory span procedure involved the presentation of digits at a rapid rate, 4/s and unpredictable length. As soon as the list ended, as many items as possible were to be recalled from the end of this list. It is not possible to encode and update the digits fast enough for them to be rehearsed, so the best strategy is to passively listen. The tone change detection task, analogous to the array task in the present experiments, presented two sequences of tones, followed by a response indicating whether one tone was different in the last sequence compared to the first. The nonverbal nature of the stimuli in this task made them hard to rehearse. Tasks modeled on these, but also including masks to eliminate sensory memory, could be useful in combination with array tasks to determine whether the central capacity limit we have reported here for auditory and visual arrays together also apply to remembering sequences together, or one sequence and one array.

In sum, although fundamental questions about the nature of WM remain, we believe that our findings provide an important step toward identifying a type of limited-capacity WM storage that cuts across modalities and codes, in that it shows reciprocity between auditory and verbal array memory provided that modality-specific memory is masked.

References

- Allen, R.J., Baddeley, A.D., & Hitch, G.J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*, 298-313.
- Assmann, P. F. (1995). The role of formant transitions in the perception of concurrent vowels. *Journal of the Acoustical Society of America*, *97*, 575-584.
- Baddeley, A.D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*, 417-423.
- Baddeley, A. (2001). Comment on Cowan: The magic number and the episodic buffer. *Behavioral and Brain Sciences*, *24*, 117-118.
- Baddeley, A., & Hitch, G.J. (1974). Working memory. In G. Bower (ed.), *Recent advances in learning and motivation*, Vol. VIII. New York: Academic Press.
- Baddeley, A.D., & Logie, R.H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, U.K: Cambridge University Press. (pp. 28-61).
- Bonnel, A.-M., & Hafter, E.R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics*, *60*, 179-190.
- Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*. *47*, 191-196.
- Broadbent, D.E. (1958). *Perception and communication*. New York: Pergamon Press.
- Brooks, L.R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*, 349-368.

- Brungart, D., Simpson, B., Ericson, M., & Scott, K. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, *110*, 2527-2538.
- Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1235-1249.
- Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of Acoustic Society of America*, *25*, 975--979.
- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*, 1086-1095.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, *104*, 163-191.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, U.K.: Cambridge University Press. (pp. 62-101)
- Cowan, N. (2001). The magical number four in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-114.
- Cowan, N. (2005a). *Working memory capacity*. New York: Psychology Press.

- Cowan, N. (2005b). Working memory capacity limits in a theoretical context. In C. Izawa & N. Ohta (eds.), *Human learning and memory: Advances in theory and application*. Mahwah, NJ, US: Lawrence Erlbaum Associates (pp. xvi-317).
- Cowan, N., Chen, Z., & Rouder, J.N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, *15*, 634-640.
- Cowan, N., Elliot, E. M., Saults, J. S., Morey, C., Mattox, S., Hismjatullina, A., and Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42-100.
- Cowan, N., Elliott, E.M., Saults, J.S., Nugent, L.D., Bomb, P., & Hismjatullina, A. (2006). Rethinking speed theories of cognitive development: Increasing the rate of recall without affecting accuracy. *Psychological Science*, *17*, 67-73.
- Cowan, N., Johnson, T.D., & Saults, J.S. (2005). Capacity limits in list item recognition: Evidence from proactive interference. *Memory*, *13*, 293-299.
- Cowan, N., Lichty, W., & Grove, T.R. (1990). Properties of memory for unattended spoken syllables. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 258-269.
- Cowan, N., & Morey, C.C. (in press). How can dual-task working memory retention limits be investigated? *Psychological Science*.
- Crowder, R. & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*, 365-373.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450-466.

- Darwin, C., Turvey, M., & Crowder, R. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 3, 255-267.
- Drullman, R. & Bronkhorst, A. W. (2000). Multichannel speech intelligibility and talking recognition using monaural, binaural, and three-dimensional auditory presentation. *Journal of the Acoustical Society of America*. 107, 2224-2235.
- Fisk, J.E., & Sharp, C.A. (2003). The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition*, 52, 364-381.
- Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review*, 2, 145-173.
- Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four... or is it two? *Memory*, 12, 732-747.
- Greenberg, S.N., & Engle, R.W. (1983). Voice change in the stimulus suffix effect: Are the effects structural or strategic? *Memory & Cognition*, 11, 551-556.
- Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology*, 37, 92-106.
- Halford, G.S., Baker, R., McCredden, J.E., & Bain, J.D. (2005). How many variables can humans process? *Psychological Science*, 16, 70-76.
- Halford, G.S., Maybery, M.T., & Bain, J.D. (1988). Set-size effects in primary memory: An age-related capacity limitation? *Memory & Cognition*, 16, 480-487.
- Halford, G.S., Wilson, W.H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, 21, 723-802.

- Hughes, R.W., Vachon, F., & Jones, D.M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 736-749.
- Johnson, N.F. (1978). The memorial structure of organized sequences. *Memory & Cognition*, *6*, 233-239.
- Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*, *36*, 138-202.
- Kane, M.J., Bleckley, M.K., Conway, A.R.A., & Engle, R.W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169-183.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. E. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217.
- Lange, E.B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language*, *53*, 513-531.
- Lee, M. (2001). Multichannel auditory search: Toward understanding control processes in polychotic auditory listening. *Human Factors*, *43*, 328-342.
- Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-81.
- Manning, S.K. (1987). Attentional control of visual suffix effects. *Bulletin of the Psychonomic Society*, *25*, 423-426.

- Massaro, D.W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 79, 124-145.
- Massaro, D.W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America*, 58, 1059-1065.
- McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, 38, 225-252.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 817-835.
- McElree, B., & Doshier, B.A. (2001). The focus of attention across space and across time. *Behavioral and Brain Sciences*, 24, 129-130.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent variable analysis. *Journal of Experimental Psychology: General*, 130, 621-640.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56-60.
- Moray, N. (1969). *Listening and attention*. Middlesex, England: Penguin Books.
- Moray, N., Bates, A., & Barnett, T. (1965). Experiments on the four-eared man. *Journal of the Acoustical Society of America*. 38, 196-201.
- Morey, C.C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, 11, 296-301.

- Morey, C.C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 703-713.
- Nairne, J.S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251-269.
- Norman, D.A. (1969). Memory while shadowing. *Quarterly Journal of Experimental Psychology*, *21*, 85-93.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 411-421.
- Oberauer, K. (2005). Control of the contents of working memory—A comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 714–728.
- Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*, 197-214.
- Oberauer, K., & Göthe, K. (2006). Dual-task effects in working memory: Interference between two processing tasks, between two memory demands, and between storage and processing. *European Journal of Cognitive Psychology*, *18*, 493-519.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369-378.
- Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398-422.
- Posner, M.I., & Peterson, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25-42.

- Postle, B.R., Ferrarelli, F., Hamidi, M., Feredoes, E., Massimini, M., Peterson, M., Alexander, A., & Tononi, G. (2006). Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in the prefrontal, but not posterior parietal, cortex. *Journal of Cognitive Neuroscience*, *18*, 1712-1722.
- Rostron, A.B. (1974). Brief auditory storage: Some further observations. *Acta Psychologica*, *38*, 471-482.
- Routh, D.A., & Davison, M.J. (1978). A test of precategorical and attentional explanations of speech suffix interference. *Quarterly Journal of Experimental Psychology*, *30*, 17-31.
- Rudkin, S.J., Pearson, D.G., & Logie, R.H. (2007). Executive processes in visual and spatial working-memory tasks. *Quarterly Journal of Experimental Psychology*, *60*, 79-100.
- Salamé, P., and Baddeley, A. (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.
- Sanders, A.F., & Schroots, J.J.F. (1969). Cognitive categories and memory span. III. Effects of similarity on recall. *Quarterly Journal of Experimental Psychology*, *21*, 21-28.
- Sirevaag, E.J., Kramer, A.F., Coles, M.G.H., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica*, *70*, 77-97.
- Slak, S. (1970). Phonemic recoding of digital information. *Journal of Experimental Psychology*, *86*, 398-406.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74* (Whole No. 498.)
- Todd, J.J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751-754.

- Treisman, M., & Rostron, A.B. (1972). Brief auditory storage: A modification of Sperling's paradigm. *Acta Psychologica*, *36*, 161-170.
- Tulving, E., & Patkau, J.E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology*, *16*, 83-95.
- Turvey, M.T. (1973). On peripheral and central processes in vision: inferences from an information processing analysis of masking with patterned stimuli. *Psychological Review*, *80*, 1-52.
- Verhaeghen, P., Cerella, J., & Basak, C. (2004). A Working-memory workout: How to expand the focus of serial attention from one to four items, in ten hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1322-1337.
- Vogel, E.K., & Machizawa, M.G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 749-751.
- Vogel, E.K., Woodman, G.F., & Luck, S.J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1436-1451.
- Wheeler, M.E., & Treisman, A.M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48-64.
- Woodman, G.F., & Vogel, E.K. (2005). Fractionating working memory: Consolidation and maintenance are independent processes. *Psychological Science*, *16*, 106-113.
- Xu, Y., & Chun, M.M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91-95.

Yi, D.-J., Woodman, G.F., Widders, D., Marois, R., & Chun, M.M. (2004). Neural fate of ignored stimuli: Dissociable effects of perceptual and working memory load. *Nature Neuroscience*, 7, 992-996.

Author Notes

This research was supported by NIH grant R01 HD-21338 awarded to N.C. We thank Matt Moreno and Mike Carr for excellent assistance. Address correspondence to either author at the Department of Psychological Sciences, University of Missouri, 207 McAlester Hall, Columbia, MO 65211. E-mail: SaultsJ@missouri.edu (J.S.S.) or CowanN@missouri.edu (N.C.).

Table 1

Accuracy Means and Standard Deviations for Experiments 1-4.

Memory Load	Visual Set Size	Trial Type	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Unimodal Memory Condition						
Visual	4	Visual Change	.94 (.10)	.97 (.05)	.93 (.09)	.98 (.07)
	4	No Change	.96 (.08)	.92 (.09)	.97 (.06)	.97 (.07)
Visual	8	Visual Change	.71 (.22)	.70 (.13)	.63 (.19)	.72 (.21)
	8	No Change	.90 (.12)	.85 (.16)	.83 (.09)	.81 (.17)
Auditory	4	Auditory Change	.69 (.19)	.64 (.14)	.55 (.16)	.61 (.20)
	4	No Change	.89 (.10)	.86 (.14)	.82 (.15)	.73 (.22)
Auditory	8	Auditory Change	.70 (.16)	.70 (.13)	.54 (.14)	.61 (.21)
	8	No Change	.84 (.22)	.83 (.12)	.79 (.16)	.76 (.20)
Bimodal Memory Condition						
Both	4	Visual Change	.87 (.12)	.88 (.14)	.80 (.19)	.81 (.20)
	4	Auditory Change	.67 (.14)	.65 (.16)	.40 (.15)	.46 (.26)
Both	4	Visual No Change	.92 (.11)	.86 (.12)	.83 (.14)	.82 (.16)
	4	Auditory No Change	.87 (.12)			
Both	8	Visual Change	.62 (.22)	.65 (.21)	.57 (.17)	.58 (.23)
	8	Auditory Change	.63 (.19)	.63 (.15)	.43 (.16)	.50 (.24)
Both	8	Visual No Change	.87 (.15)	.81 (.15)	.77 (.13)	.73 (.17)
	8	Auditory No Change	.86 (.11)			

Note. N=24 for Experiments 1 and 3, N=23 for Experiment 2, and N=48 for Experiment 4.

Separate entries are shown for visual and auditory no-change trial types in the bimodal condition for Experiment 1 because probes only included stimuli in one modality. Only one combined entry is shown for no-change trial types in bimodal conditions for Experiments 2, 3, and 4 because probes always included stimuli in both modalities and thus did not differ in this respect. Separate entries for change trials are shown for all experiments because a change only ever occurred in one modality, visual or auditory. Standard deviations in parentheses.

Table 2

Estimated Capacity Means and Standard Deviations for Experiments 1-4.

Memory Load	Visual Set Size	Change Modality	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Unimodal Memory Condition						
Visual	4	Visual	3.62 (0.52)	3.57 (0.40)	3.60 (0.47)	3.74 (0.36)
	8	Visual	4.90 (2.17)	4.42 (1.75)	3.63 (1.90)	4.18 (1.79)
Auditory	4	Auditory	2.32 (0.91)	1.98 (0.84)	1.50 (0.97)	1.37 (0.91)
	8	Auditory	2.15 (0.93)	2.16 (0.72)	1.32 (0.81)	1.34 (0.78)
Bimodal Memory Condition						
Both	4	Visual	3.27 (0.68)	2.97 (0.65)	2.52 (1.01)	2.50 (0.93)
	8	Visual	3.87 (2.08)	3.67 (2.01)	2.75 (1.81)	2.49 (1.77)
	4	Auditory	2.15 (0.73)	2.03 (0.68)	0.90 (0.69)	1.00 (0.84)
	8	Auditory	1.97 (0.91)	1.73 (0.79)	0.81 (0.68)	0.88 (0.81)

Note. N=24 for Experiments 1 and 3, N=23 for Experiment 2, and N=48 for Experiment 4.

Capacity estimates were calculated for each set size and condition separately for each participant and then averaged across participants (using the formula recommended in Cowan et al., 2005). In Experiment 1, capacity estimates in bimodal blocks were calculated using separate visual and auditory hit and false alarm rates based on the modality of probe stimuli. In Experiments 2, 3, and 4, capacity estimates for each bimodal block were calculated using separate visual and auditory hit rates based on the modality of the changed stimulus and the average correct rejection rate, because no-change trials did not differ by modality. Standard deviations in parentheses.

Table 3

Estimated Capacity Means and Standard Deviations for Each Memory Delay and Mask Delay in Experiments 5.

Memory Load	Memory Delay	Change Modality	Mask Delay		
			600 ms	1000 ms	2000 ms
Unimodal Memory Condition					
Visual	3 s	Visual	3.54 (1.12)	4.12 (1.10)	4.67 (1.11)
	4 s	Visual	3.67 (0.77)	4.04 (1.24)	4.08 (1.80)
Auditory	3 s	Auditory	1.67(1.22)	2.13 (1.13)	1.04 (1.34)
	4 s	Auditory	1.67 (1.01)	1.71 (1.43)	1.50 (1.12)
Bimodal Memory Condition					
Both	3 s	Visual	2.83 (1.68)	3.13 (1.20)	2.75 (1.18)
	4 s	Visual	2.13 (1.52)	2.67 (1.57)	2.50 (1.32)
	3 s	Auditory	0.88 (1.42)	1.38 (1.32)	1.29 (1.44)
	4 s	Auditory	0.83 (1.38)	0.88 (1.31)	1.13 (1.13)

Note. N=36 total; N=18 for each memory delay. Capacity estimates were calculated for each condition and participant and then averaged across participants (using the formula recommended in Cowan et al., 2005). Capacity estimates for each bimodal block were calculated using separate visual and auditory hit rates based on the modality of the changed stimulus and the average correct rejection rate, because no-change trials did not differ by modality. Standard deviations in parentheses.

Figure Captions

Figure 1. Basic procedures for the bimodal condition in each of the five experiments. Characters outside of the rectangles represent spoken stimuli, which differed in voice according to the loudspeaker location (for Experiments 1-4) and varied in duration as described in the text. The different typefaces of the digits represent different voices. For Experiment 4, only one of two orders of the memory sets is shown; an auditory-visual order was used on other trials. In the illustration of Experiment 5, the list of three different intervals represent mask delays of 600, 1000, of 2000 ms, respectively, and, for the mask-to-probe interval, the two lists represent intervals for memory delays of 3000 or 4000 ms.

Figure 2. Capacity estimates for all five experiments using the method of Cowan et al. (2005, Appendix A). For each experiment, two white bars represent the estimated visual capacities and two gray bars represent the estimated auditory capacities, for the unimodal (left) and bimodal (right) memory conditions. The striped bar shows the sum of the visual and auditory capacities in the bimodal memory condition and represents the estimated total capacity when both modalities are to be remembered. Experiment 5 capacities were averaged across both (3- and 4-s) memory delays and across both asymptotic (1000- and 2000-ms) mask SOAs. Error bars represent standard errors of the mean.

Figure 1

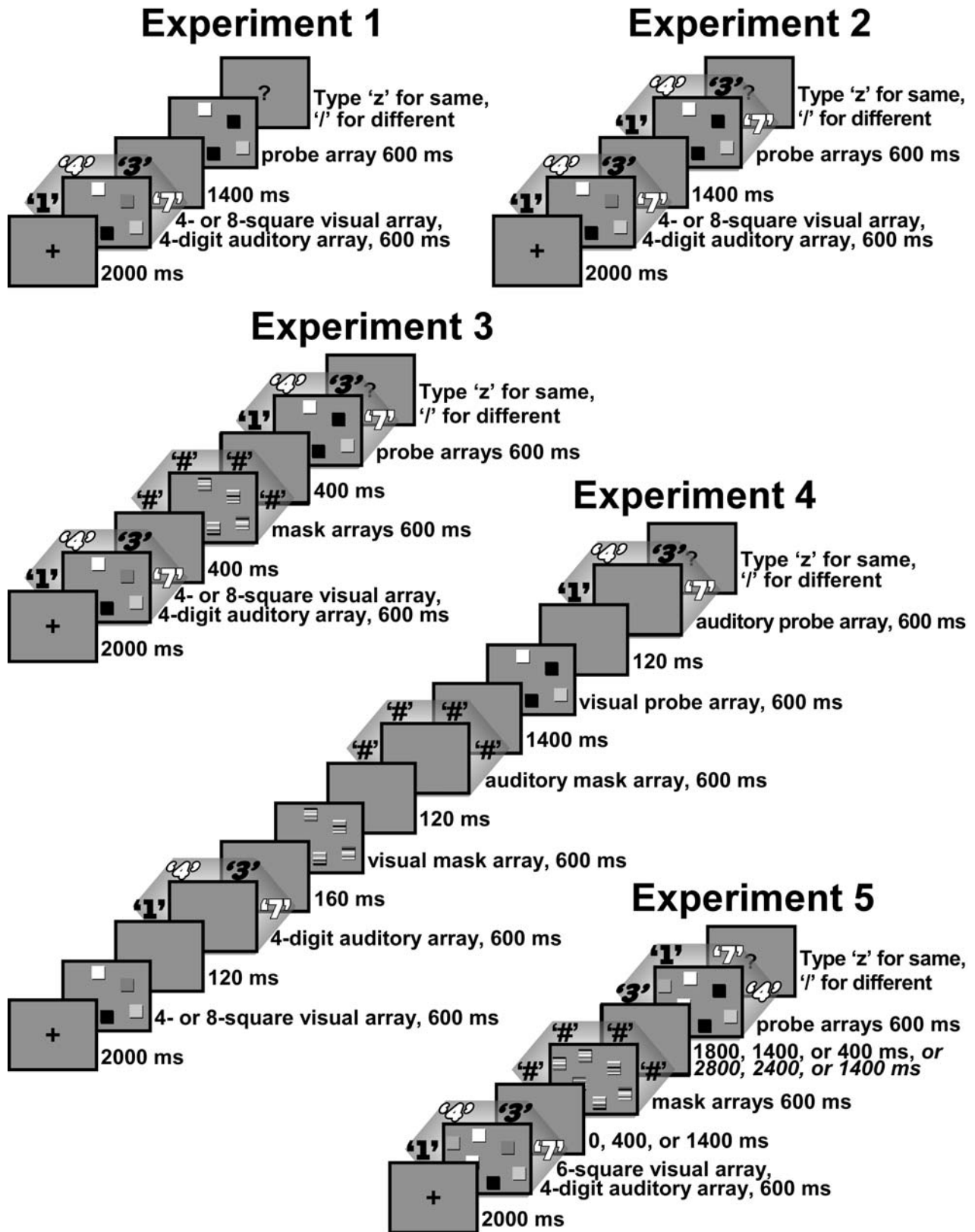


Figure 2

