Review

Resource allocation models of auditory working memory

Sabine Joseph[a,b,*], Sundeep Teki[c], Sukhbinder Kumar[d,e], Masud Husain[f,g],
Timothy D. Griffiths[d,e,**]

[a]Institute of Cognitive Neuroscience, University College London, UK
[b]Institute of Neurology, University College London, UK
[c]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3QX, UK
[d]Wellcome Trust Centre for Neuroimaging, University College London, London, UK
[e]Institute of Neuroscience, Medical School, Newcastle University, Newcastle, UK
[f]Department of Clinical Neuroscience, University of Oxford, UK
[g]Department of Experimental Psychology, University of Oxford, UK

ARTICLE INFO

ABSTRACT

Auditory working memory (WM) is the cognitive faculty that allows us to actively hold and manipulate sounds in mind over short periods of time. We develop here a particular perspective on WM for non-verbal, auditory objects as well as for time based on the consideration of possible parallels to visual WM. In vision, there has been a vigorous debate on whether WM capacity is limited to a fixed *number of items* or whether it represents a *limited resource* that can be allocated flexibly across items. Resource allocation models predict that the precision with which an item is represented decreases as a function of total number of items maintained in WM because a limited resource is shared among stored objects. We consider here auditory work on sequentially presented objects of different pitch as well as time intervals from the perspective of dynamic resource allocation. We consider whether the working memory resource might be determined by perceptual features such as pitch or timbre, or bound objects comprising multiple features, and we speculate on brain substrates for these behavioural models.

This article is part of a Special Issue entitled SI: Auditory working memory.

*Corresponding author at: Institute of Cognitive Neuroscience, Alexandra House, 17 Queen square, WC1N 3AR London, UK. Fax: +44 20 7813 2835.
**Corresponding author at: Institute of Neuroscience, The Medical School, University of Newcastle upon Tyne, NE2 4HH Newcastle upon Tyne, UK. Fax: +44 191 208 5227.
E-mail addresses: sabine.joseph@ucl.ac.uk (S. Joseph), tim.griffiths@ncl.ac.uk (T.D. Griffiths).

Contents

1.   Introduction

Every day we are required to make sense of a complex acoustic world comprising multiple auditory objects (Griffiths and Warren, 2004). Our understanding of this world depends critically on being able to hold objects in mind over seconds to appreciate both constant and changing aspects of these stimuli. Such a mechanism for retaining auditory objects constitutes a form of working memory: a cognitive mechanism for maintenance and manipulation of a limited amount of information over a short time period (Baddeley and Hitch, 1974; Cowan, 2008).

An influential model of WM suggests distinct storage for visual objects and verbal material (Baddeley and Hitch, 1974). We consider here a broader category of auditory perceptual objects that are non-verbal and are not necessarily associated with a semantic label. Baddeley actually suggested at one point that the episodic buffer (Baddeley, 2000) that was added

to his multi-component model of WM acts as temporary store of bound features across different modalities. The present synthesis addresses the mechanism for auditory WM alone, and examines possible parallels between WM for visual and auditory objects.

Comparing models of WM for visual and auditory objects is only justified if the object concept itself can be applied equally to the two modalities. Both visual and auditory objects can be considered as perceptually coherent wholes that can be distinguished from other stimuli (Griffiths and Warren, 2004). Visual objects have two dimensions at the level of the retina, while auditory objects can be considered as having dimensions of frequency and time at the cochlea. Beyond these early sensory representations, both types of objects comprise a number of perceptual features, like colour and shape for visual objects, or pitch and timbre for auditory objects, that are analysed in high-level sensory cortex. We consider here the mechanism for the maintenance of information about objects in working memory, beyond immediate
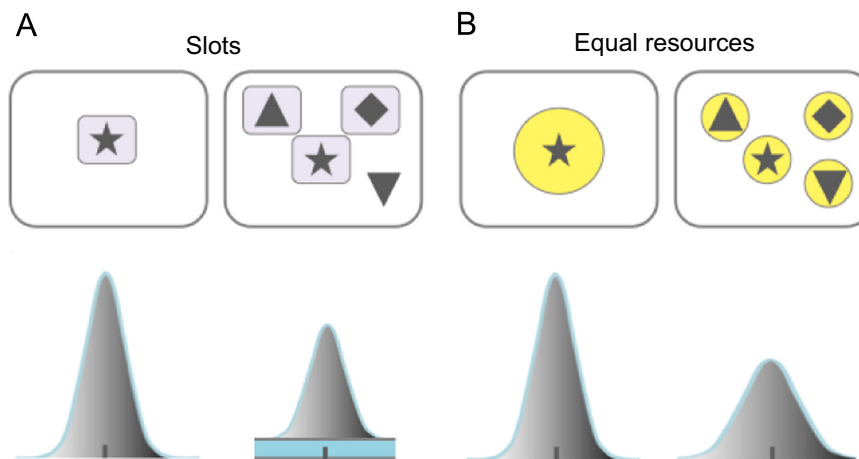
Fig. 1 – Models of working memory (A) slot model: each visual item is stored in one of a limited number of independent slots (here: WM capacity is limited to 3 slots), where each item is represented at a high resolution, indicated by a narrow error distribution around the true value of the item probed (left panel). When the capacity limit of 3 slots is exceeded, not all of the information can be stored. Thus, the slot model predicts that when probed on the item/s, which cannot be stored, one has to guess its/their identity. Both types of responses (1. recall of an item, which can be represented; and 2. recall of an item, which cannot be represented) can be described as a mixture of high-precision responses (right panel: blue component) and random guesses (green component). (B) Equal resources: Resource models of WM predict that a limited representational medium is shared out between items, where the number of items, which can be represented, is unlimited. Importantly, the precision with which each item can be stored depends on the amount of resource allocated to it. If resources are distributed equally across items, error variability (width of the error distribution usually captured by a normal distribution) increases continuously with an increase in the number of items (memory load; comparison of the error distribution for one vs. four items). Figure adapted from: Ma et al. (2014).

sensation and perception: a process that likely involves distributed cortical mechanisms.

We have specifically tested whether a recent resource model of WM for visual objects (Wilken and Ma, 2004; Bays and Husain, 2008; Ma et al., 2014), (Fig. 1B) might also apply to auditory working memory. Compared to traditional 'slot' models of visual WM that posit a fixed capacity limit for the number of objects that can be held in memory (Luck and Vogel, 1997; Zhang and Luck, 2008), (see Fig. 1A for further predictions made by the model), the new visual WM model is based on a flexible computational resource that can be shared between any number of objects (Fig. 1B). The central idea is that there is a limited WM resource, which can be shared out across any number of items. The model predicts that while a single item is represented at a high mnemonic resolution (precision), the addition of further items comes at a cost. The greater the number of items the less precisely each item is represented, as it receives a smaller share of the limited WM resource. This is a critical distinction from slot models.

Internal stimulus representations are noisy (due to random fluctuations) and the level of noise increases with the number of stimuli held in mind (represented by increasing width of error distribution, Fig. 1). To test whether the resource or the slot model explains information storage in working memory best and to measure the level of representational noise, both the method of change detection (in combination with the method of constant stimuli) as well as the method of adjustment can be applied. However, using the method of change detection to test for all-or-none storage (either detecting a change or no change) without varying the size of change parametrically is problematic, as a measure that is at least continuous to a certain extent (or ideally fully continuous as is the case with the method of adjustment) is needed to obtain measures of WM precision (see Kumar et al. 2013). The measure of precision allows distinction between models of WM.

Additionally, the resource model predicts that our limited resource can be flexibly allocated depending upon task demands (Wilken and Ma, 2004; Bays and Husain, 2008; Ma et al., 2014), where performance stays above chance even when the item limits predicted by traditional models are exceeded. This hypothesis has been tested for simultaneous (Bays and Husain, 2008; Bays et al., 2009) as well as sequential visual objects (Gorgoraptis et al., 2011; Zokaei et al., 2011). Here we describe studies that examine evidence for the resource model as applied to auditory objects. To anticipate, work on sequential objects with different pitch supports a form of resource allocation (Kumar et al., 2013). Research on more complex auditory objects with several features suggests that object rather than feature might be the form in which auditory WM is encoded (Joseph et al., 2015a), and investigation of the storage of time intervals suggests that a form of resource allocation might also be applied to WM for time (Teki and Griffiths, 2014). We focus here on behavioural studies and conclude with a brief speculation on the underlying WM substrates.

2. Working memory for sequences of tones

There have been surprisingly few attempts to examine capacity limits underlying auditory WM for tones (Li et al., 2013; Kumar et al., 2013) or other material that is difficult to verbalize or visualize (Golubock and Janata, 2013).

Previous studies used paradigms based on the method of change detection for sequences of tones. At recall, either a single probe tone is played and participants have to decide whether this was contained in the initial set, or a repeated complete sequence is judged to be the same or different. Essentially, either paradigm measures whether an item has been held in WM or not in a binary (all-or-none) fashion. Measured this way, auditory WM capacity has typically been found to be lower than visual WM capacity: for example, Morey et al. (2011) showed higher capacity for colours (3–4 items) than for tones (1–2 items). Other accounts suggest that WM capacity for tones and other sounds is highly limited, as participants were unable to report serial order of a sequence containing 3–4 sounds at 200 ms duration (Warren and Obusek, 1972).

Prosser (1995) found capacity to be limited to 1–2 tones with a short retention interval of 1 s, and to only 1 tone with a long retention interval of 7 s. Based on this study, Li et al. (2013) designed a further change detection paradigm to investigate WM capacity limits for tones in which, on average, participants were able to retain up to two tones. WM capacity estimates were based on a measure widely used in visual WM: Cowan's K (hits minus false alarms), (Cowan, 2001). The fixed capacity or 'slot' model predicts that Cowan's K or the number of items held in WM increases with memory load and peaks at a stable plateau as soon as a capacity limit is reached (here two tones) and subsequently drops to near chance performance with increasing number of items. Li et al. (2013) examined memory for tone sequences, using a single probe tone (their experiment 1) and found that WM performance decreased with an increase in memory load, with no evidence for a plateau in the pattern of Cowan's K values even at the memory load of 6 tones.

Although these studies are discussed in terms of the fixed WM capacity account, the pattern of results does not support its underlying predictions. Instead, their findings can be interpreted in terms of a resource model, which predicts a sharp drop in performance even when a single item is added to a previous one because a limited resource is now being shared between two items instead of only one. The report of Li et al. (2013) actually shows such a decline from memory load 1 to load 2, and a further decline with the addition of more items up to the highest memory load used, as would be predicted by a resource model.

Although categorical change detection tasks can be used to assess WM in more informative ways than simply measuring fixed capacity limits, e.g. by varying the degree of change (Bays and Husain, 2008; Rouder et al., 2008; Wilken and Ma, 2004), the problem with this approach is that detecting a change does not imply perfect recollection of an item. Nor does detection failure necessarily mean total absence of memory. To overcome this limitation, in the auditory domain Kumar et al. (2013) employed as their memory index a

response method, which operates in analogue fashion over a continuous scale. Using a pitch-matching task (Fig. 2A) they measured the variability of recall around the true stimulus value, where response error reflects the deviation between the target tone frequency and response frequency adjusted by the participant on continuous scale. On this basis a measure of recall precision (reciprocal of standard deviation of response error) can be obtained, reflecting the fidelity with which information is represented in memory.

In the study by Kumar et al. (2013), participants listened to pure tones in sequences of variable length and were probed on their memory for one of the tones (Fig. 2A). Participants were asked to reproduce the pitch of a tone presented at one position in the sequence (either first, second, etc.). They did so by adjusting a response dial so that the pitch they heard through headphones matched their memory. Recall precision was computed for different memory loads (sequence length). The results revealed a clear decline in recall precision as the number of tones in the sequence increased (Fig. 2). Importantly, even adding a single tone to a previous tone held in memory produced a significant drop in precision (Fig. 2C). Such a fall in WM precision cannot be explained on the basis
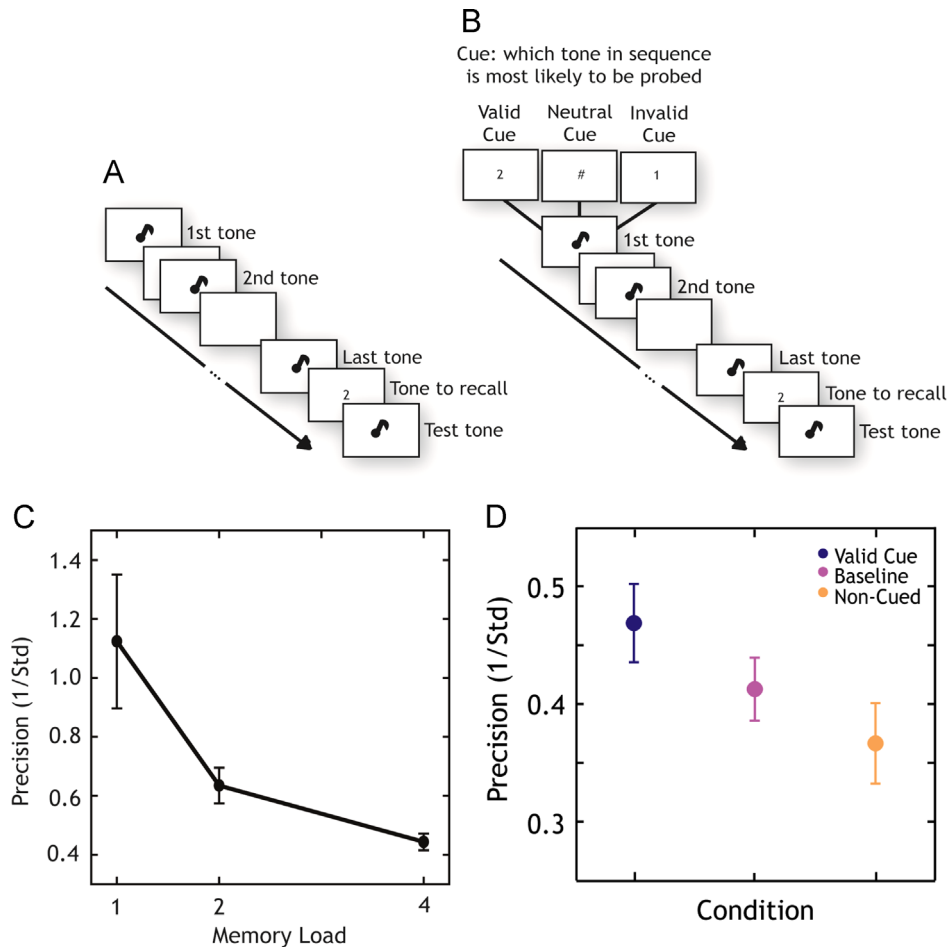
Fig. 2 – Working memory for tone sequences (A) Pitch-matching task: Subjects were presented with a sequence of tones (e.g. sample sequence with tone 1:550 Hz, tone 2:710 Hz, ..., Last tone: 670 Hz). The sequences comprised of 1, 2 or 4 tones. After the test tone sequence, a number appeared on the screen, indicating the target. A randomly-chosen probe stimulus (e.g. 520 Hz) was then played, which had to be adjusted to match the pitch of the target (here: second tone with frequency of 710 Hz). (B) Cued pitch-matching task: Subjects were presented with a cue in the form of a number appearing on the screen, indicating which tone to prioritize. The test tone sequence was then played, consisting of 3 tones. Subsequently, a number appeared on the screen, indicating the target. A randomly-chosen probe stimulus was then played, which had to be adjusted to match the pitch of the target (here: second tone). On each trial, before each sequence of three tones, a visual cue (presented for 2s) indicated the serial order position of the tone most likely to be probed. On 75% of trials, the cue was a number (62.5% valid, in which the cued tone was probed; 12.5% invalid, in which one of the two non-cued tones was probed). On 25% of trials, listeners saw a neutral cue ("#" sign), which indicated that all tones in the sequence were equally likely to be probed. (C) Results from pitch-matching: mean Precision is plotted for every memory load. The plot shows how precision decreases with an increase in memory load (number of tones in the sequence). (D) Results from cued pitch-matching task: Precision for the cued tones (blue) was significantly higher than baseline (pink). In the baseline condition, where the memory resource was equally distributed across all tones in the sequence, precision was significantly higher than in the non-cued condition (orange), resulting in a significant cost for probing non-cued tones.

of a fixed capacity model, which predicts optimal performance until the capacity limit for tones is reached (Li et al., 2013; Prosser, 1995). Furthermore, pitch-matching performance remained significantly above chance for the highest memory load of four items, which cannot be explained by a fixed capacity account either, as it predicts a sharp drop in performance when the capacity limit is breached.

These results challenge the fixed item capacity account and are better described by a shared resource model of WM. This predicts that the more items held in memory, the less precisely each item can be recalled, as previously shown in studies of visual WM (Alvarez and Cavanagh, 2004; Bays and Husain, 2008; Bays et al., 2009; Gorgoraptis et al., 2011; Wilken and Ma, 2004; Zokaei et al., 2011, Ma et al., 2014).

Additionally, the authors applied a probabilistic mixture model to individual subjects' data (Bays and Husain, 2008; Zhang and Luck, 2008), separating the variability due to random guesses from the variability associated with noisy pitch representations in WM (responses to the target). In brief, the results show that Gaussian variability in recall of the target tone frequency increases across memory loads, consistent with representations being noisier with increasing number of items distributed over a fixed resource such as a neuronal pool. Moreover, the frequency of random responses was not different across memory loads, even when memory load exceeded two – the supposed capacity limit for auditory WM (Li et al., 2013). These findings are in keeping with the results of visual WM experiments where the investigators have concluded that the data are most consistent with a limited resource model (Bays and Husain, 2008).

A further prediction made by the resource model is that memory resources can be allocated in a flexible manner (Bays and Husain, 2008). Rather than considering WM as limited to a fixed storage resolution for each item that is held, the evidence suggests that memory resources can be unevenly distributed so that prioritized items are stored with enhanced precision compared to other objects. Kumar et al. (2013) next manipulated task relevance of different sequence positions by pre-cueing (Fig. 2B). They found that recall precision was highest when the tone was more likely to be probed, relative to the neutral condition where each tone in the sequence was equally likely to be probed (Fig. 2D).

Enhancing the priority of a particular tone in the sequence thereby resulted in a robust gain in recall precision, but crucially came at a cost in precision for other tones, which were less likely to be probed, analogous to results for visual WM indicating voluntary control over resource allocation according to task priorities (Bays and Husain, 2008; Gorgoraptis et al., 2011). Again, this pattern of result would not be expected on the basis of a 'quantal' architecture of WM, where every item is stored in a 'slot' which has a fixed resolution. However, if the number of items maintained in WM was held across a limited resource or neuronal pool, it becomes easier to conceptualize how devoting more resource to one item necessarily leads to less for others in WM.

3. Working memory for sequences of complex auditory objects

In contrast to understanding how objects composed of a single sensory feature (e.g., tones) are represented in WM, most objects we perceive and remember in our everyday lives are composed of multiple features. Another longstanding debate concerns the 'unit' of working memory. That is, given that capacity of the WM is limited, what is the smallest unit for measuring this capacity? From first principles, the 'unit' could be individual sensory features like frequency, perceptual features like pitch, or objects comprising linked or bound combinations of features. What is the format for WM, and what are the implications for WM capacity? Whilst some work in vision can be interpreted in terms of the maintenance of separate perceptual features (Wheeler and Treisman, 2002), other research suggests visual WM might be limited to a fixed number of objects, but not by the number of features belonging to an object (Luck and Vogel, 1997).

More recently, Oberauer and Eichenberger (2013) concluded that WM capacity is not only limited by the number of objects, but also by the number of features per object and by their mnemonic resolution strength (indexed by recall precision). Work by Fougnie et al. (2013) on the one hand supports the idea that objects are the basic unit of visual WM, but also suggests there might be independent storage of object features. A further study in vision suggests that WM is best described as a resource shared across all objects and their features (Bays et al., 2011). Using an adjustment task to measure precision of features associated with different visual dimensions (colour and spatial location), the authors found increased binding errors (participants reported the stimulus feature of a non-probed item instead of the probed item) at high memory loads due to independent response error distributions for each feature, suggesting that features are maintained in separate WM stores. Work by Hardmann and Cowan (2015) further supports the idea of independent stores for features as opposed to objects. To summarize a large body of visual data, arguments have been developed for the storage of both perceptual features and objects in visual WM. Whether one type of storage mechanism has to 'win' this debate or whether both types of visual mechanisms might exist is a moot point.

Previous hearing studies support the independent maintenance of different sound features (Clément et al., 1999; Mercer and McKeown, 2010a, 2010b; Ries et al., 2010; Semal and Demany, 1991, 1993; Starr and Pitt, 1997). The nature of the representations retained within these stores has been recently addressed by Mathias and Von Kriegstein (2014), who explored the possibility that auditory objects might be the units of memory storage as well as perception (Griffiths et al., 2009). In a selective interference paradigm, the authors showed that auditory information (distractors) presented in the delay period influenced recall of auditory objects (tones) that were also associated with spatial features: interaural time difference (ITD) and interaural level difference (ILD). Participants held one of the features (frequency, ITD or ILD) in mind followed by the presentation of interfering tones in the

delay period. Memory performance was impaired for features affecting spatial location, when the interfering tones varied in ITD or ILD, but not when they varied in frequency. The data support the common storage of spatial cues as a position percept but separate perpetual stores for pitch and position with no evidence for the storage of a common object storage based on bound position and pitch. Whether position is a necessary intrinsic property of an auditory object, however, is another moot point. Another possible cue is loudness, but studies have shown that this is a dimension over which object invariance occurs, as opposed to an object property per se (Barbour, 2011).

The necessary cues for auditory object at the fundamental level we consider here have not been fully characterised and further work is required. We considered very simple spectral, temporal and spatial cues above. An interesting new approach is based on spectrotemporal ripples that combine spectral and temporal features and might be considered the auditory equivalent of grating stimuli used in vision (Visscher et al., 2007). Like their visual counterparts, these stimuli might be considered to be 'building blocks' from which any natural stimulus could be constructed. Initial work with these stimuli supports a good correspondence between working memory for these stimuli and forms of visual grating.

Joseph et al. (2015a) further investigated whether sounds are represented as integrated objects or individual features in auditory WM. In addition, they also tested, whether the representational format influences WM capacity by manipulating memory load. Participants memorized sequences of 1–4 auditory objects, which were composed of two different stimulus features. The centroid of the narrowband spectrum of the sounds was varied as was the amplitude modulation rate in a range not associated with pitch. Participants either maintained sequences of whole objects or sequences of individual features until recall for one of the items was tested by means of two-alternative forced choice (Fig. 3A). Memory recall was more accurate when the objects had to be maintained as a whole compared to the individual features alone (Fig. 3B). The data also show a decrease in memory performance with increasing numbers of objects (Fig. 3B), replicating the pattern of results shown for auditory WM (Kumar et al., 2013; Joseph et al., 2015b).

The results by Joseph et al. (2015a) support the storage of objects in WM at some level of processing. We cannot dismiss the storage of object cues in addition, and this study suggests that as performance on the spectral dimension is better compared to the other dimension, the spectral component may represent a dominant object cue. The argument for object-level storage is based on the interference of features within objects and the existence of a feature extraction cost for individual features in WM, and suggests we might naturally remember sounds as bound objects even when asked to only memorize one of their component features. Such feature binding might serve as a mechanism to increase WM capacity or make bound objects more robust or harder to degrade. Further research in the auditory domain is needed to investigate resource allocation for features vs. objects, ideally measuring recall precision. Clarification is required regarding the existence of feature as opposed to object level

storage – or both – and the necessary features of auditory objects that are stored in auditory WM.

4. Working memory for sequences of time intervals

Natural sequences of sounds such as speech and music have a variable temporal structure in which the onset of successive sounds is demarcated by time intervals of varying length, typically in the hundreds-of-milliseconds range. These intervals are distinct from individual sound objects or sequences
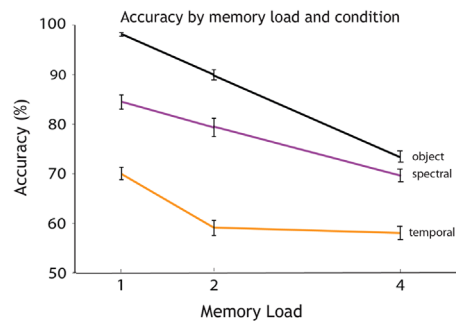

Fig. 3 – Working memory sequences of more complex objects (A) Feature vs. object task: Shown are sample trials for each experimental condition (each row illustrates one of 3 conditions). Note that the same material (2 auditory objects) is presented at encoding (identical across conditions). Each object is presented for 1s followed by an ISI of 1s. Next, a number appears onscreen for 1s, indicating which item in the sequence gets probed (here: 2 for 2nd item in the sequence). A final object is then presented and subjects have a maximum of 2s to decide, whether the object or feature of interest is the same or different from the item tested (here: 2nd item). In the spectral condition (1st row) subjects only focus on the spectral feature (in purple). In the temporal condition (2nd row) subjects only focus on the temporal feature (in yellow). In object condition, they encode the object as a whole (both features in combination). (B) Results: accuracy varies by memory load and experimental condition. Overall accuracy (percentage correct) for every memory load (1, 2, and 4 auditory objects presented within a sequence). The plot shows how accuracy decreases with an increase in memory load for each experimental condition: single feature spectral condition (in rose), single feature temporal condition (in orange) and object condition (in black).

of these and, from first principles, we would not necessarily expect the encoding of sound intervals and rhythmic patterns to rely on the same mechanisms as the encoding of individual sound objects. We consider above WM for sensory features of sound objects (such as frequency and intensity and interaural difference) and perceptual features (such as pitch, timbral dimensions, and spatial location). The encoding of time intervals between such sounds in WM has received comparatively little attention. In this section, we review behavioural work on temporal memory, especially in the context of the resource allocation model of WM (Bays and Husain, 2008; Ma et al., 2014).

Previous work on memory for time intervals has mostly focused on tasks based on detecting change in the absolute duration of a single interval (Broadway and Engle, 2011). The task requires the listener to indicate whether a reference interval is same/different or shorter/longer than the standard interval. Although studies based on single intervals have been crucial for the development of theoretical models of time perception and interval timing (e.g. Gibbon et al., 1984: for a review see Allman et al., 2014), they are limited as there is no variation in the number of intervals (or memory load) and the temporal structure of the sequences. As opposed to discrete sound features like pitch, time intervals in a sequence are serially correlated, and it is not straightforward to estimate a capacity limit as in the case for most visual and auditory features. Although a few studies have used an isochronous sequence of sounds in time perception and production tasks (e.g. Keele et al., 1989; Ivry and Hazeltine 1995), only one interval is effectively used due to the repeated presentation of a standard interval.

Recently, Teki and Griffiths (2014) examined the problem of memory for time and developed a novel paradigm to measure recall precision for representations of time intervals in WM. Inspired by the work described above on resource allocation models of WM, the authors tested whether memory resources for intervals of time in rhythmic sound sequences might be flexibly allocated between all intervals in a sequence. Fig. 4A shows a schematic of the basic paradigm: a sequence of clicks (up to four intervals) is presented with a mean inter-onset interval drawn from a range of 500–600 ms. At the offset of the sequence, a number is displayed which indicates the time interval for which the duration is to be reproduced. After a variable delay period, another click is presented which signifies the start of the interval to be reproduced. The task of the participant is to
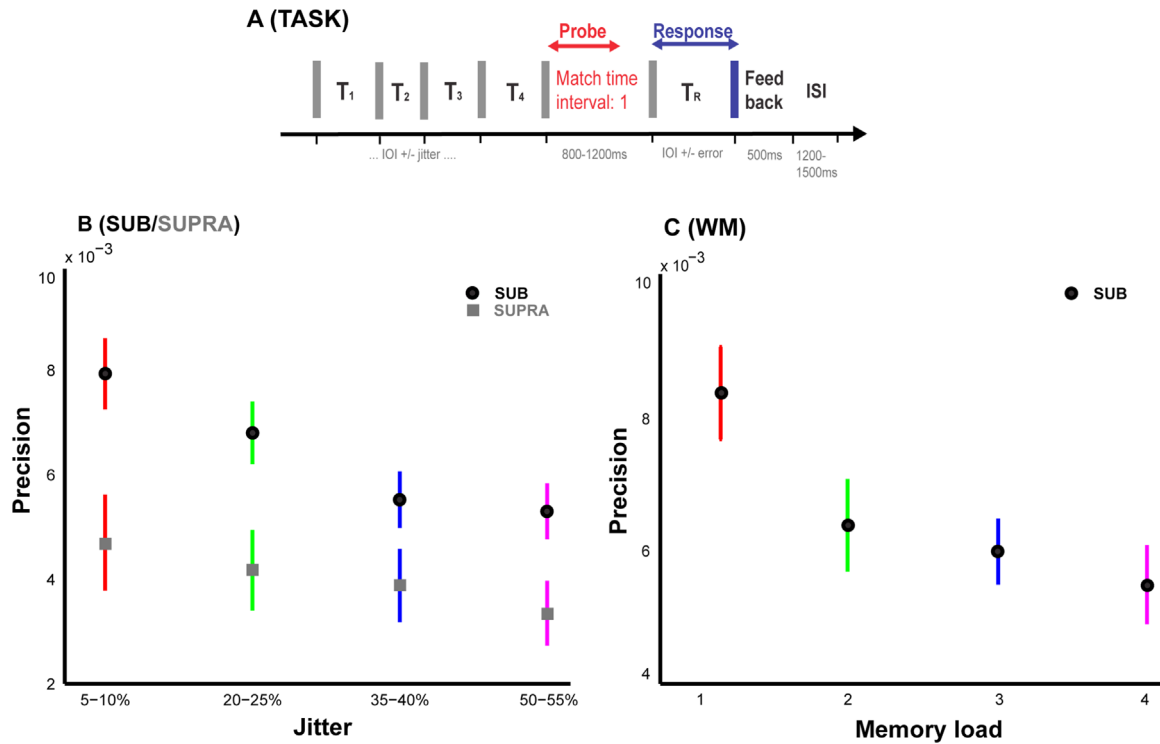
Fig. 4 – Working memory for time intervals. (A) Stimulus and task: listeners are presented a sequence of time intervals (four intervals in Experiment 1 and 2; and 1, 2, 3, or 4 intervals in Experiment 3) separated by clicks. A visual message is used to display the probe interval to be remembered and reproduced at the offset of the last click in the sequence. After a variable delay period, listeners hear another click, which signifies the start of the interval to be reproduced by pressing a button when they think that duration equal to the probed interval has elapsed. Feedback or the difference between the duration of the reproduced and the probed interval is presented after each trial. **(B) Precision vs. temporal regularity in Experiments 1 and 2:** Precision or the inverse of standard deviation of the error responses is plotted for the four different levels of temporal jitter [5–10% (red), 20–25% (green), 35–40% (blue), 50–55% (pink)]. Data with the mean indicated by black and grey circles is from Experiments 1 and 2 respectively. **(C) Precision vs. working memory load in Experiment 3:** precision is plotted as a function of the number of intervals which was the variable of interest in Experiment 3. The intervals were presented at any of the four jitter levels as in Experiments 1 and 2.

press a button at a point in time, which corresponds to his/her memory of the probed interval.

Although the task may appear to be difficult, with little training, most participants were able to execute the task well. In accordance with previous studies based on resource models of WM, performance was analysed in terms of the precision of the timing error responses, taken as the difference between the reproduced duration and the actual duration of the probed interval. The memory load of the sequence was varied through two manipulations. In the first experiment, the temporal regularity of the sequences was manipulated through introduction of temporal jitter of varying amount (from 5% to 55%). As the sequence becomes more irregular, the intervals become more dissimilar thereby increasing the effective number of intervals to be encoded. For mean interval size of 500–600 ms, a significant effect of jitter was observed, i.e. the precision of time matching performance declined with increasing amount of jitter as shown in Fig. 4B.

Teki and Griffiths (2014) also conducted a variation of this experiment with intervals twice as long as in the first study (here, 1000–1200 ms). However, no effect of jitter on the retention and reproduction of the longer supra-second intervals was observed in this case, where performance was worse compared to the task based on sub-second intervals. These results suggest that memory for an interval of time depends on the temporal context in which the intervals are presented, and like perception of time, memory is better for regular compared to irregular sequences of intervals (Teki et al., 2011, 2012).

In another experiment, Teki and Griffiths (2014) varied the number of intervals, from one to four. This task is most similar in design to the other studies based on the resource allocation model, and, in agreement with the previous results, they also found that precision of time representation in the WM significantly declined with increasing number of intervals in the sequence (Fig. 4C).

Overall, the results of Teki and Griffiths (2014) suggest that flexible allocation of WM resources could also apply to time where this is not encoded by a dedicated sensory processing system, unlike in the case of visual or auditory features. The question of whether the findings are consistent with an auditory timing resource or whether time is represented supramodally is an open one. Previous work by Grahn et al. (2011) on beat induction suggests primacy for timing in the auditory compared to the visual domain. The results from the time matching experiments show similar trends to the studies in vision and audition (Bays and Husain, 2008; Bays et al., 2009; Gorgoraptis et al., 2011; Kumar et al., 2013; Joseph et al., 2015a, 2015b) and indicate that flexible allocation of limited WM resources may contribute to the representation of the multiple objects we experience in complex and dynamic natural scenes.

5.   Conclusions and further questions

We have considered evidence for a model of auditory WM based on the use of a fixed resource that can be shared between a number of objects and allocated flexibly according to task. We do not dismiss the possibility that there might be separate stores for different perceptual features, but we have presented findings that suggest WM mechanisms exist for *bound objects* comprising multiple features. What those features might be requires further work: for example, we considered above whether spatial position is a necessary attribute of an auditory object stored in WM. Immediate questions concern the neural substrate for the WM resource and the neural mechanism for feature binding. It is premature to attempt a detailed synthesis and we briefly speculate here on possibilities in the broadest terms.

Work with musical stimuli suggests a mechanism for the maintenance of pitch that requires auditory cortex and inferior frontal cortex, which is right lateralised (Zatorre et al., 1994). For auditory memory in general a compelling idea is that the resource we consider as a behavioural entity above depends on interactions between these brain regions. Such an interaction might depend on interacting oscillations in frontal and auditory cortex for which there is preliminary evidence from direct human neurophysiological recordings (Kumar et al., 2014). One possibility is that low frequency oscillations in frontal cortex might 'drive' ongoing auditory activity during WM.

There is also preliminary evidence that the hippocampus is involved in auditory WM (Kumar et al., 2014). This is interesting in view of a suggested link between WM and long-term memory that might involve hippocampus (Cowan, 2008). Baddeley (2000) actually suggested at one point that the episodic buffer that was added to his multi-component model of WM that acts as temporary store of bound features across different modalities is dependent on the hippocampus. Work in the visual domain also suggests that visual feature binding in WM might depend on hippocampus (Pertzov et al., 2013) and we are interested in exploring whether feature binding in auditory working memory might also depend on the hippocampus.

In terms of the neural substrates encoding WM for time, there are preliminary data supporting a role for sub-cortical motor areas including the striatum and the cerebellum (Teki and Griffiths, 2013), which also mediate perception of time (Teki et al., 2011, 2012). Additionally, the inferior parietal cortex was also observed to encode memory for time as a function of load supporting earlier studies showing load-sensitive activity in visual working memory tasks (Todd and Marois, 2004; Vogel and Machizawa, 2004; Vogel et al., 2005; Ma et al., 2014).

These models have in common the idea that auditory working memory requires a critical interaction between the auditory cortex and other areas. The resource postulated above in behavioural terms is unlikely to be any form of simple storage mechanism in one area as opposed to a product of dynamic interactions between multiple areas. This system requires further clarification.

Acknowledgements

REFERENCES

Allman, M.J., Teki, S., Griffiths, T.D., Meck, W.H., 2014. Properties of the internal clock: first- and second-order principles of subjective time. Annu. Rev. Psychol. 65, 743–771, http://dx.doi.org/10.1146/annurev-psych-010213-115117.

Alvarez, G.A., Cavanagh, P., 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol. Sci. 15 (2), 106–111.

Baddeley, A.D., 2000. The episodic buffer: a new component of working memory?. Trends Cogn. Sci. 4 (11), 417–423.

Baddeley, A.D., Hitch, G., 1974. In: Bower, G.A. (Ed.), Working Memory, in the Psychology of Learning and Motivation: Advances in Research and Theory. Academic Press, New York.

Barbour, D.L., 2011. Intensity-invariant coding in the auditory system. Neurosci. Biobehav. Rev. 35 (10), 2064–2072, http://dx.doi.org/10.1016/j.neubiorev.2011.04.009.

Bays, P.M., Catalao, R.F.G., Husain, M., 2009. The precision of visual working memory is set by allocation of a shared resource. J. Vis. 9 (10), 7.1–11, http://dx.doi.org/10.1167/9.10.7.

Bays, P.M., Husain, M., 2008. Dynamic shifts of limited working memory resources in human vision. Science 321 (5890), 851–854, http://dx.doi.org/10.1126/science.1158023.

Bays, P.M., Wu, E.Y., Husain, M., 2011. Storage and binding of object features in visual working memory. Neuropsychologia 49 (6), 1622–1631, http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.023.

Broadway, J.M., Engle, R.W., 2011. Individual differences in working memory capacity and temporal discrimination. PloS One 6 (10), e25422, http://dx.doi.org/10.1371/journal.pone.0025422.

Clément, S., Demany, L., Semal, C., 1999. Memory for pitch versus memory for loudness. J. Acoust. Soc. Am. 106 (5), 2805–2811.

Cowan, N., 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. 24 (1), 87–114 (discussion 114–185).

Cowan, N., 2008. What are the differences between long-term, short-term, and working memory?. Prog. Brain Res. 169, 323–338, http://dx.doi.org/10.1016/S0079-6123(07)00020-9.

Fougnie, D., Cormiea, S.M., Alvarez, G.A., 2013. Object-Based Benefits Without Object-Based Representations, 142(3), pp.621–626. http://doi.org/10.1037/a0030300.

Gibbon, J., Church, R.M., Meck, W.H., 1984. Scalar timing in memory. Ann. N. Y. Acad. Sci. 423, 52–77.

Golubock, J.L., Janata, P., 2013. Keeping timbre in mind: working memory for complex sounds that can't be verbalized. J. Exp. Psychol. Hum. Percept. Perform. 39 (2), 399–412, http://dx.doi.org/10.1037/a0029720.

Gorgoraptis, N., Catalao, R.F.G., Bays, P.M., Husain, M., 2011. Dynamic updating of working memory resources for visual objects. J. Neurosci.: Off. J. Soc. Neurosci. 31 (23), 8502–8511, http://dx.doi.org/10.1523/JNEUROSCI.0208-11.2011.

Grahn, J.A., Henry, M.J., McAuley, J.D., 2011. FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. NeuroImage 54 (2), 1231–1243, http://dx.doi.org/10.1016/j.neuroimage.2010.09.033.

Griffiths, T.D., Warren, J.D., 2004. What is an auditory object? Nat. Rev. Neurosci. 5 (11), 887–892, http://dx.doi.org/10.1038/nrn1538.

Griffiths, T.D., Kumar, S., Von Kriegstein, K., Overath, T., Stephan, K.E., Friston, K.J., 2009. Auditory object analysis. In: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences. MIT Press, Cambridge, MA, pp. 367–381.

Hardmann, K., Cowan, N., 2015. Remembering Complex Objects in Visual Working Memory: Do Capacity Limits Restrict Objects or Features? 2(41), pp. 325–347. http://dx.doi.org/10.1037/xlm0000031.

Ivry, R.B., Hazeltine, R.E., 1995. Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. J. Exp. Psychol. Hum. Percept. Perform. 21 (1), 3–18.

Joseph, S., Kumar, S., Husain, M., Griffiths, T.D., 2015a. Auditory working memory for objects vs. features. Front. Neurosci. 9, 13, http://dx.doi.org/10.3389/fnins.2015.00013.

Joseph, S., Iverson, P., Manohar, S., Fox, Z., Scott, S.K., Husain, M., 2015b. Precision of working memory for speech sounds. Q. J. Exp. Psychol. 2006, 1–19, http://dx.doi.org/10.1080/17470218.2014.1002799.

Keele, S.W., Nicoletti, R., Ivry, R.I., Pokorny, R.A., 1989. Mechanisms of perceptual timing: beat-based or interval-based judgments? 251–256. http://doi.org/256. http://dx.doi.org/10.1007/BF00309261.

Kumar, S., Joseph, S., Pearson, B., Teki, S., Fox, Z.V., Griffiths, T.D., Husain, M., 2013. Resource allocation and prioritization in auditory working memory. Cogn. Neurosci. 4 (1), 12–20, http://dx.doi.org/10.1080/17588928.2012.716416.

Kumar, S., Gander, P., Joseph, S., Halpern, A.R., Husain, M., Nourski, K.V., Oya, H., Kawasaki, H., Howard, M.A., Griffiths, T.D., 2014. Haemodynamic pattern analysis and direct electrical recording of human brain activity during working memory for tones. Poster session presented at Soc Neurosci, Washington.

Li, D., Cowan, N., Saults, J.S., 2013. Estimating working memory capacity for lists of nonverbal sounds. Atten. Percept. Psychophys. 75 (1), 145–160, http://dx.doi.org/10.3758/s13414-012-0383-z.

Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. Nature 390 (6657), 279–281, http://dx.doi.org/10.1038/36846.

Ma, W.J., Husain, M., Bays, P.M., 2014. Changing concepts of working memory. Nat. Neurosci. 17 (3), 347–356, http://dx.doi.org/10.1038/nn.3655.

Mathias, S.R., von Kriegstein, K., 2014. Percepts, not acoustic properties, are the units of auditory short-term memory. J. Exp. Psychol.: Hum. Percept. Perform 40 (2), 445–450, http://dx.doi.org/10.1037/a0034890.

Mercer, T., McKeown, D., 2010a. Updating and feature overwriting in short-term memory for timbre. Atten. Percept. Psychophys. 72 (8), 2289–2303, http://dx.doi.org/10.3758/APP.72.8.2289.

Mercer, T., McKeown, D., 2010b. Interference in short-term auditory memory. Q. J. Exp. Psychol. 63, 1256–1265, http://dx.doi.org/10.1080/17470211003802467.

Morey, C.C., Cowan, N., Morey, R.D., Rouder, J.N., 2011. Flexible attention allocation to visual and auditory working memory tasks: Manipulating reward induces a tradeoff. Atten. Percept. Psychophys. 73, 458–472.

Oberauer, K., Eichenberger, S., 2013. Visual working memory declines when more features must be remembered for each object. Memory Cognit. 41 (8), 1212–1227, http://dx.doi.org/10.3758/s13421-013-0333-6.

Pertzov, Y., Miller, T.D., Gorgoraptis, N., Caine, D., Schott, J.M., Butler, C., Husain, M., 2013. Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. Brain: J. Neurol., 136; 2474–2485, http://dx.doi.org/10.1093/brain/awt129.

Prosser, S., 1995. Aspects of short-term auditory memory as revealed by a recognition task on multi-tone sequences. Scand. Audiol. 24 (4), 247–253.

Ries, D.T., Hamilton, T.R., Grossmann, A.J., 2010. The effects of intervening interference on short-term memory for sound

location as a function of inter-comparison interval. Hear. Res. 268, 227–233, http://dx.doi.org/10.1016/j.heares.2010.06.004.

Rouder, J.N., Morey, R.D., Cowan, N., Zwilling, C.E., Morey, C.C., Pratte, M.S., 2008. An assessment of fixed-capacity models of visual working memory. Proc. Natl. Acad. Sci. USA 105 (16), 5975–5979, http://dx.doi.org/10.1073/pnas.0711295105.

Semal, C., Demany, L., 1991. Dissociation of pitch from timbre in auditory short-term memory. J. Acoust. Soc. Am. 89 (5), 2404–2410.

Semal, C., Demany, L., 1993. Further evidence for an autonomous processing of pitch in auditory short-term memory. J. Acoust. Soc. Am. 94 (3 Pt 1), 1315–1322.

Starr, G.E., Pitt, M.A., 1997. Interference effects in short-term memory for timbre. J. Acoust. Soc. Am. 102 (1), 486–494.

Teki, S., Griffiths, T.D., 2014. Working memory for time intervals in auditory rhythmic sequences. Front. Psychol. 5, 1329, http://dx.doi.org/10.3389/fpsyg.2014.01329.

Teki, S., Griffiths, T.D., 2013. Working memory for time interval structure. Poster session presented at Soc Neurosci, New Orleans.

Teki, S., Grube, M., Griffiths, T.D., 2012. A unified model of time perception accounts for duration-based and beat-based timing mechanisms. Front. Integr. Neurosci. 5, 90, http://dx.doi.org/10.3389/fnint.2011.00090.

Teki, S., Grube, M., Griffiths, T.D., 2011. A unified model of time perception accounts for duration-based and beat-based timing mechanisms. Front. Integr. Neurosci. 5, 90, http://dx.doi.org/10.3389/fnint.2011.00090.

Teki, S., Grube, M., Kumar, S., Griffiths, T.D., 2011. Distinct neural substrates of duration-based and beat-based auditory timing. J. Neurosci.: Off. J. Soc. Neurosci. 31 (10), 3805–3812, http://dx.doi.org/10.1523/JNEUROSCI.5561-10.2011.

Todd, J.J., Marois, R., 2004. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428 (6984), 751–754, http://dx.doi.org/10.1038/nature02466.

Visscher, K.M., Kaplan, E., Kahana, M.J., Sekuler, R., 2007. Auditory short-term memory behaves like visual short-term memory. PLoS Biol. 5 (3), e56, http://dx.doi.org/10.1371/journal.pbio.0050056.

Vogel, E.K., Machizawa, M.G., 2004. Neural activity predicts individual differences in visual working memory capacity. Nature 428 (6984), 748–751.

Vogel, E.K., McCollough, A.W., Machizawa, M.G., 2005. Neural measures reveal individual differences in controlling access to working memory. Nature. 438 (7067), 500–503.

Warren, R.M., Obusek, C.J., 1972. Identification of temporal order within auditory sequences. Percept. Psychophys. 12, 86–90.

Wheeler, M.E., Treisman, A.M., 2002. Binding in short-term visual memory. J. Exp. Psychol. Gen. 131 (1), 48–64.

Wilken, P., Ma, W.J., 2004. A detection theory account of change detection. J. Vis. 4 (12), 1120–1135, http://dx.doi.org/10.1167/4.12.11.

Zhang, W., Luck, S.J., 2008. Discrete fixed-resolution representations in visual working memory. Nature 453 (7192), 233–235, http://dx.doi.org/10.1038/nature06860.

Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P.M., Husain, M., 2011. Precision of working memory for visual motion sequences and transparent motion surfaces. J. Vis. 11 (14), http://dx.doi.org/10.1167/11.14.22-2.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci.: Off. J. Soc. Neurosci. 14 (4), 1908–1919.