Binding of verbal and spatial features in auditory working memory

Murray T. Maybery a,*, Peter J. Clissa, Fabrice B.R. Parmentier b,c,a, Doris Leunga, Grefin Harss a, Allison M. Fox a, Dylan M. Jones d,a

a School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
b School of Psychology, University of the Balearic Islands, Spain
c School of Psychology, University of Plymouth, England, United Kingdom
d School of Psychology, Cardiff University, Wales, United Kingdom

A R T I C L E   I N F O

Article history:
Received 22 July 2008
Revision received 25 February 2009
Available online 31 March 2009

Keywords:
Short-term memory
Working memory
Binding
Auditory memory
Verbal memory
Spatial memory

A B S T R A C T

The present study investigated the binding of verbal identity and spatial location in the retention of sequences of spatially distributed acoustic stimuli. Study stimuli varying in verbal content and spatial location (e.g. V1S1, V2S2, V3S3, V4S4) were followed by a recognition probe stimulus. A critical test of the binding or integration of the verbal and spatial features of the study stimuli comprised a comparison of intact probes that preserved the association of those features (e.g. V2S2 or V3S3) with recombined probes (e.g. V2S3 or V2S3) that used verbal and spatial features from study items, but in new combinations. A series of five experiments showed evidence of the binding of sound identity and location information for both verbal stimuli (spoken letters) and artificial non-verbal stimuli. While binding tended to be stronger for the more recent items of the sequence, there was consistent evidence of the retention of associations of features for the early sequence items, suggesting durability of binding of auditory features over time (at least 5.5 s) and despite the interpolated processing of other stimuli. Both spatial and verbal recognition judgments were affected by the association of verbal and spatial features when the test procedure required attention to the two classes of information. However, when participants were able to focus attention on one class of information and ignore the other, spatial recognition judgments showed an advantage for intact probes compared to recombined probes, whereas verbal recognition judgments did not. The results are discussed with reference to the primacy of identity and location in the representation of sounds in working memory.

Introduction

A substantial focus of research on working memory (WM) has been to identify different systems for the retention of different types of information. A popular embodiment of this modular view is the theoretical model proposed by Baddeley and Hitch (1974; Baddeley, 1986, 1996), which distinguishes systems for the short-term retention of verbal and visuo-spatial information. While this division between verbal and visuo-spatial WM has been a dominant theme in empirical and theoretical work (see Miyake & Shah, 1999), some researchers have argued that this differentiation is insufficient and further division of WM systems is warranted. For instance, with visual information, memory for location has been dissociated from memory for color (Vuontela, Rämä, Raninen, Aronen, & Carlson, 1999), duration (Halbig, Mecklinger, Schriefers, & Friederici, 1998), and shape or pattern (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Mecklinger, 1998). With speech in the auditory domain, memory for voice (speaker identity) has been dissociated from memory for word (Stevens, 2004) and location (Rämä et al., 2004), and memory for phonemes has been dissociated from memory for location (Ahveninen et al., 2006). With non-speech sounds, dissociations have been reported for...
Research supporting differentiated WM systems has typically used pairs of tasks where each requires retention of a single class of information (e.g., items varying in verbal content for one task and in spatial location for another). Although these tasks provide useful information on the processes responsible for encoding and retrieving isolated features, they may fail to capture additional processes critical to the coherent representation and retention of the multi-featured visual and auditory stimuli that occupy everyday experience. We typically experience these complex stimuli as unitary objects or events rather than as disaggregated sets of features, so critical additional processes may be responsible for encoding and retaining links between features. Noting where objects are located in space, keeping track of who said what, and associating voices with faces are illustrations of the central importance of conjoining elements of our experience. Binding refers to the linkage of features within or across information streams (Treisman, 1999). Binding processes are claimed to be fundamental to a broad range of psychological phenomena, including object-based perception, episodic memory, and the creation of coherent action sequences (Mesulam, 1998). The central focus of the research reported in this paper is on the binding of verbal and spatial information in auditory WM. Despite the centrality of the distinction between these two classes of information in modular models of WM, there is limited evidence on whether verbal and spatial features are represented in memory in any interdependent way. Our major aim is to develop and test several alternative accounts of how the verbal and spatial features of sounds could be represented in WM. Since there has been limited investigation of binding in research on auditory WM, we first consider research on binding in other research fields that is pertinent to developing hypotheses as to how the multiple features of acoustic stimuli might be associated in WM.

There has been an extensive investigation of interactions among multiple dimensions of information using simple perceptual or classification tasks. Some of this research has used dimensions drawn from a single perceptual modality, such as the word and color dimensions in investigating Stroop interference, whereas other research has examined dimensions drawn from different modalities, such as with the auditory and visual presentation of speech information in investigating the McGurk effect (for reviews, see Calvert, Spence, & Stein, 2004; MacLeod, 2005; Welch & Warren, 1986). A particular issue investigated in the perception literature has been how the multiple features of a stimulus are integrated (Quinlan, 2003; Treisman & Gelade, 1980). In contrast to this extensive literature on binding in perception, the investigation of whether different classes of information are linked in WM is more recent and limited in scope. In research on visuo-spatial WM, there is evidence of inter-dependence in the retention of features such as shape, color and location (Elsley & Parmentier, submitted for publication; Jiang, Olson, & Chun, 2000; Olson & Marshuetz, 2005; Wheeler & Treisman, 2002). In the auditory domain, there is evidence that the recognition of spoken words is facilitated when the words are presented at test in the same voice with which they had been presented at study (Bradlow, Nygaard, & Pisoni, 1999; Craik & Kirsner, 1974; Goh, 2005; Sheffert, 1998). Thus there is some evidence of links between visuo-spatial features in the representation of visual objects and between auditory features in the representation of sounds.

However, given the key division between verbal and spatial information in many theoretical accounts of WM, there has been surprisingly little investigation of whether there is any conjoint representation of these two classes of information when stimuli vary in both verbal content and spatial location. Three neuropsychological studies have provided some evidence consistent with the binding of these features for visually presented stimuli (Campo et al., 2005; Prabhakaran, Narayan, Zhao, & Gabrieli, 2000; Wu, Chen, Li, Han, & Zhang, 2007). In each of these studies, participants attempted to retain verbal and spatial information concurrently. Information was presented in either bound displays (with each letter or word presented in one of the to-be-remembered locations) or in unbound displays (with the letters or words presented centrally, surrounded by the marked locations). Different patterns of neural activity for bound displays compared to unbound displays were reported using fMRI (Prabhakaran et al., 2000), MEG (Campo et al., 2005), and EEG synchronization (Wu et al., 2007), consistent with unique processes involved in binding verbal and spatial representations.

Importantly, Prabhakaran et al. (2000) provided more direct evidence of the binding of verbal and spatial information by comparing performance for two types of recognition probes, which we label intact and recombined. Assume the to-be-remembered verbal (V) and spatial (S) features for a bound display are V1S1, V2S2, V3S3, and V4S4. Intact probes present verbal and spatial features which had co-occurred in one of the study stimuli, that is, these probes repeat one of the study items in intact form (e.g., V2S2 or V3S3). Recombined probes also present verbal and spatial features from the study display, but the two features have occurred in different memory stimuli (e.g., as with V2S3 or V3S2). Under Prabhakaran et al.’s procedure, a positive recognition judgment is required if both the verbal and the spatial feature were part of the study display. This condition is satisfied for both the intact and recombined probes, since both contain “old” verbal and spatial features. Critically, if the verbal and spatial features are not bound in memory, with, instead, independent representations held in separate verbal and spatial memory systems, then no advantage for intact over recombined probes should be observed. However, if the verbal and spatial features of individual memory stimuli are bound in representation, then an advantage in recognition should be observed for intact probes, for which a single bound memory representation will provide sufficient evidence for a positive recognition response, compared to recombined probes, for which two bound memory representations would need to be interrogated to enable a positive response. Thus the comparison of intact and recombined probes provides a strong test of binding.

Prabhakaran et al. (2000) reported advantages in recognition accuracy and latency for intact over recombined probes, consistent with the binding of each letter to its
location. However, it is not clear that the visual presentation used by Prabhakaran et al. obligated reliance on verbal and spatial representations. First, there was nothing to mandate that the display could not be retained in a visual form. Visual WM can hold up to four objects comprised of multiple features such as color, shape and spatial location (Luck & Vogel, 1997; Wheeler & Treisman, 2002). Thus participants may have been able to remember the bound study displays as arrays of visual objects. Prabhakaran et al. (2000) attempted to encourage phonological coding of the letters and limit any reliance on visual coding by using upper-case letters for study displays and lower-case letters for probes. However, this control does not ensure that the visual form differed for all the letters (e.g., depending on the font, letters in pairs such as Ww, Xx, and Zz may differ only in scale). Importantly, Logie, Della Sala, Wynn, and Baddeley (2000) have shown that memory of the graphical form of visually presented letters can influence their verbal recall. Further, even if the change in letter case obligated some translation of the letters in the study display to phonological codes, this translation could have been delayed until after the retention interval, and so the binding between letters and locations could have been carried primarily by visuo-spatial representations. These considerations indicate that the performance advantage for intact over recombined probes reported by Prabhakaran et al. (2000) may reflect, at least partly, visually based encoding and recognition processes rather than the binding of verbal and spatial WM representations.

Possible mechanisms of binding and symmetric or asymmetric associations between features

Three basic mechanisms have been proposed for the binding of features (for reviews, see Hommel, 2004; Robertson, 2003; Treisman, 1999). One possibility is that binding is achieved through the synchronous firing of cell assemblies that code the individual features of an object (see, e.g., Gray, 1999). Alternatively, conjunctions of features could be coded in “cardinal” cells that receive input from cells processing the individual features (see, e.g., Ghose & Maunsell, 1999). The third possibility is that the integration of multiple features is achieved using cells in prefrontal cortex that provide pointers to the features represented in specialized (more posterior) networks (Ruchkin, Grafman, Cameron, & Berndt, 2003). These general mechanisms proposed for feature integration provide a backdrop for considering more specific accounts of binding in WM.

Mindful of evidence of interactions between classes of information purportedly handled by different subsystems of the Baddeley and Hitch WM model, Baddeley (2000, 2001) revised the model to include an additional component, the episodic buffer. This limited-capacity buffer uses unitary multi-dimensional codes (Repovš & Baddeley, 2006) that provide an interface to the other WM stores, the phonological loop and visuo-spatial sketchpad. By appealing to unitary multi-dimensional representations in the buffer, some of the evidence of binding reviewed above (e.g., the results of Campo et al., 2005, and Prabhakaran et al., 2000) could be accommodated. However, outcomes from several studies of visual WM would require elaboration of Baddeley’s (2000, 2001) revised model. In these studies, an asymmetric pattern of influence of certain features on the retention of others has been observed. In particular, recognition judgments of the color or shape of visual objects were facilitated if the objects presented for recognition retained their locations from the study phase, whereas recognition judgments for location were typically not influenced by whether other features such as color or shape were retained from study to test (Elsley & Parmentier, submitted for publication; Jiang et al., 2000; Olson & Marshuetz, 2005).

In reviewing accounts of feature binding for visual stimuli, Elsley and Parmentier (submitted for publication) distinguished between models that assume full integration of features and models for which there are asymmetric links between features. For full integration models, the encoding of one feature is associated with the encoding of another, and vice versa. Instead, for asymmetric models, the encoding of one feature may implicate, and be integrated with, the encoding of a second feature, whereas it may be possible to encode the second feature without associating it with the first (Elsley & Parmentier, submitted for publication). A model of this latter type was suggested by Jiang et al. (2000). They proposed that when a display of visual objects is presented, a representation of the spatial configuration of the objects is formed immediately, and other features of the objects, such as their shape and color, are then represented with reference to this context (cf., Kahneman, Treisman, & Gibbs, 1992; Treisman & Gelade, 1980). Under Jiang et al.’s (2000) proposal it is possible to encode the spatial configuration of an array of objects without including the other features. This explains why recognition memory for spatial locations is not affected by changes in visual features from study to test while changes in spatial locations affect recognition for visual features (Elsley & Parmentier, submitted for publication; Jiang et al., 2000; Olson & Marshuetz, 2005).

In relation to auditory WM, it is feasible that, here too, the spatial configuration of a sequence of sounds provides a primary representation, with other features of auditory objects, such as verbal identity and voice, being mapped to locations in the configuration. Evidence for the primacy of spatial information (at least in the temporal sense) for the encoding of sounds was provided by Ahveninen et al. (2006). They reported that a region anterior to Heschl’s gyrus responded selectively to the identity of phoneme sounds, whereas a region posterior to the gyrus responded selectively to the sounds’ locations. Critically, there was evidence that activity in the posterior (“where”) region occurred much earlier than activity in the anterior (“what”) region, consistent with the proposition that location may be primary in the representation of auditory objects, just as it is in the representation of visual objects.

However, spatial location may not be as central to audition as it is to vision. In developing the concept of an auditory object with reference to that of a visual object, Kubovy and Van Valkenburg (2001), Van Valkenburg and Kubovy (2003) argued that whereas space and time are indispensable attributes for a visual object, pitch and time are the indispensable attributes for an auditory object. They con-
tended that a difference in pitch rather than a difference in location is critical to distinguishing two sounds. Thus aspects of the identity of sounds such as pitch and duration may be more fundamental dimensions than spatial location in the representation of auditory objects. Although verbal sounds such as spoken letters are complex in structure with respect to pitch and duration, it is conceivable, given Kubovy and Van Valkenburg’s theoretical analysis, that verbal identity is primary in the representation of these sounds, with their locations coded with reference to their identities.

In summary, several distinct theoretical positions are pertinent to the present investigation of memory for the verbal and spatial information conveyed in a sequence of sounds. First, it is conceivable that each class of information is represented independently of the other, in subsystems like the phonological loop and visuo-spatial sketchpad, and so no evidence of binding of verbal identity and spatial location will be observed. Alternatively, the verbal and spatial features of each sound might be bound together in a unitary multi-dimensional code such as is assumed for the episodic buffer. If this form of coding were to provide full integration of features, then recognition of each feature should be influenced by whether the other feature is retained from study to test. (Note that this position based on fully integrated multi-dimensional codes need not be the only possibility for the episodic buffer or Baddeley’s revised WM model more generally – see the General Discussion). Another theoretical position, consistent with research on visual WM, is that location is primary in the representation of sounds, with this feature influencing recognition of verbal identity, but without verbal identity influencing the recognition of location. Finally, based loosely on the theory of Kubovy and Van Valkenburg, verbal identity could be primary in representing sounds, with identity influencing recognition of location, but without location influencing the recognition of verbal identity.

The present study

One complication noted for the interpretation of the Prabhakaran et al. (2000) study was that the visual presentation of letter stimuli need not have obligated the generation of verbal (phonological) codes for those items. According to the Baddeley and Hitch model, whereas visually presented verbal information must be recoded from a visual to a phonological code before it can enter the phonological loop, acoustic verbal information enters the loop directly (see, e.g., Repovš & Baddeley, 2006). Thus, one advantage of the use of speech stimuli in the present study is that the generation of phonological codes for these items can be assumed under the Baddeley and Hitch model. In order to convey spatial as well as verbal information, our speech stimuli were presented from different loudspeaker locations. There is good evidence that the location of a sound is coded automatically in non-primary auditory cortex (Ahveninen et al., 2006). Cortical projections from primary auditory cortex appear to divide into separate acoustic identity (“what”) and spatial location (“where”) pathways (Ahveninen et al., 2006; Rauschecker & Tian, 2000; Romanski et al., 1999; Zatorre, Bouffard, Ahad, & Belin, 2002). Scott (2005) has argued that there is increasingly sophisticated processing of speech along the “what” pathway, and cortical regions implicated in verbal WM lie adjacent to this pathway. Thus using speech sounds distributed in space has the advantage that the automatic coding of both verbal identity and spatial location is expected.

The first of the experiments reported below used the sequential delivery of spoken letters from different locations to ensure the encoding of phonological and spatial representations, and intact and recombined probes were employed in an adaptation of the Prabhakaran et al. (2000) recognition procedure. Interest centered on whether an advantage in recognition would be observed for the intact probes, consistent with some integration or association in WM of the verbal and spatial features of the sounds. In Experiment 2, non-verbal sounds were used to examine the generality of binding between identity and location features for acoustic stimuli.

The response contingencies for the recognition task used in Experiments 1 and 2 are such that an affirmative response is made to a probe if both of its features (the sound’s identity and location) had been present in the memory set. Although an advantage under this method for the recognition of intact probes compared to recombined probes indicates some association in the retention of the identity and location of the sounds, the method does not reveal the nature of that association. More particularly, it does not reveal whether the recognition of identity is affected by an association with spatial location, whether the recognition of spatial location is affected by an association with identity, or whether there is symmetry in the influence of each feature on recognition of the other. Differentiating between these patterns of effects is critical to addressing whether there is full integration of features or whether either spatial location or identity is primary in the representation of sounds.

To address this limitation of the methodology of the first two experiments, Experiments 3–5 required recognition judgments that were focused on either verbal identity or spatial location. Use of intact and recombined probes under these conditions permitted evaluation of whether the influences of the verbal and spatial features on each other are symmetric or asymmetric. In Experiments 3 and 4, each participant focused on either verbal information (while ignoring spatial information) or on spatial information (while ignoring verbal information). Thus any observed advantage for intact over recombined probes in these circumstances would indicate obligatory coding of the non-attended feature and its association with the attended feature. In the final experiment, trials assessing verbal recognition and spatial recognition were intermixed, with the type of recognition judgment required indicated by a visual cue presented between the study stimuli and the probe stimulus. This intermixing forced encoding of both features of the study items. Of central interest was whether a symmetric pattern of influences of the two features on each other would be observed when attention was directed to the two, in contrast to a particular asymmetric pattern of effects that was observed in Experiments 3 and 4.
**Experiment 1**

The major aim of Experiment 1 was to establish whether there is integration of verbal and spatial information in WM with the acoustic delivery of stimuli. This was achieved using an array of eight loudspeakers aligned in azimuth around the participant (Fig. 1). On each trial, a study sequence consisting of four letters delivered from four loudspeakers was followed by a silent retention interval, then the presentation of a recognition probe, comprising a single letter from a single loudspeaker. The participant made a yes/no judgment using decision rules analogous to those used in Prabhakaran et al. (2000). A yes response was required if the probe comprised a letter from the study sequence presented from a location used in the sequence, irrespective of whether they were coincident in the sequence. Two probes types required a yes response (see Fig. 2). Intact probes comprised a letter from the study sequence delivered from its original location. Recombined probes comprised a letter from the sequence delivered from a location used to deliver a different letter in the study sequence.

If the only representation of the letters and locations is in independent verbal and spatial systems like the phonological loop and visuo-spatial sketchpad, then participants should show equivalent levels of proficiency in responding to the intact and recombined probes. However, if integrated multi-featured representations of auditory objects are maintained in some form in WM, then intact probes should be recognized with greater ease than recombined probes. This is because an intact probe should match precisely the multi-featured representation of one of the study stimuli whereas a recombined probe should provide only a partial match to the representations of two study stimuli.

To ensure that participants based recognition judgments on both letters and locations, three types of probe requiring no responses were also used (see Fig. 2): old identity-new location probes (a letter from the study sequence presented from a location that was not part of the sequence); new identity-old location probes (a letter not from the study sequence presented from a location that was from the sequence); and new identity-new location probes (neither the letter or location were from the sequence).

Under the method of visual presentation used by Prabhakaran et al. (2000), study stimuli were presented simultaneously. Replication of this aspect of their method would be problematic in the auditory domain, hence our use of sequential presentation of the study stimuli. Allen, Baddeley, and Hitch (2006) investigated the binding of shape and color features for four visual stimuli presented sequentially, and reported that binding was much more pronounced for the last stimulus in the series. They argued that the encoding and retention of further visual stimuli disrupted the fragile binding of features of earlier stimuli. With this in mind, a subsidiary aim of the first experiment was to investigate whether any association between the verbal and spatial features of our auditory stimuli would show a similar fragility under sequential presentation. This was done by comparing levels of performance for the intact and recombined probes as a function of serial position. Since the recombined probe draws features from two study stimuli, it was not possible to compare the two probe types at individual serial positions. However, we were able to compare intact and recombined probes for when the probe features were drawn either from the first two items or from the last two items of the sequence.

**Method**

**Participants**

The 24 participants (mean age = 24.5 years, SD = 7.8 years) were psychology students at the University of Western Australia who reported normal hearing and normal or corrected-to-normal vision. Comparable samples from the same population were used in subsequent experiments.

**Apparatus**

Stimuli were presented and responses recorded using MetaCard software running on a Pentium-II PC fitted with an eight-channel Events Electronic Darla sound card. The sound card was connected to eight Yamaha speakers (YST M20DSP) located in azimuth around the participant with 36° spacing (at angles of −126°, −90°, −54°, −18°, +18°, +54°, +90° and +126°, where 0° represents the direction the participant faced). Speakers were 1.3 m from the participant, and at a height of 1.1 m (approximately head height for the seated participant). To maximize auditory clarity, testing was conducted in a semi-anechoic room with a carpeted floor and sound absorbing curtains (Shinn-Cunningham, 2001).

**Auditory stimuli**

Chosen to be acoustically distinct and minimally confusable in WM (Conrad, 1964; Hull, 1973), the stimuli com-

---

Fig. 1. The arrangement of the eight loudspeakers around the participant, and an example study sequence.
prised the consonants d, h, j, l, n, q, s, and y spoken in a female voice. Each letter was recorded at a sampling rate of 44.1 KHz and edited to a duration of 450 ms. Stimuli were presented at 64 dB SPL (A-weighted) at the participant’s head location, being 29 dB above broadband background noise.

Design and procedure

Experiment 1 was a within-subjects design with participants completing 144 trials. Two sets of stimuli were constructed. Each contained 36 intact and 36 recombined probes, and 24 old identity-new location, 24 new identity-old location and 24 new identity-new location negative probes. The probes within each category were constructed to balance the use of features from the study stimuli. For instance, for the recombined probes the two features were drawn equally often from the six pairs of serial positions for the study stimuli. Stimuli for each trial were selected without repetition from the pool of eight letters and eight loudspeaker locations. Selection was random except for the following restrictions. First, the eight locations were used with equal frequency in each probe class, as were the eight letters. Also, with regard to the selection of speaker locations, we took into account evidence that variation in path complexity affects performance on visual spatial tasks (Parmentier, Elford, & Maybery, 2005). Accordingly, mean path length (the sequential distance between loudspeakers used in the sequence), and the number of path crossings was controlled across probe types. Finally, letter sequences were chosen to avoid common acronyms and words (e.g. sly). The 144 trials were divided into blocks of 36 trials, with the overall proportions of the five probe types (3:3:2:2:2) preserved within each block, but otherwise probe type was distributed randomly. Participants were assigned randomly to one of the two stimulus sets.

Each trial began with a silent period of 2000 ms, followed by the sequential delivery of the four letters, each from a different loudspeaker location, with an SOA of 1500 ms, and then by a 2500 ms silent retention interval before the recognition probe was presented. The participant had 3000 ms to respond via a handheld box. The next trial started 2000 ms after a response or the end of the response window. Allocation of the yes and no responses to the left and right response keys was counterbalanced across participants. Participants were tested in the dark to limit visual coding of locations and were asked to remain still to minimize interference with spatial memory (Lawrence, Myerson, Oonk, & Abrams, 2001; Smyth, 1996). Detailed verbal instructions and a training task comprising 10 trials placed equal emphasis on retaining the locations and letter identities of the sounds. Participants were asked to respond as quickly as possible while remaining accurate.

Results

Two measures of performance were analyzed: percentage of responses correct (accuracy) and median response time (RT) for correct responses. Comparisons of performance for the intact and recombined probes were of critical interest. For completeness, additional analyses compared performance for the three types of negative probe. In this and subsequent experiments, the Green-
house-Geisser correction was used whenever the assumption of sphericity was not met.

Comparisons of intact and recombined probes
As shown in Fig. 3 (Panels A and B), participants were both more accurate, \(F(1,23) = 25.23, \text{MSE} = 32.51, p < .001, \eta^2_p = .513\), and faster, \(F(1,23) = 25.83, \text{MSE} = 15720.33, p < .001, \eta^2_p = .523\), at detecting intact probes relative to recombined probes, demonstrating binding of the verbal-identity and spatial-location features.

Comparisons among negative probe types
Substantial variation in error rates, \(F(1.13,26.04) = 114.11, \text{MSE} = 66.22, p < .001, \eta^2_p = .832\), and RTs, \(F(1.18,27.20) = 50.79, \text{MSE} = 37320.67, p < .001, \eta^2_p = .688\), were also observed across the three types of negative probes. Follow-up pair-wise comparisons were conducted using a Bonferroni correction. Old identity-new location probes were the most difficult type, showing lower rates of accuracy and longer RTs than both new identity-old location probes – \(F(1,23) = 103.81, \text{MSE} = 57.98, p < .001, \eta^2_p = .819\), for accuracy, and \(F(1,23) = 40.80, \text{MSE} = 37105.62, p < .001, \eta^2_p = .640\), for RT – and new identity-new location probes – \(F(1,23) = 137.35, \text{MSE} = 49.43, p < .001, \eta^2_p = .857\), for accuracy, and \(F(1,23) = 77.50, \text{MSE} = 23654.20, p < .001, \eta^2_p = .771\), for RT (see Table 1 for means). New identity-old location and new identity-new location probes did not differ significantly in levels of accuracy or RT (Table 1). Thus a mismatch in identity between the probe and study stimuli permitted an accurate and rapid negative recognition response, but this was not the case for a mismatch in location.

Binding as a function of serial position
Given that Allen et al. (2006) reported reduced binding for items early in a sequence of visual items compared to later items, subsidiary analyses investigated whether the strength of association between the verbal and spatial features of our auditory stimuli changed as a function of serial position. To do this, we split the intact and recombined probes into those where the probe features were drawn from the first two sequence items, and those where the features were drawn from the last two sequence items (consequently, only a subset of the data for recombined probes was included). As can be seen in Table 2, accuracy was higher for the intact probes relative to the recombined probes irrespective of the serial positions of the study items from which the probes’ features were drawn. This was confirmed in a 2 (probe type: intact, recombined) × 2 (serial position: 1 and 2, 3 and 4) ANOVA. The main effect of probe type was significant, \(F(1,23) = 7.97, \text{MSE} = 104.88, p = .01, \eta^2_p = .257\), as was the main effect of serial position, \(F(1,23) = 5.27, \text{MSE} = 112.71, p = .031, \eta^2_p = .187\), indicating that performance improved when the probe involved re-

---

**Fig. 3.** Mean accuracy (percent correct) and mean of median RTs for correct responses as a function of probe type for Experiment 1 (Panels A and B) and Experiment 2 (Panels C and D). Error bars are 95% within-subjects confidence intervals (Loftus & Masson, 1994).
The integration of verbal identity and spatial location in rapidly than recombined probes. This is consistent with probes were recognized both more accurately and more across the sequence, and a larger RT advantage for the in-

Discussion

The critical finding from Experiment 1 is that intact probes were recognized both more accurately and more rapidly than recombined probes. This is consistent with the integration of verbal identity and spatial location in the representation of spatially distributed verbal sounds. The advantage in accuracy for intact probes relative to recombined probes was observed for information from the beginning as well as from the end of the sequence. This indicates that some degree of association of the verbal and spatial features can be maintained for at least the 5.5 s between the second study stimulus and the probe, and that the association can survive ‘over-writing’ by subsequent sequence stimuli. However, the serial-position analysis of RTs suggested that the strength of the association between the features diminishes to some extent as a function of either time or the encoding of other sounds.

Regarding the negative probes, the difficulty of the old identity-new location probes relative to both the new-identity-old location and new identity-new location probes indicates that the verbal identity of the sounds carried more weight in recognition judgments than did their locations. Use of a letter from the study stimuli in the old identity-new location probes may have impeded the appropriate negative response in this case, or alternatively, use of a novel letter in the new identity-old location and new identity-new location probes may have facilitated a negative response in these cases. Participants typically reported that the verbal information was easier to encode and rehearse than the spatial information, so this may be why the former class of information carried more weight in influencing recognition judgments.

Experiment 2

Experiment 1 provided evidence that an association is formed between the identity and location of a sound in a task requiring the encoding of both features, and that this association persists for at least 5.5 s. The use in the first experiment of highly familiar verbal features (letters) may explain both (1) the greater influence of verbal identity relative to spatial location on judgments of negative probes, and (2) the durability of the binding of the two features. The aim of Experiment 2 was to investigate whether the binding effect observed in Experiment 1 could be generalized to non-verbal sounds, that is, unfamiliar artificial sounds designed to not directly engender familiar verbal codes.

There is evidence that non-verbal forms of acoustic stimuli can be retained in WM, as with the retention of music (Pechmann & Mohr, 1992; Salamé & Baddeley, 1989) and more elementary stimuli varying in either pitch or loudness (Clément et al., 1999). Cowan (1984, 1995, 1999) proposed that auditory sensory memory comprises short- and long-duration stores. The long-duration store, with a temporal persistence of up to 30 s, is said to retain a broad range of auditory stimuli in a sensory-specific form. Accordingly, it should retain more auditory feature

Table 1
Mean accuracy (percent correct) and mean of median RTs (in ms) for correct responses for the three types of negative probe for Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Negative probe type</th>
<th>Accuray</th>
<th>RT</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old identity, new location</td>
<td>75.87</td>
<td>1181.29</td>
<td>790.44 70.26</td>
</tr>
<tr>
<td>New identity, old location</td>
<td>98.26</td>
<td>826.08</td>
<td></td>
</tr>
<tr>
<td>New identity, new location</td>
<td>99.65</td>
<td>790.44</td>
<td></td>
</tr>
</tbody>
</table>

Experiment 1

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>75.87</td>
<td>1181.29</td>
</tr>
<tr>
<td>98.26</td>
<td>826.08</td>
</tr>
<tr>
<td>99.65</td>
<td>790.44</td>
</tr>
</tbody>
</table>

Experiment 2

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.19</td>
<td>1354.81</td>
</tr>
<tr>
<td>92.19</td>
<td>1209.79</td>
</tr>
<tr>
<td>93.40</td>
<td>1135.10</td>
</tr>
</tbody>
</table>

Table 2
Mean accuracy (percent correct) and mean of median RTs (in ms) for correct responses from Experiments 1, 2, 3 and 5 for intact and recombined probes as a function of the serial positions tested.

<table>
<thead>
<tr>
<th>Serial positions 1 and 2</th>
<th>Intact</th>
<th>Recombined</th>
<th>Serial positions 3 and 4</th>
<th>Intact</th>
<th>Recombined</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>90.51</td>
<td>.0527</td>
<td>96.30</td>
<td>89.58</td>
<td>.527</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>1078.17</td>
<td>1144.06</td>
<td>922.02</td>
<td>1157.58</td>
<td>790.44</td>
<td></td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>72.69</td>
<td>.031</td>
<td>88.89</td>
<td>78.47</td>
<td>.187</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>1479.11</td>
<td>1477.48</td>
<td>1096.46</td>
<td>1490.37</td>
<td>96.81</td>
<td></td>
</tr>
<tr>
<td>Experiment 3 – spatial recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>88.54</td>
<td>.031</td>
<td>93.23</td>
<td>91.15</td>
<td>.127</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>1054.31</td>
<td>1188.27</td>
<td>1036.61</td>
<td>1157.56</td>
<td>48.16</td>
<td></td>
</tr>
<tr>
<td>Experiment 5 – verbal and spatial recognition combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>93.98</td>
<td>.010</td>
<td>97.92</td>
<td>93.06</td>
<td>.256</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>1095.00</td>
<td>1473.52</td>
<td>979.35</td>
<td>1241.14</td>
<td>4.65</td>
<td></td>
</tr>
</tbody>
</table>
information (pitch, loudness, temporal structure) than is suggested for the phonological coding utilized in the phonological loop (Baddeley, 2000). Importantly, Cowan (1995) does not specifically exclude the possibility that this long-duration store can maintain multi-featured (acoustic identity-spatial location) representations. Cowan’s (1999) model also highlights the possibility that the binding evident in Experiment 1 reflects processes in the long-duration auditory store more closely tied to the auditory percept than to more abstract memory tokens such as are implied for the phonological loop. Furthermore, one of the neurological mechanisms proposed to underpin binding, namely the synchronicity of neural activity across distinct assemblies of neurons (e.g., Gray, 1999) also allows that binding need not be restricted by the type of features being bound.

To explore the possibility that the binding observed in Experiment 1 reflects an auditory store amenable to a range of acoustic information, the letter stimuli of Experiment 1 were replaced with non-verbal synthetic sounds in Experiment 2. Aside from the change of stimuli, Experiment 2 matched Experiment 1 in structure. If the binding of identity and location for acoustic stimuli found in Experiment 1 is restricted to familiar verbal material, an advantage in performance for intact compared to recombined probes should not be evident for the novel non-verbal sounds used in Experiment 2. Alternately if a recognition advantage for intact probes is still found in Experiment 2, this would suggest that the results for both experiments may reflect the influence of a memory system open to the association of a range of acoustic identity features with the location feature.

Method

The changes to the method of Experiment 1 were as follows. First, a new sample of 24 participants (mean age = 20.9 years, SD = 4.6 years) was recruited. Second, the stimuli comprised eight 450 ms duration sound bursts produced using pure-tone and frequency-modulation synthesizes options in Sound Forge software (Version 4.5). These sounds were either two-tone alternations (using either 300 and 600 Hz or 1200 and 2400 Hz) with either 4, 8 or 16 transitions during the 450 ms interval (see Zatorre & Belin, 2001) or composites of frequency- and amplitude-modulated sine and saw-tooth functions. Stimuli were delivered at 71 dB SPL (A-weighted). Finally, three rather than two sets of stimuli were constructed, and each set was assigned randomly to eight participants.

Results

Comparisons of intact and recombined probes

Means for accuracy and RT for the intact and recombined probes are shown in Fig. 3 (Panels C and D). As for Experiment 1, participants were both more accurate, \( F(1,23) = 30.27, MSE = 40.20, p < .001, \eta_p^2 = .568 \), and faster, \( F(1,23) = 47.35, MSE = 12480.14, p < .001, \eta_p^2 = .673 \), at detecting intact probes compared to recombined probes. These differences represent evidence that binding of a sound’s identity and location is present in the retention of artificial sounds as well as speech.

Comparisons among negative probe types

The artificial sounds of Experiment 2 yielded differences in both error rates, \( F(2,46) = 118.02, MSE = 44.53, p < .001, \eta_p^2 = .837 \), and RTs, \( F(2,46) = 13.67, MSE = 21906.03, p < .001, \eta_p^2 = .373 \), for the three types of negative probes (see Table 1) that replicated the differences reported for the letter sounds of Experiment 1. Once again, old identity-location probes were the most difficult type, showing lower accuracy and longer RTs than both new identity-location probes – \( F(1,23) = 141.94, MSE = 52.84, p < .001, \eta_p^2 = .641 \), for accuracy, and \( F(1,23) = 14.40, MSE = 17521.30, p = .001, \eta_p^2 = .385 \), for RT – and new identity-location probes – \( F(1,23) = 142.86, MSE = 57.73, p < .001, \eta_p^2 = .861 \), for accuracy, and \( F(1,23) = 20.11, MSE = 28807.15, p < .001, \eta_p^2 = .466 \), for RT. Further, no significant differences were found between new identity-old location and new identity-new location probes. Thus for non-verbal sounds (Experiment 2) as well as spoken letters (Experiment 1), probes with a new identity rather than a new location fostered accurate and rapid negative recognition responses.

Binding as a function of serial position

As for Experiment 1, performance on the intact and recombined probes was analyzed as a function of the serial positions of the study items from which the probe features were drawn. The \( 2 \) (probe type: intact, recombined) \( \times 2 \) (serial position: 1 and 2, 3 and 4) ANOVA confirmed that accuracy was higher for intact probes than for recombined probes, \( F(1,23) = 11.18, MSE = 212.64, p = .003, \eta_p^2 = .327 \), and for sequence positions 3 and 4 than for sequence positions 1 and 2, \( F(1,23) = 26.09, MSE = 227.90, p < .001, \eta_p^2 = .531 \) (see Table 2 for means). The absence of a significant interaction, \( F(1,23) < 1 \), indicates an intact probe advantage in accuracy at both the beginning and end of the sequence. However, the same analysis conducted for RTs revealed a main effect of probe type, \( F(1,22) = 19.78, MSE = 44744.90, p < .001, \eta_p^2 = .473 \), a main effect of sequence position, \( F(1,22) = 12.51, MSE = 62841.47, p = .002, \eta_p^2 = .363 \), as well as a probe type by serial-position interaction, \( F(1,22) = 17.88, MSE = 50303.57, p < .001, \eta_p^2 = .448 \). RTs were significantly shorter for intact probes relative to recombined probes at positions 3 and 4, \( F(1,23) = 41.13, MSE = 44092.54, p < .001, \eta_p^2 = .641 \), but not at serial positions 1 and 2, \( F(1,22) < 1 \) (see Table 2). Thus, in concert with the results of Experiment 1, these subsidiary analyses showed an advantage in accuracy for intact over recombined probes across the entire sequence, but an RT advantage only for stimuli from the end of the sequence.

It is noteworthy that only a minority of participants reported attempting to verbally label or classify the artificial sounds, and the patterns of performance for these partici-

---

4 One participant was excluded from this analysis because no correct responses were made for the six trials for repaired probes with features drawn from the first two serial positions of the sequence.

5 Three participants reported establishing some classification of sounds into broad categories (e.g. ‘high’ vs. ‘low’ sounds, ‘brass’ sounds versus tones), while two reported some attempt to individually name the sounds. However, the patterns of performance for these individuals did not vary from the patterns observed for the rest of the sample. Consistent with Experiment 1, the majority of participants (21 of 24) reported using some form of spatial trajectory visualization for the to-be-remembered loudspeaker locations.
Participants were consistent with the patterns observed for the rest of the sample.

Discussion

Consistent with the results of Experiment 1, intact probes were recognized more accurately and with greater rapidity than recombined probes. Further, there did not appear to be any diminution in binding with the shift from verbal to non-verbal stimuli: For both accuracy and RT, the effect size for the comparison of intact and recombined probes was at least as large in Experiment 2 as it was in Experiment 1. Thus binding of the identity and location features of sounds in WM is not restricted to verbal stimuli. Also in line with the results of the first experiment, the advantage in accuracy for intact over recombined probes was present irrespective of whether probe features were drawn from serial positions 1 and 2 or 3 and 4, whereas the RT advantage was restricted to probes that tested sequence positions 3 and 4. Hence the durability of associations between the identity and location features of sounds does not seem to depend on the use of familiar sounds carrying representations in long-term memory. Outcomes of comparing performance across the three types of negative probe were also consistent across Experiments 1 and 2. Once again the new identity-old location probe was easier than the old identity-new location probe, consistent with a sound's identity carrying greater weight than its location in influencing responses. Overall, the parallel outcomes for Experiments 1 and 2 support the view that WM provides for the encoding, retention, and integration of multiple features of sounds, with the presence of familiar verbal content not essential for binding to take place.

Experiment 3

The recognition probe methodology adapted from Prabhakaran et al. (2000) and used in Experiments 1 and 2 requires attention to both the identity and spatial location of each item in the study sequence, and responding to the probe stimulus requires a decision that incorporates information from the two dimensions. Arguably this methodology provides conditions that encourage the co-registration of the identity and location features of the memory stimuli. More significantly, since the recognition judgment is made with reference to both identity and location, it is possible that the advantage observed for intact relative to recombined probes reflects the coding of one feature with reference to the other, but not the reverse. For instance, the location of a sound might be coded with reference to its identity, but not vice versa. This would mean that location recognition, but not identity recognition, would be facilitated when identity-location pairings were retained, rather than perturbed, from study to test. Facilitation of recognition of either feature would be sufficient to yield an advantage in responding to intact probes relative to recombined probes. Thus it is not possible to differentiate symmetric association of the two features (each coded with reference to the other) from asymmetric association (one encoded with reference to the other, but not vice versa) using this methodology. Accordingly, a more direct investigation of the pattern of influence of each feature on recognition of the other is warranted, and was undertaken in Experiments 3–5.

In Experiment 3 we investigated whether any evidence of binding of verbal identity and location can be found when the participant focuses on one feature only. Participants were instructed either (1) to focus on the letters, ignore the spatial locations, and make recognition judgments with respect to whether the probed letter had been part of the study sequence, or (2) to focus on the spatial locations, ignore the letters, and then decide whether the location of the probe was among those studied. Using this methodology it will be possible to establish whether there is any obligatory association of the two features, and if so, whether the association is symmetric or asymmetric. The critical indicator of whether one feature is encoded with reference to the other is whether an advantage is found for intact probes compared to recombined probes when recognition is tested for the first of these features.

Two contrasting proposals, each assuming asymmetric associations between features, were advanced in the introduction. First, by extension from research on visual WM, it was argued that spatial location may be a primary feature in the representation of auditory objects, with a sound's identity encoded with reference to its location. This should then mean that location could be encoded without reference to verbal identity, but encoding verbal identity should obligate encoding location. Consequently, verbal but not spatial recognition should show an advantage in performance for intact probes relative to recombined probes. Alternatively, motivated by Kubovy and Van Valkenburg's (2001) analysis of the indispensable attributes for visual and auditory objects, it was argued that, instead, a sound's identity may be a primary feature that anchors the encoding of its location. As a consequence, spatial but not verbal recognition should exhibit facilitated performance for intact probes relative to recombined probes. It is also conceivable for the conditions of Experiment 3, in which participants focus on one feature only (letters or locations), that dedicated verbal and spatial WM systems are utilized, and no advantage for intact probes relative to recombined probes is observed for either form of recognition. Logically, the final possibility is that there is obligatory integration of verbal and spatial information in the representation of sounds in a multi-modal system, and these integrated representations lead to a symmetric pattern of influence in that an intact-recombined difference is observed for both recognition tasks.

While the comparison of intact and recombined probes remains the critical litmus test of whether symmetric or asymmetric associations are formed between the two features, two other comparisons of probe types can provide evidence on potential interactions in processing the verbal and spatial information. (See Fig. 2 for the responses required for the five probe types under either verbal or spatial recognition. Note that the old identity-new location and new identity-old location probes require different responses for the two recognition tasks.) In particular, the design of Experiment 3 allows us to examine how recognition performance for one feature is influenced by the famil-
arity of the irrelevant feature. For example, in verbal recognition, both recombined probes (old identity-old location) and old identity-new location probes require a positive response. However, these two probe types differ insofar as the first involves an old location whereas the second does not. Thus, to the extent that the familiarity of the irrelevant location feature influences verbal recognition, recombined probes should be easier than old identity-new location probes. Similarly, for spatial recognition, familiarity of the irrelevant identity feature should favor recombined probes over new identity-old location probes.

On the other hand, the familiarity of an (old) irrelevant feature could impede responses to negative probes. For verbal recognition, new letter-old location probes might be more difficult than new letter-new location probes, and for spatial recognition, old letter-new location probes might be more difficult than new letter-new location probes. In summary, the comparison of intact and recombined probes provides a critical test of binding and of its nature (symmetric or asymmetric), whereas comparisons of other probe types allow an assessment of other interactions between verbal identity and spatial location in maintenance or retrieval.

Method

Participants
Half of the 64 students (mean age = 19.87 years; SD = 5.90 years) were randomly assigned to the letter recognition task, with the remainder completing the location recognition task.

Apparatus, stimulus materials, design and procedure

These aspects of the method were as for Experiment 1, with the following exceptions. First, a unique stimulus set was created for each participant. Second, for half the participants the instructions emphasized focusing on the letters, ignoring the spatial locations, and basing recognition judgments exclusively on whether the probed letter had been part of the memory sequence, whereas for the remaining participants the instructions emphasized focusing on the locations, ignoring the letters, and basing recognition exclusively on whether the probed location had been part of the memory sequence.

Results

Comparisons of intact and recombined probes

ANOVAs with the independent-groups factor recognition task (verbal, spatial) and the repeated-measures factor probe type (intact, recombined) were conducted. The analysis of accuracy yielded main effects of recognition task, $F(1,62) = 22.72$, $MSE = 82.20$, $p < .001$, $\eta^2_p = .268$, and probe type, $F(1,62) = 16.39$, $MSE = 13.24$, $p < .001$, $\eta^2_p = .209$. There was also a significant interaction of the two factors, $F(1,62) = 7.29$, $MSE = 13.24$, $p = .009$, $\eta^2_p = .105$. As can be seen in Fig. 4, accuracy was higher for verbal than for spatial recognition. More significantly, whereas accuracy was substantially higher for intact probes compared to recombined probes for spatial recognition, $F(1,31) = 14.43$, $MSE = 20.89$, $p = .001$, $\eta^2_p = .318$, no significant difference in accuracy for the two probe types was observed for verbal recognition, $F(1,31) = 2.16$, $MSE = 5.59$, $p = .152$, $\eta^2_p = .065$.

Fig. 4. Mean accuracy (percent correct) and mean of median RTs for correct responses as a function of probe type for verbal recognition (Panels A and B) and spatial recognition (Panels C and D) in Experiment 3. Error bars are 95% within-subjects confidence intervals.
Table 3
Mean accuracy (percent correct) and mean of median RTs (in ms) for correct responses for two types of positive probe and two types of negative probe for spatial and verbal recognition in Experiment 3.

<table>
<thead>
<tr>
<th>Type of recognition</th>
<th>Type of positive probe</th>
<th>Type of negative probe</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial recognition</td>
<td>Recombined (old identity, old location)</td>
<td>Old identity, new location</td>
<td>86.55 ± 2.82</td>
</tr>
<tr>
<td>RT</td>
<td>1119.27 ± 1176.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td>82.68 ± 4.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1176.56 ± 46.57</td>
</tr>
<tr>
<td>Verbal recognition</td>
<td>Recombined (old identity, old location)</td>
<td>Old identity, new location</td>
<td>95.92 ± 2.82</td>
</tr>
<tr>
<td>RT</td>
<td>963.55 ± 966.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td>97.66 ± 1.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26.38 ± 37.27</td>
</tr>
</tbody>
</table>

Similar outcomes were observed in the analysis of RTs, with significant main effects of recognition task, \(F(1,62) = 4.05, \text{MSE} = 114893.48, p = .045, \eta^2_p = .061\), and probe type, \(F(1,62) = 9.98, \text{MSE} = 5675.14, p = .002, \eta^2_p = .139\), and a significant interaction, \(F(1,62) = 6.98, \text{MSE} = 5675.14, p = .010, \eta^2_p = .101\). As shown in Fig. 4, RTs were shorter for verbal compared to spatial recognition judgments. Importantly, spatial recognition judgments showed a substantial advantage in RTs for intact probes over recombined probes, \(F(1,31) = 10.88, \text{MSE} = 8783.13, p = .002, \eta^2_p = .260\), however, no such advantage was observed for verbal recognition judgments, \(F(1,31) < 1\). Thus when recognition could be focused on a single dimension, there is consistent evidence that spatial location was encoded with reference to verbal identity, whereas the encoding of verbal identity was independent of location.

Other comparisons among probe types

Further evidence that verbal information affected spatial recognition judgments is that recombined probes (old identity-old location) were responded to more accurately than new identity-old location probes, \(F(1,31) = 7.92, \text{MSE} = 30.16, p = .008, \eta^2_p = .203\) (see Table 3). The same comparison on RT yielded \(F(1,31) = 3.15, \text{MSE} = 16679.10, p = .086, \eta^2_p = .092\), with the trend being for faster responses for recombined probes than for new identity-old location probes. The comparison of the two types of negative probe for spatial recognition – old identity-new location and new identity-new location – was significant for accuracy, \(F(1,31) = 4.87, \text{MSE} = 64.37, p = .035, \eta^2_p = .136\), but not RT, \(F(1,30) < 1\). The negative spatial recognition responses were easier for the new identity-new location probes, in which the verbal information is consistent with a negative response, than for the old identity-new location probes, in which the verbal information conflicts with a negative response.

For verbal recognition, the comparison of the recombined probes (old identity-old location) and old identity-new location probes, both of which require a positive response, was significant for accuracy, \(F(1,31) = 4.19, \text{MSE} = 11.51, p = .049, \eta^2_p = .119\), but, surprisingly, it was the recombined probes that showed the lower rate of accuracy (Table 3). This comparison was not significant for RT, \(F(1,31) < 1\) (Table 3). In comparisons for the two negative probe types for verbal recognition – new identity-old location and new identity-new location – although the difference in accuracy was not significant, \(F(1,31) = 2.02, \text{MSE} = 6.57, p = .165, \eta^2_p = .061\), the difference in RT was significant, \(F(1,31) = 8.56, \text{MSE} = 7928.36, p = .006, \eta^2_p = .216\). Once again the direction of this effect was unexpected (see Table 3), with RTs being shorter for the new identity-old location probes (where location information favors an incongruent positive response) than for the new identity-new location probes (where the location information is consistent with a negative response).

Binding as a function of serial position

Since spatial and verbal recognition showed contrasting patterns of effects in the full data set, subsidiary analyses investigating serial position were conducted separately for the two forms of recognition. Each ANOVA had a 2 (probe type: intact, recombined) \times 2 (serial position: 1 and 2, 3 and 4) design. For spatial recognition, there was a nonsignificant trend for accuracy to be higher for intact probes than for recombined probes, \(F(1,31) = 3.29, \text{MSE} = 94.96, p = .079, \eta^2_p = .096\), and although accuracy was higher for serial positions 3 and 4 than for serial positions 1 and 2, \(F(1,31) = 6.30, \text{MSE} = 16674.74, p = .018, \eta^2_p = .169\), this factor did not interact with probe type \((F < 1); \text{see Table 2 for means})\). On the other hand, RTs for spatial recognition showed a substantial main effect of probe type, \(F(1,31) = 13.03, \text{SE} = 399901.43, p = .001, \eta^2_p = .296\), whereas neither the effect of serial position or the interaction was significant \((Fs < 1)\). As can be seen in Table 2, intact probes were responded to faster than recombined probes for serial positions 1 and 2 as well as...
for serial positions 3 and 4. The corresponding analyses for verbal recognition did not yield any significant effects. Thus, consistent with the analyses based on the full data set, an advantage for intact probes over recombined probes was found only for spatial recognition. However, the influence of binding on spatial recognition was present for items early as well as late in the sequence, suggesting some resilience in the retention of letter-location associations in auditory WM.

**Discussion**

Experiments 1 and 2 demonstrated an advantage in processing intact compared to recombined probes, consistent with some association of the identity and location features in the representation of each sound. However, the precise nature of this association (symmetric or asymmetric) and whether it depends on attending to both features could not be established using the methodology of those experiments. Experiment 3, in which participants focused exclusively on either verbal or spatial recognition, enabled an assessment of whether the non-attended feature is encoded in obligatory fashion and influences retention of the attended feature. For spatial recognition, intact probes were responded to with greater accuracy and shorter latencies than recombined probes. Thus recognition of a location was facilitated if the probe reinstated the letter that was used in presenting that location as part of the study sequence, implying that the spatial location of a sound is encoded with reference to its verbal identity. However for verbal recognition, no reliable differences in accuracy or latency were observed for the intact and recombined probes. Thus retention of the verbal identity of a sound appears to proceed without reference to the sound’s location, at least when attention can be dedicated to the verbal information.

Subsidiary analyses of serial position showed evidence of binding of the two features affecting spatial recognition for the first two items of the sequence as well as for the last two, confirming the robustness of the associations formed between the features of auditory objects. Further evidence of the asymmetric influence of the two features on each other is that recognition of a location from the study sequence was facilitated if the location was paired with a letter from the sequence (in a recombined probe) rather than with a letter that had not been part of the study sequence. In other words, using an ‘old’ rather than ‘new’ letter in the probe stimulus facilitated recognition of an ‘old’ location. Similarly, identification of a ‘new’ location in a negative recognition probe was facilitated by use of a ‘new’ rather than an ‘old’ letter in the probe. Thus, congruence of verbal information in terms of old/new status facilitated old/new judgments of spatial information (or in other words, familiarity of the verbal content of the probe influenced spatial recognition in the manner expected).

By contrast, for verbal recognition, additional comparisons of the positive and negative probe types yielded unexpected results. First, accuracy was lower for recombined probes (an ‘old’ letter paired with an ‘old’ location) compared to old identity-new location probes. Thus congruent (‘old’) spatial information did not facilitate a positive recognition judgment for ‘old’ verbal information. Second, RTs were shorter for new identity-old location probes than for new identity-new location probes. Thus, congruent (‘new’) spatial information did not facilitate a negative recognition judgment for ‘new’ verbal information. It is not easy to provide a coherent explanation of these results. In respect of the lower accuracy for the recombined probe relative to the old identity-new location probe, it is possible that the recombined probe activates representations of the two study stimuli with which it shares a single feature, and since one of these stimuli involved a letter different to the letter of the probe, this may act against making a positive recognition judgment. The old-identity-new location probe should activate a representation of just one study item, the item with the same letter as the probe, and so a positive response would be favored under this reasoning. A similar argument could be made for the negative probes. The new identity-old location probe might prime the representation of the study item that was presented from the probe’s location, and since this item involved a letter different to the letter of the probe, a negative response would be favored. While this explanation is reasonable in isolation, it would follow that the recombined probe should also be more difficult that the intact probe (an effect that was not observed for verbal recognition). Also, it is not clear why the pairs of positive and negative probes would show effects in the opposite direction for spatial recognition. Further investigation of these unexpected effects for verbal recognition is warranted.

With reference to the theoretical positions advanced in the introduction, the results support the proposition that verbal identity is a more fundamental dimension than spatial location in the representation of auditory objects. Thus it may be possible to encode a sequence of sounds with reference to their verbal identities but not their locations, enabling verbal recognition to proceed without any influence of spatial information. Further, it may not be possible to encode the locations of sounds without reference to their identities, meaning that spatial recognition is influenced by links between verbal identity and location laid down when the study stimuli are encoded. The results therefore support an asymmetric relationship in the binding of verbal identity and location for acoustic stimuli.

**Experiment 4**

One reason for the influence of verbal identity on judgments of location observed in Experiment 3 could be that the extended sequence of four memory items places considerable demands on differentiating and retaining the four locations. Perhaps the differentiation and retention of the locations is assisted if each location is associated with the letter presented from it. It is plausible that the highly familiar letters with their distinct long-term memory representations are more easily differentiated than the locations, and so the representations of the locations may be bootstrapped to the representations of the letters to assist the differentiation of the former.

To provide an extreme test of the dependence of spatial recognition on verbal-identity information, in Experiment 4 we simplified the memory task by presenting a single memory stimulus followed by a recognition probe stimu-
lur, thereby minimizing the demand on spatial memory. Critical interest focused on spatial recognition judgments for intact probes versus probes presenting the to-be-remembered location paired with a new letter. If location recognition judgments show an advantage for the intact probes over the new identity-old location probes, then this would suggest that representations of location are tied to representations of identity even under conditions of minimal demand on memory. As in Experiment 3, half the participants were told to ignore the letters, whereas the remaining participants were told to ignore the locations. For verbal recognition, the critical comparison concerned intact probes and old letter-new location probes. Based on Experiment 3, the expectation was that verbal recognition would proceed independently of spatial information, and so no difference in performance would be observed for these two probe types.

**Method**

**Participants**
Forty-eight students (mean age = 19.68 years; SD = 1.99 years) were randomly assigned to complete either the letter or location recognition task, with half the sample assigned to each.

**Apparatus and stimulus materials**
There were two stimuli for each trial, a study stimulus and a probe stimulus, and four basic trial types representing the relationship between them. The probe stimulus could use: (1) the letter and location of the study stimulus (i.e. an intact probe), (2) the letter from the study stimulus combined with a new location (old identity-new location); (3) a new letter combined with the location from the study stimulus (new identity-old location); or (4) a new letter and new location (new identity-new location). A unique set of stimuli for 132 trials (eight practice, 124 test) was constructed for each participant, with each set of four successive trials using one trial each of the four types described above, with these trials presented in random order. Letters and locations were chosen at random, with the restriction that all letters and locations were used approximately equally often across each stimulus set.

**Design and procedure**
For half the participants, instructions emphasized focusing on the letter from the study stimulus, ignoring its spatial location, and basing the recognition judgment exclusively on the match between the letters presented in the study and probe stimuli, whereas for the remaining participants, instructions emphasized focusing on the location of the study stimulus, ignoring the letter, and basing the recognition judgment exclusively on the locations presented in the study and probe stimuli. The procedure for each trial was that a 2 s visual ready warning was followed by the study and probe stimuli presented with a 1500 ms SOA. A 2 s delay followed the participant’s response prior to onset of the next trial. For each recognition condition, two trial types required a positive response and two required a negative response.

**Results**

The critical comparison of probe types concerns the two types of positive probe, intact probes versus probes that match the study stimulus on the attended dimension, but introduce new information on the non-attended dimension (old identity-new location probes for verbal recognition and new identity-old location probes for spatial recognition). ANOVAs included the between-groups factor recognition task (verbal, spatial) and the repeated-measures factor probe type (intact, alternative positive probe). For percent correct, the main effect of recognition task was not significant, \( F(1,46) = 2.64, \) \( MSE = 28.14, \) \( p = .11, \) \( \eta_g^2 = .03, \) indicating similar levels of accuracy for verbal and spatial recognition under this simplified recognition task. However, the main effect of probe type was significant, \( F(1,46) = 12.30, \) \( MSE = 6.03, \) \( p = .001, \) \( \eta_g^2 = .21, \) as too was the interaction between recognition task and probe type, \( F(1,46) = 10.55, \) \( MSE = 6.03, \) \( p = .002, \) \( \eta_g^2 = .187. \) Fig. 5 shows a clear contrast in the pattern of accuracy scores for verbal and spatial recognition. For verbal recognition, accuracy is essentially equivalent for the intact probes and the alternative positive probes, \( F(1,23) < 1, \) whereas for spatial recognition, accuracy was substantially higher for the intact probes compared to the alternative positive probes, \( F(1,23) = 13.54, MSE = 10.16, p = .001, \) \( \eta_g^2 = .37. \)

Very similar outcomes were observed for the RT data. In this case, the main effect of recognition task yielded \( F(1,46) = 3.91, \) \( MSE = 7130.01, \) \( p = .054, \) \( \eta_g^2 = .078, \) with latencies tending to be shorter for verbal recognition compared to spatial recognition (see Fig. 5). Once again, the main effect of probe type was significant, \( F(1,46) = 30.03, \) \( MSE = 3696.60, \) \( p < .001, \) \( \eta_g^2 = .395, \) but qualified by a significant recognition task by probe type interaction, \( F(1,46) = 11.14, MSE = 3696.60, \) \( p = .002, \) \( \eta_g^2 = .195. \) With RTs, as for accuracy, the difference in performance for the intact probes and alternative positive probes was restricted to spatial recognition judgments (see Fig. 5). For verbal recognition, RTs were similar for the intact and alternative positive probes, \( F(1,23) = 2.46, MSE = 3442.73, \) \( p = .13, \) \( \eta_g^2 = .097. \) In contrast, for spatial recognition, RTs were substantially shorter for the intact probes compared to the alternative positive probes, \( F(1,23) = 36.38, MSE = 3950.48, p < .001, \) \( \eta_g^2 = .613. \)

The final set of analyses compared the two types of negative probe: those where new information was presented for both the attended and non-attended dimensions (new identity-new location probes) and where new information was presented for the attended dimension but old information was presented for the non-attended dimension (i.e. new identity-old location probes for verbal recognition and old identity-new location probes for spatial recognition). In an analysis of accuracy involving the factors recognition task and negative probe type, there was a significant main effect of recognition task, \( F(1,46) = 7.92, MSE = 20.56, p = .007, \) \( \eta_g^2 = .147, \) with more accurate performance for verbal compared to spatial recognition (see Table 4). The main effect of negative probe type was not significant, but the interaction of the two factors yielded \( F(1,46) = 6.35, MSE = 6.40, p = .015, \) \( \eta_g^2 = .121 \) (Ta-
ble 4). However, follow-up analyses showed that the difference between the two negative probe types was not significant for either verbal recognition, $F(1,23) = 3.29$, $MSE = 2.23$, $p = .083$, $\eta^2_p = .125$, or spatial recognition, $F(1,23) = 3.77$, $MSE = 10.58$, $p = .065$, $\eta^2_p = .141$. The trend was for higher accuracy for the new identity-new location probe compared to the other negative probe for spatial recognition, but for the accurate rates to be in the reverse direction for verbal recognition. When a recognition task x negative probe type ANOVA was conducted on RTs, none of the effects was significant.

**Discussion**

In Experiment 4, participants attended exclusively to either a single letter or location. In line with the results of Experiment 3, spatial recognition was affected substantially by whether the positive probe stimulus preserved the verbal information present in the study stimulus. In contrast, verbal recognition was not affected by whether the positive probe preserved the spatial location used in the study stimulus.

The major outcomes of Experiments 3 and 4 – effects on spatial recognition of whether the verbal information presented in study stimuli is preserved in recognition probes, but the absence of comparable effects on verbal recognition – are consistent with the primacy of the representation of verbal identity for auditory objects. If the principal differentiation of auditory objects is with respect to verbal identity (‘what’), with location (‘where’) encoded with reference to verbal identity, judgments of verbal identity could proceed independently of spatial location, whereas judgments of spatial location would depend on whether verbal identity is reinstated or not. This proposal that ‘what’ has primacy over ‘where’ in auditory WM contrasts with the contention that ‘where’ has primacy over ‘what’ in visual WM (Elsley & Parmentier, submitted for publication; Jiang et al., 2000).

**Experiment 5**

The results of Experiments 3 and 4 suggest an asymmetry in the association of verbal identity and spatial location when attention can be focused on one of these features: The verbal identity of a sound can be encoded in WM without reference to its location, whereas encoding its location obligates binding to its verbal identity. Experiment 1, in requiring recognition decisions based on both letter and location, did not allow examination of whether binding was symmetric or asymmetric under conditions in which attention was directed to both features. Accordingly, there is a gap in our understanding of the association between verbal and spatial features in the retention of sounds: When attention is directed to both dimensions of information, is binding symmetric or asymmetric? In other words, does the relationship between features depend on whether one or both features are attended to explicitly?

To address this question, Experiment 5 used a procedure like that of Experiment 1 (four study stimuli followed by a retention interval and then a recognition probe) ex-
cept that participants were not informed of the nature of the recognition task until after the presentation of the to-be-remembered stimuli. At that time, a visual cue – LETTERS or LOCATIONS – was presented to focus the participant’s upcoming judgment on either verbal or spatial recognition. The rationale was that because the two visual cues occurred with equal frequency but in an unpredictable fashion from trial to trial, participants would be forced to encode both the letters and locations for the study stimuli. Therefore, attention at encoding was directed to the two features, but recognition was focused on a single feature. Critical interest is in whether, under these encoding conditions, the pattern of performance for intact and recombined probes for verbal and spatial recognition show evidence of symmetry or asymmetry in feature binding.

Method

Participants
Thirty-six students (mean age = 19.88 years, SD = 2.39 years) participated in this experiment.

Auditory stimuli, apparatus, design and procedure
As indicated above, the major change to the procedure relative to Experiment 1 is that a visual recognition cue – LETTERS or LOCATIONS – was presented between the last memory item and the recognition probe (i.e. midway through the retention interval). The cue focused the participant on one of the two features – verbal identity or spatial location – in making a recognition response to the auditory probe. The experiment used a 2 (recognition task: verbal, spatial) × 5 (probe type: intact, recombined, old identity-old location, new identity-old location, new identity-new location) repeated-measures design.

Ten practice trials and 120 test trials were constructed anew for each participant. Each successive set of 10 trials comprised one trial for each cell of the design, presented in random order. Thus there were 12 test trials for each cell in the design. As in previous experiments, the 12 trials within a cell were constructed to balance any relevant stimulus characteristics, for instance the 12 intact probes for each recognition task used the study items from the four serial positions equally often as the probe stimulus, and the letters and locations used for the recombined probes were drawn from all the possible pairs of study items with equal frequencies. The letters and loudspeaker locations were selected for each trial under the constraints described in Experiment 1.

Each trial began with a 2000 ms visual warning signal, followed by the presentation of a sequence of four letters, each from a different loudspeaker, using an SOA of 1500 ms. Next, a silent interval of 1500 ms occurred before the presentation of a visual cue (LETTERS or LOCATIONS) in the center of the computer screen, followed 1500 ms later by the onset of the auditory recognition probe. When the cue was LETTERS, the recognition judgment was based on whether the probe comprised a letter that was part of the study sequence, whereas when the cue was LOCATIONS, the judgment was based on whether the probe comprised a location used in the study sequence. Written and verbal instructions placed equal emphasis on attending to and remembering the letters and locations.

Results

Comparisons of intact and recombined probes
An ANOVA conducted on accuracy with the factors recognition task (verbal, spatial) and probe type (intact, recombined) provided a significant main effect of recognition task, $F(1,35) = 11.31$, $MSE = 64.86$, $p = .002$, $\eta^2_p = .244$, with verbal recognition more accurate than spatial recognition (Fig. 5). There was also a significant main effect of probe type, $F(1,35) = 9.23$, $MSE = 56.92$, $p = .004$, $\eta^2_p = .209$, with accuracy higher for intact probes compared to recombined probes (Fig. 6). Critically, this effect of probe type was not modified by an interaction with recognition task, $F(1,35) = 1.95$, $MSE = 41.82$, $p = .171$, $\eta^2_p = .043$. The corresponding analysis of RTs revealed a single significant effect, the main effect of probe type, $F(1,35) = 10.35$, $MSE = 34929.53$, $p = .003$, $\eta^2_p = .288$. Shorter latencies for intact probes compared to recombined probes were observed for both verbal and spatial recognition (see Fig. 6). The interaction of recognition task and probe type yielded $F < 1$. Thus the analyses of intact and recombined probes showed symmetric effects in that verbal as well as spatial recognition was facilitated when probes retained the associations of verbal and spatial features from study stimuli.

Other comparisons among probe types
For spatial recognition, recombined probes (involving an old letter and old location) were responded to more accurately than new identity-old location probes, $F(1,35) = 4.84$, $MSE = 203.92$, $p = .034$, $\eta^2_p = .122$ (see Table 5 for means), while the same comparison for RT was not significant, $F(1,35) = 2.45$, $MSE = 116540.10$, $p = .126$, $\eta^2_p = .066$. Thus when spatial recognition probes presented

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean accuracy (percent correct) and mean of median RTs (in ms) for correct responses for two types of negative probe for spatial and verbal recognition in Experiment 4.</td>
</tr>
<tr>
<td>Spatial recognition</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>RT</td>
</tr>
<tr>
<td>Verbal recognition</td>
</tr>
<tr>
<td>RT</td>
</tr>
</tbody>
</table>

Each trial began with a 2000 ms visual warning signal, followed by the presentation of a sequence of four letters, each from a different loudspeaker, using an SOA of 1500 ms. Next, a silent interval of 1500 ms occurred before the presentation of a visual cue (LETTERS or LOCATIONS) in the center of the computer screen, followed 1500 ms later by the onset of the auditory recognition probe. When the cue was LETTERS, the recognition judgment was based on whether the probe comprised a letter that was part of the study sequence, whereas when the cue was LOCATIONS, the judgment was based on whether the probe comprised a location used in the study sequence. Written and verbal instructions placed equal emphasis on attending to and remembering the letters and locations.
old’ locations, judgments were more accurate when the probes contained congruent (‘old’) rather than incongruent (‘new’) verbal information. The comparison of the two types of negative probe for spatial recognition – old identity-new location and new identity-new location – also showed evidence of the influence of verbal information. This comparison was significant for both accuracy, $F(1,35) = 20.72$, $MSE = 190.70$, $p < .001$, $\eta^2_p = .372$, and RT, $F(1,35) = 21.33$, $MSE = 41428.28$, $p < .001$, $\eta^2_p = .379$ (see Table 5 for means). The negative spatial recognition responses were facilitated for the new identity-new location probes, in which the verbal information was consistent with a negative response, compared to the old identity-new location probes, in which the verbal information conflicted with a negative response.

In contrast to these results for spatial recognition, for verbal recognition, the comparison of the recombined probes and old identity-new location probes was not significant for either accuracy or RT; larger $F(1,35) = 1.96$, $MSE = 21295.697$, $p = .170$, $\eta^2_p = .053$. Similarly, the two negative probe types for verbal recognition – new identity-old location and new identity-new location – did not differ significantly in either the analysis of accuracy or the analysis of RTs ($Fs < 1$).

**Binding as a function of serial position**

Given that both verbal and spatial recognition showed superior performance for intact compared to recombined probes in the full data set, the subsidiary analyses investigating serial position used a 2 (recognition task: verbal, spatial) x 2 (probe type: intact, recombined) x 2 (serial position: 1 and 2, 3 and 4) ANOVA design. The analysis of errors did not yield any significant effects, arguably because of reduced power due to the limited number of trials available for these analyses. However, the RT data showed shorter latencies for intact probes relative to recombined probes, $F(1,35) = 13.90$, $MSE = 530966.29$, $p = .001$, $\eta^2_p = .284$, and the main effect of serial position was also significant, $F(1,35) = 6.73$, $MSE = 324027.19$, $p = .014$, $\eta^2_p = .161$, with shorter RTs observed for probes that used information from the more recent serial positions (see Table 2). None of the interactions was significant ($Fs < 1$), indicating that the RT advantage for intact probes was present for early as well as late items in the sequence.

**Discussion**

Experiment 5 employed conditions in which both the verbal and spatial information had to be encoded but the recognition judgment was focused on just one class of information. Under these circumstances, verbal as well as spatial recognition judgments showed an advantage in both accuracy and RT for intact probes compared to recombined probes. Thus, requiring attention to both the letters and locations in the study sequence meant that associations between the verbal and spatial features of individual study stimuli were retained in memory, and reinstating these associations at test facilitated verbal as well as spatial recognition. Subsidiary analyses of serial position showed that the advantage in RT attendant on using intact probes was present for the early as well as the late serial positions. Despite evidence from the intact and recomb-
bined probes that each class of information affected recognition of the other, there was evidence from other comparisons of probe types to suggest that verbal information exerted a stronger influence on spatial recognition than did spatial information on verbal recognition. Thus recognition of ‘old’ locations was facilitated if the positive probes presented congruent verbal information (an ‘old’ letter) rather than incongruent verbal information (a ‘new’ letter). Similarly, negative recognition judgments for ‘new’ locations were facilitated if the negative probes presented a congruent ‘new’ letter rather than an incongruent ‘old’ letter. Similar effects were not observed for verbal recognition.

**General discussion**

Although the localization of speech in space and the extraction of verbal content from it are very common concurrent processes, there has been limited investigation of whether the enduring representations resulting from these processes are associated in memory. The principal objective of this series of experiments was to investigate the nature of any associations between the identity and location of auditory stimuli in WM.

Using a task in which participants encoded and judged both the verbal identity and spatial location of auditory stimuli, Experiment 1 showed a substantial advantage in ease of recognition for intact relative to recombined probes, consistent with the binding of verbal and spatial features in the representation of speech sounds. This indicates that participants did not rely exclusively on independent verbal and spatial WM stores. Experiment 2 also demonstrated an advantage in recognition for intact over recombined probes, but this time for non-verbal sounds. Thus a binding effect was observed whether auditory stimuli afforded familiar verbal codes or not. The integration of identity and location features appears to be a general property of the retention of auditory information.

Analyses taking into account the temporal position of the features involved in the recognition task revealed that binding for auditory objects persisted for at least 5.5 s (the interval between the second study stimulus and the probe), and, more significantly, was not eliminated by proactive or retroactive interference (Cowan, 1995; Mondor & Zatorre, 1995). In itself this is an important finding because it indicates that binding is not restricted to the current focus of auditory-spatial attention and immediate perception. The durability of feature binding for auditory stimuli contrasts with the evidence from Allen et al. (2006) of the fragility of binding for visual objects. As noted earlier, Allen et al. reported that the binding of shape and color features was much more pronounced for the last of four serially presented visual stimuli (using a 500 ms SOA and a 900 ms retention interval). They argued that the encoding and retention of further visual stimuli disrupts the fragile binding of features of earlier stimuli. Supporting this interference argument, Elsley and Parmentier (submitted for publication), using simultaneous presentation of three visual stimuli, reported no diminution of binding for shape and location with an increase in the unfilled retention interval of up to 4 s, the longest of the intervals they tested. Thus the limited available evidence (garnered under quite different experimental methods) points to greater resilience of the auditory system relative to the visual system in retaining associations between features under interference.

Experiments 3 and 4 provided evidence of an asymmetric relationship between features when attention was directed to a single feature. When recognition was focused on the verbal information conveyed by spatially distributed letter sounds, performance was not advantaged when positive probes reinstated letters from the locations they had occupied as study items. On the other hand, when recognition was focused on the spatial information for the same stimuli, performance was superior when positive probes comprised a location conveyed with the same letter sound as was used for this location at study, rather than a different letter. This asymmetric pattern of effects, with recognition of location affected by identity information, but recognition of identity unaffected by location informa-
tion, is consistent with the proposal that identity is primary in the representation of a letter sound, with the sound's location bootstrapped to its identity.

The outcomes for Experiments 3 and 4 contrast with outcomes from studies of visual WM. For example, using a method similar to that used in Experiment 4, Olson and Marshuetz (2005) reported that recognition of the identity of a single visual stimulus (an abstract shape or face) was made more difficult if its location was changed, rather than remained the same, from study to test (see Jiang et al., 2000, for similar evidence). These outcomes are consistent with Jiang et al.'s (2000) argument that the spatial configuration of a display of objects provides a reference frame for encoding features associated with the identity of the objects (e.g. shape and location). The reversed relationship between "where" and "what" features in the visual and auditory modalities suggests that location is not a prominent feature in all circumstances.

Further extending the limited knowledge of binding in auditory WM, Experiment 5 revealed that in circumstances in which both the verbal-identity and spatial-location features are encoded but only one is tested, verbal recognition is influenced by spatial information and vice versa. This suggests that, in contrast to what was observed in single-feature encoding conditions (Experiments 3 and 4), binding is symmetric under dual-encoding conditions. Thus the nature of the links formed between the features of sounds depends on how attention is allocated in encoding the stimuli. The contrast in results between Experiments 3 and 5 shows that the task-relevance of features influences the nature of their binding. The same conclusion was reached by Hommel (1998, 2004; Hommel & Colzato, 2004) in investigating binding of features in visual perception. His research demonstrated binding not only among stimulus features (e.g. color and location) but also between stimulus features and response features (e.g. whether a left or right key-press is made). Further, the strength of association of a particular feature with other features was greater if the particular feature was task-relevant (Hommel, 1998; Hommel & Colzato, 2004). Thus a general characteristic of cognition could be that the extent to which features are relevant to task performance influences the extent to which they are associated in representation. However, task-relevance is not sufficient to account for all of the variation in the binding of features for either perception or WM. Hommel and Colzato (2004) argued that a task-irrelevant feature could enter into associations with other features provided the values for the task-irrelevant feature were sufficiently salient (e.g. vivid colors). In visual WM, location influences the recognition of other features even when it is task-irrelevant (Elsey & Parmentier, submitted for publication; Jiang et al., 2000; Olson & Marshuetz, 2005), and in auditory WM, verbal identity influences the recognition of location when identity is task-irrelevant (Experiments 3 and 4). Accounting for the influence of these non-attended features will be critical in theory-building.

**Difficulty of retention and relative discriminability of features**

Before discussing implications of the results for the theoretical positions advanced in the introduction, we first address two alternative interpretations of differences observed for the retention of verbal identity and spatial location. First, perhaps the asymmetric pattern of results in Experiment 3 arose because memory for spatial location was more difficult than memory for verbal identity. Perhaps an intact-recombined difference can be demonstrated only when recognition judgments are of some critical level of difficulty. To address this possibility, we reanalyzed the key data of Experiment 3, dividing the sample completing each recognition task using a median split on overall accuracy. This between-groups accuracy factor did not enter into any significant interactions with probe type (intact, recombined) when the accuracy and RT data were reanalyzed for either verbal or spatial recognition. Thus the differences in performance for intact and recombined probes reported for spatial recognition, and the absence of such differences for verbal recognition, do not appear to be especially sensitive to task difficulty. Further, it is noteworthy that Experiment 4, using a much less demanding memory task, demonstrated a similar asymmetric pattern of influence of the two dimensions of information on each other, despite spatial and verbal recognition not differing significantly in accuracy in this case.

A related explanation of the asymmetric pattern of effects is that they stem from differences in the relative discriminability of the features, such that letters high in discriminability influence retention of locations low in discriminability, but not vice versa. In domains other than WM, there is evidence that relative discriminability affects the extent to which one dimension influences the processing of another. For instance, Melara and Mounts (1993) showed that classifying sounds as to loudness was made more difficult by introducing an orthogonal variation in pitch, and that increasing the separation of stimuli in pitch (while holding intensity differences constant) further increased the difficulty of loudness classification. Similarly, Melara and Mounts (1993) showed that the magnitude of Stroop interference depends on the relative ease of discriminating the word and color stimuli. Returning to our WM data, Experiment 5 provides the most direct evidence on the relative discriminability of the letters and locations, since this is the only experiment in which recognition of the two classes of information was tested independently, but on the same sample. To test the influence of relative discriminability, we divided participants into those who showed the larger advantages in accuracy for letters compared to locations (means of 95.1% versus 76.5%, respectively) versus those who showed the smaller such advantages (96.5% versus 93.7%, respectively). When this factor was entered into ANOVAs on accuracy and RT along with recognition task (verbal, spatial) and probe type (intact, recombined), it did not interact with probe type in any way (Fs < 1 for the critical three-way interactions). Thus the extent of influence of one feature on recognition of the other did not appear to be influenced by the relative difficulty of the two. Of course this post hoc analysis is limited in that it was not possible to remove entirely the difference in accuracy for letters versus location. A more systematic investigation in which discriminability is manipulated experimentally is clearly warranted.
Re-evaluation of theoretical proposals

While recognizing that an explanation based on relative discriminability cannot be dismissed conclusively, it is worth reconsidering the alternative theoretical proposals advanced in the introduction, in light of the reported results. None of these proposals is sufficient to account for the full set of results from Experiments 1–5, so we consider possible extensions to these proposals. Evidence of the retention of associations between the identity and location features of the sound stimuli was found in each of the experiments, and so a WM structure limited to independent verbal and visuo-spatial systems is implausible. Baddeley (2000, 2001) added the episodic buffer to the WM model to accommodate, among other phenomena, evidence of binding of different classes of information. Binding is achieved through representing events or objects in the buffer using unitary multi-dimensional codes. The revised WM model could accommodate the results of Experiments 1, 2 and 5 by assuming that identity and location features are incorporated in unitary multi-dimensional representations of the sound stimuli in the buffer. Assuming the artificial sounds of Experiment 2 were not coded phonologically, the buffer would need to be capable of retaining representations of lower-level perceptual features of sounds for at least 5.5 s and be resistant to retroactive interference during that interval. Further, the results of Experiments 3 and 4 would require additional assumptions of the revised WM model. Since Baddeley (2000) argued that the central executive can influence the contents of the episodic buffer through directing attention to particular sources of information, it could be argued that location information did not influence verbal recognition in these two experiments since location was an irrelevant feature, therefore attention was not directed to it, and so this feature was not encoded in the buffer. Alternatively, it could be argued that it was sufficient to rely on coding in the phonological loop in making verbal recognition judgments. A variant of this idea is that rehearsal may have been restricted to codes handled by the subvocal rehearsal mechanism of the phonological loop. However, further assumptions would be required in explaining why verbal information influenced spatial recognition in these experiments. Why would verbal information be incorporated in representations of the episodic buffer under conditions where only spatial information is relevant to recognition judgments? Alternatively, why would participants not rely exclusively on the visuo-spatial sketchpad if only spatial coding is required? It would seem that substantial elaboration of the revised WM model would be needed to accommodate our results.

The proposal that location is primary in representing visual objects whereas identity is primary in representing auditory objects, can explain the asymmetric patterns of results reported for visual (Elsley & Parmentier, submitted for publication; Jiang et al., 2000; Olson & Marshuetz, 2005) and auditory (Experiments 3 and 4) WM. In each domain there is evidence that the primary feature is encoded in obligatory fashion when recognition is focused on another, secondary feature, since recognition judgments are influenced by associations with the primary feature. Conversely, it seems that the primary feature can be encoded without reference to a secondary feature if the task focuses attention on the primary feature. However Experiment 5 showed an influence of the binding of verbal-identity and location features on verbal recognition. This was when both features were task-relevant at the stage when the study stimuli were encoded. Thus if identity is primary and location secondary in the representation of auditory objects, then here too additional theoretical assumptions would be required to explain how recognition of a primary feature is influenced by binding to a secondary feature. It might be the case that when there is uncertainty at the time of stimulus presentation, encoding necessarily proceeds beyond the primary feature, incorporating the secondary feature in an integrated representation. The probe might be processed to a similar extent, even when preceded by cuing of the primary feature for recognition, thus providing for the advantage of intact over recombined probes.

The results across all the experiments point to a central role for voluntary attention since different processes appear to be invoked when participants attend to and maintain both features versus when they attend to one only. When attention is directed to multiple features, voluntary processes may keep the information together, in which case an intact probe enjoys better recognition. These controlled processes may well come on top of more basic and less controlled processes that relate more to perception, for which one feature may be important in acting as an anchor for other features. So symmetric binding may reflect binding at some effortful and controlled level while asymmetric binding might reflect binding of a more primitive nature.

While we have interpreted asymmetric binding with reference to how identity may be a primary feature for sounds, an alternative explanation is that it could reflect an asymmetry in the failure of selective attention. Perhaps attention is more often inadvertently drawn to a sound’s identity when location is the focus of encoding, rather than to a sound’s location when identity is the focus. For classification tasks, Melara and Mounts (1993, 1994) interpreted interference due to stimuli varying on an irrelevant dimension in terms of selective-attention failure, and, as noted above, further argued that the relative discriminability of stimuli on the relevant and irrelevant dimensions affected the extent of this failure (and hence the degree of interference). Although the subsidiary analyses reported above did not provide evidence of an influence of relative discriminability in our data, it is possible that lapses in selective attention are influenced by other factors, such as the familiarity of the stimuli. Experiments 3 and 4 used highly familiar letters which may have drawn attention away from the arguably less-familiar locations more so than vice versa. Further research directed at identifying the precise basis of asymmetric binding is clearly warranted.

Summary and conclusions

This study demonstrated binding of the identity and location features of spatially distributed speech sounds that persists in WM for at least 5 s and despite the encoding of other sounds. The retention of associations between identity and location features was also demonstrated for
non-verbal artificial sounds. Under conditions where recognition judgments were focused on a single feature, evidence was presented that the verbal identity of a sound can be retained without reference to its spatial location, whereas retention of its location obligates an association with its verbal identity. However, when participants encoded the study stimuli uncertain of whether the recognition judgment would pertain to verbal identity or location, there was evidence that identity-location binding influenced both forms of recognition. An extension of a model based on the primacy of encoding the identity of sounds, with location then associated with identity, can accommodate these results if additional assumptions are made as to differences in processing mediated by voluntary attention. More empirical research is warranted in investigating memory for more basic features of sounds, such as pitch in combination with location (cf. Kubovy & Van Valenburg, 2001), in systematically manipulating the relative discriminability of pairs of features, and in further examining the role of attention in binding, perhaps using dual-task methods. However, recognizing the potential value of this further work should not detract from the informativeness of the present study. Compelling evidence has been presented that two fundamental classes of information, verbal and spatial, are represented interdependently in WM, not independently as modular models have traditionally assumed.

Acknowledgments

This research was supported by Grants from the Australian Research Council (DP0452312 and DP0773836) and from the Economic and Social Research Council, UK (RES-062-23-0241).

References


