University College London,

Institute of Neurology

&

Institute of Cognitive Neurosciences

Multi-Sensory Working Memory

– in Vision, Audition and Touch -

Sabine Joseph

Thesis submitted for the degree of Doctor of Philosophy

of

University College London

June 2014

I, Sabine Joseph confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Our nervous systems can perform a vast variety of cognitive tasks, many involving several different senses. Although sensory systems provide a basis for the creation of mental representations, we rely on memory to form mental representations of information that is no longer present in our external world. Focussing on the initial stage of this process, working memory (WM), where information is retained actively over a short time course, experiments included in this thesis were directed toward understanding the nature of sensory representations across the senses (vision, audition and touch).

Instead of quantifying how many items one can hold in each sensory modality (all-or-none representations), new response methods were devised to capture the qualitative nature of sensory representations. Measuring quality rather than quantity of information held in WM, has led to the re-evaluation of the nature of its underlying capacity limits. Rather than assuming that WM capacity is limited to a fixed number of items, it may be more suitable to describe WM as a resource which can be shared and flexibly distributed across sensory information. Thus it has been proposed that at low loads we can hold information at a high resolution. However, as soon as memory load is increased, there is a deterioration of the quality at which each individual item can be represented in WM.

The resource model of WM has been applied to describe processes of visual WM, but has not been investigated for other sensory modalities. In the first part of my thesis I demonstrate behaviourally that the resource model can be extended to account for processes in auditory WM, associated with the storage of sound frequency (pitch, chapter 2) and speech sounds (phonemes, chapter 3). I then show that it can also be extended to account for storage of tactile vibrational frequencies (chapter 4). Overall, the results suggest that memory representations become noisier with an increase in information load, consistent with the concept that representations are coded as distributed patterns.

A pattern may code for individual object features or entire objects. As studies in chapter 2 - 4 only looked at a single type of feature each in separation, I

next examined WM information storage for auditory objects, composed of multiple features (chapter 5). Object formation involves binding of features, which become reorganized to create more complex unified representations of previously distributed information. The results revealed a clear feature extraction cost when recall was tested on individual features rather than on integrated objects. One interpretation of these findings is that, at some level in the auditory system, sounds may be stored as integrated objects.

In a final study, using fMRI with MVPA (mulitvoxel pattern analysis), memory traces represented as distributed patterns of brain activity were decoded from different regions of the auditory system (chapter 6). The major goal was to resolve the debate on the role of early sensory cortices in cognition: are they primarily involved in the perception of low-level stimulus features or also in maintenance of the same features in memory? I demonstrate that perception and memory share common neural substrates, where early auditory cortex serves as a substrate to accommodate both processes.

Overall, in this thesis memory representations were characterized across the senses in three different ways: (1) measuring them in terms of their quality or resolution, (2) testing whether the preferred format is on the feature or integrated object level; and (3) as patterns of brain activity. Findings converge along the concept that noisy representations actively held in WM are coded as distributed patterns in the brain.

Acknowledgements

First and foremost I would like to thank Masud Husain for his invaluable support, encouragement and excellent guidance throughout all stages of my PhD. I am extremely grateful to have him as my first supervisor and would like to thank him for his patience, motivation, vision, enthusiasm and immense knowledge. I would also like to thank my second supervisor Sophie Scott who also provided outstanding guidance at any time and has given me the opportunity to get involved in public engagement with neuroscience. I would also like to acknowledge Timothy Griffiths who as a close collaborator throughout the whole of my PhD offered even more excellent support on many auditory projects. He has also been a tremendous mentor to me.

I'm also extremely grateful to everyone who has contributed to the work included in my thesis, in particular Sukhbinder Kumar, as well as Paul Iverson, Sanjay Manohar, Ben Pearson, Sundeep Teki and Andrea Halpern. I would also like to acknowledge Zoe Fox and Praveetha Patalay for advice on statistics. Furthermore, I would like to thank all current and previous members of Masud Husain's lab, in particular, Nikos Gorgoraptis, Paul Bays, Yoni Pertzov, Nahid Zokaei, Valerie Bonnelle, Trevor Chong, Mia Dong and others for creating a work environment that could not have been more enjoyable. I would also like to thank them for their suggestions, feedback inspiration and expertise, which were very helpful and have encouraged my research. It has been wonderful and extremely inspiring to me not only to be a part of Masud's lab, but also the whole community at Queen Square, where I greatly value my friendship with Sundeep Teki and Anna Jafarpour.

It would not have been possible to write this doctoral thesis without the help and support of my parents Monika and Manfred Joseph and siblings Christine, Katharina and Benjamin Joseph and the rest of my family. I would also like to thank all my kind friends, to only some of whom it is possible to give particular mention here. Last but not least, a special thanks to London Surfers and the UCL Make Space community.

Publications related to this thesis

Chapter 2 has been published in Cognitive Neuroscience (Kumar and Joseph* et al., 2013).

* Joint first authors

Contents

Table of figures

Abbreviations

ACC	anterior cingulate cortex
CFC	cross frequency coupling
CM	contextual modulation
BOLD	blood oxygenation-level dependent
DLPFC	dorso-lateral pre-frontal cortex
DMPFC	dorso-medial pre-frontal cortex
ECOG	electrocorticography
EEG	Electroencephalography
FFA	fusiform face area
FIT	feature-integration theory
fMRI	functional magnetic resonance imaging
IFC	inferior frontal cortex
IPL	intra parietal lobule
IPS	intra parietal sulcus
IT	inferior temporal
LIP	lateral intraparietal
MEG	Magnetoencephalography
MT (V5)	medial temporal (extra striate visual area)
MTL	medial temporal lobe
MVPA	multi-variate pattern analysis
PAC	phase-amplitude coupling
PET	positron emission tomography
PFC	pre-frontal cortex
PM	pre motor cortex
PO	parietal operculum
PPC	posterior parietal cortex
RF	receptive field
S1	primary somatosensory cortex
S2	secondary somatosensory cortex
SPL	superior parietal lobule
SMG	supra marginal gyrus
SSTP	steady-state topographical probe
SSVEP	steady- state visual evoked potential
TMS	transcranial magnetic stimulation
V1	primary visual cortex
V4	extrastriate visual cortex
VLPFC	ventro-lateral pre-frontal cortex
WM	working memory

Chapter 1. General introduction

In every moment of our lives we perceive information through our senses. Once this moment has ceased, traces of sensory information remain in memory. Remembering means being able to access knowledge of former sensory experiences (Gale, 2005), which is a core instance of the general, flexible capacity to think about past events.

We can perform a vast variety of cognitive tasks such as reading a book, engaging in a conversation, listening to music or playing an instrument. For each task we rely on memory, which helps guide our behaviour (Baddeley, 1986). Although some actions rely more on a particular sense than another, the accomplishment of the vast majority of cognitive tasks engages multiple senses and the memories associated with them.

Sensory systems provide a basis for the creation of mental representations and memories of our external world. In 1682, the natural philosopher R. Hooke described memory as a 'repository of ideas' in which information is laid down on the 'coils' or 'spirals' of the brain. Since then the scientific study of the nervous system has significantly increased and the problem of how sensory information is retained in the brain has stimulated a great deal of research.

At an early stage of information processing, representations are believed to be sensory specific, whereas at a later stage, information from different senses is bound into unified percepts (Kaas, 1997; Treisman & Gelade, 1980). As a result, information is not only represented but also enters memory stores. However, to date it remains controversial at which processing stage information is actually 'recorded'. Throughout the history of philosophy and experimental psychology, metaphors for memory processes depended on the latest technology of recording devices, in a given era: ranging from the *camera obscura* to the tape recorder and then digital computers.

Once we accumulate experiences of the perceptual world, we can access and reproduce this information. One might think that such processes occur in a similar fashion to the above recording devices; and some models of primate memory developed in classical Artificial Intelligence for example do rely on the analogy with random-access memory systems of digital computers. However, the tangible nature of memory representations cannot be fully captured by such analogies. This is due to the transient nature of representations and the influence of other mental processes causing potential distortions of representations. Thus, memory is the immediate, noisy and sometimes erroneous knowledge of something past, influenced by additional knowledge and sensory input.

The attempt to understand the qualitative and dynamic nature of memory representations is a relatively new field of study in cognitive psychology and neuroscience. This approach originated in the domain of *vision* and the aim of this thesis is also to understand the fundamental memory-processes involved in *audition* and *touch.* Specifically, can theories that attempt to explain the nature of memory representation in vision also be extended to other senses?

My focus here is on immediate information storage – *short-term memory* which may be considered the storage aspect of *working memory* (WM). WM can be defined as the capacity to hold and manipulate task related information in mind in the absence of an external stimulus (Baddeley, 2003; Cowan, 2008). WM encompasses a number of processing stages including: (1) the stage of information encoding (external stimuli are present) followed by (2) maintenance (holding information in mind when external stimuli are absent), before (3) retrieval is tested at the stage of memory recall (Baddeley, 2003; 2007). In addition, working memory may involve manipulation or updating of stored information.

different
definitions
STM vs.
WM?

Although it is accepted that information in WM is directly linked to our sensory systems (vision, audition, touch and others), it is unclear whether information perceived with each sense is processed in a common memory store or in separate systems (Lehnert & Zimmer, 2008a; 2008b; Cowan, 2001). A fundamental feature of WM is its limited capacity (Miller, 1956) but, although studied extensively, this topic has become highly controversial in recent years. The research presented in my thesis aims to understand the nature of limits and mechanisms of WM across the senses and addresses the following key questions:

- What is the nature of WM representations of sensory information across the senses?
- Can WM processes across the senses be described by a common mechanism?
- What limits WM capacity across the senses?
- How might WM representations be characterized?
- Is the same neural circuitry involved in perception of sensory information also involved in representing the same information in WM?

repetition? common format vs basic format

1.1 Working memory stores: for different sensory modalities and attributes

Standardized WM tests are used across a wide population of patients and healthy participants. Most involve remembering lists of items – verbal (e.g. spoken words, letters (Sperling, 1960) or numbers) or visuospatial (e.g. Corsi block tapping task) – to compute memory span: the number of items that can be recalled correctly. In contrast to paradigms using free recall, methods from behavioural psychophysics have been widely employed to investigate how basic stimulus features (visual: colour or contrast; auditory: pitch or loudness; and tactile: vibrational frequency or somatosensory location) are memorised.

Psychophysics refers to a range of methods used to study perceptual or here memory processes associated with a given sensory system by measuring the effect of systematically varied stimulus properties (along one or multiple dimensions) on a subject's experience and response made to the presented information (Bruce, Green & Georgeson, 2003). Applied to WM, the most common method used to test for information storage is *change detection*: subjects have to detect a difference (or no difference) between two sets of information, where the first one is presented at the stage of memory encoding, separated by a variable delay period, before the second comparison set is presented at memory recall. More recently, as we shall see, methods of adjustment – which require participants to reproduce from memory a feature of the remembered item – have been introduced to challenge some of the inferences made from earlier change detection studies.

1.1.1 Change detection as a quantitative measure of WM

The change detection or discrimination task was first developed to measure perceptual thresholds: the point of stimulus intensity at which a subject can detect the presence of a stimulus. It was then extended to measure WM processes: holding a set of stimuli in mind over several seconds in the absence of a physical stimulus. The change detection task used to measure WM performance is often referred to as a variant of the *delayed-match to sample* task (see Figure 1-1 for an illustration of this task for different types of sensory information).

On a visual version of this task, a number of objects (e.g., coloured squares) are presented and have to be maintained in memory. Then, a probe display appears and the participant has to decide, whether the display at test differs from the previous one (Luck & Vogel, 1997). This task provides a quantitative measure of information storage: it can be used to determine the threshold needed for storage, as well as the duration of a time period over which information can be maintained. It can also be used to determine capacity limits by manipulating the factor of memory load (number of visual objects displayed). The paradigm has been used for a wide range of stimulus material across the senses, including visual (Luck & Vogel, 1997; Cowan, 2001; Alvarez & Cavanagh, 2004; Awh et al., 2007; Zhang & Luck, 2008), auditory (Saults & Cowan, 2007; Fougnie & Marois, 2011) and tactile information (Bancroft, Hockley & Servos, 2011; Gallace et al., 2008; Auvray, Gallace & Spence et al, 2011).

Evidence from visual psychophysics

Using the delayed-match to sample task, visual WM has been examined for stimulus attributes, such as size, spatial orientation, contrast, colour and motion. Such information is retained for a number of seconds, where each feature may be represented in a separate memory store (Magnussen et al, 2000). The separation of stores hypothesis is based on the observation of distinct memory decay functions for different features. For example, spatial frequency (Magnussen et al, 2000) and orientation (Magnussen et al, 1998; Vogels & Orban, 1986) can be maintained for several seconds without hardly any forgetting, while contrast and texture decay

one feature could be more dominant, initially encoded at high res, whereas other less salient features can never be encoded as high res unless cued rapidly. However, the separation of stores hypothesis remains controversial, as it is unclear, whether stimulus attributes were matched across procedures (Pasternak & Greenle, 2005) and adjusted for storage thresholds across features.

Change detection: delayed-match to sample tasks

Figure 1-1 *Delayed-match to sample task for different types of sensory information*

A) Visual information: subjects are presented with a memory array of coloured squares, followed by a delay period. At test subjects have to indicate, whether the memory array matches the final one (same) or whether one of the colours has changed (Luck & Vogel, 1997).

B) Auditory information: subjects are presented with a single tone at a given pitch (here: sampled from a range of 500-2000Hz). After maintaining this tone throughout the delay period, a second tone is presented at recall. Subjects have to make a same/different judgement (Clement, Demany & Semal, 1991).

C) Tactile information: monkeys encoded a single base frequency of a mechanical vibration (vibrotactile flutter). There were two possible base frequencies (20 and 34 Hz), which were held in mind over a variable delay period (1.5 – 4.5 s). At test, a comparison with a second vibration was made. This frequency different from one of the base frequencies by 8 Hz. Monkeys had to decide, whether the test frequency was higher or lower than the base (Romo et al., 1999).

Additional manipulations to the task include the presentation of interfering or to be ignored stimulus material in the delay period. The delay period can be of variable length in order to measure memory decay over time. Furthermore, the factor of memory load can be manipulated by presenting multiple items at encoding (e.g. here: 4 items on the visual task in contrast to a single item on the auditory and tactile tasks). Additionally, the size of change between information at encoding and at test can be manipulated.

On another version of the delayed-match to sample task, interfering or masking stimulus material is presented in the delay period (see Figure 1-2). Such task irrelevant material is selected as either of the same or distinct in nature from the memoranda, where each type of stimulus affects memory performance in a different way. For example, when retaining information about the speed and direction of motion, an interfering stimulus only leads to a decrease in performance, if it moves at the same speed as the memoranda, irrespective of its direction of motion (Magnussen & Greenlee, 1992; Blake, Cepeda & Hiris, 1997). This result suggests that speed and direction of motion are maintained in separate memory stores. Furthermore, the effectiveness of interfering material depends on when it is presented during the delay period (Lalonde, & Chaudhuri, 2002).

similar vs dissimilar, similar info will exhaust resources faster

Similarly, it has been suggested that separate memory stores exist for combinations of other stimulus attributes: e.g. spatial frequency and orientation (Magnussen et al, 1991), as well as colour and spatial orientation (Vuontela et al, 1999). Vuontela et al. presented colour or location distractors in the delay period, when either type of information also had to be maintained in memory. Location distractors interfered with location stimuli only and the same result was observed when both stimuli were composed of colour only. In another version of the delayed-match to sample task, memory load is manipulated, with the amount of information presented varying from trial to trial. Above a certain number of items, memory recall falls and, on this basis, estimates of memory capacity have been obtained. For vision, Luck & Vogel found that WM is limited to 3-4 integrated objects, defined by colour and spatial orientation (1997), (see Figure 1-1A for the paradigm used).

Although many vision researchers agree with their claim of WM capacity being fixed to a specific number of items (Cowan, 2001; Pashler, 1988; Phillips, 1974; Luck & Vogel, 1997), item-limits are affected by both stimulus complexity and change detection thresholds. Thus, it was observed that fewer items could be stored when stimuli become more complex (Alvarez & Cavanagh, 2004). Moreover, item-limit estimates are dependent on the magnitude of change deployed (Salmela, Maekelae & Saarinen, 2010, Salmela, Laehde & Saarinen, 2012). Most visual change

perceptual or memory thresholds?

detection paradigms measure memory performance using the same supra-threshold change in stimuli on ‘difference trials’ equally for all memory loads. But stimulus selection and the magnitude of change between stimuli presented at encoding and at test affects memory performance (e.g. performance is at ceiling for low memory loads) and likewise influences estimates about item-limits. For example, bringing the difference closer to the threshold directly reduces memory performance and prevents the occurrence of ceiling effects found at low memory loads (Keshvari; van den Berg & Ma, 2013).

Evidence from auditory psychophysics

Methods similar to vision have been employed to study auditory WM for a variety of stimulus features, such as pitch (Massaro, 1970; Wickelgren, 1969, Deutsch, 1970, 1972, 1973), loudness (Clement et al, 1999 (see Figure 1-1B for the task used)), spatial location (Anourova et al, 1999, Clarke, Adriani & Bellmann, 1998), complex sounds (Golubock & Janata, 2012), verbal material (Conrad, 1964; Conrad & Hull, 1964; Gupta, Lipinski & Aktunc, 2005; Drenowski & Murdock, 1980; Haberlandt et al, 2005) and speech sounds (Pisoni, 1973; Pisoni & Tash, 1974). As in experiments on visual WM, auditory WM studies predominantly use different versions of the delayed match to sample task (e.g. presenting distractors in the delay period (Deutsch, 1972, 1973) or varying the length of the delay period (Wickelgren, 1969, Deutsch, 1970)) to investigate the nature of memory for sounds.

It has been shown that humans can not only discriminate accurately between pitch and loudness (Posner, 1967), but it has also been suggested that either feature is maintained in a separate memory store (Clement et al, 1999), as discrimination thresholds for pitch and loudness decay at different rates. Anourova et al (1999) used an interference paradigm to test for a dissociation between auditory WM for pitch and location. They found a memory-load dependent segregation of stores due to selective auditory distraction on either task, which was only present for the lowest memory load of 1 item, but not for 2 items. Further interference paradigms showed that pitch and timbre are stored differently (Hall &

???

why not?

Wieberg, 2003; Krumhansl & Iverson, 1992; Semal & Demany, 1991, 1993; Starr & Pitt, 1997).

The feature of timbre, a multidimensional attribute defining the quality of a sound, makes up a sound's identity more than any other feature such as pitch, loudness or duration. Timbre is the perceptual attribute, which allows one to distinguish among tones equated for pitch, loudness and perceived duration (ANSI, 1973), and each musical instrument has a characteristic timbre at a given pitch. As in pitch delayed match to sample tasks, there was a cost when the timbre of the probe differed from the timbre of the memorandum (Crowder, 1989; Melara & Marks, 1990). Timbre comprises multiple acoustic feature dimensions (e.g. frequency or amplitude modulation (FM/AM), complexity, periodicity etc. (Grey, 1977; McAdams & Cunible, 1992; Krumhansl, 1989; McAdams et al., 1995; Samson, Zatorre & Ramsay, 1997; Lakatos, 2000; Marozeau et al., 2003), which can be difficult to measure separately. Only recently, WM has been studied for complex sounds, using the Euclidian 3D model of 'timbre space', where dimensions refer to attack (rise time from sound onset), spectral centroid (the center and mass in the frequency domain) and spectral flux (spectro-temporal characteristic: e.g. consonant-vowel transition), (Golubock & Janata, 2012). Such abstract auditory objects were presented in sequences composed of 2-6 items at variable delays, and the investigators concluded that WM capacity is limited to only 1-2 items (Golubock & Janata, 2012).

Li, Cowan & Saults (2013) obtained similar capacity estimates for WM for less complex sounds - pure tones (pitch). In contrast to the study of non-verbalizable sounds, analogous capacity estimates of 1-2 items have been found for verbal stimulus material: e.g. digits (Saults & Cowan, 2007; Fougnie & Marois, 2011). A range of further studies have used sequences of spoken digits (Morgan, Chambers & Morton, 1973), letters (Conrad, 1964; Conrad & Hull, 1964), syllables (Gupta et al., 2005) or words (Drenowski & Murdock, 1980; Haberlandt et al, 2005) as verbal memoranda. However, a number of other investigations by Cowan's team (Chen & Cowan, 2005; 2009; Cowan, Chen & Rouder, 2004) estimated verbal WM capacity to be greater, at 3-5 items or chunks.

how were stimuli presented again?

Not only memory researchers have used verbal material to investigate the nature of WM. Speech scientists have also been interested in this domain. A number of studies addressed WM on the phonemic level, where acoustic speech properties were carefully considered. While verbal WM studies have used letters, with consonants and vowels being part of the same stimulus set, speech scientists have studied both types of phonemes in separation. They showed that vowels and consonants are not only perceptually distinct, but may also be represented in separate memory stores, based on the observation that vowels decay faster than consonants (Pisoni, 1973; Pisoni & Tash, 1974). The findings of Shouten & van Hessen (1992) have also pointed to different types of memory stores for stop consonants and vowels in relation to their perceptual distinctiveness and underlying difference in acoustic cues: vowel perception is much less categorical than the perception of consonants.

In sum, previous work on auditory WM using psychophysics has mainly focussed on capacity limits for individual sound features. The findings have been taken to favour the existence of specialized sub-stores for a variety of features within a modality specific memory system.

Evidence from tactile psychophysics

As for the study of visual and auditory WM, tactile WM is often studied using the method of change detection (delayed match-to-sample task) for a variety of stimulus materials perceived via receptors located in the skin of the body surface, the hand or fingertips. In the study of memory for touch an important distinction is drawn between microgeometric stimulus properties (e.g. regarding fine textural detail: roughness), macrogeometric properties (an object's shape, length, area, curvature, size) and spatial properties (location of stimuli), (Gallace & Spence, 2009), where each class is processed in a different brain region (O'Sullivan, Roland & Kawashima, 1994; Roland, 1987; Roland, O'Sullivan & Kawashima, 1998).

which region?

Early studies on tactile WM using psychophysics showed that tactile information (e.g. macrogeometric properties: 3D shapes and location; as well as duration of stimulation) can be maintained for up to 80s, even when interfering material is presented in the delay period (Gilson & Baddeley, 1969; Sullivan & Turvey, 1972). In contrast, microgeometric properties (vibrational frequency) cannot be maintained for the same duration. Using a vibrotactile memory task it has been shown that discrimination accuracies were at ceiling for short delays, then decreased rapidly within 5 seconds of the delay period, before performance remained constant reaching a plateau (Sinclair & Burton, 1996). Similarly, monkeys can hold vibrotactile information over many seconds (Hernandez et al, 1997).

'Vibrotaction' is important for texture perception and identification: we perform a spatial time-frequency analysis of a given surface when we stroke a finger over it. A number of studies investigated the effect of interfering stimulus material presented in the delay period on memory performance (Bancroft &, Servos, 2011; Bancroft, Servos, Hockley, 2011; Bancroft, Hockley & Servos, 2011). As a mechanism the authors suggest feature overwriting: they found that the distractor frequency (when closer to the probe than to the target) partially overwrites the frequency information stored in WM. A wide range of studies on vibrotactile memory used a similar task – with and without interfering stimuli – in combination with a variety of neuro-imaging techniques (Romo et al, 1999 (see figure 1.1.1 C for the task used); Preuschhof et al, 2006; Sörös et al, 2007; Hegner et al, 2010; Spitzer, Wacker & Blankenburg, 2010; Spitzer & Blankenburg, 2011; Haegens et al., 2010). The underlying brain regions and networks will be discussed in section 1.3 while oscillations characterizing WM will be considered in 1.4.

closer in time as presented in a sequence?

A major limitation of all of the above studies on vibrotaction is that they only tested WM for a *single* frequency. Indeed, it has been suggested that this modality may only be able to store a representation of a single stimulus because the presentation of a distractor in the delay period of a change detection paradigm caused strong interference with the maintenance of a single vibration (Bancroft et al., 2011; Bancroft & Servos, 2011) and because the memory trace decays rapidly (Sinclair & Burton, 1996, Hernandez et al, 1997). However, in a subsequent study by the same authors (Bancroft et al., 2012) it has been shown that at least two vibrations can be maintained in WM at above chance performance.

In contrast to capacity limits for microgeometric tactile stimulus properties, capacity limits on macrogeometric properties have been more widely explored and are not as limited. While a number of studies presented stimuli at different locations on the body surface (Gallace et al., 2008; Auvray, Gallace & Spence, 2011; Alluisi, Morgan & Hawkes, 1965; Geldard & Sherrick, 1965), others (Auvray et al., 2011) presented stimuli to the fingertips. In the experiments reported by Gallace et al (2008) and Auvray et al (2011) participants had to memorise up to six somatosensory locations. Depending on the report procedure employed, they were able to remember 3 items (full report) or up to 6 items (partial-report).

difference between full and partial report?

Extensions and limitations of change detection

A major limitation of change detection, the task employed by studies reviewed in this section so far, is that responses made by subjects are binary. On a given trial they either detect a change or not. Therefore, the task cannot quantify at high granularity the *quality* of a memory representation. A number of studies on visual and auditory WM extended the traditional change detection task to include multiple stimulus levels (multiple sizes of change) on different trials, where the probe differs from the target along a number of values sampled from the dynamic range (near threshold, further away etc.).

look up definition of dynamic range

In the study by Bays and Husain (2008) subjects had to either report the direction of displacement of a visual object or make an orientation judgement about the rotation of a stimulus. Between encoding and test the stimulus could differ in magnitude of displacement or degrees of rotation. Items were remembered most precisely at the largest size of change between probe and target and at the lowest memory load. Overall, the task showed a continuous decline in memory performance with an increase in memory load and as a function of the size of change. In audition, varying the magnitude of change has also been employed as a response measure for pitch WM (Ries & DiGiovanni, 2007, 2009).

Although the traditional method of change detection can be extended to measure precision of recall in this way, this measure is still limited as stimulus values are sampled from a predefined range, where the possible size of change between target and probe stimulus defines potential stimulus properties. Hence, although researchers claim to study the ability to store parametric stimulus values, the extended method of change detection is also limited as the outcome is still a binary measure. What change detection cannot capture is that even when one fails to make a correct response, a representation of the target stimulus may still exist. It may simply be of a lower resolution than a stimulus reported correctly.

Additionally interfering stimuli may degrade the quality of a memory representation, which is difficult to capture. For example, it has been shown that we encode irrelevant information in WM even when explicitly instructed to ignore distractors (e.g., for vibrotactile information, Bancroft & Servos, 2011; Bancroft, Hockley & Servos, 2011). Hence, it would be more informative, instead of capturing whether a stimulus is represented or not, to obtain its mnemonic resolution using truly continuous measures.

1.2 Method of adjustment as a qualitative measure of WM

Due to these concerns about discrimination tasks in the visual domain, continuous measures of recall – WM precision – have been developed, designed to capture mnemonic resolution. These are continuous and analogue in nature rather than discrete and binary (Wilken & Ma, 2004; Bays, Catalo & Husain, 2009). In the study by Bays et al. participants were presented with a stimulus display at memory encoding consisting of multiple visual objects of different colours and at different spatial locations (2009). At test, one of the objects had lost its colour and participants had to reproduce the missing value by making a selection from a continuous colour space (colour wheel), (see figure 1.2 A). On each trial, the deviation between the actual target colour and the reported value was obtained as a measure of response error. Recall precision was calculated as 1/STD of overall response error.

Investigating the fidelity of memory representations with continuous measures of precision, it has been shown that visual WM may not be limited in capacity to a fixed number of items. Instead, it might best be considered as a limited resource, but without any limit on the number of items it can store. The WM resource can be flexibly allocated to objects defined by different features in the visual scene (Wilken & Ma, 2004; Bays et al., 2009; Fougnie, Asplund & Marois, 2010; Anderson, Vogel & Awh, 2011; Gorgoraptis et al., 2011; Zokaei et al., 2011; Salmela et al. 2012). More specifically, the way the resource is shared out across visual information is reflected in a monotonic decrease in recall precision with an increase in memory load (see figure 1.2B). A major goal of this thesis is to apply such continuous measures of memory precision to other sensory modalities: audition (see chapters 2 &3) and touch (see chapter 4).

Figure 1-2 *Precision of visual WM obtained with an adjustment task*

A) Adjustment task: subjects are presented with a sample array of multiple coloured squares (memory encoding). After a blank period, a test array was presented in which the location of a randomly selected sample item was highlighted (here: black frame). Subjects reported the remembered colour corresponding to this particular location by clicking on a colour wheel (Zhang and Luck, 2008; Bays, Catalo & Husain, 2009; Gorgoraptis et al., 2011).

B) Result: precision as the function of the number of items to be memorised at the stage of memory encoding. Precision is defined as the reciprocal of the standard deviation of the error in subject responses (zero indicates chance performance). Error bars indicate SEM. A power law was fitted to the results relating precision to the fraction of resources available per item (Bays, Catalo & Husain, 2009).

1.3 Representations in WM

In sensory brain regions, neurons respond selectively to preferred stimulus features, with information represented in topographical feature maps (Kaas, 1997; 2000). Such mapping has been demonstrated in the visual (retinotopy: Engel at al., 1994; Sereno, McDonald & Allman, 1994; Tootell et al., 1997; Baseler, Morland & Wandell, 1999; Dumoulin & Wandell, 2008); auditory (tonotopy: Formisano et al., 2003; Humphries, Liebenthal & Binder, 2010; Da Costa et al., 2011) and somatosensory system (somatotopy: Mountcastle, 1957; Killackey, Rhoades & Bennett-Clarke, 1995; Sanchez-Panchuelo et al., 2010). In contrast to perceptual representations formed in the presence of a stimulus, selective neural responses are still found in its absence. A prevailing view is that WM storage is accomplished via sustained neural activity, which can be regarded as a mnemonic code used to maintain representations of external stimuli in WM (Fuster & Alexander, 1971; Fuster, 1990). Evidence for stimulus specific delay activity has been associated with WM processes across the senses, which will be reviewed here, before discussing the representational format (e.g. features vs. objects etc.) stored in WM in more detail.

1.3.1 Evidence from neurophysiology in human and non-human primates

Sustained neural activity measured during the memory delay period using electrophysiological techniques was first identified by extracellular recordings in the non-human primate (Fuster & Alexander, 1971; Niki, 1974; Funahashi, Bruce & Goldman-Rakic, 1989). Many authors have argued that WM representations might be subserved by persistent spiking activity in the cortex (Funahashi et al., 1989; Miller, Erickson & Desimone, 1996).

In the visual system

Persistent delay activity has been found in the inferotemporal (IT) cortex in monkeys, a higher-level area part of the ventral visual stream (“what pathway”), involved in processing complex shapes (Ungerleider & Pasternak, 2004). IT neurons showed delay-period activity, when monkeys actively maintained visual objects defined by colour and shape (Fuster, 1990; Miller & Desimone, 1993; Miyashita & Chang, 1988; Chelazzi et al., 1998). Delay-period activity has also been found in area V4 of extrastriate visual cortex, a lower-level area of the ventral visual stream associated with the maintenance of colour and motion (Ferrera, Rudolph & Maunsell, 1994).

Delay-period activity has also been found in the dorsal visual stream (“where pathway”), e.g. in area 7a and lateral intraparietal cortex (LIP) associated with remembering spatial locations on the delayed saccade task (Constantinidis & Steinmetz, 1996) or for motion direction (Ferrera et al., 1994). While Ferrera et al. (1994) failed to find such sustained activity during a delay period in area MT (middle temporal area also know as V5), Bisley et al. (2004) have reported activity in this region. Over a delay period of 1.5s they found a short burst of firing, followed by suppression and final reactivation in firing before probe stimulus presentation, revealing the presence of reliable signals of directional motion.

Importantly, the study by Bisley et al. (2004) shows that an area usually involved in the perception of motion can also be active during memory maintenance of motion stimuli. Further support for such shared neural hardware across perception and memory has been found in early visual cortex (V1), (Super, Spekreijse & Lamme, 2001). The authors recorded multi-unit activity (MUA) in V1 during the delay period, when monkeys made a saccade to a remembered target location (spatial WM). MUA increased in a group of neurons whose receptive fields (RFs) covered the target presented during the delay, i.e. long after the target had been extinguished. Thus, this particular set of neurons in V1 holds a representation of the upcoming target in WM. The same neurons hold more noisy representations when the target is presented in a different location not covered by their RFs. In this case, however, the target is represented with greater fidelity by another group of neurons.

In the auditory system

Sustained delay-period activity has also been reported in the auditory system in both humans and non-human primates. Gottlieb, Vaadia and Abeles (1989) recorded single-unit activity in auditory cortex, where firing rates of many neurons reflected the sound frequency of the encoded pure tone when the sensory stimulus was no longer present. The involvement of auditory cortex in WM has been supported by a Magnetoencephalography (MEG) study in humans, which investigated memory decay for tone loudness in auditory cortex and showed that echoic memory directly reflects the decay of the physiological activation trace in this early auditory region (Lu, Williamson & Kaufmann, 1992). Their observation was based on the auditory ERP component N100, which decays exponentially with time. Another MEG study (combined with fMRI) revealed load-dependent effects of brain activity in secondary auditory cortex (A2), superior parietal cortex and frontal cortex (Grimault et al., 2010).

In the somatosensory system

As in the visual and auditory systems, neurons coding WM processes also exist in the somatosensory system. Findings in monkeys show that persistent delay activity is an intrinsic property of SI (primary somatosensory cortex) neurons, where a network subserves tactile WM for haptic objects (Wang et al., 2012). This concept is additionally supported by previous findings on WM for other tactile features: surface textures (Zhou & Fuster, 1996) and vibrations in the range of flutter (Salinas et al., 2000). Vibrotactile flutter is also represented in SII (secondary somatosensory cortex), (Salinas et al., 2000), medial premotor cortex (MPC) and PFC, (Romo et al., 1999; Romo & Salinas 2003; Preuschhof et al, 2006; Spitzer et al, 2010; Wang, Bodner & Zhou, 2013).

The role of PFC is special for memory maintenance, as neural delay period activity is systematically modulated only in this region, where mean responses typically increase monotonically with stimulus frequency. In other words, neural discharge rates vary systematically as a function of the encoded stimulus frequency (Romo et al., 1999) This specific neural code describes how representations of vibrotactile flutter held in WM vary with frequency. Hence, a continuous parameter (frequency) is represented in a parametric fashion (monotonic discharge). Additionally, it has been shown that similar variations in firing rate occur on the level of S1 during flutter stimulation, where responses can be described by a linear function (Salinas et al., 2000). Thus, the interpretation has been put forward that monotonic encoding of vibrational frequency recorded in PFC neurons are likely to originate from inputs in lower-level somatosensory areas (see: Romo and Salinas, 2003; for a review postulating a continuous downstream network between multiple sensory areas). Selective discharge during the memory delay period has also been found in area 5 of posterior parietal cortex (PPC), which has been implicated to play a role in WM for tactile object identity (Koch & Fuster, 1989).

1.3.2 Decoding WM in humans

A specific pattern of distributed neural delay-period activity reflects a neural code of a memory representation associated with a particular item held in WM. Using pattern recognition techniques from the field of machine learning with neuro-imaging data acquired with functional magnetic resonance imaging (fMRI), recent studies have shown that the identity of a stimulus held in visual WM can be decoded from delay period activity recorded in early visual regions (e.g., V1-V4, MT), (Xing et al., 2013; Harrison and Tong, 2009; Serences et al., 2009; Riggall and Postle, 2012). Multi-variate pattern analysis (MVPA) can be applied in the absence of sustained delay-period activity (derived from uni-variate data analysis) to identify which particular stimulus of a given set has been held in mind.

not a requirement

Some of the above studies used stimulus features, which are perceptually represented in early visual areas with a retinotopic organization, such as orientation (Harrison and Tong, 2009; Serences et al., 2009), stimulus contrast (Xing et al., 2013) and motion (Riggall and Postle, 2012). As WM representations could be successfully decoded from early visual areas including V1-V4 and MT, the same neural circuitry mediating visual perception seems to also be involved in memory maintenance of stimuli. However, one study showed that even non-retinotopically represented visual features could still be decoded during memory maintenance from early visual areas (colour: Serences et al., 2009). Finally, Christophel, Hebart and Haynes (2012) decoded complex stimuli (defined by a multitude of features: shape, color, and spatial orientation) successfully from posterior parietal cortex, but could not show an involvement of frontal or early visual regions.

In contrast to visual WM, a single decoding study has attempted to predict the identity of pure tone in auditory memory (Linke, Vicente-Grabovetsky & Cusack, 2011). The authors could classify whether the frequency content was represented in auditory cortex during the delay period, but there were a number of problems with this investigation, which will be discussed in more detail in chapter 7. The study I

should be 6

present in chapter 7 aims to decode frequency-specific patterns (high vs. low pure tones) from auditory memory. To the best of my knowledge, no decoding study exists to date to identify the contents of somatosensory WM.

1.3.3 Towards a representational format: features vs. objects

Although these more recent findings suggest that WM representations are coded as distributed patterns of activity in the visual system and most of the evidence suggests that a pattern codes for individual features (Harrison and Tong, 2009; Xing et al, 2013), there may also be a code for stimuli represented as objects (Christophel et al., 2012). The mechanisms underlying binding of features into higher-level representations (objects) remains to be understood in light of the processing hierarchy of a given sensory system. Considering the special role of PFC in WM, many PFC neurons show activity reflecting stimulus features of a particular sensory modality (Romo et al, 1999, Miller et al, 1996; Fuster, 1999; Rainer, Asaad & Miller, 1998; Funahashi, Bruce & Goldman-Rakic, 1989; Constantinidis, Franowicz & Goldman-Rakic, 2001). There have also been some claims that PFC may be functionally segregated by sensory modality and stimulus features, with PFC neurons also integrating information across sensory modalities (Fuster, Bodner & Kruger, 2000) as well as features within a modality (Rao, Rainer & Miller, 1997). The integration of features (binding) is considered to result in the formation of more complex object representations or higher-level percepts (Robertson, 2003).

According to feature-integration theory (FIT) features of e.g. a coloured square are represented in individual feature maps (e.g. one map for colour and another map for its orientation in space). And only at a later processing stage, features from both dimensions are integrated (bound) to form coherent wholes (objects), (Treisman & Gelade, 1980). Although FIT was proposed for perception, an analogous problem arises for WM processes. Does WM hold individual features or bound representations? The representational format has implications on memory capacity: how much information can be retained in WM?

In vision, it has been argued on the basis of binding errors observed on a change detection task that features in different dimensions are maintained in distinctly different memory stores (Wheeler & Treisman, 2002). In contrast, Luck and Vogel (1997) showed that visual WM may be limited by the number of objects, but not by the number of features per object. However, Oberauer & Eichenberger (2013) have recently demonstrated that WM capacity is not only limited by the number of objects, but also by the number of features per object and by the mnemonic resolution strength (precision) at which they are represented. Further studies in vision suggests that there is no upper object limit, as processes in WM are best described as a resource shared out across all objects and their features (Wilken & Ma, 2004; Bays, Wu & Husain, 2010; Bays, Husain & Ma, 2014). Bays et al. (2010) reported that there were increased binding errors at high memory loads due to independent response error distributions for each feature, suggesting that features are maintained in separate WM stores.

The study presented in chapter 5 aims to clarify which of the above principles found in vision can be applied to auditory WM. How is auditory information held in WM represented? Is it as individual features or bound wholes (objects)? With regards to the somatosensory system, little is known about the format of representations. However, one study suggests that short-term representations are maintained in a finite set of feature units, such as prefrontal neurons (Bancroft, Hockley & Servos, 2011).

1.4 Brain regions, pathways and networks associated with WM storage

The brain regions acting as WM stores for modality specific (visual, auditory and tactile) and supra-modal information together with current knowledge of processing pathways and networks in humans will be reviewed here. The evidence is based on results from functional neuro-imaging (fMRI and positron emission tomography (PET)) as well as lesion studies including virtual lesions induced by local cooling or transcranial magnetic stimulation (TMS).

As has already been pointed out in the previous section on ‘representations in WM’ (see 1.3), regions representing information held in WM include lower level early sensory areas and higher-level frontal and parietal areas. A working hypothesis suggested by Linden (2007) is that sensory cortices have the highest degree of modality specificity, where higher-level frontal and parietal areas are only partly modality specific. Frontal and parietal areas are also supra-modal and act as a hub where modal representations converge across sensory specific processing streams (e.g. visual and auditory). Interestingly, along the lines of sensory specific processing hierarchies found in the brain Linden also suggests that the involvement of an area (from early sensory to higher level) varies with the degree of abstraction or complexity of a stimulus held in mind. For example, the more basic the sensory feature, the larger the degree of involvement of early sensory regions compared to regions further at the processing hierarchy (frontal and parietal).

1.4.1 Visual information storage

In order to study the neural correlates associated with the maintenance of basic visual features (spatial frequency) in WM, Greenlee, Magnussen & Reinvang (2000) used fMRI with a delayed discrimination task and found an activity increase in PFC, PPC and occipital cortex. Another fMRI study using a similar design to investigate WM for visual orientation found increased activation in visual cortex during the delay period (Pessoa et al., 2002). From visual cortex, it is believed that information travels further up along either one of two pathways depending on the type of information perceived. It has been proposed that the ventral stream or ‘what’ pathway from primary visual cortex to inferior temporal regions is associated with the detailed analysis of visual objects defined by colour and shape, whereas the analysis of an object’s spatial location, orientation or movement is performed along the ventral stream or ‘where’ pathway to posterior parietal cortex (Mishkin & Ungeleider, 1982).

Based on connectivity patterns in monkey PFC, it has been argued that this division continues to anterior portions of the brain, although such a proposal has been contested. PFC can be subdivided into dorso-lateral pre-frontal cortex (DLPFC) and ventro-lateral pre-frontal cortex (VLPFC), (Goldman-Rakic, 1996). It has been shown that DLPFC is densely connected with visuo-spatial processing areas and the posterior parietal cortex (PPC), while VLPFC is more closely linked to inferior temporal areas processing shape and identity of visual objects (Miller, 2000). Such functional subdivisions of visual (form and identity) and spatial stores in PFC based on dissociations are supported by some findings in patients (Della Salla et al., 1999) and data from selective interference tasks in healthy subjects (Tresch, Sinnamon & Seamon, 1993). However, some recording studies from monkey PFC have demonstrated that in addition to 'what' and 'where' related activity, many neurons within the same region are tuned to both the object and its location (Rao, Rainer & Miller, 1997).

should be ventral stream?

whole 2 paragraphs are somewhat confusing

A local cooling study of the inferior-temporal (IT) cortex in monkeys (Fuster, Bauer & Jervey, 1985) and several lesion studies (Frey & Petrides, 2000; Walsh et al., 2000) confirm that this region is involved in visual WM for coloured objects as part of the dorsal stream. Further, a double dissociation of virtually induced lesions using TMS has been reported. Oliveri et al. (2001) delivered single-pulse TMS during a short delay period to bilateral parietal regions, causing selective interference with a spatial WM task. Additionally, stimulation to bilateral temporal regions interfered selectively with memory performance on an object task, where object identity was defined by abstract patterns. Interestingly, stimulation of the anterior DLPFC led to disruption of memory performance on both tasks.

Although Mottahgy et al. (2002) used stimulus material different from Oliveri et al. (2001), they could reproduce their results with repetitive TMS delivered to DLPFC during a spatial and face identification WM task. Performance was impaired on both tasks, while the authors caused selective WM impairments after stimulation of VLPFC with WM for faces, in contrast to selective interference after stimulation to dorso-medial pre-frontal cortex (DMPFC) on a spatial WM task. Therefore, there is some evidence that PFC appears on the one hand to be segregated into content specific stores (DMPFC: object identity; DMPFC: spatial WM). On the other hand some findings also suggest that PFC contains a region which does not display content specificity (DLPFC: general visual representations or attention-based processing).

1.4.2 Auditory information storage

Several neuro-imaging studies investigated WM for pitch using tone-sequence comparison tasks with PET (Zatorre, Evans & Meyer, 1994) and fMRI (Gaab et al., 2003). Both studies found memory-related activity in associative auditory cortex, lateralized to the right superior temporal lobe, while pitch judgements were associated with increased activity in the right frontal lobe. Thus, studies on the basic auditory feature of pitch implicate a right-lateralised system involving auditory cortex and infero-lateral frontal cortex. Another study on pitch memory shows that activity in auditory cortex is highest early in the delay period, whereas regions in supra-marginal gyrus (SMG) and parts of the cerebellum are activated later on. Thus, each region may play a different role in processing information in memory (Mathiak et al., 2004).

With respect to different roles for different stages of memory-related processes, a PET study by Griffiths et al. (1998) demonstrated findings consistent with the view that while posterior auditory cortex is involved in memory encoding, more anterior regions are involved in retrieval mechanisms. Activity in auditory cortex and inferior frontal cortex (IFC) is associated with WM maintenance (Griffiths et al., 1998; Overath et al., 2007). Additionally, Overath et al. (2007) showed that in the posterior encoding area of auditory cortex, activity increases are related to the complexity of pitch sequences.

Similarly to the visual system, information-processing streams have also been identified in the auditory modality (Rauschecker & Tian, 2000). The dorsal or ‘where’ pathway serves auditory spatial information processing, leading from auditory cortex to inferior frontal cortex via the posterior temporal and parietal lobe. The ventral or ‘what’ auditory object (involved in the identification of a sound) pathway extends from auditory cortex to IFC via the rostral part of the superior temporal lobe. Support for this view of dual streams of information processing has come from lesion studies in humans, demonstrating a double dissociation between abilities associated with locating sounds in space and identifying them (Adriani et al, 2003; Clarke & Thiran, 2004; Clarke et al., 2002), as well as by neuro-imaging studies in healthy controls (Maeder et al., 2001; Warren & Griffiths, 2003).

Imaging has additionally revealed that both pathways play a role in processing information in WM (Belger et al., 1998; Courtney et al., 1996; Ventre-Dominey et al., 2005). A meta-analysis shows that the majority of neuro-imaging studies using spatial WM tasks were associated with activity in posterior parietal regions, while object tasks have been associated with activity in inferior temporal regions (Wager & Smith, 2003). The activations found are in line with those in both visual processing streams. In addition, activations in DLPFC and VLPFC have been reported using auditory WM tasks (Lehnert & Zimmer, 2008), where DLPFC processes spatial and VLPFC is involved in information about object identity (Courtney et al., 1996; Sala, Raemae & Courtney, 2003; Ventre-Dominey et al., 2005). However, as in the visual system, alternatively this functional difference could also be due to executive processes, which differ across mnemonic tasks (D'Esposito et al., 1998; Wager & Smith, 2003).

1.4.3 Somatosensory information storage

A number of neuro-imaging studies investigated the ability to hold tactile stimuli in mind (Burton & Sinclair, 2000). PET studies found increased activity in S2 and DLPFC during a WM task for vibrations (Klingberg, Kawashima & Roland, 1996) or shapes (Bonda, Pertrides & Evans, 1996). An event-related fMRI study identified a fronto-parietal circuit (PFC, frontal operculum and anterior parietal cortex) to be associated with delay activity on WM discrimination task for tactile objects (Stoeckel et al., 2003). Activations for tasks involving the encoding of tactile objects and visual information seem to overlap, resulting in supra-modal information storage as suggested by Pasternak & Greenle, 2005.

A more recent fMRI study aimed to identify the neural correlates of different stages characterizing vibrotactile WM (Preuschhof et al, 2006). The investigators reported that encoding was associated with activity in S1 and ventral pre-motor cortex (PM); maintenance or delay-period activity with activity in PM, VLPFC and intra parietal lobule (IPL); and making a response decision on a discrimination task with activity in S1, S2, PM, lateral PFC and IPL. They suggest that the activation network found in humans differs from activations associated with a similar task in monkeys. A subsequent fMRI study aimed to identify brain regions involved in tactile attention and WM for vibrotactile flutter (Sörös et al, 2007). These authors found that attentional and WM networks are partly overlapping. Finally, Hegner et al. (2010) compared WM storage across two types of stimulation: tactile patterns vs. vibrotactile flutter. They found that cortical processes differ among types of information, as somatosensory areas including the right intra parietal sulcus (IPS), SMG and parietal operculum (PO) were more activated in the pattern compared to the frequency task.

TMS has also been used to understand the neural basis of WM for vibrations, with stimulation over area S1 at two different time points during a delay period of 1.5 s duration (Harris et al., 2002). When applied early in the delay period to the contralateral side, discrimination thresholds increased, whereas when applied later it had no effect. As supported by monkey physiology studies (Zhou and Fuster, 1996, 2000; Super et al., 2001) showing that neuronal activity can be maintained during the delay period between two stimuli, the authors concluded that S1 stores vibrations early, in a critical time window of 300-600 s after stimulus encoding. Overall, there is evidence that the same neural networks that are involved in initial sensory processing of tactile information might also be involved in representing the same information in WM.

1.4.4 Supra-modal information storage

The processing and storage of sensory stimuli in a single domain (e.g. tactile) cannot always be separated from the processing of other types of information, especially where information may not be purely unimodal in nature (Gallace & Spence, 2009). For example, Gallace & Spence (2008) suggest that spatial features of tactile objects are not purely tactile in nature, as they may have a visual imagery component to it when explored by tactile palpation. Therefore, different unimodal areas interact at different levels along the hierarchy of information processing with amodal processing components to store sensory information. Such a complex neuro-cognitive architecture suggests that unimodal sensory inputs, associated with different sensory processing streams, ultimately converge to form supra-modal representations, potentially at the level of PFC (e.g., see Linden, 2007).

The fMRI study by Ricciardi et al. (2006) compared a tactile with a visual WM task and found that similar fronto-parietal networks were involved in spatial WM tasks in both sensory modalities. The network includes the PPC, DLPFC and anterior cingulate cortex (ACC). The authors concluded that such shared neural substrates are involved in the formation and storage of supra-modal representation in WM. A recent tri-modal (visual, auditory and tactile) study found evidence for supra-modal WM representations based on oscillatory EEG activity across discrimination tasks (Spitzer & Blankenburg, 2012). Related studies will be reviewed in the next section.

1.5 Oscillations characterizing WM

WM research is concerned with how groups of neurons represent and sustain sensory information in the absence of external sensory inputs. A possibility is that neural oscillations as measured with electroencephalography (EEG), electrocorticography (ECOG) and MEG reflect neural codes describing information held in WM (Roux & Uhlhaas, 2014). And in particular such rhythmic neural activity at different frequencies is associated with different WM stages (e.g. gamma oscillations and memory maintenance: Tallon-Baudry et al., 1998; Roux et al, 2012; Medendorp et al., 2007), memory load-specific modulations (Howard et al., 2003; Roux et al., 2012; Palva et al., 2010; 2011; van Vugt et al., 2010; Linden et al., 2012) and protection of task relevant information (held in WM) from irrelevant distractors (Herrmann, Munk & Engel, 2004; Kaiser et al., 2007; Sauseng et al., 2009; Haegens et al., 2010). Some studies have also presented evidence that rhythmic activity can also be entrained in order to tag WM contents (Silberstein, 1995).

1.5.1 WM delay activity by frequency-band

Enhanced amplitude and synchrony of oscillatory activity and specific modulations associated with different memory loads have been found in the theta (4-7Hz), alpha (8-13Hz) and gamma (30-200Hz) frequency ranges (Roux & Uhlhaas, 2014). Tallon-Baudry et al. (1998) first established a link between oscillations in the gamma range (here: 24-60Hz) and maintenance of information in WM, as the authors found a sustained increase in gamma oscillations using EEG with a visual delayed-match to sample task. Similar findings have been reported for visual WM for spatial stimulus features (Roux et al, 2012; Medendorp et al., 2007) as well as auditory (spatial sounds: Kaiser, Heidegger & Lutzenberger, 2008; Lutzenberger et al., 2002; auditory patterns: Kaiser et al, 2003) and tactile stimuli (mechanical vibrations: Haegens et al., 2010). In addition a number of studies indicate a parametric relationship between WM load (number of items held in mind) and the amplitude of gamma-band oscillations (Roux et al., 2012; Howard et al., 2003; van Vugt et al., 2010; Palva et al., 2010; 2011).

Parietal and pre-frontal cortices represent key regions in the WM network, to which gamma-band activity has been localized showing WM load dependent modulations (Roux et al., 2012; Palva et al., 2010; 2011; Linden et al., 2012). Furthermore, ECOG data shows that gamma-band oscillations are modulated by WM load in parietal and frontal cortex (Howard et al., 2003) and the hippocampus (van Vugt et al., 2010). A further functional relationship has been established between particular gamma frequencies (Kaiser et al., 2008) as well as single-trial fluctuations in gamma activity (Roux et al., 2012, Polania, Paulus & Nitsche, 2012) and the maintenance of distinct items in WM.

Alpha oscillations recorded in the delay-period may not be directly involved in any of the memory stages (encoding, maintenance and retrieval), but may play an active role in protecting task-relevant information from distracting task-irrelevant information. Alpha-band oscillations have been demonstrated during WM maintenance in several sensory modalities (visual: Herrmann et al., 2004; auditory: Kaiser et al., 2007; tactile: Haegens et al., 2010), with WM load increases being associated with increased alpha amplitude reported by Leiberg, Lutzenberger and Kaiser (2006), although Gevins et al. (1997) found otherwise. However, it seems difficult to disentangle, whether load dependent effects, reflect memory or attentional processes.

The study by Sauseng at al. (2009) was designed to measure oscillations associated with attention, separate from WM processes. Task-relevant and task-irrelevant items were indicated by a cue, which pointed either left or right, making information in one of the two hemi-fields relevant, while the non-indicated side became irrelevant. The number of irrelevant items presented in the task-irrelevant hemi-field modulated alpha-band activity. Additionally, repetitive TMS at alpha frequency (10Hz) over the task-irrelevant hemisphere caused an enhancement of WM capacity.

Finally, theta-band oscillations are less well understood in the context of WM, although it has been proposed that they characterize capacity limits (Jensen & Lisman, 1998; Jensen & Tesche, 2002). Theta-band activity has been recorded in the delay period and associated with WM maintenance. An ECOG study demonstrated this activity to be present in several cortical regions; it systematically varied with WM load (Raghavachari et al., 2001). An MEG study reported a parametric increase in frontal theta activity with WM load (Jensen & Tesche, 2002). A more complex relationship for theta activity has been established in a network of hippocampal, frontal and parietal regions, linked to WM capacity (Moran et al., 2010). Finally, a recent meta-analysis concluded that theta activity occurs predominantly in tasks where multiple items are encoded sequentially. In contrast, alpha oscillations are more likely to be observed in tasks where visual or spatial information is presented simultaneously (Roux & Uhlhaas, 2014).

1.5.2 Phase synchronization in WM

Phase synchronization refers to the synchronization of oscillatory patterns of oscillatory phases between different brain regions. It has been proposed that such means of communication between brain areas supports processes in WM (Fell & Axmacher, 2011). A number of EEG studies report that theta phase synchronization between PFC and temporal lobe occurs during memory encoding and retrieval (Sauseng et al., 2004), while it persists during the maintenance period (Sarnthein et al., 1998; Serrien, Pogosyan & Brown, 2004). Additionally, several studies describe a link between theta phase synchronization and WM load, reflected in increased theta coherence between frontal and temporal-parietal areas when WM load was increased (Payne & Kounios, 2009). The amount of theta coherence also predicts memory capacity (Kopp, Schhroeger & Lipka, 2006).

Not only theta, but also oscillations in the beta and gamma range show phase synchronization in WM. Coherence between frontal and parietal regions is enhanced in these frequency ranges during WM maintenance (Lutzenberger et al., 2002; Babiloni et al., 2004). Results from an ECOG study revealed a sustained enhancement of beta phase synchronization among different areas in visual cortex, when holding complex shapes in mind. Further inter-regional associations mediated by beta phase synchronization have been found between the fusiform face area (FFA) and the medial temporal lobe, where coherence was increased with WM load. Beta phase synchronization in visual cortex was also accompanied by gamma phase synchronization in MTL (Axmacher et al., 2008). Phase synchronization within the same or across frequency ranges may mediate the communication between different brain regions, where each region forms a local assembly holding stimulus representations.

1.5.3 Cross-frequency coupling as codes for WM

Some investigators have proposed that cross-frequency coupling (CFC) occurs between two frequency ranges as a means of neural communication between neuronal ensembles. This idea has found support by many studies, showing that a slower rhythm occurs simultaneously with a faster one during WM (Sohal et al., 2009; Tallon-Baudry et al., 1998, Kaiser et al., 2008). There are different forms of CFC: according to phase-amplitude coupling (PAC) the amplitude of a higher frequency is modulated by the phase of a slower rhythm. Other means of communication are amplitude to amplitude coupling and phase locking (Roberts et al., 2013), described by power and phase co-fluctuations between multiple frequencies.

Theta as well as alpha oscillations are often accompanied by gamma frequencies (Haegens et al., 2010, Palva et al., 2011; Jokisch & Jensen, 2007). One way for neural oscillations to interact is to form nested assemblies in a hierarchical system (Lisman & Jensen, 2013). Using ECOG, Axmacher et al. (2010) reported that the maintenance of multiple items in WM is accompanied by load dependent theta-gamma coupling in the hippocampus. This is in line with EEG findings by Holz et al. (2010) on increased theta-gamma coupling on a visual delay match-to sample task. Both studies assume that sequentially presented items are represented in cycles of gamma oscillations coordinated through an underlying theta rhythm. In this respect, the number of gamma cycles *nested within a cycle of theta* is what limits memory capacity (Axmacher et al., 2010). Additionally, gamma-band oscillations are believed to represent a generic mechanism for the representation of individual items in WM independent of sensory modality and content. To the best of my knowledge, the potential role of alpha-gamma coupling in WM has not been investigated yet.

1.5.4 Frequency entrainment: tagging WM contents

Rhythmic neuronal activity can also be evoked by an external stimulus which appears after stimulation and is phase-locked to it (Hermann, 2001). For example, neurons in visual cortex respond to visual flicker at the frequency of the flickering light. In this form of sensory entrainment, such visual stimuli presented at a given frequency entrain periodic neural responses at the same frequency in visual cortex (Rager & Singer, 1998; Hermann, 2001; Williams et al, 2004), known as steady-state visual evoked potentials (SSVEP) as measured with EEG (e.g. Silberstein, 1995). In the context of WM, the SSVEP is used as a steady-state “topographical probe” (SSTP), (Silberstein et al., 1995); it is used as a frequency tag combined with a visual memory task (Silberstein et al., 2001; Perlstein et al., 2003).

The EEG signal propagation associated with the task can be measured indirectly, when comparing the SSVEP frequency propagation observed with the activity associated with a control task (Vialette et al., 2010). Fundamental power differences between tasks may reflect differential engagement of neural activity in brain regions involved WM (Silberstein et al., 2001; Perlstein et al., 2003). For example, as SSTPs propagate through different areas of the brain, the memory load can be estimated from the decrease of SSVEP amplitude.

in
(missing)

Using a visual delay match to sample task, Perlstein et al. (2003) measured SSVEP responses to a 10Hz flicker (alpha-range). The authors showed that SSVEPs are sensitive to the active maintenance of information held in WM (reproducing results by Silberstein et al. (2001), and that WM-related activity can be localized to PFC. Furthermore, task-related SSVEP in this region correlated with WM performance and memory load, associated with selective SSVEP amplitude modulations. A 13Hz SSVEP served as a frequency tag in a WM study by Ellis, Silberstein and Nathan (2006), in which subjects had to memorise the spatial location of dots. In this experiment, responses in frontal areas decreased with an increase in task difficulty (memory load manipulated as increased n-back positions). Thus, SSVEPs may index WM processes associated with maintenance of visual items.

1.6 Multi-sensory models of WM

At present, the dominant model of WM incorporating at least the auditory and visual modalities is the multicomponent model, originally proposed by Baddeley & Hitch (1974), (see Figure 1-3 which shows the original version of their model). This conceptual framework has been very influential, providing a fundamental basis for the interpretation of results from various studies (behavioural, neuro-physiological, neuro-psychological and neuro-imaging), which appear to support the idea that memory is the combined function of multiple modules. In contrast, another influential model describing WM is the embedded-processes model of WM by Cowan, which does not assign a separate module to store a particular type of modality specific sensory information (Miyake & Shah, 1999).

1.6.1 The multi-component model of WM

The multi-component or modal model encompasses two working memory stores (phonological loop and visuo-spatial sketchpad) and their fundamental properties are their limited capacity and their functional independence. Historically, both assumptions are based on studies using dual task paradigms in healthy subjects and on studies in patients with focal brain lesions leading to selective memory impairments. This line of research has led to the distinction between a store for visual information (visuo-spatial sketchpad), separate from a store for auditory information (phonological loop).

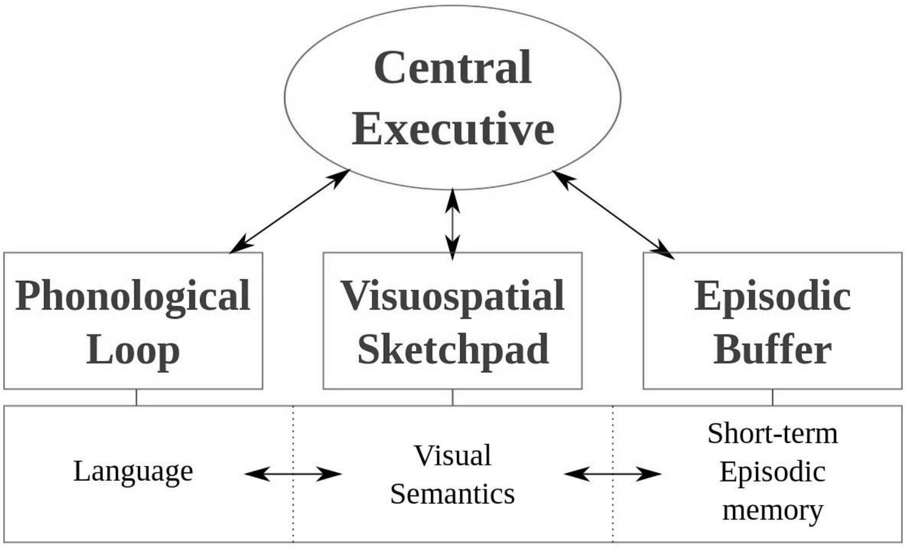


Figure 1-3 *Multicomponent model of working memory*

The multicomponent model of working memory proposed by Baddeley & Hitch (1974) is composed of two separate memory stores. The visuo-spatial sketchpad stores visual information (e.g. about shapes, colour, location, movement etc.) and the phonological loop stores auditory information (e.g. verbal or non-verbal material). Auditory information can be refreshed by phonological rehearsal to prevent decay of information stored in the visual sketchpad. The central executive represents another functional subunit, which is associated with processes of selective attention, inhibition control, regulation, coordination, switching between tasks, retrieval strategies (Baddeley, 2004). The episodic buffer was added to the model later on (Baddeley, 2000), which integrates information across short-term stores to form perceptual whole and access information stored in long-term memory.

Adding a degree of complexity to the multicomponent model, the visual store has been further subdivided to process information about visual objects (linked to the 'what' pathway (Goodale & Milner, 1992; Mishkin & Ungeleider, 1982) separately from spatial features of visual objects (associated with the 'where'- pathway (Goodale & Milner, 1992; Goodale & Humphrey, 1998), (see section 1.4 for the neural correlates associated with both pathways). In an analogous fashion, the auditory store has been further subdivided; where one subdivision processes non-spatial auditory information and the other subdivision processes spatial sounds. A specialization into domain specific 'what/where' pathways has also been proposed for audition (Rauschecker & Tian, 2000), which found further support by lesion studies (Adriani et al., 2003; Clarke & Thiran, 2004; Clarke et al., 2002).

In addition to different memory stores the central executive represents another functional subunit, which is associated with processes of selective attention, inhibition control, regulation, coordination, switching between tasks, retrieval strategies etc. (Baddeley, 2004). Thus, there is an integral relationship between a control system (central executive), which accomplishes the deposition and removal of information from short-term stores and the storage buffers themselves. This close level of interaction between short-term stores via such control processes enables effective mental information processing (WM book chapter ref). While the original multi-component model was only composed of three components reviewed so far (central executive, visuo-spatial sketch pad and phonological loop), a further component was added in a revised version of this model - the episodic buffer (Baddeley, 2000). This component links information across domains to form integrated wholes of different types of sensory information (i.e. visuo-spatial, verbal and time-sequencing such as the order in which information was presented). The episodic buffer also has links to long-term memory and semantic meaning (Baddeley, 2000).

Although the scope at which Baddeley's model is currently considered as the leading brain mechanism is enormous, there is also a wide body of evidence contesting it. Many studies, which directly compare auditory and visual memory processes, have shown that instead of regarding working memory as separated across the senses, it is better described as a single amodal resource. This claim is based on the repeated finding that visual information held in working memory from interferes with auditory information storage and vice versa (Brooks, 1968; Sperling & Speelman, 1970; Baddeley & Lieberman, 1980; Cowan, 1995; 2001; Nairne, 1990; Lehnert & Zimmer 2006; Smith & Jonides, 1997; Morey & Cowan, 2004, 2005; Saults & Cowan, 2007). In contrast, others claim that working memory tasks draw on at least partially but not fully dissociable storage systems (Fougnie & Marois, 2011). Thus, memory stores for vision and audition may overlap, contradicting an entire dissociation between stores across the senses.

1.6.2 The embedded-processes model of working memory

In contrast to the multicomponent model, the embedded-processes model (Cowan, 1995; Miyake & Shah, 1999) puts less emphasis on the separation between processes associated with individual sensory modalities. Instead, WM represents the cognitive faculty in which information is kept active in order to solve complex tasks. Information held in WM is accessed via a hierarchically organized system, which includes the focus of attention. Attention is needed in order to actively hold traces in WM, where the faculty of WM is embedded in long-term memory. Traces held in a temporarily active state represent information stored in WM. The focus of attention is capacity limited, but can hold any type of sensory information (Miyake & Shah, 1999).

The embedded-processes model suggests that visual and auditory information are held in a common store, where information from each sensory modality overlaps entirely (Cowan, 2001). This idea has been tested directly in a number of dual-task studies, which aimed to establish “trading relations” (interference) between visual and auditory modalities. In the study by Saults & Cowan (2007), subjects had to memorise visual (coloured squares) and auditory information (spoken digits) in combination or in separation (unimodal control). If remembering both types of information is based on a central capacity, the total number of items that can be remembered together (visual and auditory information in combination) should be equal to the number of items that can be remembered from either modality alone. Dual-task costs occurred for both modalities and the number of visual and auditory items combined did not exceed unimodal capacities. Although this result was not achieved in all individual experiments, the authors concluded that WM is limited by a central capacity (focus of attention): the simultaneous storage of information from both modalities, since visual and auditory information compete to be represented by a limited cognitive resource.

A number of multi-sensory studies using dual tasks (audio-visual, but not in combination with tactile information) found results supporting the concept of a common WM store or resource across the senses for visuo-spatial and verbal information (Morey & Cowan, 2004; 2005; Baddeley & Lieberman, 1986; Brooks, 1968; Nairne, 1990; Cowan's 1995; 2001; Sperling & Speelman, 1970; Lehnert & Zimmer 2006; Smith & Jonides, 1997). In contrast to the finding of conflict between visual and auditory tasks, no such interference was found in a number of studies using dual tasks for visuo-spatial and verbal information (Cocchini et al., 2002; Scholl & Xu, 2001; Luck & Vogel, 1997; Smith & Jonides, 1997), therefore supporting the separation of stores hypothesis associated with the multi-component model.

1.7 Models explaining capacity limits of WM

As discussed in the previous section memory stores, whether considered to be a single entity shared across sensory modalities or multiple entities separated by modality, have capacity limits. This section will focus on models, which have been designed to specifically explain such limits. In vision, there is a large controversy among theories accounting for capacity limits of working memory. The *slot model* assumes that working memory capacity is limited to a fixed number of items (e.g. visual objects; see Figure 1-4A), (Luck & Vogel, 1997). However, capacity limits may not be fixed, but dependent on the type of information stored and memory task employed (Alvarez & Cavanagh, 2004; Linden, 2007; Brady, Konkle & Alvarez, 2011). Moreover, the *resource model* of WM (see Figure 1-4B) proposes that we have one memory resource, which can be shared out across an unlimited amount of information, but at a significant cost in the fidelity with which we store individual items (Wilken & Ma, 2004; Bays & Husain, 2008; Bays et al, 2009). In contrast to this *continuous* resource model (see Figure 1-4B), a further *discrete* or *quantized* resource model (see Figure 1-4C) has been proposed, where every slot is involved in information storage at every memory load, (Zhang & Luck, 2008).

Each model makes a number of predictions about how one performs on a given memory task, where for example one has to memorize a varying amount of information such as visual objects, which differ in colour or in spatial orientation.

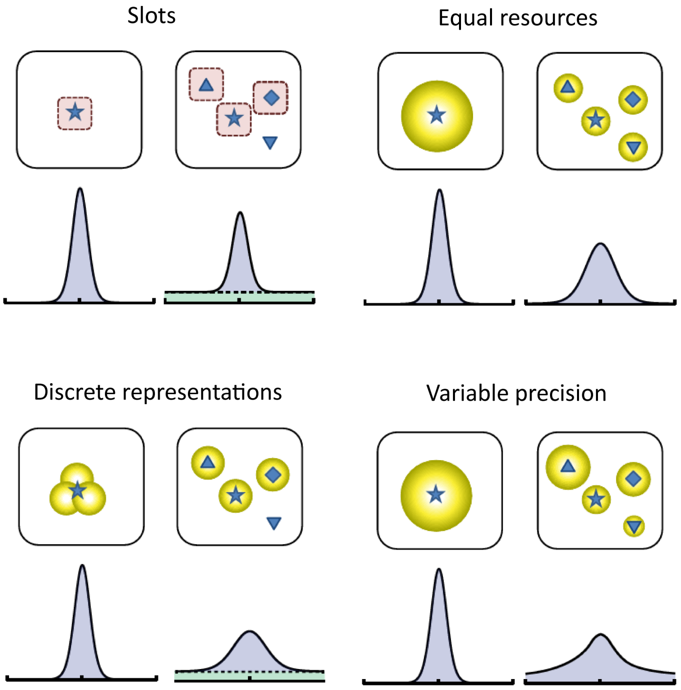


Figure 1-4 *Models explaining capacity limits in visual working memory*

(A) According to the slot model of WM each visual item is stored in one a single slots. The number of slots available (here: 3) limits WM capacity, 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 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) 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). (C) According to discrete representation models, the WM medium is divided into a fixed number of units (in a similar manner to the slot model). In contrast to the slot model, these slots are shared out across items. For a memory load of a single item (see left panel) the units combine to produce a high-resolution memory of a given item. In contrast, at higher memory loads (above the number of units available; see right panel), all items are either represented by all units in combination or none at all, predicting responses made to the target item to be composed of a mixture of low-resolution recall and random guesses. This distribution differs from is different from those in A and B. (D) Variable precision models predict that WM precision varies from trial to trial and item to item. Variability centred at the mean of the actual value of an item to be recalled decreases with an increase in memory load as a result of a limited resource. The model suggests that recall errors are composed of an infinite mixture of distributions (assumed normal) of different widths. Variable precision may stem from variability in resource allocation dependent on bottom-up factors (adopted from Ma, Husain & Bays, 2014).

According to the original *slot model*, (see Figure 1-4A) one can store 3-4 items perfectly at a high resolution (Luck & Vogel, 1997). However, as soon as the capacity limit is reached, no information about further items can be stored and the performance on those items can be explained by chance.

The *resource model* (see Figure 1-4B) makes fundamentally different predictions about performance on the same task, assuming that our memory capacity is not limited to a fixed number of items. According to this model, WM is indeed a limited resource but without an item limit. It can be divided across items, allows storage of a potentially unlimited number of items, but with worsening resolution as the load increases. Crucially, the resource model predicts that visual WM performance remains above chance, even when the slot model's capacity limit of 3-4 objects is exceeded. The resource model has been shown to account for a number of different features in vision, such as colour, orientation and motion (Bays & Husain, 2008; Bays et al, 2009; Gorgoraptis et al., 2011; Zokaei et al., 2011).

Recently, the resource model of working memory has received a lot of interest. This has led to a debate around how the memory resource is essentially shared across items, aiming to explain what proportionate amount of this medium is allocated to each individual item. While the *equal precision model* predicts that the resource is allocated evenly across all items (see Figure 1-4B), (Bays & Husain, 2008), the *variable precision model* (see Figure 1-4D) predicts that the amount of resource an item receives varies randomly across items, trials and memory loads (see Figure 1-4 (Van den Berg et al., 2012; Mazyar, van den Berg & Ma, 2012)). The major goal of the studies presented in chapter 2, 3 and 4 is to test whether the resource model can be applied to sensory modalities other than vision, where no direct distinction was made between the equal and variable precision account.

As opposed to different types of *continuous* resource models, the *discrete* resource model (see Figure 1-4C) predicts that the representational medium (each slot) is shared out across all items presented at each memory load (Zhang & Luck, 2008; Ma, Husain & Bays, 2014). The authors also reported that prioritized items are represented by more slots compared to less task-relevant information. At first this model appears difficult to distinguish from the resource model without testing memory storage at higher memory loads (i.e. below capacity limits). However, the main distinction of *discrete* from *continuous* resource models is that it predicts a fixed capacity limit on the number of slots available to represent information. Thus, once a capacity limit is reached, performance for any further items, which cannot be represented, as no more slots are available, should be explained by chance (Ma, Husain & Bays, 2014).

1.8 Thesis overview

The aim of this thesis is to investigate the mechanisms underlying WM across the senses, including vision, audition and touch. The research presented here can be subdivided into three parts.

The resource model of working memory (see section 1.7) has been shown to account for WM processes associated with the storage of a number of visual features, but has not been applied to other sensory modalities. Here, the aim is to develop continuous measures (see section 1.2) assessing the fidelity of WM representations for auditory and tactile sensory modalities. First, a pitch matching study was designed to test whether the resource model can be applied to processes of auditory WM (see chapter 2). Next, an auditory WM study investigated whether this model can also account for another auditory feature: speech sounds or phonemes (chapter 3). In Chapter 4, I go on to describe a study designed to investigate the nature of WM processes in the somatosensory modality, with recall precision measured for mechanical vibrations. In these respects, the first part of the thesis aimed to test whether the resource model of WM can be extended to the auditory and tactile modality by means of continuous measures of WM memory recall (precision).

In the second part of the thesis a study designed to test for the representational format of auditory information held in WM is presented (chapter 5). Are sounds composed of multiple features represented in WM as coherent wholes (objects) or individual features? The findings of this study revealed that memory recall was more accurate when the object had to be maintained as a whole object compared to its individual features alone.

In the final part of the thesis, I present an fMRI study, which was aimed to decode the contents of auditory memory for pitch (chapter 6). The results demonstrate that a number of auditory regions can retain specific auditory information about this basic auditory feature held in memory in the absence of an external stimulus. The same neural circuitry involved in perception of tones is also involved in their maintenance in auditory memory.

The general discussion (chapter 7) draws together the findings from all of the above studies and suggests directions for future research including techniques employed here as well as other methods outlined in the general introduction.

Chapter 2. Flexible resource allocation and prioritization of auditory working memory for pitch

2.1 Abstract

A predominant view of WM considers its capacity to be limited to a fixed number of items. However, recent shared resource models of WM have challenged this ‘quantized’ account under the use of measures of recall precision. Although this conceptual framework can explain processes of visual WM and has been tested for several visual features, it remains to be established whether it also accounts for auditory WM.

A novel pitch matching task was developed to probe participants’ memory of pure tones in sequences of variable length (memory load), and quantified their recall precision. Crucially, this provides an index of the variability of a memory representation around its true value, rather than a binary “yes/no” recall measure typically used in change detection paradigms. The results from this study (experiment 1) show that precision of auditory WM varies with both memory load and serial order. The results from experiment 2, where the factor of task-relevance of individual items was manipulated, show that auditory WM resources can be prioritized to cued tones, improving precision of recall, but with a concomitant cost to other items, consistent with the resource model of WM.

2.2 Introduction

On standardized tests of auditory WM such as those measuring memory span for digits, letters or words items are usually remembered in sequences (Waugh & Norman, 1965). Using these measures, performance is assessed in a categorical, purely binary fashion: On a given trial, an item is either remembered or not. The same principle also applies to change detection tasks (delayed-match to sample, see Figure 1-1), which have been widely used to assess auditory WM (Deutsch, 1970; Jump & Ries, 2008; Lu, Williamson & Kaufman, 1992; Starr & Pitt, 1997). Based on such measures, it has been proposed that auditory WM capacity is extremely limited. Previous studies suggested that the capacity limit of auditory WM is fixed to a limited number of items: equal to one or two (Fougnie & Marois, 2011; Saults & Cowan, 2007), compared to three to four items for visual WM (Luck & Vogel, 1997).

Although binary change detection tasks can be modified to assess WM in more informative ways than only estimating fixed capacity limits (Bays & Husain, 2008; Rouder et al., 2008; Wilken & Ma, 2004), the remaining problem with this approach to measuring WM is that detecting a change does not necessarily infer perfect remembrance of an item; nor does detection failure mean that a memory representation is entirely absent. Instead of using a binary (recall/no recall) procedure, the current study employed a continuous response method as our index for memory performance. This method operates in analogue fashion, over a continuous scale. The variability of recall around the true stimulus value was measured in order to estimate the precision of WM under different memory loads (number of tones played within a sequence).

A likewise approach for visual WM has shown that the resolution with which items are represented is not fixed but changes with memory load. In other words, the fidelity with which each item is represented can vary depending on task demands (Anderson et al., 2011; Bays & Husain, 2008; Bays et al., 2009; Fougnie et al., 2010; Wilken & Ma, 2004). For instance, if an item is prioritized by pre-cueing and it becomes more likely to be probed, it is recalled more precisely but at a cost in memory strength for other, less task-relevant (here: uncued) items (Bays & Husain, 2008; Gorgoraptis et al., 2011).

Based on these observations some authors proposed that although WM is indeed highly limited, its capacity is not fixed to holding a set number of items (Bays & Husain, 2008; Bays et al., 2009; Fougnie et al., 2010; Wilken & Ma, 2004). Instead, WM is better considered to be a resource that can be dynamically shared across items, as well as a resource that can be allocated to prioritized ones (Bays & Husain, 2008; Gorgoraptis et al., 2011). This novel conception of WM has recently been shown to explain results for sequential as well as simultaneously presented visual stimuli. Using both types of stimulus displays at encoding, a drop in precision with an increase in memory load could be captured well by a power function (Gorgoraptis et al., 2011; Zokaei et al., 2011). However, it remains to be determined whether this new perspective on WM can be extended to sensory modalities other than vision.

Here we ask whether the resource model can also account for auditory working memory for sequences of pure tones. Surprisingly, although there have been numerous studies of pitch WM (Clarke, Adriani & Bellmann, 1998; Clément, Demnay, & Semal, 1999; Deutsch, 1970, 1972a, 1972b, 1974; Gosselin, Jolicoeur, & Peretz, 2009; Grimault et al., 2010; Jump & Ries, 2008; Linke et al., 2011; Lu et al., 1992; Massaro, 1970; Mukari, Umat, & Othman, 2010; Pechmann & Mohr, 1992; Ries & DiGiovanni, 2007, 2009; Ruusuvirta, Wikgren, & Astikainen, 2008; Semal & Demany, 1991; Starr & Pitt, 1997; Wickelgren, 1969; Williamsom & Stewart, 2010), no previous study has used a pitch-matching paradigm to measure recall precision as a function of memory load.

In the present study, subjects listened to sequences of pure tones at variable length and their memory for one of the tones was probed on each trial, as indicated by its serial position in the sequence. Importantly, their response was continuous rather than binary: they matched the pitch of a tone by turning a hand-held dial by making an adjustment as closely as possible to their memory of the probed tone (method of adjustment in psychophysics). The precision of WM for pitch at each serial position was calculated as the inverse of the standard deviation of response error. This pitch-matching paradigm provides a measure of the variability of a memory representation around its true value.

In experiment 1, we examined how working memory precision varies with memory load (number of tones in the sequence). Test tone sequences consisted of 1-4 tones, which were randomly sampled from a range of one octave (500-1000 Hz). We also investigated sources of error in listeners' responses using a probabilistic model that has been applied to visual working memory (Zhang & Luck, 2008; Bays et al., 2009; Fougnie et al., 2010; Gorgoraptis et al., 2011; Zokaei et al., 2011), which allows us to look at responses directed at the target separately from random guesses. In experiment 2, we then examined whether working memory resources can be flexibly allocated to a prioritized tone by cueing its task relevance, just as for visual stimuli (Bays & Husain, 2008; Gorgoraptis et al., 2011).

Concerning experiment 1, we found that the precision of memory for pitch varied with the memory load (number of items in the sequence) as well as on the item's position in the sequence. Experiment 2 showed that our working memory resource can be flexibly allocated to prioritized tones in the sequence resulting in a gain in precision as opposed to a cost in precision for less task relevant information. Our results support the shared resource model for working memory for pitch and show that measuring precision as an index of working memory provides new insights into the dynamic nature of memory allocation in auditory working memory representations.

2.3 Methods

2.3.1 Participants

26 listeners with normal hearing and no musical training (16 female, mean age: 23.5 years, age range: 18-31) participated in this study after providing written informed consent to procedures approved by the local ethics committee. 11 subjects took part in experiment 1 and 15 took part in experiment 2.

2.3.2 Stimulus and Apparatus

Stimuli consisted of pure tones, logarithmically sampled at random from a range of 500-1000Hz. Auditory stimuli were created at a sampling rate of 44.1 kHz in Matlab 6.5 (Mathworks Inc.) and presented using Cogent (http://www.vislab.ucl.ac.uk).

Sounds were delivered binaurally through headphones (Sennheiser HD 380 pro) in a soundproof testing room. The stimulus duration was 500ms with an ISI of 500ms. The beginning and end of each trial were indicated by text appearing onscreen. There was a minimum separation of 2 semitones between any two tones of the sequence. The probe tone was randomly selected from the same 500-1000Hz range, and participants could make a response in the range of 250-2000Hz by adjusting a dial (Power Mate, Griffin Technology).

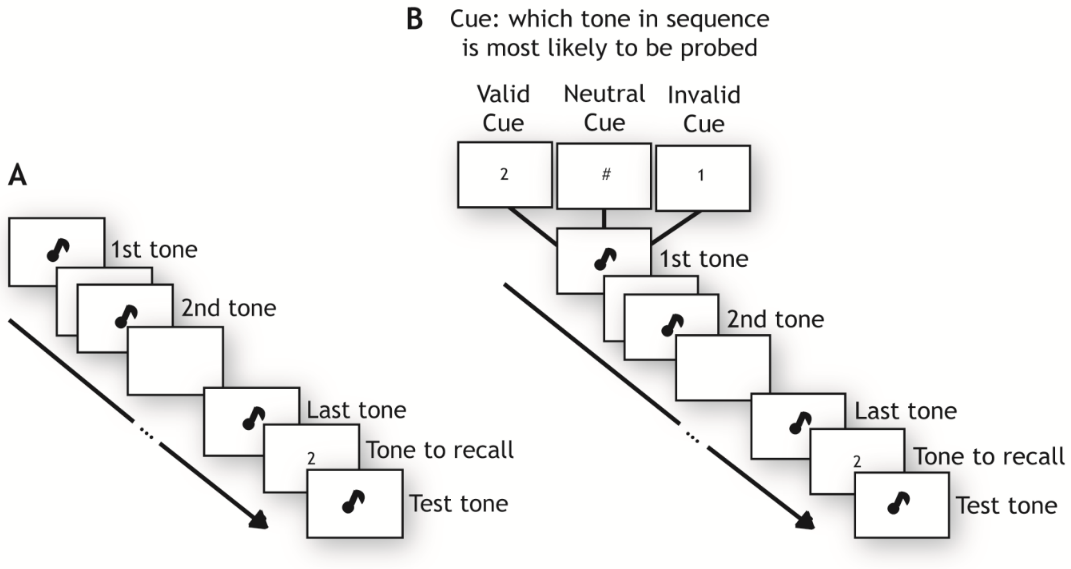


Figure 2-1 *Experimental Paradigm*

(A) Sample sequence for Experiment 1: Subjects were presented with a sequence of tones (e.g. 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) Sample sequence for Experiment 2: 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).

2.3.1 Design and Procedure

Experiment 1

On each trial, subjects listened to a sequence of variable length: 1, 2 or 4 tones (see Figure 2-1A). At the end of each sequence recall of one of the tones was probed, indicated by a number on the screen, e.g. 2 for second tone. A randomly selected probe tone was then played, whose pitch had to be adjusted to match the tone probed, using the dial. Subjects were required to perform the matching task within a maximum response window of 20s. Each subject completed 6 blocks of 48 trials each with equal number of trials for each memory load.

Experiment 2

To examine how WM precision varies when a tone in a sequence is made more task relevant, we used a variant of the previous experiment (Figure 2-1B). On each trial, before each sequence of 3 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 un-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. All conditions were randomly interleaved and each subject completed 6 blocks of 48 trials.

2.3.2 Data analysis

Continuous responses

The deviation between the target tone frequency and the response made by the subject was calculated to obtain a measure of response error (ϵ, in semitones) on each trial as follows,

$$\epsilon = 12 * log_2\left(\frac{f_m}{f_a}\right)$$

where f_m is the measured or matched frequency (i.e. response given by the subject) and f_a is the actual frequency. The raw error is illustrated by histograms in figure 8.

Precision was calculated as the reciprocal of the standard deviation of response error (P = 1/σ), just as in visual experiments (e.g. Gorgoraptis et al., 2011). The precision was calculated separately for each subject, memory load, serial position (experiment 1) and experimental condition (experiment 2). In order to describe the relation between precision, P, and number of items to be encoded, N, a power law of the form $P \propto N^K$, where K is a power law exponent, was fitted using maximum likelihood estimation. Each subject listened to varying numbers of tones in a sequence (memory loads) and each tone appeared at a different serial position (e.g. memory load 1 has serial position 1 and memory load 2 has serial positions 1 and 2 etc.). We fitted unadjusted and adjusted multilevel models to account for the fact that we had repeated measurements from individuals at different positions using different memory loads (Laird & Ware, 1982). Tests were performed in Stata 11.2 software (StataCorp LP, College Station, Texas) and all p-values <0.05 were considered significant, using 2-tailed tests.

Probabilistic model fitting

To examine sources of error contributing to performance, the distribution of responses was analysed using a mixture model, as previously described for visual WM (Zhang & Luck, 2008; Bays et al, 2009). Here we use a model which assumes that errors in memory recall can be decomposed into two separate components: (1) responses centred at the probed tone (target tone) and (2) random responses (guessing) unrelated to the target. The first component can be captured by a Gaussian distribution centred at the target frequency and we modelled the second component describing random guesses as a uniform distribution across the range of one octave. A third component explaining confusion errors was not added to the mixture model as no peak was found around the non-target frequencies in the raw error histograms.

not two

The model is described as:

$$P(\epsilon) = \alpha N(0,\sigma) + (1-\alpha)\frac{1}{24}$$

where ϵ is the error computed as in equation (1); $N(0,\sigma)$ is the normal distribution with zero mean and standard deviation of σ; α is the probability of reporting the correct target value and $(1-\alpha)$ is the proportion of all other responses, or random guessing in this model (Figure 2-2).

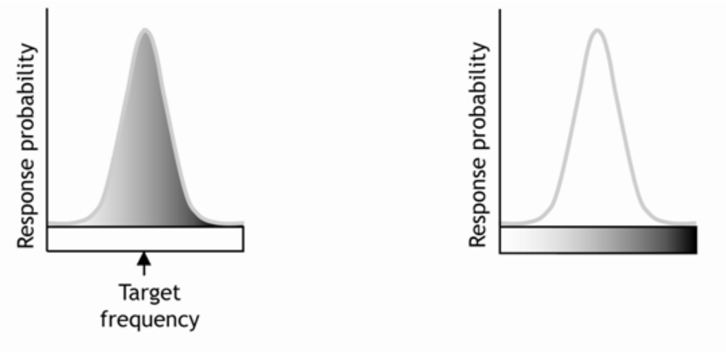


Figure 2-2 *Probabilistic modelling of response distribution*

Error of subjects' responses. Subject responses were decomposed into two components, illustrated by the shaded regions. The first component (left panel) captures responses directed at the target as a Gaussian normal distribution centred on the target frequency. The second component (right panel) captures random responses unrelated to any of the test tone frequencies as a uniform distribution.

Maximum likelihood estimates (Myung, 2003) of the parameter α were obtained separately for each subject and memory load in Experiment 1 using an expectation-maximization algorithm (Figure 2-2).

2.4 Results: experiment 1

Effects of memory load on pitch matching

The precision with which listeners performed the pitch matching was assessed for different sequence lengths (memory loads; Figure 2-3A) and all serial positions within each sequence (Figure 2-3B). As the distribution of precision values across the group was positively skewed, a log transform was applied to precision values, which was entered as the outcome in the multilevel models.

The results for the unadjusted analysis show that log-precision declined significantly with memory load (global p < 0.0001). The mean WM log-precision was 43% lower for 2 items compared to 1 item (95% CI: 63 % to 23 %). Thus there was a significant drop in log-precision even when the number of items to be maintained in WM was increased from 1 to 2, i.e., below the capacity limit of 2 items assumed previously for auditory stimuli (Saults & Cowan, 2007, Fougnie & Marois, 2011). Precision was 73% lower for 4 items compared to 1 item (95% CI: 91 % to054 %), demonstrating a further drop in memory resolution with the addition of more items.

Secondly, there was an effect of serial position (a tone's order in a sequence) on log-precision of recall (global $p < 0.0001$): it was 35% lower for the first item in the sequence (95%CI: 58 % to 12 %) compared to trials on which the last item in the sequence was probed (Fig.2B), indicating a recency effect. Precision was 53% (95% CI: 76% to 30 %) lower and 25% (95% CI: 43 % to 6%) lower for the third and second items in the sequence, respectively, compared to the last item. The unadjusted analysis shows that both factors, memory load and serial position, were significantly associated with log-precision.

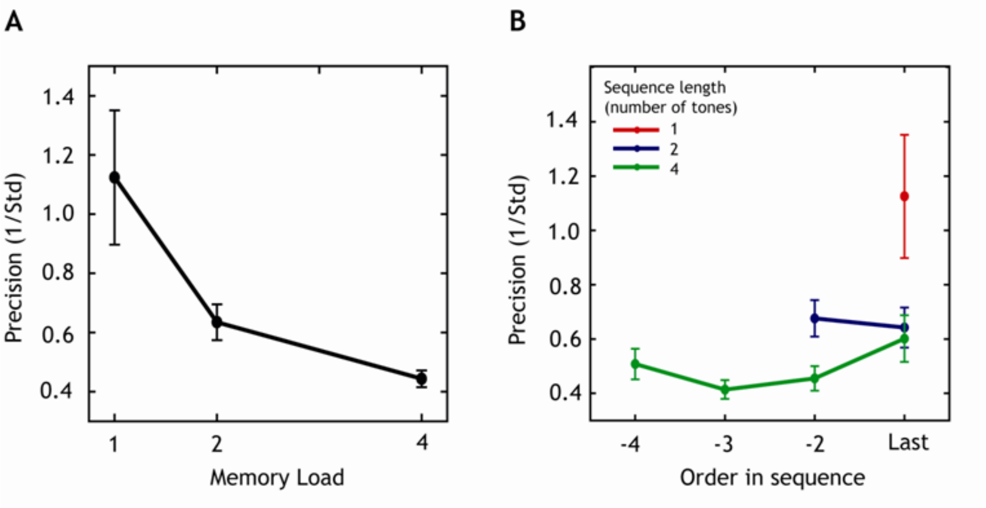


Figure 2-3 *Precision of recall varies with total memory load and serial order*

(A) Overall mean precision for every memory load. The plot shows how precision decreases with an increase in memory load (number of tones in the sequence). Errorbars represent one SEM. (B) Mean precision plotted against order in the sequence for different memory loads, denoted by different colours. Errorbars represent one SEM.

We performed an adjusted analysis including both factors, memory load and serial position, in the model in order to assess the independent effects of both variables. In this model, the effect of memory load remained significant ($p < 0.0001$; decline in precision of 39% for load 2 compared to load 1 (95% CI: 60 % to -18 %; decline in precision of 63% from load 1 to load 4 (95%CI: 85 % to 42 %). However, the effect of serial position on log-precision was no longer significant ($p = 0.18$). The interaction effect was marginally significant ($p = 0.07$), suggesting that within a given memory load serial position has an influence on log-precision. Importantly, pitch matching performance was significantly better than chance for every combination of serial order and memory load ($t(10) > 2.2$, $p < 0.05$).

A power function ($P \propto N^k$) was fitted to the scaled precision values, an analysis previously used for visual experiments (Bays & Husain, 2008; Gorgoraptis et al., 2011). Precision values were scaled with respect to the precision for sequences with only one tone in each of the experimental conditions. The function effectively quantifies the relation between the precision and the number of items (N) to be encoded. The value of k was estimated to be to -0.53.

what does it mean compared to vision?

2.5 Discussion: experiment 1

To investigate the fidelity of auditory WM representations for pitch of pure tones played in sequences, we analysed the precision with which listeners matched the pitch of a target tone, rather than asking them whether a tone was higher/lower, or same/different from such a target. An influential model, developed on the basis of change detection measures, assumes that our working memory capacity is limited to a fixed number of items, which has been estimated in auditory change-detection tasks to be two items or less (Saults & Cowan, 2007; Fougnie & Marois, 2011). Our results challenge this quantized account of WM and are better described by a shared resource model of WM.

The shared resource model predicts that the more items that are held in memory, the less precisely each item can be recalled, as has been shown for visual working memory (Wilken & Ma, 2004; Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Bays et al, 2009; Gorgoraptis et al. 2011; Zokaei et al., 2011). We demonstrated a clear decline in precision of working memory as the number of tones in the sequence increased (Figure 2-3A), indicating that memory capacity is in fact highly limited. Importantly, however, even adding a single tone to a previous tone held in memory was sufficient to produce a significant drop in precision.

Such a fall in WM precision cannot be explained on the basis of a fixed capacity model, which predicts optimal performance until the capacity limit is reached (Luck and Vogel, 1997; Cowan, 2001; Fougnie & Marois, 2011; Saults & Cowan, 2007). We also observed that the pitch matching performance remained significantly above chance for the highest memory load of 4 items. This result cannot be explained by a fixed capacity account either, which predicts a sharp drop in performance when the fixed limit is exceeded. However, both results are consistent with a resource model in which a limited resource has to be shared between items held in WM (Bays & Husain, 2008).

When assessing the serial position independently of memory load no significant effects of order were found. Thus, each tone within a sequence of a given length (load 2 or 4) was recalled with equal precision. These results show that the equal precision model might explain our results best when compared to the variable precision model (Figure 1-4). The equal precision model predicts that our memory resource is equally distributed across information, which corresponds to our observations, although there is a slight trend towards edge effects (primacy and recency). The first item in load 2 and load 4 conditions is recalled with slightly higher precision compared to the following ones (trend towards primacy). There is also a trend towards a recency effect for memory load 4, where the last item is recalled with higher precision compared to the previous tones. Primacy and recency effect, although non-significant, can be best explained as edge-effects according to the principles of Gestalt grouping. To fully distinguish between different resource accounts (equal or variable precision) a more detailed trial by trial analysis is needed, which shall be addressed by future research.

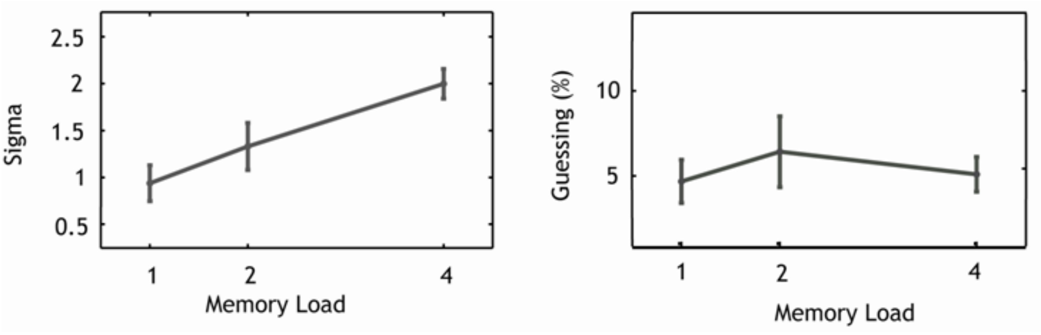


Figure 2-4 *Probabilistic model results*

Model components for each memory load. The variability of responses is expressed by the parameter sigma (σ) describing the standard deviation from the actual target frequency. Fitted precision increased with an increase in the number of tones in the sequence (left panel). The proportion of random responses (guessing percentage) is constant across memory loads (right panel), also showing that the proportion of responses directed at the target does not decrease with an increase in memory load. Errorbars represent one SEM.

2.6 Results: probabilistic model

A probabilistic model to investigate the sources of error in pitch sequences

As the number of tones in the sequence increased, responses centred on the target became increasingly variable, indicated by a significant increase in the standard deviation σ of their distribution across memory loads (global p < 0.0001; increase in sigma of 39% from load 1 to load 2 (95% CI: 17% to 62%); increase in sigma of 106% from load 1 to load 4 (95%CI: 83% to 129%). But, importantly, there was no difference in the proportion of random responses $(1-\alpha)$ across memory loads (global p = 0.59). The decrease in precision with memory load was therefore not simply due to increased guessing, but increased variability of representation of the probed (target) pitch in memory.

2.7 Discussion: probabilistic model

In order to confirm that the increase in variability with memory load on this paradigm was not due to random guesses made by the subjects, we fitted a probabilistic mixture model to individual subjects' data (Zhang & Luck, 1997; Bays & Husain, 2008). In this model, the variability due to random guesses was separated from the variability associated with noisy pitch representations in WM (responses to the target). The principles underlying the model (Figure 2-2) originated in several studies of visual WM (Zhang & Luck, 1997; Bays et al., 2009; Fougnie et al., 2010; Gorgoraptis et al., 2011; Zokaei et al., 2011). There was a significant increase in Gaussian variability in recall of the target tone frequency (described by the concentration parameter (σ) across memory loads, indicating that as the number of tones increases within a sequence, responses centred on the target become increasingly variable. Thus, as the amount of information held in auditory WM increases, the memory representation for each tone becomes noisier (expressed in a decrease in precision).

By contrast, the frequency of random responses was not different across memory loads, even when the number of items exceeded 2 (the supposed capacity limit for auditory WM). This finding is consistent with the principles of the resource model account applied previously to visual WM (Bays et al., 2009; Wilken & Ma, 2004; Fougnie et al., 2010; Gorgoraptis et al., 2011; Zokaei et al., 2011). In contrast, the fixed capacity model predicts that the addition of a further item above the maximum number of items that can be represented will produce an increase in random guesses. Our result clearly contradicts fixed capacity models and shows that working memory can be better described as a limited resource.

Mixture models used in previous studies on visual working memory (Bays & Husain, 2008; Bays et al. 2009; Gorgoratis et al, 2011; Zokaei et al, 2011) also included a third component to account for misbinding (confusion errors), capturing responses to the non-target values. We first considered adding this third component to the model. However, when plotting the raw error histograms of response error directed at the non-target frequencies no peak was found around those values. Thus, we did not find any evidence in the data indicating the occurrence of misbinding. We, therefore, did not include this component in the model. Confusion errors have been more common on auditory working memory tasks when two succeeding items in a sequence sound most similar to each other (Drewnowski & Murdock, 1980; Wickelgren, 1966). However, by incorporating a minimum separation of 2 semitones between any two tones in the sequence, this manipulation might have prevented the occurrence of misbinding errors.

2.8 Results: experiment 2

Memory precision depends on task relevance

In contrast to experiment 1 where all tones were equally likely to be probed and equally relevant to the task, the factor of task relevance was manipulated in Experiment 2: subjects were informed which tone in the sequence was most likely to be probed by a visual cue prior to the onset of each pitch sequence (figure 1). This manipulation allowed us to study if the memory resource can be directed dynamically to a particular tone. A mixed effects multilevel analysis revealed a significant main effect of condition on precision (global $p < 0.0001$). Precision was 5% (95% CI: 1% to 9%) higher in the valid and 5% (95% CI: -9% to -1%) lower in the invalid conditions with respect to baseline. Thus, there was a significant gain for the valid cue compared to baseline and a significant cost for the invalid cue compared to baseline.

We also investigated, whether the relative benefit of the cue is dependent on a tone's serial position (Figure 2-5).

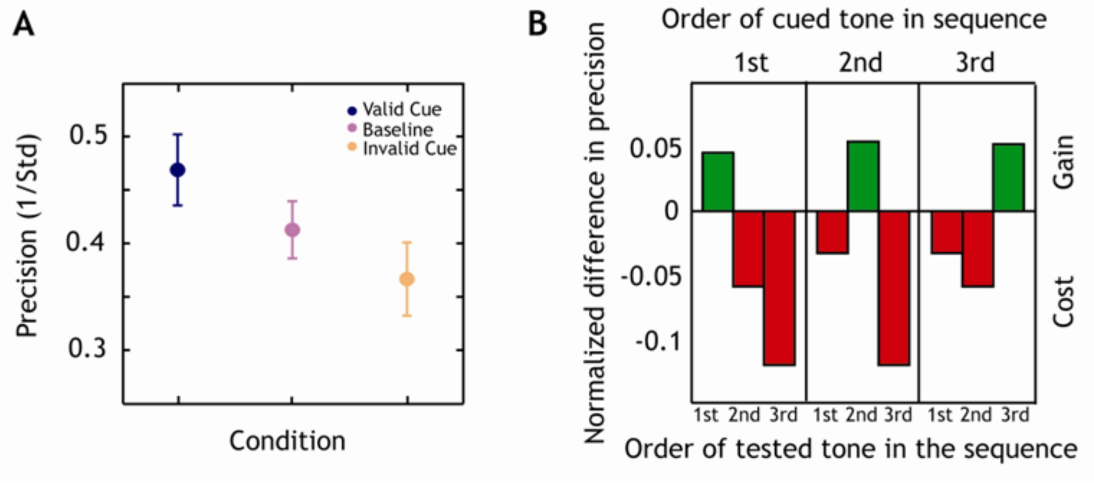


Figure 2-5 *Effects of task relevance on memory precision*

(A) Precision for 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. Errorbars represent one SEM. (B) The fractional difference in precision between the trials of the cued condition and the baseline condition for each tone in the sequence. There was a gain in memory precision for the cued tones (green) and a cost for the uncued tones (red), which were both observed at all serial positions.

The fractional difference in precision (P) between the valid or invalid target (PT) and baseline (PB) conditions for each serial position in the sequence was computed, as in Gorgoraptis et al, 2011:

$$P = (PT - PB)/(PT + PB)$$

The relative gain (positive values in figure 12) was similar for the tones indicated by the valid cue at every serial position (t(14) < .31, p > .75). There was no significant difference between the relative cost (negative values) for the tones in the invalid condition at each serial position (t(14) < 1.2, P > .22) . However, the relative gain was significantly different from the relative cost, when collapsing across serial positions (t(14) = 4.04, p < .001). Thus, with improved precision for a cued item there was corresponding decreased precision for un-cued items.

2.9 Discussion: experiment 2

WM performance depends on the ability to encode and maintain selected information most relevant to the task (Awh et al., 2006; McNab & Klingberg, 2007). Manipulating the factor of task relevance of different sequence positions by pre-cueing, we found that precision was highest when the tone was most likely to be probed, relative to the neutral condition where each tone in the sequence was equally likely to be probed (Figure 2-5). Enhancing the priority of a particular tone in the sequence thereby resulted in a clear benefit in precision, but came at a cost in precision for other tones, which were less likely to be probed, analogous to results for visual WM (Bays & Husain, 2008; Gorgoraptis et al., 2011). These finding show working memory resources may be allocated flexibly according to priorities. Measuring the fidelity of auditory WM representations not only provides new insights into the dynamic nature of memory for pitch, but also shows unifying principles that exist across visual and auditory modalities.

Pre-cueing can be regarded as an attentional task manipulation, which influences how attention is directed at a particular tone in the sequence. The process of directing attention is believed to lead to an enhancement of representations in working memory (Olivers et al, 2011 for a review), which are strengthened by pre-cueing and thus carry less noise as reflected in our results. This mechanism can be regarded as communicating material between the cognitive faculty of attention and working memory. According to Baddeley's model the attentional process is associated with the central executive and auditory working memory with the phonological store (Figure 1-3). However, it has been debated whether attention and memory may reflect the same process. One of the reasons for a common resource pool across both faculties, is that memory retrieval might reflect a form of selective attention to such internal representations stored in memory (Badre et al., 2005; Wagner et al., 2005).

We argue that attentional processes are at hand during memory encoding and briefly afterwards reflected in delay activity (Anderson et al, 2011), as the cued target stimulus has to be selected rapidly, where after its representation remains in a heightened state ('focus of attention'; Oberauer, 2002). At the stage of memory retrieval the previously attended stimulus has to be reproduced by activating representations from auditory working memory. It may be argued that memory underlies attentional control mechanisms (Chun & Turk-Browne, 2007). In contrast, Lewis-Peacock et al (2012) provide neural evidence for a distinction between working memory and 'the focus of attention' using multivariate pattern analysis of brain activity recorded in event-related fMRI. They show that the information content of the delay period activity corresponds to attention rather than the full contents of working memory. Additionally, attention and memory can be distinguished on the basis of events occurring at different points in time, where attention is a brief perceptual process at action when stimuli are presented and stimuli remain in the focus of attention 1-2s after stimulus presentation. In contrast, representations are active in working memory several seconds after stimuli disappear (Oberauer, 2002).

2.10 General Discussion

In a novel pitch matching paradigm, we show that precision of auditory working memory for pure tones varies with memory load. Fitting a probabilistic mixture model to our results shows that responses come in two varieties. While the majority of responses are directed at the target, a small proportion is made at random. The mixture model results and findings from experiment 1 contradict a fixed capacity account, limited to a set number of items, and suggest instead that processes in auditory working memory are better described by a shared resource.

The results from experiment 2 further support the resource model, showing that our resource is flexibly weighted depending on a tone’s task relevance. Measuring precision as an index of working memory as previously applied to vision also captures the dynamics of memory representations in audition. Thus, the resource model no longer only describes working memory mechanisms in vision, but also shares common properties with processes in auditory working memory.

Comparison to vision

A major commonality across vision and audition is that precision decreases with an increase in memory load, where especially the drop in precision from information load of 1 to 2 items is most pronounced. This finding across the senses strongly contradicts a fixed capacity account, where the item limit for either sense has not been reached yet. As predicted by the slot model performance remains optimal up to this point. In general, item limits are estimated to equal 3-4 items in vision and 2 or fewer items in audition. Nevertheless, a fundamental problem with the item-limit account is that the number of slots available seems to differ across stimulus material tested. For example in vision, while the capacity limit for coloured squares is 3-4 items (Anderson et al., 2011; Zhang & Luck, 2008; Todd & Marois, 2004; Luck & Vogel, 1997), the limit for faces and polygons is 2 items (Saults & Cowan, 2007;

Fougnie & Marois, 2011). Accordingly in audition, item limits vary across material tested, where limits are higher for digits compared to letters (Brener, 1940; Crannel & Parrish, 1957) and the limits also differ for abstract- and pseudo-words (Brener, 1940; Frick, 1988).

Thus, precision may appear as a more universal measure to make comparisons across stimulus materials used and across sensory modalities tested. However, the fallacy may be that precision is scaled to the stimulus space tested. Where a number of visual studies used circular stimulus spaces from which stimulus properties, such as colours or orientations (Bays et al., 2009, Gorgoraptis et al, 2011; Bays et al., 2011; Burnett Heyes et al., 2012) were sampled and along which stimuli could be matched; the current study did not use a circular space. Stimuli were drawn from a range of pitch values (500 Hz to 1000 Hz) and the matching dimension exceeded this range by one octave on either side (250 Hz - 2000 Hz) to prevent subjects from finding the edges of the stimulus space, which could have encouraged the use of memory strategies. A solution to the problem of precision being scaled to a particular stimulus space is to normalize scores to the precision of the lowest memory load of 1 item making results from different studies comparable across material and sensory modality tested.

somewhat comparable, but still not fully comparable

Furthermore, in vision serial order effects have been predominant in a number of studies. For the different types of visual stimulus material tested (colour, orientation, motion), the last item in the sequence has been remembered more precisely than the previous ones (Gorgoraptis et al., 2011; Zokaei et al., 2011). However, there was only a small trend towards a recency effect for the memory load of 4 tones. Therefore, it remains to be determined whether the finding in audition holds for other types of auditory stimulus material as well. For example, auditory working memory studies for speech (verbal material) usually show a strong primacy and recency effect reflected in a U-shaped recall curve (Gupta et al., 2005; Baddeley, 1986; Conrad & Hull, 1964; Levy, 1971). Measuring precision of recall, one may speculate that the difference in serial order effects across modalities may be related to the time taken to match a tone compared to response times associated with matching a visual probe to a target. Pitch matching takes around 4-

5 seconds, where matching an oriented bar takes around one second or less. As serial order or 'edge effects' (Botvinick & Plaut, 2006) may reflect temporal distinctiveness (Glenberg & Swanson, 1986), those may become less pronounced when the duration of matching is increased.

Methods and types of memory processes

The shared resource model proposes that memory representations are noisy and that the amount of noise varies with information load. To measure this amount of representational 'noise' for pitch memory, we used the method of adjustment in the form of a pitch matching paradigm. Although previous attempts have been made to obtain a 'continuous' response measure for pitch with the method of constant stimuli (MOS), (Ries & DiGiovanni, 2007; 2009), pitch matching by adjustment has several advantages over MOS. Instead of sampling tones from a predefined and therefore limited number of pitch values (frequencies), where the possible size of change between the test and probe stimulus defines potential stimulus properties, we aimed at obtaining a truly continuous response measure to allow precision to be better characterized.

Using the method of adjustment allowed us to sample from a wider range of stimuli than can be used with MOS, and allowed subjects to respond with any pitch within this range (and beyond). Pitch increments were sampled below the threshold for pitch discrimination. Consequently, a pitch adjustment of the probe tone can be made at any step along the continuum where a response is recorded. Measuring the pitch matching performance using this method allows us to acquire very reliable estimates of the resolution of memory. Estimating the resolution with which information is stored, rather than assuming all-or-non storage, is crucial if we are to gain deeper insights into the mechanisms underlying working memory.

A further methodological difference between the current study and previous work relates to the type of memory retrieval process tested. Pitch discrimination studies using change detection measures test for recognition memory, first introduced by Wickelgren (1969). Recognition memory can be considered to be a passive form of memory retrieval process (Mohr & Linden, 2005, Cornoldi et al

2000). However, tasks other than change detection require the subject to reproduce or manipulate information held in working memory. In our pitch matching paradigm, subjects had to reproduce a tone from memory as accurately as possible. The memory retrieval process underpinning the reproduction of information may be different from simply recognizing whether information is present or absent.

2.11 Conclusion

The current study shows that the resource model of working memory as previously applied to vision can also account for processes in auditory working memory for pitch. Thus, using precision as an index of working memory provides new insights on the dynamic nature of immediate visual as well as auditory information storage. This provides a novel framework along which neural data on internal auditory memory representations can be interpreted, where representational noise depends mainly on information load.

Chapter 3. Precision in auditory working memory for speech sounds

3.1 Abstract

Memory for speech sounds is a key component of models of verbal working memory (WM). But how good is verbal WM? Most investigations assess this using binary report measures to derive a fixed number of items that can be stored. However, recent findings in visual WM have challenged such 'quantized' views by employing measures of recall precision with an analogue response scale. WM for speech sounds might rely on both continuous and categorical storage mechanisms. Using a novel speech matching paradigm, we measured WM recall precision for phonemes. Vowel qualities were sampled from a formant space continuum. A probe vowel had to be adjusted to match the vowel quality of a target on a continuous, analogue response scale. Crucially, this provided an index of the variability of a memory representation around its true value, and thus allowed us to estimate how memories were distorted from the original sounds.

Memory load affected the quality of speech sound recall in two ways. First, there was a gradual decline in recall precision with increasing number of items, consistent with the view that WM representations of speech sounds become noisier with an increase in the number of items held in memory, just as for vision. Based on multi-dimensional scaling (MDS), the level of noise appeared to be reflected in distortions of the formant space. Second, as memory load increased, there was evidence of greater clustering of participants' responses around particular vowels. A mixture model captured both continuous and categorical responses, demonstrating a shift from continuous to categorical memory with increasing WM load. This suggests that direct acoustic storage can be used for single items, but when more items must be stored, categorical representations must be used.

may be difficult to understand at this point

3.2 Introduction

Tests of auditory/phonological working memory (WM) typically measure recall performance for lists of items (Baddeley, 2007). Tests of WM for speech often use verbal stimuli such as spoken digits (Morgan, Chambers, & Morton, 1973), letters (Conrad, 1964; Conrad & Hull, 1964), syllables (Gupta, Lipinski, Abbs, & Lin, 2005) or words (Drewnowski, & Murdock, 1980; Haberlandt, Lawrence, Krohn, Boer, & Thomas, 2005). Such measures have provided key evidence for models of verbal WM and their dysfunction in brain disorders, including developmental language conditions (e.g. Gathercole & Baddeley, 1993; Klingberg, 2008; Alloway & Gathercole, 2012).

In these traditional methods used to measure verbal span, performance is assessed in a purely categorical/binary fashion: an item is either recalled correctly or not at all. This principle also applies to a wide variety of commonly employed experimental tasks, used to measure verbal WM, for example, n-back (Sörqvist, Stenfelt, & Rönnberg, 2012), change detection (Crowder, 1981) or delayed-match to sample paradigms (Saults, & Cowan, 2007; Fougnie, & Marois, 2011). On the basis of such measures of absolute judgement, capacity limits are commonly estimated as a *fixed number of items*.

Such constant capacity limits vary enormously depending upon the type of material employed in the task, as well as with other experimental design parameters. For example, some studies estimated verbal WM capacity for spoken letters to be only two or even fewer items (Fougnie, & Marois, 2011; Saults, & Cowan, 2007). However, even when we fail to recall an item, it is possible that we might still have a memory representation of it rather than no representation at all. This representation might simply be of a lower resolution compared to an item recalled correctly.

In the visual domain, such concerns have led to development of measures of precision of WM that are continuous and analogue in nature rather than discrete and binary (Wilken, & Ma, 2004; Bays, & Husain; 2008, Bays, Catalao, & Husain, 2009; Ma, Husain & Bays, 2014). Indeed, research using these new methods has led to the proposal that visual WM might not be limited in capacity to a fixed number of items but might instead be better considered to be a highly limited resource, one that can flexibly be allocated to objects in the visual scene, without an upper limit to the number of items that can be retained (Wilken, & Ma, 2004; Bays, & Husain; 2008, Bays et al., 2009; Fougnie, Asplund, & Marois, 2010; Gorgoraptis, Catalao, Bays, & Husain, 2011; Zokaei, Gorgoraptis, Bahrami, Bays, & Husain, 2011).

According to this *resource model*, if no item has priority to-be-remembered, the fidelity of an object's representation is inversely proportional to the number of objects stored in visual WM, with the resource being divided between all the items in memory (Bays, & Husain, 2008; Bays et al., 2009, Gorgoraptis et al., 2011). Thus, as the amount of information is increased, each visual object is remembered less precisely, but crucially without a fixed-item limit. Such an account has been shown to describe well WM performance for several visual object features, including colour, orientation (Bays, & Husain, 2008; Bays et al., 2009, Gorgoraptis et al., 2011) and motion (Zokaei et al, 2011).

may be too strong

However, there is no *a priori* reason to believe that similar principles would apply to speech sounds, particularly since speech might be considered to be ultimately 'categorical' in terms of encoding of information (Harnad, 1987; Liberman et al., 1967; Holt & Lotto, 2010). In fact, a critical question in working memory is the extent to which items are stored in "stimulus space", or in a more abstract or categorical form. Computational models of WM (Pascanu & Jaeger, 2011; Compte et al. 2000; Bays 2014) typically use homogeneous arrays of neurones to maintain stimulus representations, such that the possible representations are evenly distributed across stimulus space, forming a 'line attractor' (Burak & Fiete 2012). This may be suitable for highly contrived experimental domains, such as line orientation or motion direction. However, for most real-life stimuli, an efficient encoding strategy, such as categorical encoding, is likely to be employed before WM storage. Even in the special "continuous" domains previously used, colour categories and horizontal/vertical lines may enjoy privileged or categorical representation. In such cases, point attractors, or some more complex circuit structure, may be more appropriate (Amit & Brunel 1995; Molter, Salihoglu & Bersini 2007).

Under some circumstances, we can remember the sensory properties of stimuli, but under other circumstances, an abstracted, categorical representation is recalled (Pisoni 1973, Massaro & Cohen 1983). A key question is then, under what circumstances do we remember by categories, as opposed to along a sensory continuum. The domain of speech sounds allows us to probe this directly, as a speech sound can be remembered both acoustically, and also by its identity as a vowel (Schouten, Gerrits & van Hessen 2003). To address this the fidelity – or precision – of memory representation was measured for phonemes. A novel response method was designed, which operates in analogue fashion, over a continuous scale. This allowed us to measure the variability of verbal recall around its actual stimulus value (speech sound), as well as to identify whether responses were biased by categorical representations

Previous studies have used continua to test for *perceptual* discrimination between phoneme pairs (Fry Abramson, Eimas, & Liberman, 1962; Pisoni, & Tash, 1974; Macmillan, Goldberg, & Braida, 1998; Pisoni, 1973; Gerrits, & Shouten, 2004; Shouten, & van Hessen, 1992). In these experiments, speech sound continua typically consisted of a fixed set of stimuli. For example, Fry et al. (1962) synthesized thirteen stimuli by dividing the space between three vowels (/I/, /Ɛ/, /æ/). Similarly, Schouten and van Hessen (1992) used natural sound recordings to generate spectral interpolations between a set of three vowels (embedded in a context of consonants) or stop consonants only. Thus, stimulus material was limited: (1) to a rather narrow stimulus range and (2) its underlying acoustic dimension. In the above studies, however, the response method was again binary.

In contrast, other studies manipulated acoustics and used a much larger stimulus range (spanning the entire F1/F2-space) in combination with a continuous response method to perceptually map vowels in a multi-dimensional vowel space (Evans, & Iverson, 2004; Iverson, Smith, & Evans, 2006; Iverson, & Evans, 2007). In the current study, we synthesized speech sounds in a similar fashion and used an adopted response method (uni-dimensional path) but now to measure speech sound recall for vowels.

Here, isolated vowels were studied rather than vowels with consonants. Both types of phonemes are perceptually distinct, but vowels are usually perceived as more continuous than consonants (Fry et al, 1962; Shouten, & van Hessen, 1992). It has also been proposed that they are represented differently in memory, as consonants decay faster than vowels (Pisoni, & Tash, 1974; Pisoni, 1973). Shouten and van Hessen (1992) claim that there are two different types of memory stores associated with either type of phoneme. Additionally, there is neuroimaging evidence that vowels and consonants might be processed in distinct neural systems (Carmazza, 2000). Based on these findings, as well as aiming for a continuous sampling and response method, the paradigm focussed on WM of speech sounds for vowels only.

Variable length sequences of phonemes were used, with each phoneme composed of a vowel sound sampled from a continuous scale (based on the formant frequency (F1/F2) space), and always paired with the same consonant /d/. Participant's memory for one of the phonemes in the sequence was tested, where the target was indicated by its serial order position. The same continuous scale used to create stimuli also served as a vector along which the sound of a probe vowel could be adjusted by participants.

Importantly, this response method allows us to record a continuous rather than a binary response, where the probe is adjusted to match the vowel sound of the actual target value. The deviation between the target and response made by the subject provides a simple measure of response error, from which memory precision is calculated as the inverse of the error's standard deviation. This novel phoneme-matching paradigm provides an index of the quality or fidelity of phonological memory representations.

The current study investigated how WM precision varies with memory load (number of phonemes within a sequence) and serial order (phoneme position within a sequence). In addition, we applied multi-dimensional scaling (MDS) to the results to map responses made by participants back onto the continuous vowel space. Furthermore, a probabilistic model was applied that has previously been used to model data from visual WM (Zhang, & Luck, 2008; Bays et al., 2009;

Fougnie, Asplund, & Marois, 2010; Gorgoraptis et al., 2011) to investigate sources of response error: the extent to which these are related to increased variability in response around the probed item (*continuous response*), or the proportion of responding to a *vowel category*, as well as random guessing.

3.3 Methods

3.3.1 Participants

21 native English speakers with normal hearing and no musical training (10 female, mean age: 25 years) participated in the experiment after providing written informed consent approved by the local ethics committee.

3.3.2 Stimuli and apparatus

Stimuli consisted of VC syllables, composed of a vowel and the consonant /d/. Vowel qualities were sampled at random from a circular 1D acoustic space. This stimulus continuum was constructed based on the formant frequency (F1-F2) space for vowels of Southern British English (see Figure 3-1). The figure depicts average values of the 1st and 2nd formant frequencies for a group of speakers of the given accent.

Using Klatt and Klatt (1990) synthesis, a circular path was overlaid on the diagram and stimuli were synthesized along this dimension; Iverson, Smith and Evans (2006) used a similar method for speech sound synthesis (except for the /h/). The total set of 360 stimuli was generated to form a 360-step continuum. This method allowed us to construct a continuous ‘vowel stimulus space’, which served as a 1D acoustic dimension from which stimuli were selected. At the same time this space also served as a dimension along which subjects could make a response.

Stimuli were presented using Cogent (http://www.vislab.ucl.ac.uk) via Matlab 6.5 (Mathworks Inc.) at a sampling rate of 44.1 kHz. Sounds were delivered binaurally through headphones (Sennheiser HD 380 pro) in a soundproof testing room. The stimulus duration was 300 ms, followed by an ISI of the same length. Next, an auditory mask was presented, which was generated using signal processing in Praat (http://www.fon.hum.uva.nl/praat/) by layering all 360 stimuli. The amplitude of the mask was adjusted and approximated the intensity of memoranda (speech sound stimuli). Each mask was followed by another ISI of 300 ms in length. The mask was added in order to prevent the memory strategy of chunking (e.g. connecting multiple syllables to form words (Miller, 1956)), and to ensure that subjects store each syllable as a single unit.

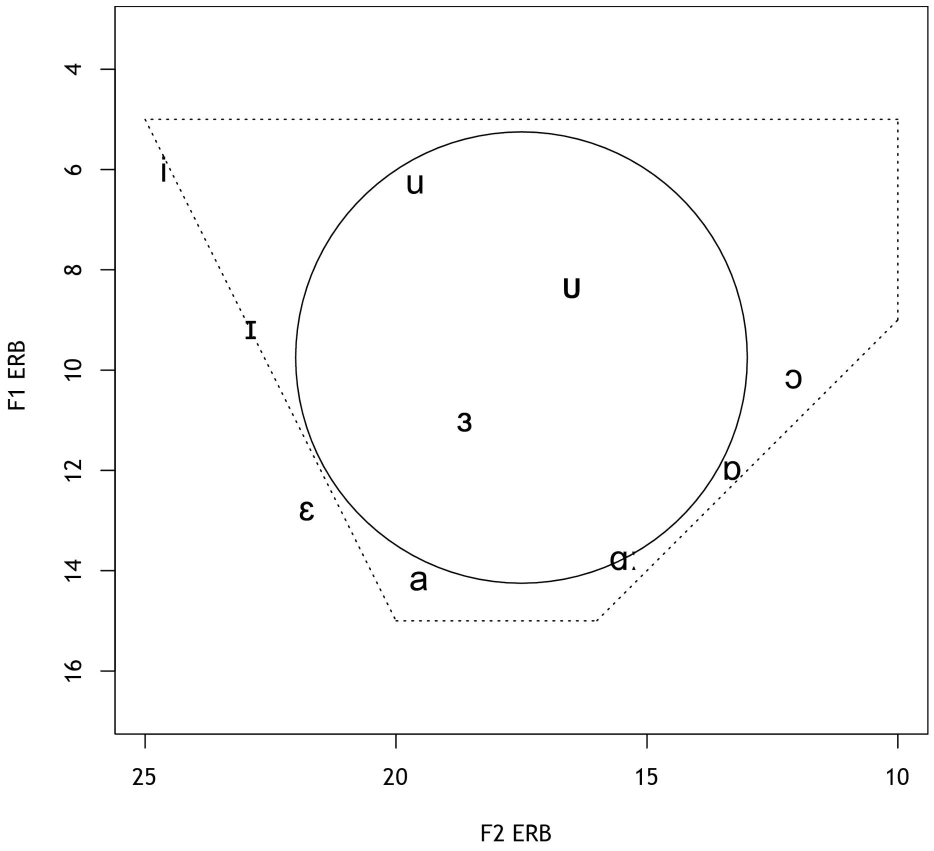


Figure 3-1 ***Formant frequency & vowel stimulus space***

F1-F2 space for vowels of Southern British English, showing average values of both formant frequencies for a group of speakers of the given accent. The circle forms the ‘vowel stimulus space’, which served as a sampling and matching dimension.

The beginning of each trial was indicated by text appearing on screen, followed by a further visual message, indicating the number of speech sounds to be memorized (e.g. “Memorize 4”). There was a minimum separation/distance of 10 steps (degrees) between any two speech sounds in the sequence.

The vowel quality of the probe sound was randomly selected from the same circular sampling dimension. The probe was played as a continuous sound stream, where the playback of each individual speech sound was looped. Participants could make a response by adjusting a dial (PowerMate, Griffin Technology): in this way the participant was able to update the vowel quality of the probe in real-time in order to approximate the vowel quality of the target by dialling through the vowel stimulus space.

3.3.3 Design and Procedure

In order to familiarize participants with the task and to reach a stable performance level, each participant completed 2 training-blocks of 50 trials each. On each trial, subjects memorized a single speech sound, which was followed by a mask (Figure 3-2A). Their memory recall for this particular syllable was then probed. A randomly selected probe syllable was played as a continuous sound stream, where the playback of this particular speech sound was looped. The vowel quality of the probe sound was then adjusted by the participant to match the speech sound probed using the dial.

In this manner, participants searched through the circular 1D vowel stimulus space (see Figure 3-1) to find the vowel quality that matched the target speech sound best. Once they were certain of their selection, they made a button press to confirm that this particular speech sound resembles the one held in memory best. They were required to perform the matching task within a maximum response window of 25s. The start of the next trial was initiated by either making a button press to indicate response selection or when 25s elapsed.

Upon completion of the training-blocks, participants continued with the main experiment, which consisted of 6 blocks of 48 trials each with equal number of trials per memory load (16 trials per block per memory load). Within each block, trials selected from different memory loads were randomly interleafed. On each trial, they listened to a sequence of speech sounds of variable length; containing 1, 2 or 4 syllables (Figure 3-2B).

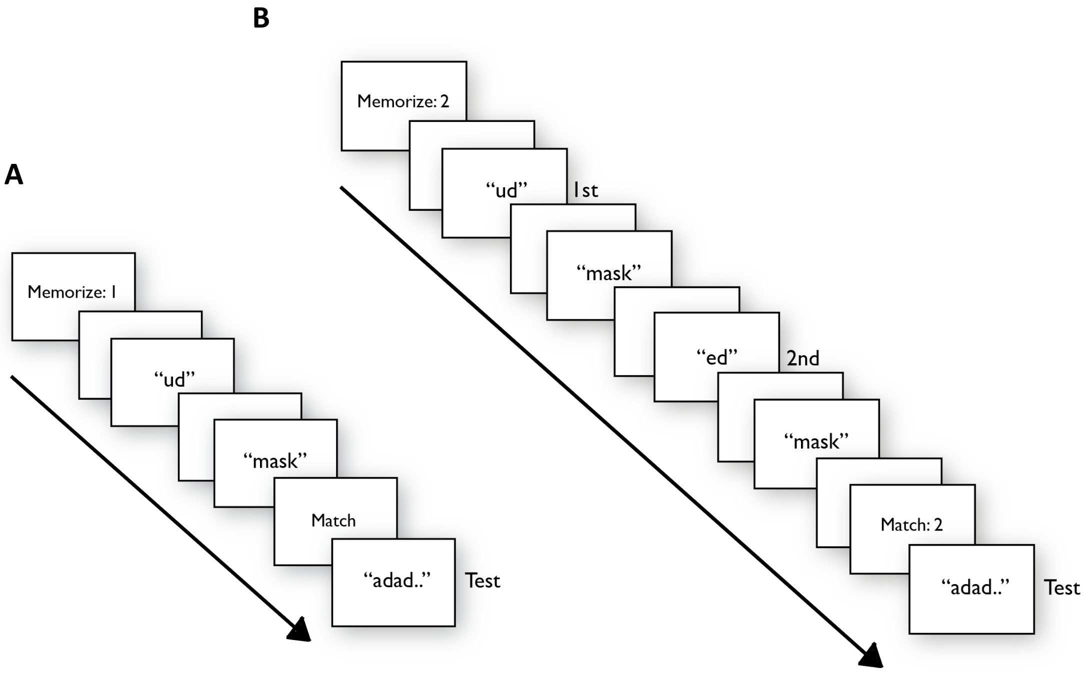


Figure 3-2 *Experimental Paradigm*

(A) Sample sequence for training: A message appeared onscreen indicating that subjects would have to memorize a single phoneme. Next a randomly sampled single phoneme (e.g. vowel quality at 3 degrees and consonant 'd') was played. This phoneme was followed by a mask. Next, participants were instructed to perform the vowel matching. A final randomly sampled probe phoneme was presented in repetition forming a continuous sound stream (e.g. starting at vowel quality 120 degrees and consonant 'd'). The probe had to be adjusted to match the vowel quality of the previous phoneme. (B) Sample sequence for main experiment: A message appeared onscreen indicating the number of phonemes to be memorized. Next, subjects were presented with a sequence of phonemes (e.g. phoneme 1: vowel quality 3 degrees and consonant 'd'; and phoneme 2: vowel quality 120 degrees and consonant 'd'). Each phoneme was followed by a mask. After the memoranda were presented, the instruction to match was presented also indicating the target number (here: second phoneme). A randomly-selected probe phoneme was then played, which had to be adjusted to match the vowel quality of the target (2^{nd} phoneme).

Each speech sound within any given sequence was followed by a mask (i.e. for a sequence of memory load 2 the order of events was as follows: 1^{st} speech sound, 1^{st} mask, 2^{nd} speech sound, 2^{nd} mask). At the end of each sequence, recall of one of the speech sounds was probed, indicated by a number on the screen, e.g. 2 for second speech sound. A randomly selected probe sound was then played in the same way as described for the training blocks. The vowel quality of the probe sound was then adjusted to match the speech sound probed, as above. Again, subjects were required to perform the matching within 25s.

3.3.4 Data Analysis

As the parameter space for speech sounds was circular, the angular deviation between the vowel quality of the target sound and the response made by the subject was calculated to obtain a measure of error (Ɛ, in radians) on each trial,

$$\mathcal{E} = wrap\,(VT - VM)$$

where *'VM'* is the measured or matched vowel quality (i.e. response given by the subject) and *'VT'* is the actual vowel quality of the target. The MATLAB *wrap* function (http://www.sobell.ion.ucl.ac.uk/pbays/resources.htm) was used to ensure that the measure of error remains in the given stimulus range.

Mean scores of the absolute response error (in radians) were obtained for each memory load and serial position. Precision was calculated as the reciprocal of the circular standard deviation (Fischer, 1993) of response error, just as in visual experiments (e.g. Bays et al., 2009; Gorgoraptis et al, 2011). It was calculated separately for each subject, memory load and serial position.

Each subject listened to a varying number of speech sounds, presented in a sequence (memory load: 1, 2 or 4 sounds) and each speech sound appeared at a different serial position (e.g., memory load 1 has serial position 1 and memory load 2 has serial positions 1 and 2 etc.). As the distribution of precision values across the group was positively skewed, a log transform was applied to precision values.

For statistical analysis, precision values were subjected to ***multilevel modeling*** (Laird & Ware, 1982)**.** Unadjusted and adjusted multilevel models were fitted, allowing repeated measurements from individuals at different serial positions with different memory loads to be taken into account (Laird & Ware, 1982). Tests were performed in Stata 13.1 software (StataCorp LP, College Station, Texas) and all p-values < .05 were considered significant, using 2-tailed tests.

3.3.5 Estimating level of chance performance

The level of chance performance was derived from a permutation test, which was carried out in the following way. First, each participant’s actual response values were randomised. Next, the distance between a given randomized response and its corresponding target value was calculated for each trial. This analysis was run over 1000 iterations per participant. The values from all iterations were averaged. On this basis we derived an individual chance threshold for each participant. Finally, the mean of all individual chance thresholds was taken to obtain an overall level of chance performance (absolute response error = 2.72 rad). This method has the following advantage: the measure of chance performance is based on each individual subject’s response distribution rather than on a uniform distribution. Task performance (response error) was tested against chance, using a one-sample t-test.

shuffled

3.3.6 Multi-dimensional scaling analysis

In order to relate speech sound representations held in WM to actual British vowels, we applied a multi-dimensional scaling (MDS) algorithm to our data (Iverson & Kuhl, 1994). In brief, this allowed responses made by participants to be mapped back onto our vowel stimulus space, and determine how memory load and the phonetic representations of the listeners may shrink and stretch perceptual distances between stimuli (e.g., Iverson & Kuhl, 2000; Iverson et al., 2003; 2008).

To compute the MDS space, the stimulus-response data were sorted into 24 overlapping stimulus-response bins that each spanned a 45-degree range (e.g., 0-45 degrees, 15-60 degrees, etc.). Such a matrix is potentially affected by response bias (e.g., more responses for prototypic vowels), so this was controlled by applying the Similarity-Choice Model (e.g., Nosofsky, 1985) to calculate bias-free similarity coefficient values between bins. These values were converted to distances, much like *d’* within detection theory (e.g., Macmillan & Creelman, 1991), by calculating the square root of the log similarity coefficient.

Classical (metric) multidimensional scaling (Gower, 1966) was used to plot these bin points into a two-dimensional space for each memory load. Given that MDS solutions are invariant across rotation and translation, the three MDS plots were centred at the mean of the points and rotated to maximise to correspondence between plots, such that the configurations under each memory load could be more easily compared.

3.3.7 Modelling sources of Error in Participants' Responses

Scatterplots (Figure 3-4) show the relationship between responses made to target stimuli. While a large part of responses were closely scattered along the diagonal, which can be characterized as continuous, we also observed clustering of responses in particular regions along the diagonal. Clusters may form as responses are biased by categorical representation. Visualizing this relationship let us to the examine different sources of error contributing to memory performance. The distribution of responses was analysed using a mixture model for information that is both continuous and categorical. The model we used here proposes that response errors can be decomposed into three components:

(1) ***Continuous responses*** directed at the vowel quality of the *target vowel*. Scatterplots describing the relationship between responses made to target stimuli show that a large part of responses are closely scattered along the diagonal, with responses closely matching the target (Fig. 4). In the model, the first component captures continuous responses made in response to different targets, with a Gaussian distribution centred on the vowel quality of the target (probed item), where the target is remembered with precision σ. (Fig. 6a).

(2) ***Categorical responses*** directed at the target category (the category the vowel quality of the target falls in (Fig. 6B)). Scatterplots of target vs response (Fig. 4) also show clustering of responses in particular regions along the diagonal. For this type of response we assume that the acoustic representation is lost and participants only remember the category a speech sound belongs to.

The second component captures such categorical responses, where each category is modelled as a 'square of confusion', assuming that all vowels within this region sound alike, if the continuous memory representation is lost. The categories were determined in a data-driven manner as the best-fitting category boundaries for each subject individually.

(3) ***Random responses*** **(guesses)** unrelated to the target.

The third component was modelled as a uniform distribution across the range of the entire stimulus space.

The model is described as:

$$P\ (response|target) = \alpha N(response - target, \sigma)$$

$$+\ \beta(responses\ \epsilon[b_i, b_i + 1])\frac{1}{(b_i + 1 -\ b_i)}$$

$$+\ (1 - \alpha - \beta)\frac{1}{2\pi}$$

where *P* is the probability of responding to a target based on three fitted parameters: α, β and σ. α is the probability that the target is recalled perceptually (***continuous response***), where N is the normal distribution and standard deviation σ of response error. β is the probability that the target's ***category*** is recalled and b_i is the position of the category boundary immediately preceding the target (in radians). The three category boundaries b_1, b_2, b_3 ϵ[0,2π] are also fitted to maximise the likelihood of the model. (1 – α - β) is the proportion of the remaining responses, captured by the model as ***random guesses***. Maximum likelihood estimates (Myung, 2003) of the each parameter were obtained separately for each participant and memory load using a expectation-maximization algorithm.

We therefore obtained measures of how often each subject relied on an acoustic representation (alpha), the acoustic representation's fidelity (sigma), and how often they relied on their categorical representation (beta), under each memory load.

3.4 Results

3.4.1 Effects of memory load and serial position on recall precision

The precision with which listeners performed phoneme-matching was assessed for different sequence length (memory loads; Figure 3-3A) and all serial positions within a sequence (Figure 3-3B) using multilevel modelling. Unadjusted and adjusted multilevel models were fitted, so that repeated measurements from individuals at different serial positions with different memory loads could be analyzed (Laird & Ware, 1982).

The results for the *unadjusted analysis* show that log-precision declined significantly with memory load (global $p < 0.00001$). Mean WM precision was 18% lower for two items compared to a single item (95% CI: 4% to 33%). Thus there was a significant drop in precision even when the number of items to be maintained in WM was increased from one to two, i.e., below the capacity limit of two items assumed previously for auditory stimuli (Golubock & Janata, 2013; Fougnie & Marois, 2011; Saults & Cowan, 2007). Precision was 39% lower for four items compared to one item (95% CI: 25% to 52%), demonstrating a further drop in memory resolution with the addition of more items.

Secondly, there was an effect of serial position (a speech sound's order in a sequence) on precision of recall (global $p < 0.00001$): It was 19% higher for the last item in the sequence (95% CI: 6% to 33%) compared to trials on which the first item in the sequence was probed (Figure 3-3B), indicating a recency effect. Precision was 25% lower (95% CI: 8% to 41%) and 3% lower (95% CI: -11% to 17%) for the second and third items in the sequence, respectively, compared to the first item, indicating a primacy effect in the sequence of 4 items.

The observed order effects are driven by differences in precision observed for memory load 4. The unadjusted analysis shows that both factors – memory load and serial position – were significantly associated with precision. Analysis of the impact of memory load on the most recent item in the sequence (last position) across memory loads reveals that precision was 17% lower (95%CI: 2% to 31%) at memory load 2, and 19% lower at load 4 (95%CI: 5% to 34%) compared to load 1 respectively (global p = 0.015).

As part of the multi-level model (Laird & Ware, 1982), we also performed an *adjusted analysis* including both memory load and serial position in order to assess the independent effects of each variable. In this model, the effect of memory load remained significant (p = 0.0007; decline in precision of 26% from load 1 to load 4 (95% CI: 40% to 11%). The effect of serial position on precision (p < 0.0001) and the interaction (p = 0.02) between load and serial position both also remained significant. These results demonstrate that within a given memory load serial position significantly influences precision of recall.

Performance was significantly above chance for every combination of memory load and serial order (t(6) > 2.168, p < 0.001), indicating that some information was stored about every speech sound within the sequence. We also visualized the relationship between the response made by the participant (in degrees) and the vowel quality of the target (also given in degrees in the circular 1D acoustic space we used). In line with the finding of a decrease in precision due to an increase in variability in responding, there is more scatter from a diagonal when memory load is increased.

On the one hand, memory representations reduce in quality with an increase in load. On the other hand, they appear to become less continuous (more categorical) with as more information has to be retained. This suggests that as load is increased, responses made by participants are more influenced by representations of the discrete identities of the speech sounds (speech categories) stored in the mental lexicon. Further analysis is reported below to verify this interpretation.

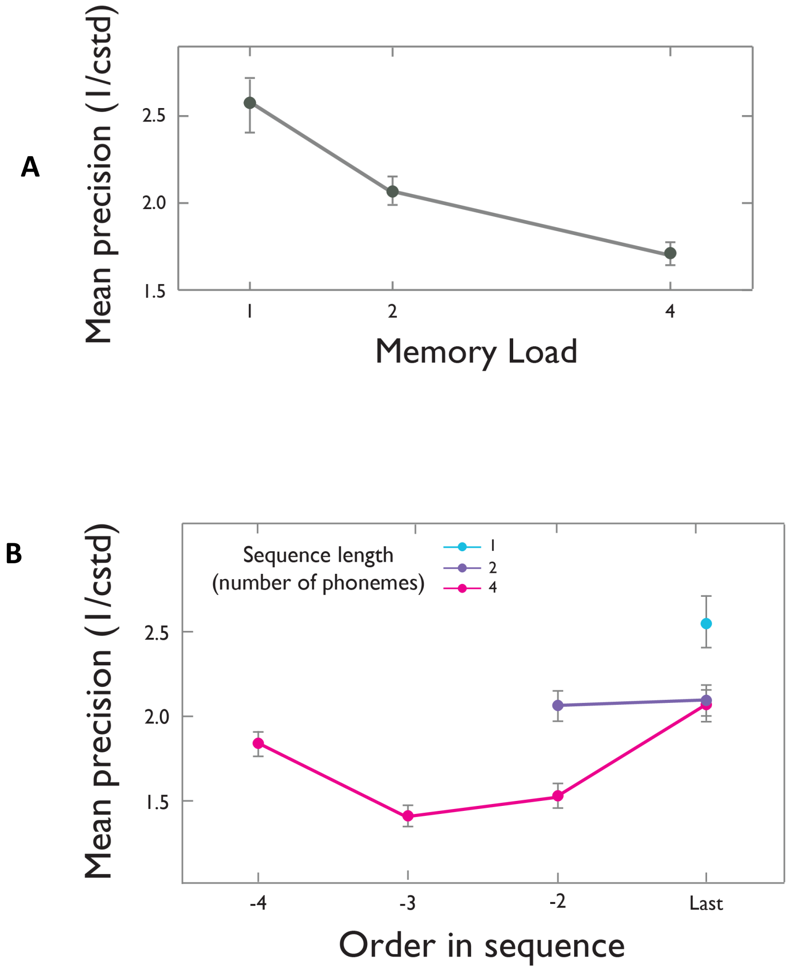


Figure 3-3 *Precision of recall varies with total memory load and serial order*

(A) Overall mean precision by memory load. The plot shows how precision decreases with an increase in memory load (number of phonemes within a sequence). Error bars represent one SEM. (B) Mean precision plotted by order in the sequence for different memory loads, denoted by different colours. Error bars represent one SEM.

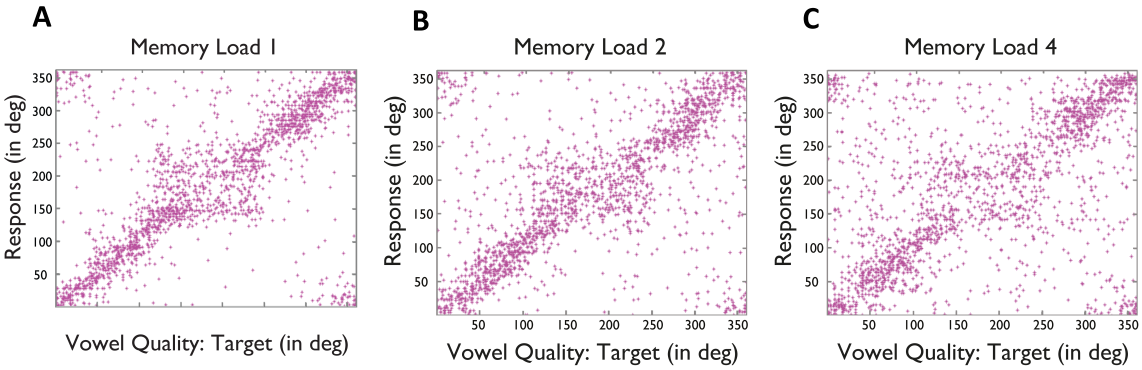


Figure 3-4 *Target – response relationship across memory load*

For memory load 1 (A), load 2 (B) and load 4 (C), each figure contains the data points of all trials obtained from the entire group of 21 participants. There is a strong correlation between the quality of target vowel and participants' matching responses. In addition to clustering around the diagonal (corresponding to the region around the vowel /u/), there was also evidence of clustering of responses in particular regions along the diagonal which might correspond to responses made to the category a speech sound belongs to. Clusters at the top left and bottom right corner of each scatterplot, near 0^{o} or 360^{o} are due to the circular nature of the vowel space.

3.4.2 Multi-dimensional scaling analysis

MDS (see Methods) allowed us to map responses made by participants back onto the vowel space we designed for the purpose of our study (Figure 3-1). In this way, we determine which regions appear as distorted due to representations becoming noisier with an increase in information load. We can also map out the locations along the vowel space, where any response clusters emerge. MDS solutions for a given memory load were computed to map memory representations back onto the stimulus space (Figure 3-5). There was an overall shrinkage of the vowel space as a whole when memory load increased. This observation is in line with our previous results, which revealed that memory representations become noisier with an increase in load, where fine-grained acoustic differences among neighbouring vowel sounds are lost.

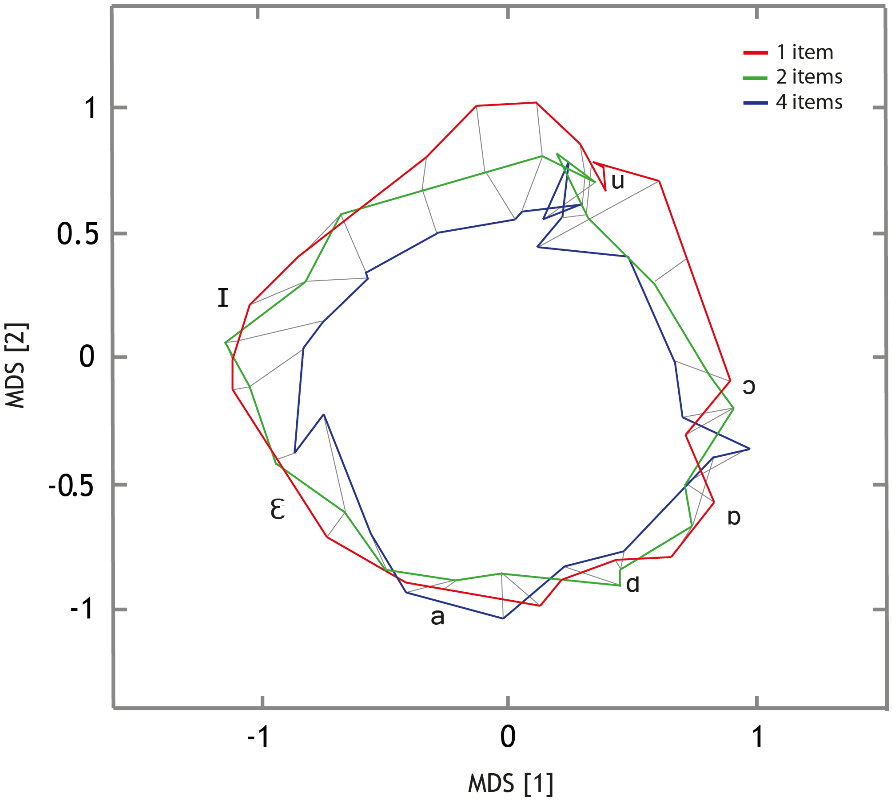


Figure 3-5 *Multi-dimensional scaling solutions (MDS)*

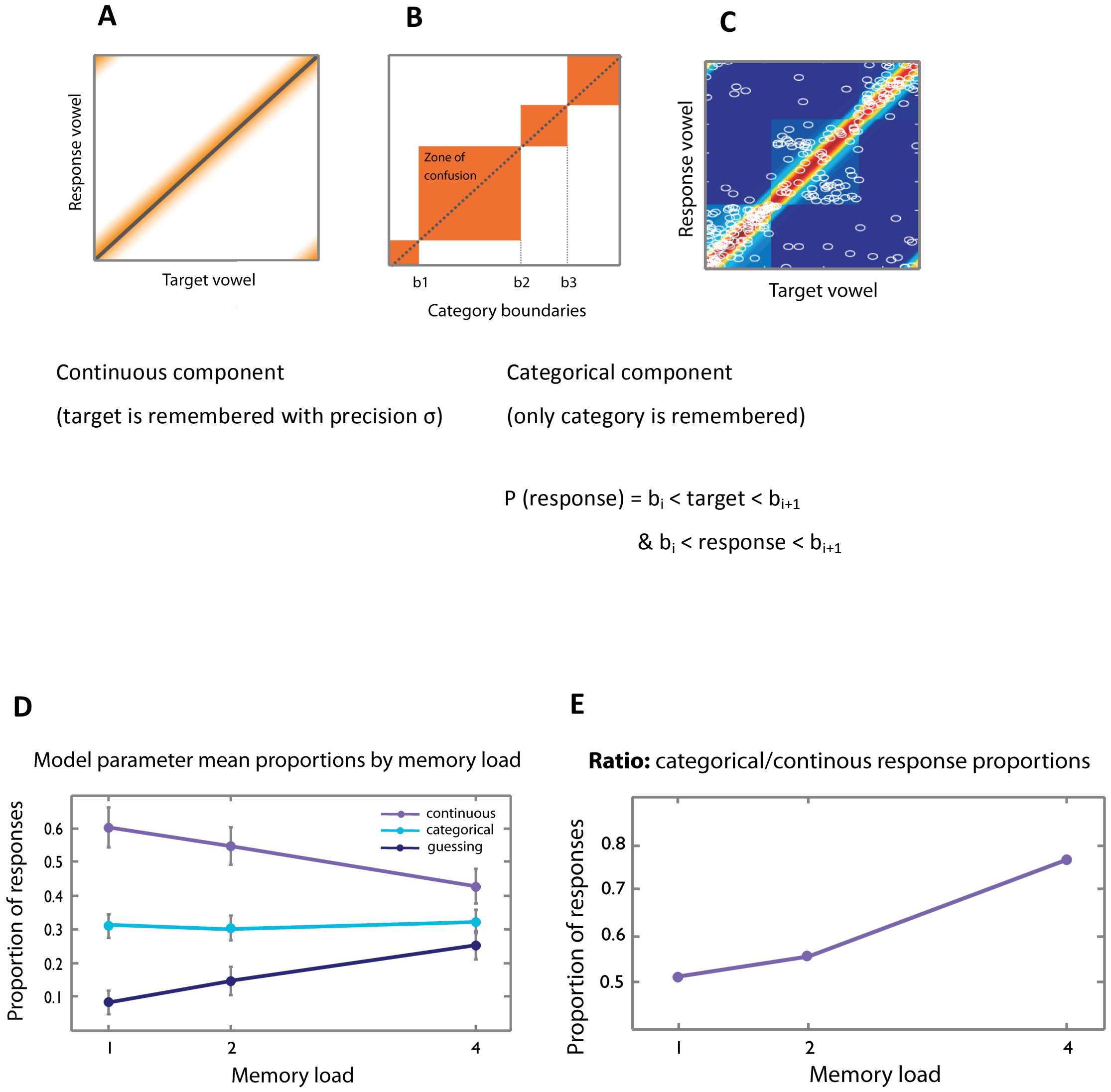
Each circle shows an individual MDS solution for the target-response data obtained for each memory load (red = load 1, green = load 2, blue = load 4). Corresponding points connected by grey lines, and phonetic symbols indicate the position of British English vowel categories. The solutions were centred at the origin based on the average position of the points, and rotated to maximise the comparability of the MDS solutions. There was an overall shrinkage of the vowel space as a whole when memory load is increased and a particularly strong shrinking of the space around /u/.

As also suggested by the linear regression analysis, the shrinking in the MDS solutions was not uniform across the vowel space. When memory load increased, there was a particularly strong shrinking of the space around the back vowel /u/, as well as some shrinking of the space toward the high-front corner of the space. That is, distances between individual points were reduced and the overall space collapsed toward the centre in these regions. In contrast, areas of the vowel space that were more dense in terms of English phonological vowels were relatively unaffected by memory load. MDS analyses captured the actual data well, as there was a significant correlation between MDS model fits and distances between targets and responses directly derived from the data (load 1: $R^2 = 0.87$, load 2: $R^2 = 0.88$, , load 4: $R^2 = 0.83$). However, although MDS is useful to further visualize the relationship between targets and responses in relation to the vowel space, it is not possible to examine different effects of categorization and perception independently. Thus, in order to decompose responses into different types (e.g. continuous, categorical and random guesses), we applied a mixture model to our results.

3.4.3 Mixture Model

The mixture model was employed to examine *sources of error* contributing to memory performance (see Figure 3-6), as previously described for visual (Bays et al., 2009; Zhang, & Luck, 2008) and auditory WM (Kumar et al., 2013). However, the components employed in the model here differ from previous studies because we model in categorical responses here. As the number of speech sounds increased within a sequence, responses directed at the target (probed item) became increasingly variable. This is indicated by a significant decrease in the mean proportion of responses captured by the *continuous* response component across memory loads (global $p < 0.0001$; decrease in the proportion of responses of 6% from load 1 to load 2 (95% CI: 13% to 1%); and a decrease in the proportion of responses of 18% from load 1 to load 4 (95% CI: 25% to 11%), (see Figure 3-6C, purple line).

Although, there was no difference in the absolute proportion of *categorical* responses (global $p > 0.05$), (see Figure 3-6C, cyan line), the ratio between categorical and continuous responses, increased with memory load (see Figure 3-6D). Thus, there was a greater reliance upon categorical responses compared to continuous ones with increasing WM load. Finally, the proportion of *guessing* increased with an increase in memory load (global $p = .001$; increase of guessing of 4% from load 1 to load 2 (95% CI: 1% to 10%), and an increase of guessing of 10% from load 1 to load 4 (95%CI: 4% to 16%), (see Figure 3-6C, blue line). Finally, there was no significant increase in the standard deviation of the Gaussian component (fitted sigma) with an increase in memory load ($p = 0.17$).



Continuous component

(target is remembered with precision σ)

Categorical component

(only category is remembered)

$$P\text{ (response)} = b_i < \text{target} < b_{i+1}$$
$$\& \; b_i < \text{response} < b_{i+1}$$

Figure 3-6 *Probabilistic model for information that is both continuous and categorical*

Response errors were decomposed into three components: (A) Scatterplots describing the relationship between responses made to target stimuli (Fig. 4) show that a large part of responses are closely aligned along the diagonal, corresponding to matching closely the target vowel. In the model, therefore, one component captures such continuous responses directed at the target vowel. It is described by a Gaussian normal distribution, centred on the vowel quality of the target (probed item), where the target is remembered with precision σ. (B) A second component captures categorical responses directed at the category a target vowel is contained in. Each category is described as a 'zone of confusion', described by a uniform distribution. A third component (not shown) captures random (guessing) responses unrelated to any of the vowels presented in the memoranda as described by a uniform distribution. (C) Example of a single subject's fitted data, showing how the sum of the continuous and categorical components captures the relationship between responses made to targets. (D) Mean proportion of responses captured by each component. While the proportion of continuous responses decreases with an increase in memory load, categorical responses remain constant and guessing increases. (E) Ratio of categorical to continuous responses (second component) increase with memory load.

3.5 Discussion

To examine the fidelity of representations in auditory WM for speech, participants were presented with sequences of speech sounds of variable length. The precision with which listeners matched the vowel quality of a probe (composed of a particular F1 & F2 value) to the vowel quality of a target was analysed. Importantly, we used an analogue response method, instead of a binary/categorical one, to obtain precision. This method allowed us to estimate WM capacity in terms of its resolution. In particular, we measured how information load influences the quality of speech sound representations in WM, as well as whether a representation can be captured as continuous or categorical.

Using this method, the results presented here demonstrate that precision decreases with an increase in memory load (Figure 3-3A). Thus, speech sound WM representations degrade in quality as information load is increased. Additionally, our findings reveal that the quality of speech sound representations is continuous at low memory loads but there may be a tendency for more categorical responses at higher loads. For a memory load of 4 items, there was also a clear recency and primacy effect. Such U-shaped serial position curves (Figure 3-3B, in pink), demonstrating that the first and last items in the sequence are remembered best, have been reported in many verbal WM studies using binary (recall/no recall) methods: for letters or digits (Baddeley, 1986; Conrad, & Hull, 1964; Levy, 1971), syllables (Gupta, Lipinski, Abbs, & Lin, 2005) and words (Haberlandt, Lawrence, Krohn, Bower, & Thomas, 2005).

The finding that memory precision for speech sounds declines as the number of items to be maintained increases is consistent with shared resource models of WM, that have been introduced for visual WM (Wilken, & Ma, 2004; Bays, & Husain; 2008). Such models predict that as the amount of information is increased, each item is remembered less precisely, which has been shown to account for processes in visual WM (Wilken, & Ma, 2004; Bays, & Husain; 2008, Bays et al., 2009; Fougnie, Asplund, & Marois, 2010; Gorgoraptis, Catalao, Bays, & Husain, 2011; Zokaei, Gorgoraptis, Bahrami, Bays, & Husain, 2011). The results demonstrate a drop in precision of WM, even when adding just a single phoneme to a previous one held in memory (Figure 3-3A). This decline in precision is difficult to explain on the basis of a fixed object capacity account, which predicts peak performance until the object capacity limit is reached (Cowan, 2001; Luck, & Vogel, 1997).

However, even when 4 phonemes were maintained in WM, a fixed item limit was evidently not reached at this stage as performance remained significantly above chance. This held for each individual item (serial position) within the sequence. Again, this observation is difficult to explain on the basis of a fixed item capacity account, which predicts chance performance as soon as the object limit is breached, as excess items cannot be represented. Instead, even at the highest memory load of 4 phonemes, at least some information was still recalled about every item.

The scatter plots of target versus response (Figure 3-4) add further detail on this representational noise (response variability). Participants' responses did not deviate far from the target when a single item was presented, (Figure 3-4A) in comparison to higher loads. Therefore, as participants respond closely to target sounds spanning the entire formant continuum, vowel perception and reproduction can be considered continuous when a single item has to be reproduced. As soon as further speech sounds were added to the sequence, the representation of items became noisier (Figure 3-4), but still with a tight zone of responses around the target, and a suggestion of increased clustering around specific phonemes. In other words, with the addition of further items, memory quality of a speech sound appears to be both weaker, but also less homogeneous over the vowel space. This observation led to further analysis, which was conducted particularly to clarify whether loss in continuity in speech sound representations might be due to the influence of *categorical* representations.

In order to relate speech sound representations held in WM to actual British vowels, we applied a multidimensional scaling (MDS) algorithm to our data. This allowed us to map responses made by participants back onto the vowel space. We were interested in determining which regions of the space appear distorted as well as where response clusters emerge along the space. We found that the vowel space shrinks as memory load increases, corresponding to an increased confusability between all vowels. Moreover, there was greater shrinking in regions where there are fewer English vowel categories, particularly near the vowel /u/. That vowel in most varieties of British English has become fronted toward /i/, such that there are no categories in the high-back region of most British English vowel spaces. There was also some shrinking of the space in the high-front corner of the space, where there were also few English vowels because our circular path missed /i/. Our listeners thus selectively made more errors with increasing memory load regions with fewer categories, and there was less shrinking along the continuum near regions that had more British English vowels. These results are thus consistent with the hypothesis that listeners relied more on the category identity of the vowels as memory load increased.

To examine different effects on responses (e.g. continuous vs. categorical) in separation, we applied a mixture model to our results to decompose different types of responses. The model confirmed that the increase in variability of memory for the target item with load was associated with responses directed at the target and speech representations becoming noisier (Figure 3-6C). This finding is consistent with the view that a limited resource has to be shared out across all items that are being stored. As more items are stored, the representation of each item becomes noisier (Wilken, & Ma, 2004; Bays, & Husain; 2008; Ma, Husain & Bays, 2014). Additionally, a gradual increase in random guesses corrupted memory for the target item, which would be inconsistent with an item limit predicting a sudden sharp increase in guessing once a capacity limit has been breached.

In the mixture model, a third component captures responses directed at a given target category. There were three categories, where each one was modelled as a ‘zone of confusion’. The proportion of categorical responses was constant across memory loads (Figure 3-6C). In this respect, categorical responding could represent a memory strategy, which is constantly available to participants; as such representations are not transient, but stored in the metal lexicon. However, the categorical response component has to be regarded in relation to other types of possible responses. It seems that participants are more likely to make use of the categorical response strategy, as the memory task becomes more demanding (with an increase in load). This is reflected in the ratio between categorical and continuous responses, where the relative proportion of categorical responding increases with load (see Figure 3-6D). Thus, as representations of the target vowel become less continuous, memory for the actual stimulus fades and participants begin to guess, whilst their memory for the stimulus category is retained. Therefore, categorical memory may be more resilient to interference from other items.

In line with our results, it has been shown previously that *perception* – rather than recall – of vowels is more continuous for low loads compared to more categorical for higher loads (Macmillan, Goldberg, & Braida, 1988; Pisoni, 1973, 1975; Repp, Healy, & Crowder, 1979; Iverson, & Kuhl, 2000). These previous observations are based on acoustically narrow stimulus dimensions and 'binary' (recalled/not recalled) response methods. However, results from those studies have been discussed in the context of a dual process theory to explain the shift from continuous to categorical memory representations (Fujisaki & Kawashima, 1971). As speech sound qualities are acoustically distinguishable from one another as variations of timbre, dual process theory suggests that short-term memory represents timbre in a temporarily fragile manner.In contrast, the mental lexicon (part of long-term memory) is considered to contain representations based on speech sound categories, which in our task subjects may have relied upon at higher loads.

While WM represents timbre of a single item acoustically at a high resolution (i.e. continuous representation, as shown here), the addition of further information leads to a loss in memory resolution of each speech sound in our study, consistent with the view that a limited resource has to be shared out across all items that are being stored. On a neural level, this might mean that for the lowest memory load, both WM and speech perception processes are recruited, possibly involving common neural substrates (Ravizza, Hazeltine, Ruiz, & Zhu, 2011). In contrast, an increase in load might in addition lead to reactivation of speech representations in LTM (Hickok, & Buchsbaum, 2003; Hulme, Maughan, & Brown, 1991). Mapping low-resolution representations onto existing categorical speech

at encoding

representations in LTM might serve as a strategy to support recall. One might therefore speculate that the results presented here suggest that as information load is increased, categorical representations in the mental lexicon are recruited to support recall as more information has to be maintained in WM.

These categorical representations, with WM as activity ‘within’ long-term memory, have two advantages. Firstly they may be more efficiently encoded, since only a limited number of discrete states are allowed. Secondly, neural models of sustained activity suggest they may be more resilient to noise, since they are encoded by point attractors rather than continuous attractors. For the first time, we are able to directly compare the properties of these two distinct kinds of memory representation, using the same stimuli and response modality.

As the paradigm currently stands for vowels, the continuous, analogue speech-matching task introduced here might provide a sensitive means of assessing verbal WM not only in healthy participants but also patients with brain disorders. Impairments in verbal WM may result from developmental language conditions (e.g. dyslexia, specific language impairments (SLI)), neurodegenerative disorders (e.g. dementia) as well as from focal lesions (e.g. dysphasia following stroke), (Brain & Michael, 2009; Brandt et al., 2003). In many studies, patients’ verbal WM capacity is usually compared to a fixed capacity limit (number of items) found in controls. Such a quantized view has proved to be very useful. However, it might be even more informative to gain insights into the *quality* of memory representations of speech – not just the number of items that might be stored – for individuals with different types of language deficit. In vision, WM precision has already been used in patients with memory disorders (Pertzov et al, 2013) while a developmental study has shown how visual WM representations become more precise with age (Burnett, 2012).

In summary, the method used here provides a means to obtain an index of the variability of verbal WM representation around its true value, very different from traditional binary (recall/no recall) measures. Increases in memory load affected the quality of speech sound representations leading to a gradual decline in precision of recall, consistent with the view that WM representations become noisier with an increase in memory load, analogous to the findings in vision (Wilken, & Ma, 2004; Bays, & Husain; 2008, Bays et al., 2009; Fougnie, Asplund, & Marois, 2010; Gorgoraptis, Catalao, Bays, & Husain, 2011; Zokaei et al, 2011). Unlike previous findings in vision, the quality with which auditory information (speech sounds) is stored is not only described by precision, but also in terms of how continuous or categorical a representation is. Thus, in addition to finding general principles that apply across WM representations for different sensory modalities, our paradigm offers a novel method of probing the interaction of continuous and categorical information storage in the brain, within a single task. This opens the way for investigation of the neural basis of these two distinctive kinds of WM, and may help bring together two hitherto separate approaches in working memory research.

Chapter 4. Remembering more than a single touch: working memory for vibrotactile sequences

4.1 Abstract

Although tactile memories play a significant role in our everyday lives, little research has addressed the limits underlying tactile WM. Previous studies suggest that our capacity is limited to the storage of a single stimulus only. Vibrotactile representations are highly fragile as they decay rapidly and are subject to interference from distractor stimuli. However, as other sensory modalities can store multiple items, the current study investigates whether we can hold more than a single touch in vibrotactile WM.

Previous investigations used the method of change detection: a single frequency (continuous parameter) is memorised and then compared to a probe. Although, it has been shown that frequency representations in WM are continuous on a neural level, responses are binary (change/no change detected), not continuous. We designed a *continuous response* method to assess the fidelity of memory representations. Participants memorized sequences of vibrotactile stimuli at different frequencies of variable lengths (1-5 items). Their memory for one of the stimuli was then tested in a frequency-matching paradigm. Deviation between target and response served as a continuous measure of response error, from which recall precision was obtained.

A significant decrease in precision was found, when one item was added to a previous one (memory load 1 vs. memory load 2), and further decline in performance up to the highest memory load of 5 items. Performance remained above chance even at the highest memory load of 5 items, i.e., above any putative item limit. The overall pattern of results is consistent with the resource model account of WM. Another interesting finding was that the resolution of receptive field size of vibrotactile units (responding to different types of vibrations in the

frequency-range of flutter and fusion) at encoding corresponds to the resolution with which information is stored and represented in WM.

4.2 Introduction

Previous studies using tactile stimuli have shown that the cognitive limitations underlying their processing may be more severe than those associated with the processing of basic stimuli in other sensory modalities (vision and audition: Gallace et al, 2008; Gallace, Tan & Spence, 2006, 2007; Gallace & Spence, 2008, 2009). In contrast to previous studies on vibrotactile WM suggesting that our capacity is limited to the storage of a single stimulus only (Bancroft & Servos, 2011; Bancroft, Servos & Hockley, 2011), here we ask, if vibrotactile WM can store multiple items.

Most previous studies only tested WM performance for a single item (Romo et al., 1999; Preuschhof et al., 2006; Hegner et al., 2010; Spitzer et al., 2010; Spitzer & Blankenburg, 2011; Haegens et al., 2010) although it is well known that other sensory systems (e.g. vision and audition) can usually store information beyond a single item. Interestingly, Bancroft, Hockley & Servos (2012) tested vibrotactile WM for two items, where memory performance remained above chance, showing that we can store 2 frequencies in WM. Can we store more than 2 items in somatosensory memory?

Evidence from other macro-geometric somatosensory stimuli (location) suggests that we can recall between 3-6 items depending on response method (Gallace et al., 2008; Auvray, Gallace & Spence et al, 2011). However, stimuli used in those studies recruit neural systems different from the ones associated with WM storage of micro-geometric stimulus properties such as mechanical vibrations (O’Sullivan, Roland & Kawashima, 1994; Roland, 1987; Roland, O’Sullivan & Kawashima, 1998; Gallace & Spence, 2009).

In humans, although a lot has been learnt about visual working memory (WM), in recent years, controversy has arisen about WM capacity: Is it limited to a small number of objects (estimated to be ~3) or is it best considered as a resource without any item-limit? Previous studies in different sensory modalities (including somatosensory/tactile, see Figure 1-3C) used the method of change detection to assess memory capacity. On each trial, one or more stimuli are presented at the stage of memory encoding. After a given delay period, a stimulus reappears and subjects have to decide whether it is the same or different from the previous one/s. As a result, responses are binary (change/no change detected).

There have been concerns about discrimination tasks in the visual domain, which, as I have discussed, have led to the development of continuous, analogue measures of WM precision, describing mnemonic resolution. (Wilken & Ma, 2004; Bays & Husain; 2008, Bays et al., 2009). Investigating the fidelity of memory representations with measures of precision, it has been shown that visual WM might best be considered as a resource without any item-limit which can be flexibly allocated to objects in the visual scene (Wilken & Ma, 2004; Bays & Husain; 2008, Bays et al., 2009; Fougnie, Asplund & Marois, 2010; Anderson et al., 2011; Gorgoraptis et al., 2011; Zokaei et al., 2011; Salmela et al. 2012). Similar findings have been obtained for auditory WM (see chapter 2 or Kumar et al, 2013, chapter 3). It remains to be established, whether tactile WM can also be described as a limited resource without an upper item limit.

A second question concerns the relationship between resource allocation and densities of tactile receptors in the skin. Two types of receptors can be found at different locations of the skin, where receptor density corresponds to receptive field (RF) size (Meissner corpsules: small RFs vs. Pacinian corpsules: large RFs (Schiff & Foulke, 1982). According to the 'duplex theory of vibration' (Katz & Gibson, 1982; Verillo, 1968; Verillo & Gescheider, 1979; Gescheider & Verillo, 1979), while Meissner corpsulses respond to low frequency vibrations in the range of 5-40Hz (Mountcastle, 1984) or up to 50Hz (Gordon, 1978), (vibrotactile flutter), Pacinian corpsules are selectively sensitive to higher frequencies between 40Hz-400Hz (Gordon, 1978) or 60-300Hz (vibrotactile fusion), (Mountcastle, 1984, Vallbo & Johansson, 1984). Hence, mechano-receptive units are well studied for frequency detection, but much less is known about the neural processes involved in the representation and maintenance of vibrations in WM. We hypothesize that the fidelity of WM representations is scaled to RF size, 'flutter' is represented at a higher fidelity than 'fusion'.

It has been shown that vibrations in the frequency range of flutter are represented in primary and secondary somatosensory cortex (SI, SII), (Salinas et al., 2000), medial premotor cortex (MPC) and prefrontal cortex (PFC), (Romo et al., 1999; Romo & Salinas 2003; Preuschhof et al, 2006; Spitzer et al, 2010), (Wang, Bodner & Zhou, 2013). The role of PFC is special for memory maintenance, as neural delay period activity is systematically modulated only in this region (Romo et al., 1999). Neural discharge rates vary as a monotonic function of the encoded stimulus frequency, describing representations of vibrotactile stimuli in WM. Hence, a continuous parameter (frequency) is represented in a parametric/continuous fashion (monotonic discharge). However, the typical discrimination task used to quantity this relationship doesn't assess WM performance in a continuous but binary manner.

On tactile discrimination tasks (delayed-match to sample), where a frequency is encoded and maintained in memory, before it is compared to a probe frequency, one is given two response options only: change or no change. Thus, instead of testing memory performance in a continuous way, it is tested in a purely binary fashion. Consequently, the task cannot quantify the quality of a memory representation, when one fails to make a correct response. Even in this case a representation of this stimulus may still exist, which may simply be of a lower resolution compared to a stimulus recalled correctly. For example, it has been shown that we encode irrelevant virbotactile information in WM even when explicitly instructed to ignore distractors (Bancroft & Servos, 2011; Bancroft, Hockley & Servos, 2011). Hence, it would be more informative, instead of capturing whether a stimulus is represented or not, to capture its mnemonic resolution.

The current study aims to test whether we can store multiple items in vibrotactile WM. How is memory performance affected by an increase in memory load? The resource model predicts a monotonous decrease in memory precision with an increase in memory load, without a discrete step that would suggest an item capacity limit. According to the resource model, if no item has priority to-be-remembered, the fidelity of an item's representation is inversely proportional to the number of items stored in WM (for visual objects: Bays & Husain, 2008; Bays, Catalao, & Husain, 2009, Gorgoraptis et al., 2011).

The present study was designed to assess the qualitative nature of memory representations in vibrotactile WM. It also addresses whether the resolution of receptive field size of vibrotactile units (responding to different types of vibrations in the frequency-range of flutter and fusion) at encoding corresponds to the resolution with which information is stored and represented in WM. A continuous response method (frequency matching) was used to quantify the variability of memory recall around its true stimulus value. The aim was to determine how an increase in information load affects recall variability, which may reflect noise in neural representations providing a framework for the interpretation of neurophysiological data.

4.3 Materials and Methods

4.3.1 Participants

13 healthy young adults (8 female, mean age 21 years, age range: 19-26) participated in this study after providing written informed consent to procedures approved by the local ethics committee. Participants were selected based on the following criteria: normal hearing and no musical training, for the reason that e.g. being able to play an instrument or having an auditory impairment may be of a tactile advantage (enhanced perceptual sensitivity and memory).

4.3.2 Stimuli

Stimuli consisted of mechanical vibrations, logarithmicallly sampled at random from a frequency range of 10-100Hz, with peak amplitude of 2mm. The stimulus duration was 1s with an ISI and a delay period of the same length. The beginning and end of each trial were indicated by text appearing onscreen. There was a minimum separation of 2 semitones between any two vibrations within the sequence. The probe stimulus was randomly selected from the same frequency range of 10-100Hz and participants could make a response in the range of 1-110Hz by adjusting a dial (PowerMate, Griffin Technology). All frequencies were delivered to the right index finger and adjustments of the probe frequency were made with the left hand using a rotary dial device.

4.3.3 Apparatus

The experimental apparatus was constructed in the following way: we used a vibration test system - shaker model V101-PA 25E (by Bruel & Kjaer), (see Figure 4-1, left). A screw was mounted on top of this device and a small rubber patch of 1 cm in diameter was glued on its head. Participants placed the tip of their right index finger on the patch. The device functions on the principles of a loud speaker and was connected to an audio amplifier. A computer was used to send sine waves (pure tones) to the amplifier. Sine waves of the desired frequency were generated in Matlab 6.5 (Mathworks Inc.) and presented using Cogent (http://www.vislab.ucl.ac.uk). To mask any residual sound from the device, participants had to wear foam earplugs and were also presented with white noise over headphones (Sennheiser HD 380 pro). Furthermore, the device was hidden from their field of view.

4.3.4 Design and Procedure

On each trial, participants were presented with a sequence of mechanical vibrations at different frequencies (see Figure 4-1). Sequences varied in length: 1, 2, 4 or 5 vibrations. At the end of each sequence recall of one of the vibrations was probed, indicated by a number on the screen, e.g. 4 for 4th vibration. A randomly selected probe vibration was then presented, whose frequency had to be adjusted to match the frequency of the vibration of the probe, using the dial.

Participants were required to perform this frequency matching task in a maximum response window of 20s. After completing 2 practice blocks of 50 trials each, where memory was probed for a single item, each participant completed 5 blocks of 80 trials each with equal number of trials for each memory load. Participants were given the option to take a break after 40 trials had been completed on each block. The experiment was split into 3 experimental sessions, 1 for the practice and 2 for the main experiment, of approximately 45 minutes each.

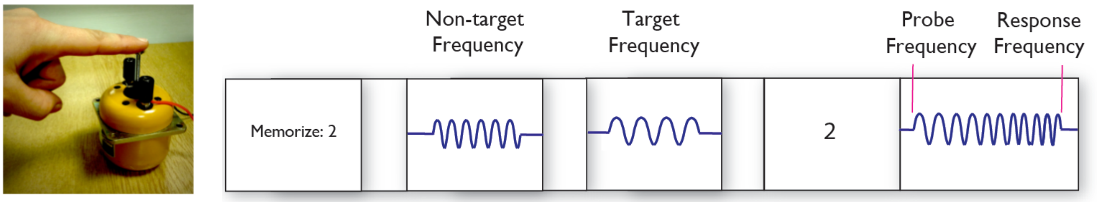


Figure 4-1 *Experimental paradigm*

Sample sequence: A message appeared onscreen indicating the number of vibrations to be memorized within a given trial (here: 2). Next, the sequence of vibrations was presented (e.g. vibration 1 at 32Hz and vibration 2 at 18Hz). The sequences were comprised of 1, 2, 4 or 5 vibrations. After all test stimuli were presented, a number appeared on the screen indicating the target. A randomly chosen probe stimulus (e.g. 15Hz) was then presented, which had to be adjusted to match the frequency of the target (here: first vibration with frequency of 32Hz).

4.3.5 Data Analysis

Response error (Ɛ, in semitones) was calculated as the deviation between the vibrational frequency of the response made by the subject and the target frequency for each individual trial. The following equation was used,

$$\mathcal{E} = 12 * log_2\left(\frac{fm}{fa}\right)$$

where f_m is the measured or matched frequency (i.e. response made by the subject) and f_a is the actual frequency of the target (stimulus quality tested). Precision (*P*) was calculated as the reciprocal of the standard deviation of response error (*P* = 1/std), just as in previous studies (e.g., in vision Gorgoraptis et al. 2011; in audition: chapter 2 or Kumar et al., 2013, chapter 3). Precision was calculated separately for each subject, memory load and serial position. Each subject was presented with sequences of vibrations of variable length (memory load), where each vibration appeared at a different serial position (e.g., memory load 1 has serial position 1 and memory load 2 has serial positions 1 and 2 etc.). As the distribution of precision values across the group was positively skewed, a log transform was applied to precision values, which was then subjected to multilevel modeling (Laird & Ware, 1982). We fitted unadjusted and adjusted multilevel models to account for the fact that we had repeated measurements from individuals at different positions using different memory loads (Laird & Ware, 1982). Tests were performed in Stata 11.2 software (StataCorp LP, College Station, Texas) and all p-values < .05 were considered significant, using 2-tailed tests. ANOVA was used to test for a difference between frequency ranges (flutter vs. fusion), after splitting the data set accordingly.

The level of chance performance was derived from a permutation test, which was carried out in the same way as described in chapter 3 (see 3.3.5 Estimating level of chance performance). The mean level of chance performance across the group amounted to 10.4 ST, SEM=0.31. Task performance was tested against this particular chance value at every memory load and serial position, using a one-sample t-test.

A power law was fitted to our data using maximum likelihood estimation to describe the relation between precision (P) and the number of items to be encoded (N),

$$P \propto N^{k}$$

where k is the power law exponent. Each participant memorized sequences of vibrations of variable length (memory loads), where each vibration appeared at a different serial position.

4.3.6 Proportions of different types of response error

In a further analysis, responses made by participants were decomposed into different categories, capturing different types of error.

Category 1: Responses directed at the target frequency
Category 2: Responses directed at the non-target frequency
Category 3: Random responses (guesses)

Responses directed at the target frequency are defined as responses centred at the target value. Only responses, which do not deviate further than ± 10.4 ST (chance performance threshold) from the centre of the target distributions fall into *Category 1*.

As participants make confusion errors and sometimes report the non-target instead of the target frequency, the second category captures responses directed at the non-target frequency. A threshold of 1 ST was used, as stimuli within a sequence were always presented at a 2 ST minimum distance. Thus, if the response error towards a non-target (deviation between the actual non-target value(s) and response made by the participant) (1.) fell below the threshold of ST, and (2.) the response error to the non-target was smaller than the response error to the target, responses on a given trial fall into *Category 2*. A given participant reports the frequency of a non-target item (misbinding), when the target number was associated with a different serial position from the one indicated. All other types of errors were classified as random guesses and fall into *Category 3*, which are neither related to the target nor non-target value.

4.4 Results

4.4.1 Effects of memory load and serial order on frequency matching

The precision with which participants performed the frequency matching was assessed for all memory loads (number of stimuli within a sequence; Figure 4-2) and all serial positions within a sequence (Figure 4-3). The results for the unadjusted analysis show that log-precision declined significantly with memory load (global $p < .00001$). The mean WM precision was 21% lower for two items compared to a single item (95% CI: 34% to 7%). As can be seen in Figure 4-2, there was a sharp drop in precision when a single item was added to a previous one (comparison between load 1 and 2). Such a significant change in mnemonic strength is difficult to explain on the basis of a fixed capacity account, which predicts peak performance until a capacity limit is reached.

Although previous research suggests that vibrotactile memory capacity is limited to one item, as representations are highly fragile, decaying rapidly (Bancroft & Servos, 2011) and subject to interference from distractor stimuli (Bancroft & Servos, 2011; Bancroft, Servos & Hockley, 2011), most investigations simply did not assess memory performance items beyond one (Salinas et al., 2000; Romo et al., 1999; Romo & Salinas 2003; Preuschhof et al, 2006; Spitzer et al, 2010), except for a single study using two items (Bancroft, Hockley & Servos, 2012). The latter reported findings in line with the results obtained here, showing that vibrotactile WM can store more than one item. In the experiment reported here, precision was 25% lower for four items compared to one item (95% CI: 37% to 13%), as well as 29% lower for five items compared to one item (95% CI: 41% to 18%), demonstrating a further drop in memory resolution with the addition of more items.

A second finding here concerns the effect of serial position (a vibrations' order in a sequence) on recall precision (global $p < .00001$): 29% lower for the last item in the sequence (95% CI: 41% to 18%) compared to trials on which the first item in the sequence was probed (Figure 4-2), indicating a primacy effect. Precision was 2% lower (95% CI: 15% to 10%) for the second item, equal (0% different, 95% CI: 12% to 12%) to the third item and 9% (95% CI: 2% to 21%) higher for fourth item in the sequence, respectively, compared to the first item. The unadjusted analysis shows that factors, memory load and serial position were significantly associated with precision.

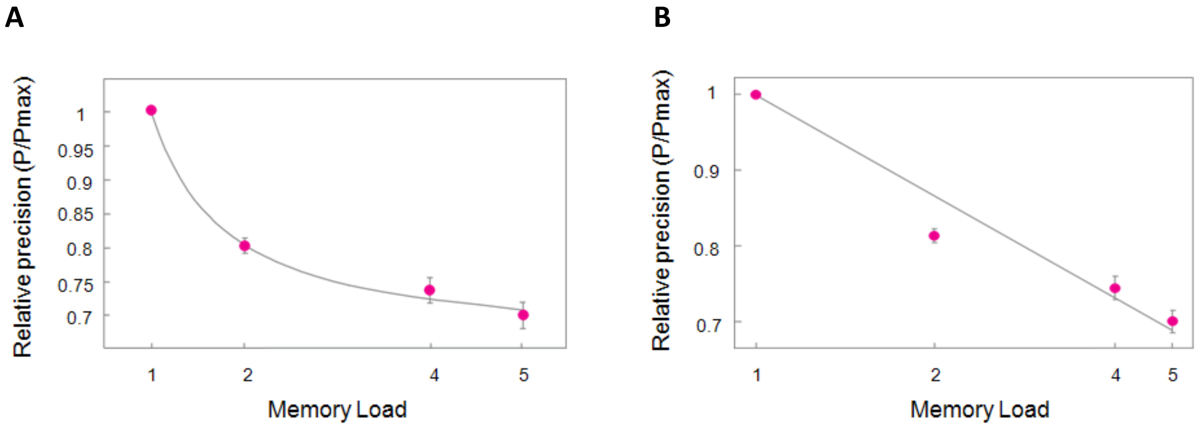


Figure 4-2 *Precision by memory load*

(A) Relative mean precision as a function of memory load (number of frequencies presented within a sequence (N)). Precision is defined as the reciprocal of the standard deviation of response error. Error bars indicate SEM. The grey line indicates the best fit to the data of a power law relating precision to the proportion of resource available to encode each frequency (1/N) for each memory load. Normalization is with respect to performance at memory load for a single frequency (N= 1). (B) Relative mean precision and memory load are plotted on a log scale (log-log plot), showing a linear relationship between both variables.

An adjusted analysis including both memory load and serial position in the model was performed in order to assess the independent effects of both variables. In this model, the effect of memory load was near significance ($p = .06$; decline in precision of 16% for load 2 compared to load 1 (95% CI: 28% to - 2%; decline in precision of 12% from load 1 to load 4 (95% CI: 24% to 1%), and a decline in precision of 15% from load 1 to load 5 (95% CI: 28% to 3%). Thus, there was a sharp drop in precision from load 1 to 2, but no further drop with the addition of further items, which may be due to the large recency effects observed for sequences of 4 and 5 items (Figure 4-3). The effect of serial position on precision remained significant ($p < 0.0001$). The interaction effect was significant ($p = .02$), suggesting that within a given memory load serial position has an influence on precision.

Next, to examine the effects of different potential frequency channels, the dataset was split into two frequency ranges (range 1: 10-50Hz corresponding to channel 1 (flutter); range 2: 51-100Hz corresponding to channel 2 (fusion)). There was a significant difference in error scores between channels (for load 1: F (1, 24) = 7.89, MSE = 21.96, p = 0.01, load 2: F (1, 24) = 6.87, MSE = 22.08, p = 0.01, load 4: F (1, 24) = 6.28, MSE = 18.95, p = 0.02, load 5: F (1, 24) = 3.44, MSE = 10.48, p = 0.08). Thus WM resolution appears to be scaled to receptive fieldsize/receptor density and their properties (e.g. slowly vs. rapidly adapting) within a given channel. Thus frequencies in the flutter range appear to be represented in WM at a higher fidelity compared to frequencies falling in the fusion range.

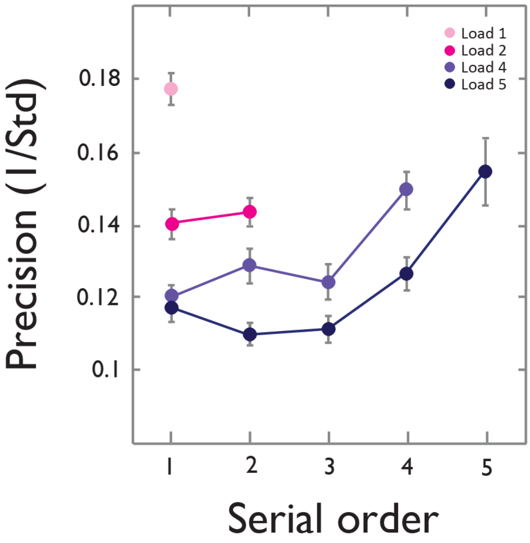


Figure 4-3 ***Precision by serial order***

Recall precision is modulated by memory load and the serial order in which frequencies appeared. Precision is plotted against order in the sequence, i.e. the serial position at which a probed item appeared in the sequence. Each coloured line represents a different sequence length (memory load). The last frequency in the sequence was remembered most precisely, while earlier items in the sequence were recalled less precisely (e.g. in particular for the highest memory loads of 4 and 5 items). Error bars represent SEM.

Finally, performance was significantly above chance for every combination of memory load and serial order (t(11) > 7.597, p < 0.001), indicating that some information about vibrotactile frequencies was stored at each serial position within each sequence.

4.4.2 Raw error histograms: target and non-target responses

Although participants reported information for each memory load and serial position above chance, we cannot exclude that their WM representations were corrupted in some way. For example, previous research has shown vibrotactile memory is influenced by other items in the sequence (inference (Bancroft, Hockley & Servos, 2011; Bancroft, Servos & Hockley, 2011)). This may lead to misbinding errors, where participants report non-target frequencies instead of target frequencies. To gain a better understanding of whether responses are directed at the target or non-target stimulus we plotted raw response error histograms (Figure 4-4) for all sequence lengths.

Each raw response error histogram is centred at the target (or non-target) frequency value and contains all trials collapsed across all participants. Response errors are deviations in either direction from the target: a negative error value is obtained, when one responds at a lower frequency than the target stimulus frequency and vice versa for positive error values. Responses made to the target are shown in pink for each memory load and serial position. As non-target frequencies within the sequence may lead to interference, participants sometimes respond to the non-target instead of the target frequency (misbinding error). Therefore, responses directed at the non-target (unprobed item) frequency were also plotted (in blue), for each memory load and combination of targets and non-targets as they occur at different serial positions within a sequence.

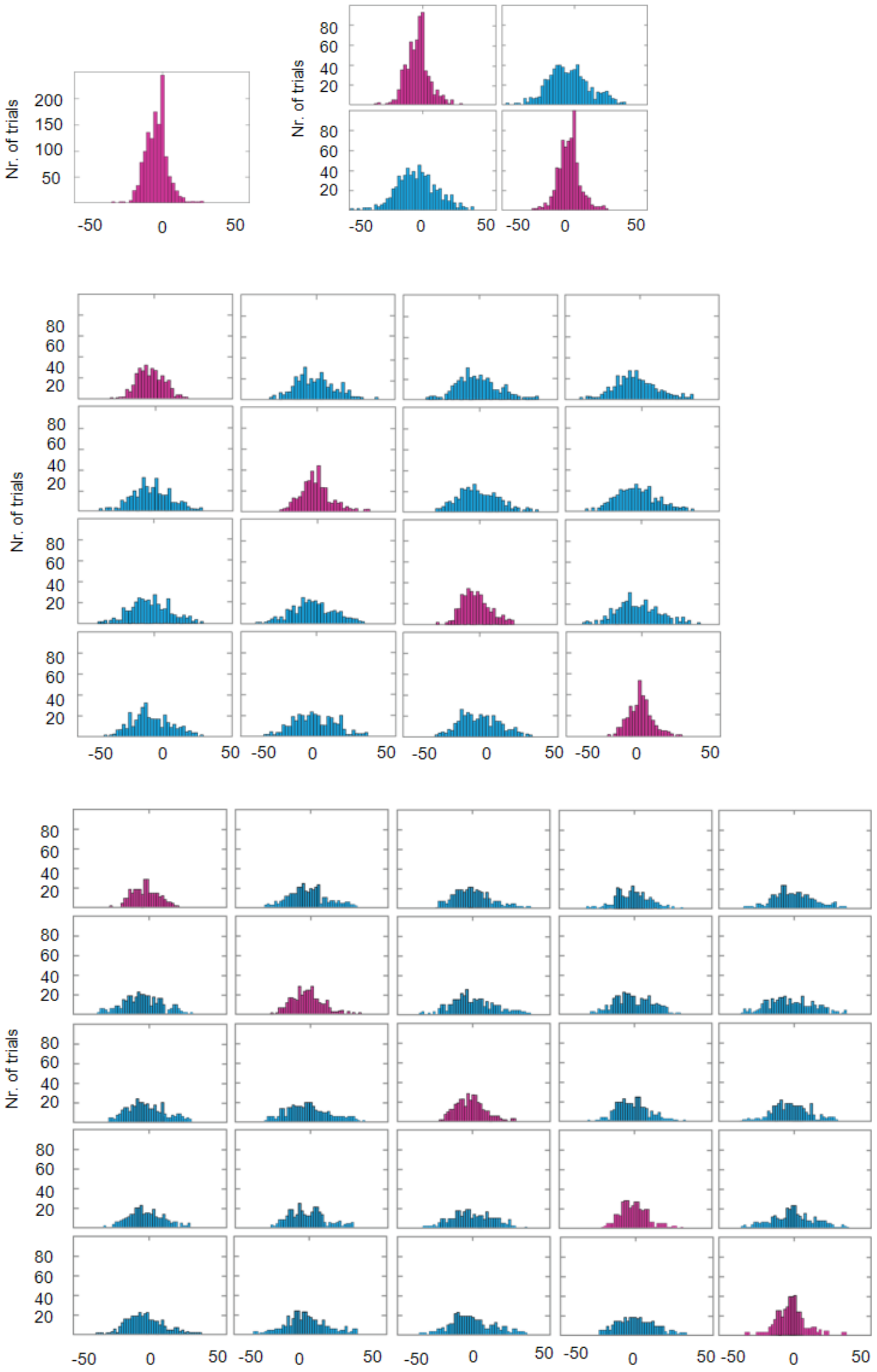


Figure 4-4 ***Raw error histograms***

Shown are the histograms of raw response error centred at the actual target (in pink) or non-target (in blue) stimulus value. The data is collapsed across all trials and participants (N13) and sorted by memory load and target – non-target serial order. The y-axis shows the total number of trials and the x-axis shows the response error in semitones as a deviation in either direction from the actual stimulus value of the target or non-target, respectively.

4.4.3 Proportions of different types of response error

Category 1: A one-way ANOVA revealed a significant main effect of memory load on the proportion of responses directed at the *target* frequency (F (3, 48) = 23.82, MSE = 0.204, p < 0.001). The proportion of responses falling into category 1 decreases with an increase in memory load (see Figure 4-5, purple line). LSD post hoc comparisons showed the following: the difference in responses falling into category 1 for memory load 1 and 2 was near significance (p=0.09). There was a significant difference between memory load 2 and 4 (p<0.001), but no difference between memory load 4 and 5 (p=0.345). Overall, there is a pattern showing that the amount of responses directed at the target frequency decreases with an increase in memory load.

Category 2: A one-way ANOVA revealed a significant main effect of memory load on the proportion of responses directed at the *non-target* frequency (misbinding), (F (3, 48) = 99.31, MSE = 0.123, p < 0.001). The proportion of responses captured by category 2 increases with an increase in memory load (see Figure 4-5, dark blue line). LSD post hoc comparisons showed a significant difference in the proportion of misbinding errors between memory load 2 and 4 (p<0.001), as well as a significant difference between memory load 4 and 5 (p=0.04).

Category 3: A final one-way ANOVA showed that there was no effect of memory load on the proportion of responding at *random* (guessing), (F (3, 48) = 1.07, MSE = 0.008, p = 0.37). Thus, the proportion of guessing was constant across memory loads (see Figure 4-5, cyan line).

The decrease in precision with an increase in memory load was therefore not due simply to increased guessing, but to increased variability of the representation of the probed target frequency in memory as well as an increased amount of interference from non-target frequencies.

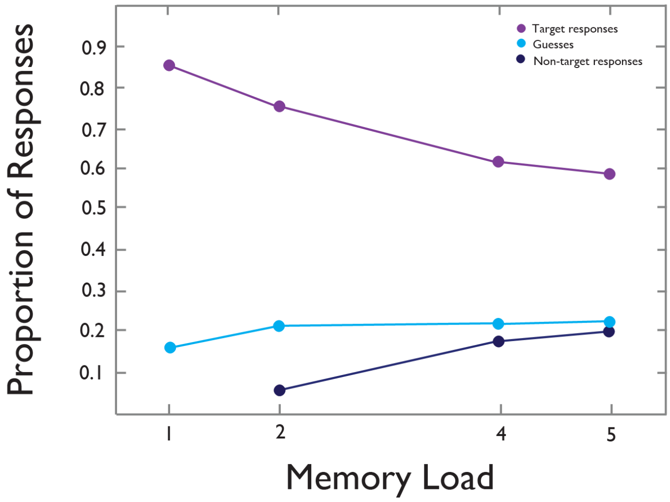


Figure 4-5 ***Types of response error***

The proportion of responses falling into either category: 1. Target responses (in purple), 2. Non-target responses (in cyan), and 3. Guesses (in dark blue) are plotted by memory load. Error bars (SEM) are not visible as they are smaller than the data points.

4.5 Discussion

The fidelity of vibrotactile WM representations was measured as recall precision in a frequency matching paradigm. The task required participants to memorize a continuous parameter – stimulus frequency, which was also assessed in a continuous fashion, using a frequency matching paradigm. By varying the number of vibrations presented within a sequence (memory load), storage of multiple items in vibrotactile WM was assessed. The results show a monotonic decline in memory precision with an increase in memory load (Figure 4-2), where performance remained above chance for the highest memory load of 5 items, suggesting that there isn't a fixed upper limit in the number of vibrations one can store in vibrotactile WM. Instead a shared resource model of WM best describes the results obtained here.

In line with previous findings in other sensory modalities, in vision (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Bays et al., 2009; Gorgoraptis et al., 2011; Wilken & Ma, 2004; Zokaei et al., 2011) and audition (chapter 2 or Kumar et al., 2013, chapter 3), the shared resource model captures the relationship between information load and mnemonic resolution strength at which information is represented. The model predicts that the more information held in working memory, the less precisely each item is represented. Here, we confirm this prediction for the sensory modality of touch, by showing a clear decline in precision, e.g. when a single vibration is added to a previous one held in working memory.

Such a sharp drop in mnemonic strength is difficult to explain on the basis of a fixed capacity account, which predicts peak performance until a capacity limit is reached (e.g. in vision: Luck & Vogel, 1997; Cowan, 2001). We also found that performance on the frequency-matching task remained above chance even for the highest memory load of 5 items, showing that capacity limits are not yet reached. Again, this result cannot be explained on the basis of a fixed capacity account, which predicts a sharp decline from near ceiling to at chance performance once capacity limits are exceeded. Hence, our results are in line with the resource model, which is shared out across items held in WM (Bays & Husain, 2008).

Critically, the results demonstrate that vibrotactile WM can store multiple items. Previous evidence suggested that this modality may only be able to store a representation of a single stimulus; as the presentation of a distractor in the delay period of a change detection paradigm, caused strong interference with the maintenance of a single vibration (Bancroft et al., 2011; Bancroft & Servos, 2011). However, in a further study by the same authors (Bancroft et al., 2012) it has been shown that at least two vibrations can be maintained in WM above chance. Here, capacity limits have not only been pushed further, where the results not only show that we can remember information about at least 5 items, but also at which mnemonic resolution strength those items are stored in WM.

The continuous response method employed here allowed quantification of the variability of memory recall around its true stimulus value. Recall variability, the resolution with which items are represented may reflect noise in neural representations, visualized by the raw response error histograms (Figure 4-4). The histograms show that recall variability increases with an increase in memory load (see target stimuli, in blue). Thus, memory representations become noisier with the addition of each item held in WM, providing a framework for the interpretation of neurophysiologic data.

What is shown in the histograms plotted in Figure 4-4 is not only the response variability around the target stimuli (in pink), but also the response variability around the non-targets (in blue), for each sequence length and serial position. As soon as a sequence contains two items (or more), participants might sometimes report the non-target instead of the target frequency. By visual inspection of the histograms, it seems that the variability of responses directed at the non-targets increases with an increase in memory load, but is constant across serial positions (possible combinations of targets and non-targets within a sequence). This means that participants were not more likely to respond to non-targets surrounding the target (e.g. target position 3 in a sequence of 5 items and non-target positions 2 and 4), but equally likely to respond to any non-target within the sequence.

However, irrespective of the serial order in which non-targets were presented, serial order had an overall effect on memory precision (Figure 4-3). For the sequence length (memory load) of 4 and 5 vibrations, the last item was remembered best compared to the previous items within the sequence (recency effect). However, there was no order effect for the sequence length of 2 items, unlike the observation of a primacy effect reported by Bancroft et al (2012) for the same memory load, where the first item was remembered best.

As evident from the raw response error histograms, there were at least 2 types of response error: (1) responses directed at the target and (2) responses directed at the non-target frequency. Additionally, a common type of error simply is (3) guessing. Thus, in a further analysis, all responses made by participants were decomposed into these three categories, each capturing a different type of response error. A decrease in the proportion of responses directed at the target (category 1) was observed with an increase in memory load (see Figure 4-5, purple line), which is in line with the overall pattern of a decline in precision with an increase in memory load (see Figure 4-2). This result confirms that the increase in response variability with memory load is due to responses directed at the target (the majority of responses) and not other types of responses, consistent with previous findings in the visual (Bays et al., 2009, Gorgoraptis et al., 2011) and auditory modality (chapter 2 or Kumar et al., 2013, chapter 3).

As more items are presented within a sequence, the larger the proportion of non-target responses (category 2: misbinding errors), (Figure 4-5, cyan line). This has previously been observed in vision for items presented sequentially (Gorgoraptis et al., 2011). Once the participant's memory gets tested for one of the items in a given sequence (once it becomes the target) all other items may be regarded as 'distractors'. The more distractors are present within a sequence, the more likely a response to one of these non-targets, instead of the target.

Inspection of category 3 type responses reveals that the proportion of guesses remains constant across all memory loads. This finding is also consistent with the resource model account of WM applied previously to visual (Bays et al., 2009, Gorgoraptis et al., 2011) and auditory WM (chapter 2 or Kumar et al., 2013). In contrast, the fixed capacity account predicts that performance should drop entirely to chance as soon as the capacity limit is exceeded. Although capacity limits for vibrotactile stimuli have not been addressed by previous research, neither a drop to chance performance – even at the highest memory load of 5 items – nor an increase in the proportion of guessing was observed here.

The results of this study not only extend the knowledge on vibrotactile WM capacity, but also provide insights in the qualitative nature of stimulus representations. On this basis, storage of vibrotactile information can be compared to other features within the same and across sensory modalities. Capacity limits have been more explored for tactile stimulus material other than mechanical vibrations such as somatosensory location. While a number of studies presented stimuli at different locations on the body surface (Gallace et al., 2008; Auvray et al., 2011; Alluisi et al., 1965; Geldard & Sherrick, 1965), others (Auvray et al., 2011) presented stimuli to the fingertips.

In the investigations of Gallace et al (2008) and Auvray et al (2011) participants had to memorise up to six somatosensory locations. Depending on the report procedure employed, participants were able to remember 3 items (full report) or up to 6 items (partial-report), demonstrating that the selection of response method is of importance. Furthermore the format of presentation is important as well: for example, Alluisi et al. (1965) found that even when only 3 stimuli were presented simultaneously over the body surface error rates exceeded 30%. However, assigning verbal labels to different body parts or fingers may aid task performance. Although verbal coding cannot entirely be excluded on the frequency matching task employed here, it does appear as more complex to label a particular stimulus frequency than a somatosensory location. Furthermore, Gallace and Spence (2009) showed that somatosensory spatial information is represented by neural systems different from the ones encoding vibrations (e.g. regions in prefrontal cortex, Rome & Salinas, 2003).

The somatic sensation of a vibration travels via two different pathways or channels, Meissner and Pacinian afferents, which are associated with different characteristics due to their receptor selectivity. While Meissner corpsulses respond to low frequency vibrations in the range of 5-40Hz (Mountcastle, 1984) or up to 50Hz (Gordon, 1978), (channel 1: vibrotactile flutter), Pacinian corpsules are selectively sensitive to higher frequencies between 40-400Hz (Talbot et al., 1968) or 60-300Hz (channel 2: vibrotactile fusion), (Mountcastle, 1984, Vallbo & Johansson, 1984), (1982; Verillo, 1968; Verillo & Gescheider, 1979; Gescheider & Verillo, 1979). Both types of receptors can be found at different locations of the skin, where receptor density corresponds to receptive field size (Schiff & Foulke, 1982).

already in
introduction!?

The analysis performed here revealed that there was indeed a significant difference in error scores between channels (range 1: 10-50Hz corresponding to channel 1; range 2: 51-100Hz corresponding to channel 2) across all individual memory loads. Therefore the findings raise the possibility that mnemonic resolution is scaled to receptive field size or density and their properties (e.g. slowly vs. rapidly adapting) within a given channel. Encoding a stimulus via frequencies encoded by Meissner corpsules was associated with a more precise stimulus representation in WM compared to a less precise representation of frequencies encoded via Pacinian corpsules.

How is vibrotactile information represented in the brain with respect to processes in WM? Neurophysiological studies in monkeys investigated delay period activity (memory maintenance) in somatosensory cortex. While Zhou and Fuster (1996) as well as Koch and Fuster (1989) found delay period activity in SI on a texture discrimination task, Romo and Salinas (2003) did not report such activity changes for vibrotactile flutter. However, Romo et al. (2002) found delay activity in SII, which seems to be involved in vibrotactile WM.

The role of PFC is special for memory maintenance, as delay activity is systematically modulated only in this region, where neural discharge rates vary as a monotonic function of the encoded stimulus frequency (Romo et al., 1999). Interestingly, findings in monkeys translate well between species (Bancroft et al., 2011), where even early studies showed similar discrimination thresholds of vibrotactile stimuli across species (e.g. in frequency and amplitude; LaMotte & Moutcastle, 1975; Talbot et al., 1967). With regards to WM, activity changes in PFC are also associated with WM maintenance in humans, using fMRI (Preuschhof et al., 2006) as well as EEG (Spitzer et al., 2010). Thus, it has been suggested that vibrotactile WM serves as a model paradigm for memory in human and non-human primates across a variety of research fields (Bancroft et al, 2011), including neurophysiology (Romo et al., 1999, Hernandez, Zainos & Romo, 2000; Romo & Salinas, 2001), computational modelling (Machens, Romo & Brody, 2005; Deco, Rolls & Romo, 2009), behavioural psychophysics (in humans: Sinclair & Burton, and monkeys: 1996, Hernandez et al, 1997) and cognitive neuroscience (Spitzer et al., 2010, Haegens et al., 2010; Hegner et al., 2010).

4.6 Conclusion

Future research on tactile WM might address the neural signatures of WM representations from a new perspective using recall precision, the mnemonic strength of a representation. Neurophysiological tools in monkeys (e.g. multi-unit recordings) might allow a more direct probe of memory precision, to determine if neural noise contributes to WM recall (e.g. see a recent theoretical model of population coding for visual WM: Bays, 2014). Non-invasive techniques such as

EEG/MEG can also be used to relate precision to patterns observed in oscillatory networks (e.g. across species: Reinhart et al., 2012). Finally, the use of fMRI with sensitive multivariate analysis techniques can be used to associate different levels of mnemonic strength with different patterns of brain activity (as in vision: Emrich et al., 2013). Selecting different regions of interest within the tactile processing hierarchy from somatosensory cortex to PFC may yield a better understanding of where and how a stimulus is represented in WM.

Chapter 5. Auditory object representations of spectro-temporal features held in working memory

5.1 Abstract

Are sounds in auditory working memory (WM) represented as integrated objects or individual features? And how does the representational format influence WM capacity? Here we address this for auditory objects composed of two fundamental dimensions of natural sounds: spectral content and temporal amplitude modulation rate. In a change detection paradigm, participants had to memorize sequences of auditory objects of variable length (1-4 items). They either maintained objects as a whole or their individual features until recall for one of the items was tested.

We found that, memory recall was more accurate when the object had to be maintained as a whole compared to its individual features alone. For individual features, performance was higher on the dimension of the spectral content of sound compared to the dimension of temporal amplitude modulation rate. One interpretation of our findings is that, at some level in the auditory system, sounds may be stored as objects, as there is an extraction cost for single features stored in WM, which is higher for temporal than spectral content.

5.2 Introduction

The auditory scene contains auditory objects, which can be considered to be fundamental perceptual units comprising combinations of sensory cues that form a coherent whole that may or may not be associated with a semantic label (Griffiths & Warren, 2004). While an object's individual features are analysed and represented separately in specialized feature maps, features are bound together to form perceptual wholes at a further level of information processing (Treisman & Gelade, 1980). Instead of addressing the perceptual level, here we ask how features of auditory objects are represented in working memory (WM). When holding auditory objects in mind for just a few seconds in the absence of an external stimulus, what is the format of representation in WM? Is it possible that auditory objects are represented in WM as bound units (objects) or as individual features in separation? Do individual object-features interfere with each other when held in WM?

Although the exact definition of auditory objects remains controversial (Griffiths & Warren, 2004), here we specifically investigate WM for auditory objects defined by different dimensions of timbre. Timbre is the quality of sound, which contributes to the identity of an auditory object more than any other dimension (e.g. pitch, loudness or duration). Timbre itself comprises multiple acoustic feature dimensions, where evidence for the perceptual dimensionality of timbre comes from studies using multidimensional scaling (MDS) to identify the perceptually most salient features (Caclin et al., 2005; Grey, 1977; Marozeau et al., 2003; McAdams et al., 1995; Samson, et al., 1997). The feature dimensions we manipulated for the purpose of the current study, temporal amplitude modulation rates and spectral content, relate to the perceptually most dominant dimensions of timbre: attack time and spectral centroid. Attack time describes the temporal envelope of a sound in terms of its rise, which we manipulated as temporal amplitude modulation rate. Another feature dimension of interest, spectral content, was manipulated, which relates to the spectral centroid, representing the centre of mass in the frequency domain.

Here, we aim to investigate how two features of auditory objects, relevant to the perception of timbre, are integrated into coherent wholes. It has been shown that auditory object formation depends on stimulus structure and attentional processes (Shinn-Cunningham, 2008). Characteristics of stimulus structure, such as common onsets and offsets, harmonic structure, and continuity of frequency over time determine object formation (Bregman, 1990; Darwin & Carlyon, 1995). Auditory objects composed of spectro-temproal features are linked together over time (auditory streaming) and this process is intricately linked with selective attention. This directly relates to the binding problem, which in the context of perception and attention, refers to how distributed neural codes representing multiple features of the perceptual scene are recombined so that one perceives the actual object (Treisman & Schmidt, 1982). Distinct neural populations would thus code for various clusters of dimensions of timbre, where any two features compete for 'representational space' in the underlying neural regions, resulting in over-writing or partially corrupted representations. It remains to be determined, which features of timbre held in WM might be extracted and maintained separately. However, ERP studies showed that sound features (intensity, duration, attack, spectral centroid and fine structure) may be represented in distinct regions of auditory cortex (Caclin et al., 2006; Giard et al., 1995).

In vision, features in different dimensions are maintained in different memory stores due to binding errors observed on a change detection task (Wheeler & Treisman, 2002). In contrast, Luck and Vogel (1997) showed that visual WM may be limited by the number of objects, but not by the number of features per object. Additionally, Oberauer & Eichenberger (2013) showed 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 (precision). A further study in vision suggests that there is no upper object limit, as processes in WM are best described as a resource shared out across all objects and their features (Bays et al., 2011). The authors found increased binding errors at high memory loads due to independent response error distributions for each feature, suggesting that features are maintained in separate WM stores.

We aim to test which of the above principles found in vision apply to auditory WM. What is the representational format of auditory objects and their features in WM? Are auditory objects maintained as bound units in a common memory store or as individual features in separate stores? And how does the representational format influence WM capacity limits? In the current study, subjects performed a within modality dual task. They had to memorize sequences of variable length (1-4 items), composed of auditory objects containing two features (spectral envelope and temporal amplitude modulation rates). Subjects had to maintain in WM the objects as a whole or their individual features until recall for one of the items was tested by change detection. Accuracy scores were compared across conditions and memory loads. We found that memory performance was best when subjects had to hold the object in mind compared to performance on the individual feature conditions. One interpretation of our findings is that, at some level of the auditory system, sounds may be stored as objects, as there is an extraction cost when recollecting single features stored in WM.

5.3 Materials and Methods

5.3.1 Participants

10 healthy young adults (5 female, mean age 24 years, age range: 19-39) participated in the main experiment and initial control experiment. Another 10 participants completed an additional control experiment (5 female, mean age 24 years, age range: 19-39). All participants provided written informed consent to procedures approved by the local ethics committee. Participants were selected based on the following criteria: normal hearing and no musical training.

5.3.2 Stimuli and Apparatus

Auditory objects served as stimuli, composed of two features. Both features are fundamental components of natural sounds, relevant to the perception of timbre. The first sound feature is described as the frequency structure or spectral content of sound (relevant to the spectral centroid as a dimension of timbre). Its centre frequency was selected from a fixed set of 8 values from the range of 250-1500 Hz.

There was a minimum separation of 29% between any two succeeding values in the set, resulting in the following set of centre frequencies: 250, 322.92, 417.12, 538.8, 695.98, 899, 1161, 1500 Hz. A narrow band noise of 1/8th of an octave was added on each side of the centre frequency, composing the spectral content of a given sound attribute. The second sound feature is defined as the temporal amplitude modulation rate of sound (relevant to the attack time as a dimension of timbre). Another fixed set of 8 values was generated from the range of 6-32 Hz with a minimum separation of 28%, resulting in the following set of values: 6, 7.62, 9.67, 12.29, 15.61, 19.83, 25.19, 32 Hz.

A sound attribute of each dimension (spectral (S) and temporal (T) content) was selected at random from each set without replacement. Attributes (S & T) were combined to form auditory objects, e.g. object 1 with spectral content of 538.8 Hz and temporal content of 25.19 Hz; and object 2 with spectral content of 1500 Hz and temporal content of 12.29 Hz. The stimulus duration was 1s with an ISI and a delay period of the same length. Auditory stimuli were generated online at a sampling rate of 44.1 kHz in Matlab 7.12.0 (Mathworks Inc.) and presented using Cogent (http://www.vislab.ucl.ac.uk). Sounds were delivered binaurally over headphones (Sennheiser HD 650) in a soundproof testing room.

5.3.3 Design and Procedure

On each trial, subjects listened to a sequence of sounds of variable length: One, two or four auditory objects; see Figure 5-1). At the end of each sequence recall for one of the objects was probed, indicated by a number appearing onscreen, e.g., 2 for second sound. Each sound within the sequence was equally likely to be probed. A final auditory object was presented as the probe sound and subjects had to decide whether it was the same or different from the target sound (second sound, in Figure 5-1). They responded by pressing a button for ‘same’ or ‘different’.

Subjects started the experiment by completing a practice block of 48 trials. Once they were familiarized with the sound attributes as well as the different experimental conditions, they completed the main experiment, consisting of 9 blocks of the same length with equal number of trials for each memory load. There were three types of experimental conditions (3 x 3 blocks), which took approximately 90 minutes to complete. In a second session subjects completed 6 blocks of two types of control conditions (2 x 3 blocks), which could be completed in 60 minutes. An additional control condition was added later on, which was tested on a new set of participants consisting of 3 blocks (duration: 30 minutes).

Experimental conditions: There were two conditions on which subjects had to memorize single attributes of auditory objects and another condition on which they memorized objects as a whole (see Figure 5-1 for an overview of all conditions). In *the spectral condition*, they had to focus on the spectral content of the objects whilst ignoring the other attribute (temporal amplitude modulation rate). Their memory for the spectral content was tested by change detection. On 'same' trials, the amplitude modulation could be different, while the spectral content was identical to the target. On 'different' trials, there was a change on both features. In *the temporal condition*, subjects focussed on the amplitude modulation rate (temporal content), whilst ignoring the other attribute (spectral content). Their memory for the attended attribute was then tested. In *the object condition,* subjects memorized the objects as a whole, forming bound percepts of both attributes. They were probed at random on either dimension. On 'same' trials there was no change, whilst on 'different' trials both features contained in the object changed.

It is important to note that stimulus presentation and probe selection was identical among all experimental conditions. A particular ink colour of text appearing onscreen was assigned to each condition (spectral condition = pink, temporal condition = green, object condition = white). Experimental blocks were randomly interleaved and subjects knew by the ink colour of text appearing onscreen which condition they were on.

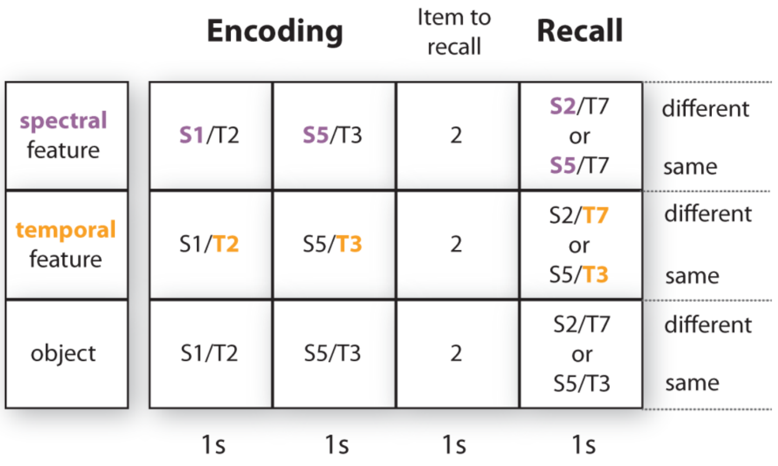


Figure 5-1 *Task and experimental conditions*

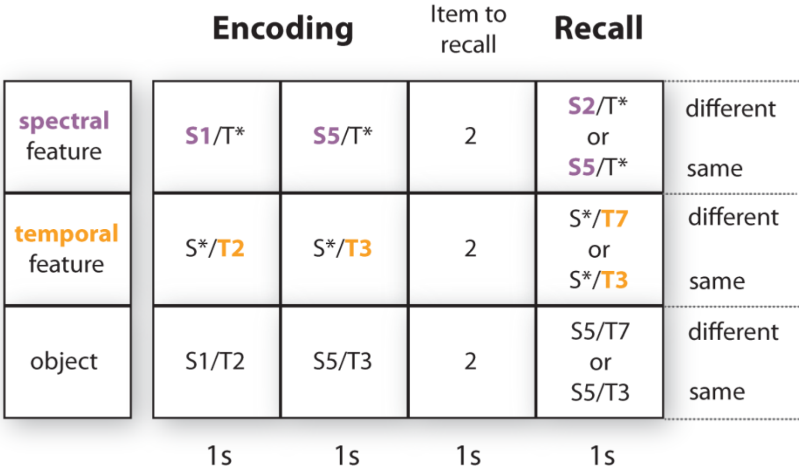
Shown are sample trials for each experimental condition (each row illustrates one of 3 conditions). Note that the same material (2 auditory objects) is presented at encoding (identical across conditions). Each object is presented for 1 second followed by an ISI of 1s. Next, a number appears onscreen 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 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).

Control conditions: There was a control condition for each of the three experimental conditions. When a single feature (e.g. spectral) has to be maintained in WM, the other irrelevant dimension (e.g. temporal) may cause interference with the relevant dimension of sound (see Figure 5-2). In the experimental condition the dimension of the irrelevant sound feature was varied at random. However, the irrelevant dimension was held constant in the control condition in order to capture the amount of interference when comparing across conditions (experimental vs. control).

The spectral control condition was identical to the experimental spectral condition, where subjects had to focus on the spectral content of the objects whilst ignoring the other attribute (temporal amplitude modulation rate). However, the amplitude modulation rate was not selected at random as in the experimental condition, but held constant at 14 Hz (middle value of the temporal stimulus range).

The temporal *control condition* was identical to the experimental temporal condition, where subjects focused on the amplitude modulation rate, while ignoring the spectral content, which was held constant at 600 Hz (middle value of the spectral stimulus range).

Importantly, the object control condition was identical to the experimental object condition: subjects memorized auditory objects as a whole. However, on different trials the probe only changed on a single feature instead of a perceptually larger change of 2 features on the experimental object condition. The object control condition was added to ensure potentially better memory performance on the experimental object compared to the experimental single-feature conditions isn't solely due to a perceptually larger change between the item to recall and item presented at test.



	Encoding		Item to recall	Recall	
spectral feature	**S1**/T*	**S5**/T*	2	**S2**/T* or **S5**/T*	different same
temporal feature	S*/**T2**	S*/**T3**	2	S*/**T7** or S*/**T3**	different same
object	S1/T2	S5/T3	2	S5/T7 or S5/T3	different same
	1s	1s	1s	1s	

Figure 5-2 *Task and control conditions*

Shown are sample trials for each control condition (each row illustrates one of 3 conditions). In the spectral control condition (1st row) subjects focus only on the spectral feature (in purple). In the temporal control condition (2nd row) subjects only focus on the temporal feature (in yellow). In the object control condition (3rd row), they encode the object as a whole (both features in combination). Both single feature control conditions differ from the single feature experimental conditions in the following way: the irrelevant feature (*) is held constant at the middle value of the corresponding stimulus range. The object control condition differs from the experimental object condition in the way that on change trials, the item presented at recall (probe) differs from the target (here: 2nd item in the sequence presented at encoding) by 1 feature, instead of 2 features.

5.3.4 Data analysis

Hypotheses regarding the effects of memory load and condition, on memory performance (accuracy) were tested by ANOVAs and LSD post hoc t-tests, as specified in the results.

5.4 Results

Effects of memory load and condition on memory performance

Memory performance was measured as accuracy (percentage correct) for each sequence length (memory load: 1, 2 or 4 sounds), all experimental and control conditions within each sequence. A two-way ANOVA was employed to test for the effects of factor 1, memory load, and factor 2, experimental condition, on accuracy. This analysis revealed a significant main effect of memory load ($F(2,32)=31.94$, $p<0.001$) and a significant main effect of experimental condition on accuracy ($F(2,32)=67.13$, $p<0.001$), as well as an interaction ($F(4,32)=2.72$, $p=0.035$), (see Figure 5-3). All post hoc comparisons between individual memory loads across conditions were significant at $p < 0.001$. Memory recall was more accurate when the object had to be maintained as a whole compared to its individual features. This result cannot be explained based on the level of difficulty, where monitoring a single feature could be easier than two. Features therefore may be represented as bound units. Participants had more accurate recollection on the dimension of the spectral envelope compared to the temporal amplitude modulation rate.

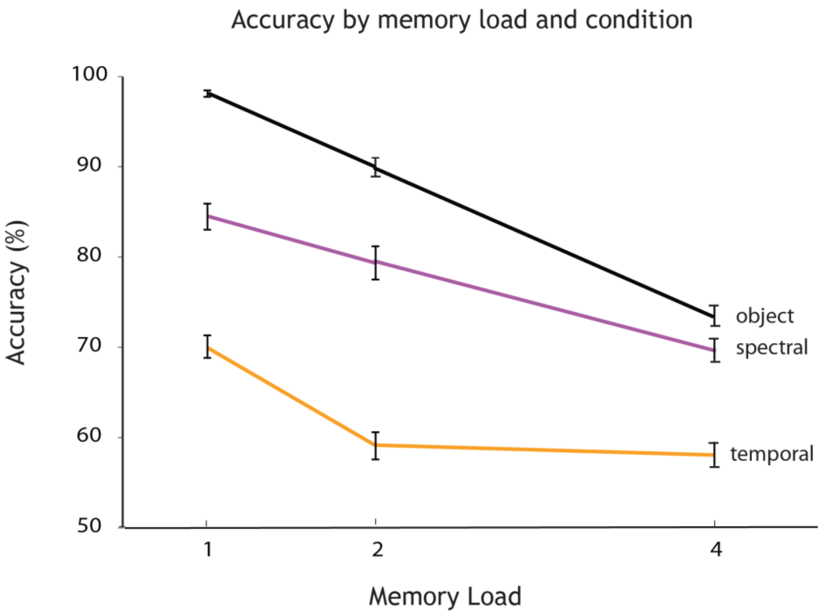


Figure 5-3 *Accuracy varies by memory load and experimental condition*

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).

Further two-way-ANOVAs were carried out to compare the effects of memory load on accuracy across experimental and control conditions. Comparing the spectral condition (see Figure 5-4, in rose) with the spectral control condition (in pink), a two-way ANOVA revealed a significant main effect of memory load ($F(2,32)=47.8$, $p<0.001$) and a significant main effect of condition on accuracy ($F(1,32)=37.85$, $p<0.001$), as well as an interaction ($F(3,32)=5.46$, $p=0.007$), (see figure 4). This comparison shows that varying the irrelevant feature at random (here: the temporal dimension of sound) leads to a decline in memory performance when tested on the spectral dimension of sound. Thus the irrelevant feature induces a feature extraction cost; the spectral dimension of sound is extracted from the object, while the temporal dimension induces a cost when varied at random.

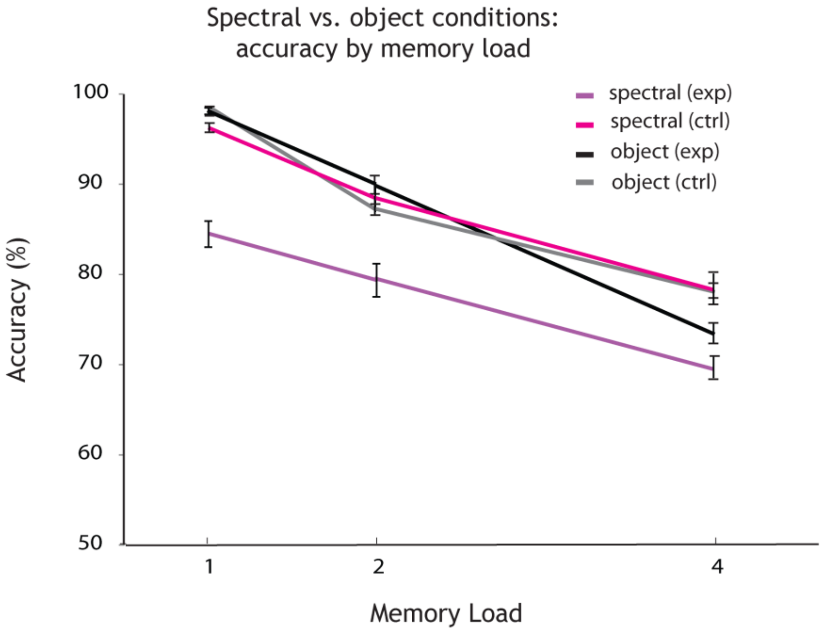


Figure 5-4 *Accuracy by memory load for single feature spectral vs. object conditions*

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 condition. Shown in rose are the results for the single feature *spectral* condition. *Spectral* is the relevant and *temporal* the irrlevant feature varied at random. In pink: again *spectral* is the relevant feature, but the irrelevant *temporal* feature is held constant (spectral control condition). The results for the object condition are shown in black (the probe changes on both features on change trials). The results for the object control condition (the probe changes by a single feature on change trials) are shown in grey.

Comparing the temporal condition (see Figure 5-5, in yellow) with the temporal control condition (in orange), a two-way ANOVA revealed a significant main effect of memory load (F(2,32)=18.46, p<0.001) and a significant main effect of condition on accuracy (F(1,32)=44.97, p<0.001), but no interaction (see figure 5). This comparison shows that varying the irrelevant feature at random (here: the spectral dimension of sound) leads to a decline in memory performance when tested on the temporal dimension of sound. Thus, the temporal dimension of sound is extracted from the object, while the other dimension (spectral) induces a cost when varied at random.

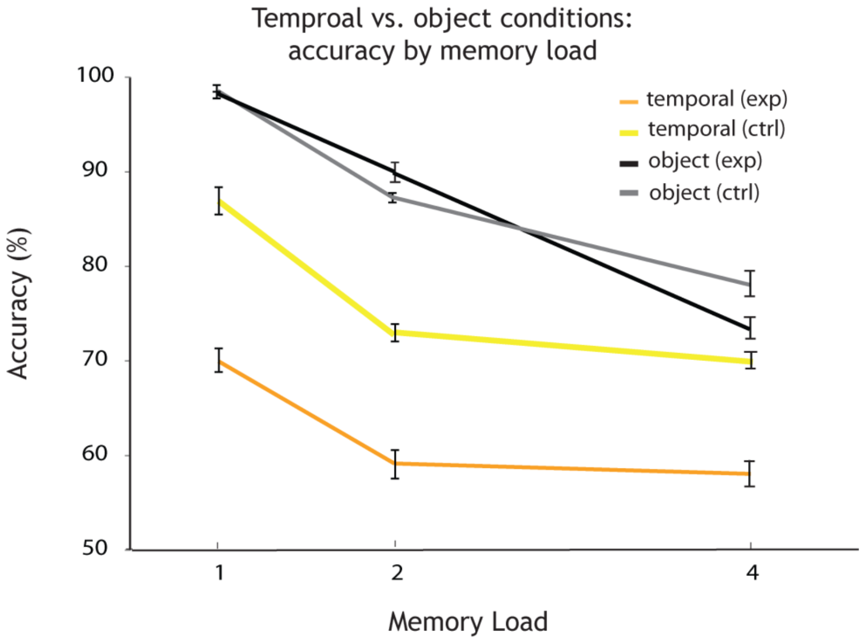


Figure 5-5 *Accuracy by memory load for single feature temporal vs object conditions*

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 condition. Shown in orange are the results for the single feature *temporal* condition. Temporal is the relevant and *spectral* the irrelevant feature varied at random. In yellow: again *temporal* is the relevant feature, but the irrelevant *spectral* feature is held constant (temporal control condition). The results for the object condition are shown in black (the probe changes on both features on change trials). The results for the object control condition (the probe changes by a single feature on change trials) are shown in grey.

Further comparisons were carried out between the object condition (Figure 5-4 & Figure 5-5: in black) and the spectral control condition (Figure 5-4: in pink). A two-way ANOVA revealed a significant main effect of memory load on accuracy (F(1,32)=49.47, p<0.001), but no main effect of condition and no interaction. There was no difference in memory performance when the object was maintained as a whole compared to maintaining only its spectral content (single feature), while the irrelevant dimension (temporal) is held constant.

We then compared the object condition (Figure 5-4 & Figure 5-5: in black) with the temporal control condition (Figure 5-5: in yellow). A two-way ANOVA revealed a significant main effect of memory load (F(2,32)=47.8, p<0.001) and a significant main effect of condition on accuracy (F(2,32)=37.85, p<0.001), as well as an interaction (F(4,32)=5.46, p=0.007). There was a difference in memory performance when the object was maintained as a whole compared to maintaining only its temporal content (single feature), while the irrelevant dimension (spectral) is held constant.

A final comparison was made between the object condition (Figure 5-4 & Figure 5-5: in black) and its corresponding object control condition (Figure 5-4 & Figure 5-5: in grey). A two-way-ANOVA revealed a main effect of memory load (F(2,20)=65.8, p<0.001), but no effect of condition (F(1,20)=0.176, p=0.667) and no interaction (F(3,20)=1.85, p=0.166). The size of change between the item tested and the item presented at recall (by 1 or 2 features) did not make difference in memory a
performance, when maintaining objects as a whole.

5.5 Discussion

We investigated the format of representations of auditory objects held in WM by analysing listeners’ memory performance when they either maintained objects as a whole or their individual features (spectral vs. temporal dimensions of sound). Memory recall was more accurate when the object had to be maintained as a whole compared to its individual features. Participants had more accurate recollection on the dimension of the spectral content compared to the other auditory dimension of temporal amplitude modulation rate (Figure 5-3). The difference in memory performance across individual feature conditions may depend on the definition of each feature dimension (spacing of stimulus range), but alternative accounts will be taken into consideration. One interpretation of our findings is that, at some level, sounds may be stored as objects, as there is an extraction cost for single features stored in WM, which is higher for temporal than spectral content.

Holding whole objects in mind

In the current study, auditory objects were composed of two basic auditory features, spectral and temporal dimensions of sounds. Why are such spectro-temporal stimuli perceived as objects? As both features are presented as overlapping in time (common onsets and offsets), they are perceptually grouped together forming coherent wholes. They seem to emanate from a single source (Alain & Arnott, 2000), while such features may also be perceived as belonging to a common auditory event or entity (Bregmann, 1990; Blauert, 1997; Hartmann, 1988). The particular features used here are relevant dimensions of natural sounds, especially important to the perception of timbre, where the brain holds mechanisms of abstraction for both the spectral and temporal dimension of sound (this is explained in more detail in chapter 6).

On the object level, performance was at ceiling for the memory load of a single item and then dropped with the addition of each further item to be held in working memory. Performance remained above chance (>50% accuracy) at the highest memory load of four auditory objects. A similar pattern was observed among object conditions and was independent of the size of change of the probe stimulus (on different trials) relative to the target stimulus: by two features (object condition) or by one feature (object control condition, see Figure 5-4). The size of change of the probe stimulus did not lead to a difference in performance across memory loads.

Holding individual features in mind: spectral envelope

Holding only the dimension of spectral content in mind, whilst the other to be ignored dimension of sound is held constant (spectral control condition), we found memory performance to equal performance in the object condition. Thus, for this particular dimension, holding an object or its feature equally draws on memory resources. Here, WM appears to be object based depending on how we define an auditory object. If an auditory object is primarily defined by a single dimension of sound (one feature), this feature can be regarded as an object similarly to an object composed of multiple features. Furthermore, comparing memory performance across conditions, on which only the dimension of spectral content had to be maintained (spectral condition: the other dimension is varied randomly, where on the spectral control condition: it was held constant), performance was better on the control compared to the experimental condition (see Figure 5-4). This comparison shows that there is a significant cost in feature extraction.

When the other (irrelevant) dimension of sound is varied randomly, it interferes with encoding of the feature of interest to a larger extent then when the irrelevant dimension is held constant. In audition, it has been suggested that object formation depends on attentional processes (Shinn-Cunningham, 2008). Relevant to perception and memory encoding, previous research in audition has shown that we can focus on a desired feature presented as part of an auditory object (Brungart & Simpson, 2002; Rakerd et al. 2006; Shinn-Cunningham, 2008; Helfer & Freyman, 2009) and that task irrelevant features influence selective attention (Maddox & Shinn-Cunningham, 2012). Thus, a certain extent of the observed cost in feature extraction may arise at the stage of memory encoding, as attention has to be allocated to extract the feature of interest. As a result, the interference caused by the irrelevant dimension may add noise to the memory representation of the feature of interest in WM.

However, at the stage of memory recall, when deciding whether the probe matches the target stimulus, the relevant feature dimension has to be extracted from the probe sound once again (spectral condition). The relative cost of extraction of the relevant dimension of sound is larger when the irrelevant dimension is varied randomly compared to when it is held constant (spectral control condition). Therefore, at the stage of memory recall, the irrelevant dimension of the probe may have an additional decremental effect on WM performance. In sum, the observed cost in feature extraction may be based on multiple sources of interference caused by the irrelevant dimension at different stages of WM (encoding, maintenance and recall). Future research using neuro-imaging methods (e.g. event-related potentials with a high temporal resolution), is needed to disentangle the time-course of events of the different stages of memory processes involved to clarify information loss in WM representations of features vs. objects.

object = passive recall; feature = less passive recall, more demanding, as, manipulation -> STM vs. WM

Holding individual features in mind: temporal content

On the individual feature level with regards to the temporal amplitude modulation rate, performance was lower when the other dimension (now spectral) was held constant (spectral control condition) compared to the object condition. This is unlike the comparison between the corresponding comparison between the spectral dimension (spectral control condition) and the object condition, where performance was equal across conditions. Although the spacing of stimuli along each feature dimension was based on perceptual thresholds and approximately equivalent across dimensions, they may not have been perfectly matched in terms of difficulty. Thus, the observed difference across conditions may reflect that it was easier to hold spectral features in mind compared to temporal ones. Only, if we can assume the level of difficulty to be matched across conditions, a possible alternative interpretation is that the spectral envelope may represent a stronger cue for object formation. In other words, the spectral dimension of sound defines the object to a larger extent compared to its temporal dimension. It may dominate object formation at encoding or override temporal representations at maintenance. An alternative explanation is that WM stores for either feature dimensions may be partly independent, where each store has a different capacity limit. Therefore, the capacity limit for the temporal dimension would be lower compared to the capacity limit found on the other dimension. However, it would be important to reproduce our findings with different stimulus spacing to verify this account.

Moreover, comparing memory performance across conditions on which temporal information has to be maintained (temporal condition: the other dimension is varied randomly vs. temporal control condition: it was held constant), performance was better on the control compared to the experimental condition (see Figure 5-5). This comparison shows that there is a significant cost in feature extraction on the temporal dimension similar to our findings for the spectral dimension (spectral condition vs. spectral control condition). The observed cost in feature extraction on the temporal dimension of sound may also be based on multiple sources of interference caused by the irrelevant dimension (spectral) at different stages of WM (encoding, maintenance and recall). The extent of this interference seems to be larger when comparing across temporal conditions (temporal condition vs. temporal control) compared to spectral conditions (spectral condition vs. spectral control), as the drop in memory accuracy is more severe when extracting the temporal feature, which may again be due to the design of the temporal stimulus dimension (not equated in terms of difficulty to the spectral dimension). Alternatively, this asymmetry in interference may also be due to uneven amounts of interference from the irrelevant feature dimension. E.g., when the spectral dimension is varied randomly (irrelevant dimension), it interferes more with the relevant feature (temporal), than the interference caused by the temporal dimension (irrelevant) on the spectral dimension (relevant).

Relation to feature integration theory

Object formation involves binding of features, which become reorganized to create more complex unified representations of previously distributed information (Treisman & Schmidt, 1982). A previous study on auditory objects composed of either 3 or 6 dimensions of timbre, showed that WM capacity increased, when the acoustic separation between the probe and test items increased, as well as the number of feature dimensions (Golubock & Janata, 2013). In relation to feature integration theory and in particular feature overwriting (distinct neural populations code for different dimensions of timbre), they showed that capacity is facilitated when stimuli have un-shared features. Thus, there is less competition between any two features for 'representational space' in corresponding neural regions. Our results are difficult to directly compare to this study, as our auditory objects were always composed of the same two dimensions, where only the relevance of features was manipulated. Apart from a single previous and the current study, little is known about the organization of auditory object features in auditory WM.

Comparison to vision

In contrast, research in vision offers a more detailed account on how object-feature information is represented in visual WM, although it remains a highly controversial topic in this sensory modality. One study in vision concluded that WM storage capacity for features is unlimited once features are combined to form objects. However, the number of objects, which can be held in WM, may be the sole factor limiting WM capacity (Luck & Vogel, 1997, Luria & Vogel, 2011, Cowan, Blume & Saults, 2013). Others showed that capacity is also limited by the number of features, as features in different dimensions are maintained in different memory stores (Stevanovski & Jolicoeur, 2011, Fougnie & Alvarez, 2011). Additionally, they can be recalled on their own even when memory of their binding/association/linkage within an object (e.g. the features as a pair forming the object (*conjunction*)) is lost (Wheeler & Treisman, 2002). In contrast, Oberauer & Eichenberger (2013) showed that WM capacity is not only limited by the number of objects and the number of features per object, but also by mnemonic resolution strength (precision). A further study in vision suggests that there is no upper object limit, as processes in WM are best described as a resource shared out across all objects and across their features (Bays et al., 2011). The authors found increased binding errors at high memory loads due to independent response error distributions for each feature, suggesting that features are maintained in separate WM stores. Future research could address this in more detail in audition by measuring recall precision.

5.6 Conclusion

For auditory WM, we find that memory performance declines with an increase in memory load, for all types of auditory information, such as objects or their features. Thus, information load draws on memory resources. We show that combining auditory features to form integrated objects, the object is the preferred format in order to store a representation successfully in auditory WM. Therefore, it seems that feature integration increases WM capacity, where object recall (memory accuracy) exceeds recall of individual features. Unlike findings in vision recall of features either equals (Wheeler & Treisman, 2002) or exceeds the number of objects that can be recalled successfully from WM (Luck & Vogel, 1997), which means that there is no cost when extracting individual features from visual WM.

Such opposing findings across sensory modalities may be due to the perceptual nature of visual vs. auditory stimuli. In vision we can form a clear representation of each feature (e.g. colour or spatial location, which is easy to verbalize); representations in auditory WM are relatively more abstract (and for stimuli employed here: difficult to verbalize (see Golubock & Janata, 2013; who used stimuli composed of different dimensions of timbre). Thus, once features are combined into objects in WM, the reverse process (un-binding or feature extraction) when tested on a particular feature at recall, comes at a cost in audition. In contrast, no equivalent findings have been obtained in vision, where a feature extraction cost has only been associated with attentional processes rather than WM (Wheeler & Treisman, 2002; Cowan, Blume & Saults, 2013). Therefore, as feature extraction draws on memory resources in audition, this does not seem to be the case for visual WM.

Our results show that once the information is unified into an auditory object, manipulating it at memory recall by extracting single features decreases memory accuracy. One interpretation of our findings is that, at some level, sounds may be stored as objects, as there is an extraction cost for single features stored in WM, which is higher for temporal than spectral content, dependent on how auditory objects are defined. However, information processing of features in WM may be more complex than this. Features are represented at different levels in the auditory systems and may only be bound and interfere with each other at some level in this hierarchy. The difference in memory performance across individual feature conditions may depend on the definition of each feature dimension (spacing of stimulus range) rather than on how memory resources are allocated differently to either feature. It remains to be tested, whether introducing different stimulus spacing along either feature dimension, has an influence on memory performance. In this way, it would be possible to determine fully, whether it is more resource demanding to hold one or the other feature in mind. However, it may also be that either feature underlies a unique function of decay: e.g. a temporal trace decays faster than a spectral one.

Future research in the auditory domain shall address memory decay and feature binding of different dimensions of sound and their underlying neural signatures, as well as aim towards an understanding of different time-periods of WM to disentangle the encoding, maintenance and recall of auditory objects and their features.

Chapter 6. Decoding memory for pure tones in auditory cortex

6.1 Abstract

It is widely believed that early sensory cortices, including auditory cortex, are primarily involved in the perception of low-level stimulus features. Additionally, it has been shown that visual cortex is also involved in the maintenance of such basic stimulus features in memory, e.g. orientation (Harrison & Tong, 2009) and contrast (Xing et al., 2013). Here, we ask whether auditory cortex is sensitive to the memory of one basic feature, the frequency of pure tones.

Participants were presented with tones and had to actively hold either a low or a high tone in mind for 16 seconds. Using functional magnetic resonance imaging (fMRI) with multivariate pattern analysis (MVPA), prediction of which tone was perceived was demonstrated in Heschl's gyrus (HG) with high classification accuracy. Activity patterns in the same and additional auditory areas (planum temporale (PT) and superior temporal gyrus (STG)) could also be used to successfully predict which tone was held in memory. Our results indicate that all of the above auditory regions are involved in maintenance of pure tones.

As a control, we scrambled the data obtained for HG resulting in at chance level classification. Moreover, classification of tones was also at chance level in somatosensory cortex, although above chance in visual cortex. We discuss possible bases for the latter. Our results demonstrate that auditory cortex (HG, PT and STG) can retain specific auditory information about basic auditory features (frequency of pure tones) held in WM, over periods of many seconds in the absence of an external stimulus.

6.2 Introduction

Controversy remains about the role of early sensory cortices in cognition: are they primarily involved in the perception of low-level stimulus features or also in higher level cognitive processes such as maintenance of the same features in memory?

In vision, evidence from fMRI studies suggests that basic stimulus features represented topographically on visual cortex during perception can also be decoded from the same regions during memory maintenance: e.g. orientation (Harrison & Tong, 2009) and contrast (Xing et al., 2013). Here we ask, whether similar principles apply to auditory memory: are early perceptual auditory processes and memory maintenance mediated by common neural circuitry? Similarly to other sensory cortices, neurons in auditory cortex are tuned to respond to specific stimulus features. Mapping of the fundamental stimulus property of frequency, tonotopy, has been demonstrated in the pathway from cochlea to cortex. Studies of human auditory cortex demonstrate multiple tonotopic maps in the auditory cortex in the superior temporal plane (Formisano et al., 2003; Humphries, Liebenthal & Binder, 2010; Da Costa et al., 2011). The existence of different patterns of activity in cortex for different frequency values during perception is therefore well established. In this study the aim was to establish whether there are distinct patterns of activity for pure tones of different frequency, when no sound is perceived, as evidence for a role of auditory cortex in memory maintenance.

A previous decoding study aimed to test whether auditory cortex shows frequency specific involvement in the maintenance of pure tones in memory (Linke et al., 2011). However, its design may not have been suitable to investigate frequency-specific changes in brain activity. Pure tone frequencies were selected from four frequency ranges, where each range is known to activate distinct areas in auditory cortex. Tones presented at encoding were always selected from two of four frequency ranges. On each trial, participants encoded sequences of two tones (e.g. 1st tone: sampled from frequency range 4; 2nd tone from range 2); they always maintained both tones in memory and then decided whether a final sequence matched those sounds. As two tones from different frequency ranges were maintained, it is not possible to derive the particular frequency range of either sound held in memory, when decoding from the maintenance period. Additionally, tones were presented in rapid succession (total duration: 1250 ms with an ISI of 500 ms) and fitting individual regressors to each tone would potentially lead to significant correlation between them, where the data fed into the classifier is dependent on how regressors are fitted in the first place.

The current study was designed to overcome such limitations and investigates in an arguably more robust manner the frequency specific involvement of pure tones in perceptual vs. memory maintenance processes. Subjects were presented with two tones, where each tone was selected from one of two frequency ranges (range 1: low tones, range 2: high tones). They were then instructed to actively maintain one of the two tones in memory. On half of the trials they maintained a low tone and on the other half of trials they maintained a high tone. After a lengthy maintenance period of 16 seconds, subjects were presented with a single probe tone, which they compared to the tone held in mind in a change detection task.

Our study was specifically designed for MVPA: (1) in order to decode, which tone was perceived by predicting activity pattern obtained from the period of the probe tone presentation; (2) and to decode, which particular tone, a low or a high tone, was held in mind by predicting activity patterns obtained from the memory maintenance period. Thus, the major goals of this study were: (1) to test whether tonotopically organized areas representing pitch are involved in the perception and maintenance of pure tones in WM; and (2) to determine whether such contents can be decoded from auditory cortex.

Using MVPA to define different patterns of activity corresponding to the perception of different tones (high vs. low), activity patterns in Heschl Gyrus (HG) could successfully predict which tone was perceived. Activity patterns in auditory regions recorded during the maintenance period could also be decoded successfully in order to predict, which tone was held in mind. Regions include Heschl Gyrus (HG), planum temporale (PT) and the whole of superior temporal gyrus (STG). The current study shows that HG is involved in perception and maintenance of pure tones in WM. Furthermore, auditory cortex (HG, PT and STG) can retain specific auditory information about basic auditory features (pitch of pure tones) held in WM, over periods of many seconds in the absence of an external stimulus.

6.3 Methods

6.3.1 Participants

12 healthy adults (8 female, mean age 29 years, age range: 19-53) participated in this study after providing written informed consent to procedures approved by the local ethics committee. Participants were paid for their participation and were selected based on the following criteria: normal hearing and no musical training. A single subject had to be excluded from further analysis, due to the occurrence of two Heschl Gyri in each hemisphere, leaving N11 (7 female, mean age 30 years, age range: 19-53) to be included in further analysis. 1 participant was left-handed.

6.3.2 Stimuli

Stimuli consisted of pure tones, logarithmically sampled at random either from a low range of 200-300Hz or high range of 2500-3000Hz. Stimuli from either range are represented in areas well separated on the tonotopic plane (Formisano et al., 2003; Humphries, Liebenthal & Binder, 2010; Da Costa et al., 2011). Auditory stimuli were generated at a sampling rate of 44.1 kHz in Matlab version R2013b (Mathworks Inc.) and presented using Cogent (http://www.vislab.ucl.ac.uk). Sounds were delivered binaurally through MRI conform in-ear headphones (Sensimetrics, S14, http://www.sens.com/s14/) at an intensity of 70dB.

6.3.3 Procedure and Experimental Design

Outside the scanner, subjects performed a pitch discrimination task (1 block of 16 trials), consisting of an adaptive procedure to assure pitch discrimination thresholds were in the normal range. This was followed by a practice block of the same memory experiment (see Figure 6-1), which was afterwards also carried out inside the scanner. This practice block contained 24 trials.

The start of a given trial of each memory block was indicated by text appearing onscreen, “Sounds to start soon”. This message was presented for 1 second and was followed by the presentation of two pure tones (memory encoding). There was always a low and a high tone contained within each set of sounds, where the order in which they appeared was randomized. Therefore the low tone could e.g. be presented at the first or second serial position and vice versa. The stimulus duration was 500 milliseconds with an ISI of 1 second. The total duration of the stimulus encoding period was 2 seconds.

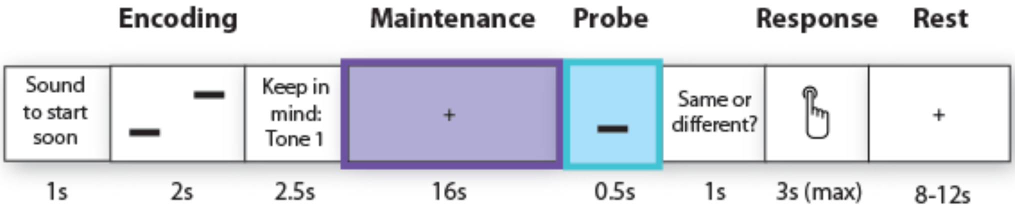


Figure 6-1 *Sample trial*

On each trial, a message indicated that sounds are going to be played, which was followed by the presentation of 2 tones at memory encoding. On the given trial a low tone is followed by a high tone, where in the given scenario, only the 1st tone (here: low) has to be actively maintained for 16 s throughout the memory maintenance period (in purple). We decoded activity patterns from the maintenance period in order to predict, which tone (here: low) was held in mind. Next, a probe tone was presented (see blue interval). We decoded activity patterns from the period of the probe tone presentation in order to predict, which tone was perceived (here: low tone). Participants had to decide, whether the probe was the same or different from the tone held in mind. They indicated their response by making a button press. This decision period was followed by a resting interval of variable length.

Next, another message appeared onscreen for 2.5 seconds, indicating which tone subjects had to actively maintain in mind. For example, “Keep in your mind: Tone 1”. Whether tone 1 or tone 2 had to be held in mind was randomized. Out of 24 trials, on half of them a low tone had to be maintained, while on the other half a high tone was held in memory. After subjects knew which tone to maintain, they had to actively keep this particular tone in mind for 16 seconds, whilst maintaining gaze on a fixation cross at the centre of the screen. Finally a probe tone was presented for 500 milliseconds. Participants had to decide, whether this tone was the same or different from the one held in mind. The message “Same or different?” appeared onscreen for 1 second followed by a button press response. Subjects were instructed to make a response as quickly as possible without making mistakes within a time-window of 3 seconds. Each trial was followed by a resting period of variable length, jittered at 8-12 seconds.

Inside the scanner, a functional localizer was followed by four experimental memory blocks of 24 trials each. Afterwards tonotopic maps were acquired for each participant in two blocks, before acquiring a structural scan. Tonotopic mapping was carried out in a similar way as described in Da Costa et al. (2011). Throughout the scanning, participants had to keep their gaze on a fixation cross, presented at the centre of the screen at all times, which was controlled by eye-tracking. We also acquired physiological measures, such as respiration and heart rate, which were modelled as physiological regressors in our GLM analysis (see MRI univariate analysis). The total duration of the experiment was approximately 2 hours (scanning time) and 15 minutes (preparation time: outside the scanner).

6.3.4 Functional Imaging/MRI data collection

All imaging data were collected on Siemens 3 tesla Quattro head-only MRI scanner (http://www.siemens.com) at the Wellcome Trust Functional Imaging Centre. Stimuli were presented in an event-related paradigm (see procedure), with an inter-trial interval (rest period) jittered between 8 and 12 seconds. MRI images were acquired continuously (3 tesla; TR, 1.1 s; TE, 15.85 ms; FA = 15 degrees; 3D sequence; slices covering the whole brain; nominal flip angle; isotropic voxel size, 2 mm; matrix size, 96 x 96). A single scan was acquired per trial (4 blocks x 24 trials), yielding 96 scans in total per participant. After the fMRI run, a high resolution (1 x 1 x 1 mm) T1-weighted structural MRI scan was acquired for each subject.

6.3.5 MRI univariate analysis

MRI data were analysed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm). Pre-processing: images were realigned to the first volume, then normalized to stereotactic space and finally smoothed with a 3D Gaussian kernel with full-width at half maximum of 8 mm. After pre-processing was complete, general linear model (GLM) was used for statistical analysis. The design matrix consisted of functions encoding the onsets of different events convolved with a hemodynamic response function (HRF) and five parametric regressors.

The following events were included in the design matrix and modelled as parametric regressors: (1) onsets of the stimulus presentation at encoding, (2) onsets of the maintenance period, (3) onsets of the probe tone presentation, (4) onsets of the decision period (button press) and (5) onsets of the resting period. The design matrix also included non-parametric physiological regressors of non-interest, such as respiration and heart rate. A high pass filter with a cut off frequency of 1/128 Hz was applied to remove low frequency variations in the BOLD signal. The GLM for each subject was estimated, the contrasts of parameter estimates for each individual subject were entered into second-level *t-tests* to form statistical parametric maps and a whole-brain random effects analysis was implemented. Estimating this GLM produced a volume of beta values for each trial of the memory task, representing overall activation in each voxel of the brain for that trial. Beta images were used as the basis for classification analyses. Classification analysis was performed on unsmoothed data.

6.3.6 Multivariate Analysis

The principal aim for MVPA is to incorporate information distributed across a set of voxels that might discriminate between two or more experimental conditions (Haynes & Rees, 2006; Norman et al., 2006). The current study was explicitly designed for multivariate pattern analysis (MVPA). Classification analysis was used to predict, whether a high or a low tone was perceived (decoding from the period of probe tone presentation, see figure 1 in blue) or whether it was held in mind (decoding from the memory maintenance period, see figure 1 in purple). MVPA was carried out to decode the type of information associated with perceptual in contrast to memory processes.

For each subject, regions of interest (ROIs) were selected for MVPA using the anatomical toolbox (Eickhoff et al., 2005) in SPM8. The following bilateral ROIs were located as HG, putamen (PT), superior temporal gyrus (STG), primary visual cortex (BA17) and somatosensory cortex (BA123). All voxels contained in each ROI were included in classifier analysis, which does not eliminate noise input from voxels that may contain white matter, empty space or gray matter that was not strongly activated by events of interest.

Beta values from each voxel for each trial were extracted from the GLM analysis of the unsmoothed data and fed into a support vector machine (SVM) classifier, using custom Matlab code centered around the LibSVM toolbox (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). Classification analysis was run on two events of interest convolved with HRF and the following parametric regressors: (1) onsets of the probe tone presentation to decode perceptual processes and (2) onsets of the maintenance period to decode memory processes. The method of linear support vector machine learning was used as ‘leave one trial out’. In this analysis, data was taken from 23 trials in order to train the classifier, where the remaining trial was used as test data. In other words, the classifier predicted to the 24th trial. Then it was rotated through all 24 trials and 4 blocks, which trial was used as test. Each trial received a score indicating the classifier’s confidence that the activity pattern on a given trial resembles a low tone or a high tone, which was either perceived or held in mind (two separate events on each trial). Then the scores for all trials were collapsed across the given classifier.

6.3.7 Further analysis

Simple summary statistics were obtained and ANOVA was carried out to describe and evaluate the behavioural data collected, as well as to compare classifier performance across ROIs. Linear regression analysis was employed to test for correlations between individual classifier scores across perceptual and memory processses. Further linear regression analyses was used to test for correlations between classifier scores and mental imagery questionnaires (auditory imagery: Bucknell Auditory Imagery Vividness Scale (BAIS); and visual imagery: Marks' Vividness of Visual Imagery Questionnaire).

6.4 Results

6.4.1 Behavioural Results

Outside the scanner, percentage correct on the pitch discrimination task was 95% (SEM=0.03) and 82% (SEM = 0.03) on the memory practice block. Inside the scanner, percentage correct on the memory blocks was 88% (SEM=0.02). Based on the behavioural data acquired inside the scanner, there were no serial order effects: there was no difference in performance when either the 1^{st} or the 2^{nd} tone was probed in the sequence (One-way ANOVA: F(1,22)=0.317, p=0.58). There was also no significant difference in performance between probing a low or a high tone (One-way ANOVA: F(1,46)=1.86, p=0.18). Mean reaction time (RT) was 903.92ms (SEM=73). Subjects made faster responses on correct compared to incorrect trials (F(1,46)=17.11, p<0.001)

6.4.2 Results from univariate analysis

In a GLM analysis we aimed to determine the brain areas in which BOLD activity varies as a function of events of interest corresponding to parametric regressors included in the design matrix (see MRI univariate analysis above). The following contrasts of interest were evaluated:

- stimulus encoding vs. rest
- maintenance vs. rest
- probe tone presentation vs. rest, and
- decision period (button press) vs. rest.

No response was observed for any of the contrasts of interests in any of the ROIs (HG, PT, STG, BA17, BA123) after family-wise error (FWE) corrections.

6.4.3 Decoding Perceptual Processes

In order to decode perceptual processes, the classifier was run to predict activity patterns acquired during probe sound presentation. In this period a single pure tone (either low or high) was presented only. Activity patterns in HG could successfully predict with a high accuracy of 80% (p=0.001), which tone was held in memory (see Figure 6-2, left panel). However, the classifier could not predict similar processes for primary visual cortex (BA17), which resulted in at chance classification (mean classification accuracy = 51%, p=0.37), see Figure 6-2 (right panel). This analysis confirms that HG is involved in the perception of pure tones, whereas BA17 is not.

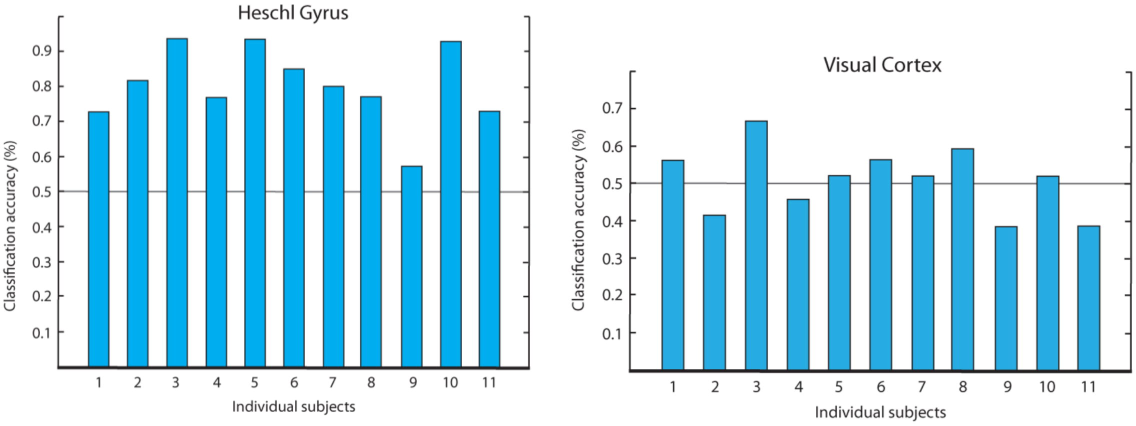


Figure 6-2 *Classifier performance for auditory perception*

The classification accuracy is shown for each participant, where the horizontal line at 0.5 indicates at chance classification. The result obtained for HG shows that the classifier can predict successfully, whether a low of a high tone was perceived. However, running the same classifier on a control ROI, here BA17 (primary visual cortex) could not predict which tone was perceived.

6.4.4 Decoding auditory memory in auditory cortex

Activity patterns in auditory areas, Heschel Gyrus (HG), Planum Temporale (PT) and Superior Temporal Gyrus (STG), acquired during the maintenance period could successfully predict which tone was held in memory (see **Error! Reference source not found.**). Mean classification accuracy for HG was at 60% (p=0.001) and at 59% (p=0.004) for PT. When extending our region of interest to include the whole of STG, the classifier could still predict, which tone was actively maintained in memory. Mean classification accuracy for STG was 59% (p=0.001). Scrambling the data for

HG, using a permutation test, resulted in at chance classification (mean=50%, p=0.44), see Figure 6-3 (top right panel).

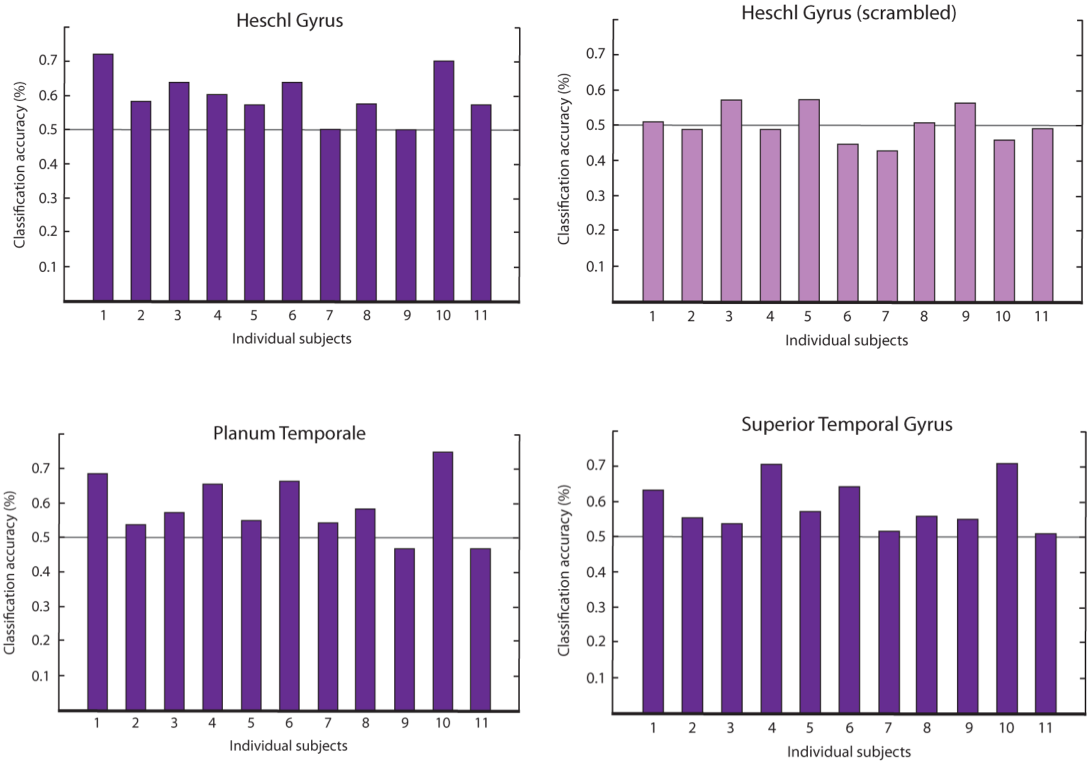


Figure 6-3 *Classifier performance for auditory memory in auditory regions*

The classification accuracy is shown for each participant, where the horizontal line at 0.5 indicates at chance classification. The results obtained for HG (top left panel), for PT (bottom right panel) and STG (bottom right panel) show that the classifier can predict successfully, whether a low of a high tone was held in mind. However, running the same classifier on scrambled data for HG (permutation test) could not predict which tone was held in mind.

For HG, linear regression analysis revealed no significant correlation between the individual classification accuracy scores obtained for perceptual (Figure 6-2, left panel) vs. memory processes (Figure 6-3, top left panel), (p>0.05).

6.4.5 Decoding non-auditory Sensory Cortices

Classifier performance dropped for activity patterns in non-auditory sensory cortices compared to auditory regions, when decoded from the maintenance period (Figure 6-4). Activity patterns in somatosensory cortex (BA123) could not be predicted by the classifier, which performed at chance (mean classification accuracy = 52%, p=0.15), see Figure 6-4 (left panel). Thus BA123 does not seem to be involved in maintenance of tones, where no motor strategies were employed to aid this process. However, activity patterns in primary visual cortex (BA17) could be predicted above chance (mean classification accuracy = 55%, p=0.007), (Figure 6-4 (right panel)). Therefore, BA17 appears to have some form of involvement in memory maintenance of the auditory information tested, perhaps in the form of visual imagery as part of a mnemonic strategy (e.g., imaging a high position in visual space for a high tone).

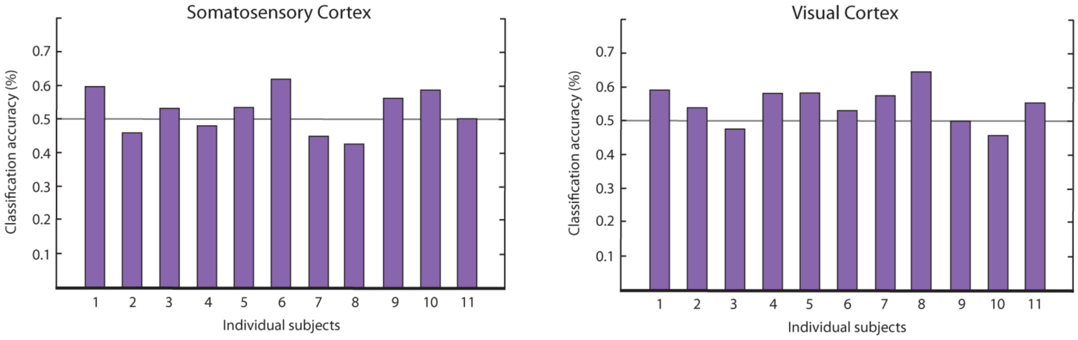


Figure 6-4 *Classifier performance for memory in non-auditory regions*

The classification accuracy is shown for each participant, where the horizontal line at 0.5 indicates at chance classification. The results obtained for somatosensory cortex (left panel), show that the classifier could not predict from activity patterns found in this region, which tone was maintained in memory. In contrast, the classifier could predict above chance, which tone was held in mind based on activity patterns found in primary visual cortex (right panel).

6.4.6 Memory strategies

Finally we performed linear regression analysis to test for correlations between classifier accuracies for different ROIs and test scores obtained from mental imagery questionnaires. The Bucknell Auditory Imagery Vividness Scale (BAIS) was used to measure auditory imagery and Marks' Vividness of Visual Imagery Questionnaire was used to assess visual imagery. There was no significant correlation between classifier performance in auditory ROIs, such as HG, PT and STG, and auditory imagery scores ($p>0.05$). Again there was no significant correlation between classifier performance in visual cortex (BA17) and visual imagery scores ($p>0.05$).

6.5 Discussion

We decoded activity patterns from perceptual as well as memory processes. Whether a low or a high tone was perceived as a probe sound could be decoded successfully with a classification accuracy of 80% from HG, which was significantly higher than the accuracy level of 60% associated with memory processes. The analysis confirms the primary involvement of HG in the perception of tones (Da Costa et al., 2011), where the decoded content is strongly linked to tonotopically organized representations of sound frequency. Additionally, we show that the same patterns decoded from HG are not only involved in perception, but also in the maintenance of tonal information in memory.

In order to directly compare across perception and memory we assume a common sensory mapping, although equivalence across such levels of processing has not been formally established, yet. The lower classification accuracy associated with memory maintenance may be due to the fact that some degree of specificity in tonotopy is lost leading to more distributed patterns of activity. This could be a result of memory traces decaying over time, where representations of tones seem especially fragile compared to other types of sensory information (Nq, Plakke & Poremba, 2009). The lower classification accuracy for memory maintenance could also mean that this period cannot be described by a single process, but represents multiple stages (Sternberg, 1999; McLeod, 2007), where information fluctuates over time.

With respect to the maintenance of tones in memory, we show that apart from HG also other auditory regions, including PT and the whole of STG form representations of memory content, as their activity patterns could be decoded successfully in all of these auditory regions. Activity patterns represent whether a low or a high tone had been held in mind. Classification accuracy is equal across auditory regions tested.

What are the neural codes underlying pattern recognition with regards to our findings? Since low and high tones are represented in distinct regions on the tonotopic plane, activity patterns found are likely to reflect tonotopic organization. The presence of at least two tonotopic maps with a mirror symmetric 'high-low-low-high' progression has been demonstrated in humans (Formisano et al., 2003; Talavage et al., 2004; Woods et al., 2009; Humphries et al., 2010; Striem-Amit, Hertz & Amedi, 2011; Da Costa et al., 2011). Da Costa el al. acquired tonotopic maps in individuals using high-resolution fMRI (at 7T) and reported that functional tonotopic maps of primary auditory cortex correspond to the anatomical shape of HG. Therefore, our interpretation of activity patterns reflecting tonotopy may best explain the results we obtained for HG, whereas in a control analysis, scrambling the data in this region resulted in at chance classification.

More specifically, as we chose pure tone stimuli from two different ranges (low range: 200-300Hz and high range of 2500-3000Hz) represented in areas well separated on the tonotopic plane (Formisano et al., 2003; Humphries, Liebenthal & Binder, 2010; Da Costa et al., 2011), the activity patterns associated with either range of tones are likely to reflect activity associated with local maxima found along the macrostructure of HG. Several tonotopy studies mapping frequency responses in HG (Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011) show a consistent 'V' or 'U' shaped configuration formed by high frequency areas (belt), surrounding/framing low frequency areas (core), (Baumann, Petkov & Griffiths, 2013). Thus, activity in one local maximum centred on lateral HG may reflect the organization of the pattern associated with the perception or memory of low tones.

In contrast, the pattern associated with high tones reflects activity found in two local maxima, which are found anterior and posterior to lateral HG. While the anterior high frequency area is predicted by the first transverse sulcus confining HG on its anterior side, the posterior high frequency area extends on the lateral side beyond the posterior bank of HG into PT, but often overlaps with HG on the medial side (Baumann et al., 2013).

Activity patterns could also be predicted successfully in another auditory region: PT. Da Costa et al. (2011) report that smaller frequency progressions were also observed in some cases posterior and anterior to the main two tonotopic maps (corresponding to fields in HG; Rivier and Clarke, 1997), consistent with previous findings at the group-level obtained with ultra-high field 7T fMRI (Formisano et al., 2003). Although these maps were less consistent across subjects (Da Costa et al., 2011), a less detailed level of tonotopic organization is preserved beyond HG. As the frequency ranges (low tones: 200-300Hz, and high tones: 2500-3000Hz) used in our study are highly distinct, the activity patterns obtained may still reflect frequency specific responses in PT. Responses to tones are weaker and comparably less specific than found in HG (Rauschecker, Tian & Hauser, 1995; Rauschecker et al. 1997, Tian et al. 2001) and may contain more complex representations of sound frequency (Warren, Jennings & Griffiths, 2005).

Our interpretations about the neural codes reflecting the decoded activity patterns are based on anatomical data supporting the frequency specific organization of auditory cortex in humans (and non-human primates). Although frequency response patterns provide a detailed account of the anatomical organization of different frequencies, it should not be neglected that there is inter-individual variability in tonotopy due to variable gyrification and adaptation to speech (Baumann et al., 2013), which affects the neural codes underlying the representations of frequency specific information.

Extending our region of interest to include not only HG and PT, but the whole of STG, activity patterns on whether a low or a high tone had been held in mind could still be decoded at the same accuracy level compared to HG and PT alone. In this respect, the activity pattern found may resemble a combination of at least two different types of neural codes: (1) patterns may in part represent tonotopy, which applies to regions of STG covered by HG and to a certain extent of PT. (2) Additionally a more complex neural code, which does not preserve tonotopy, may represent information in auditory memory, found in non-tonotopic parts of PT, as well as the remaining region of STG. Thus, the property of frequency specific tonotopic coding may be translated to a different more complex neural code, which still preserves some degree of frequency-specificity.

A recent behavioural study investigated the nature of the units of auditory memory representations (Mathias & von Kriegstein, 2013). The authors distinguish between two types of representations. The first refers to *percepts*, where for example the specificity of tonal information (sound frequency) is retained (Massaro, 1972). In vision, this would correspond to the idea that representations during mental imagery are 'image like' (depictive), where spatial relations are preserved and retinotopic areas are activated (Kosslyn & Ochsner, 1994). The second type of representation might be concerned with *acoustic properties* evoking a particular percept, where specificity is lost (Mathias & von Kriegstein, 2013).

In vision, it has been argued that mental representations are held using labels (descriptive) and spatial relations are not preserved (no retinotopic activations). In audition, Mathias and von Kriegstein (2013) conclude that percepts are stored in auditory memory. However, based on our findings both representations may exist in auditory memory: where tonotopically organized *percepts* are the units of representations in HG and parts of PT, while representations in the remaining part of PT and STG correspond to *acoustic properties* (e.g. representations of pitch). Therefore, a memory store may not contain one or the other type of representation. However, the nature of representations stored in auditory memory and its underlying neural codes depend on the levels of processing involved.

In a control analysis, we decoded activity patterns from early sensory cortices other than auditory cortex during the maintenance period. We found at chance classification for somatosensory cortex, confirming that activity patterns associated with memory maintenance are specific to auditory regions. However, we found above chance classification for visual cortex during maintenance of pure tones, but not during the perception of pure tones. Therefore the use of a given visual mnemonic strategy cannot be excluded, which may correspond to a visual-spatial mapping scheme (i.e. visual imagery of tones using spatial coding for their relative pitch height) for sound frequency information about high vs. low tones. Although pure tones presented in our study were not selected on the basis of musical notes, participants may have mapped sound frequencies to vertical locations as represented on any standard music notation system, whereby notes corresponding to higher frequencies are represented with higher spatial positions on the staff.

Previous studies have provided evidence that human brains map sound frequency automatically onto representations of space (high tones are higher in space than low tones), where both dimensions are of integral nature (Rusconi et al, 2006). Forming cross-modal associations between dimensions naturally, might explain why subjects did not specifically report the use of memory strategies in a follow-up questionnaire. Yet, visual codes represented retinotopically do not correspond to tonotopic neural codes. For instance, receptive field gradients reverse direction at the border of primary and secondary visual cortices (Dumoulin & Wasndell, 2008; Baseler, Morland & Wandell, 1999), whereas in auditory cortex there are at least two mirror-symmetric gradients in the center of auditory cortex (Woods et al., 2009; Da Costa et al., 2011; Striem-Amit et al., 2001; Langers & van Dijk, 2012).

Analogous to the current finding that auditory memory content is represented in auditory cortex, previous decoding studies have shown that the contents of visual memory are represented in visual cortex (Harrison & Tong, 2009; Xing et al., 2013, Christophel et al., 2012; Riggall & Postle, 2012). Due to the increased sensitivity and specificity of multivariate over mass-univariate techniques (Haynes & Rees, 2005; Jimura & Poldrack, 2012; Kamitani & Tong, 2005) its use can not only lead to a better characterization of information revealed in neural activity, but one can also easily misidentify a pattern of activity as informative (Todd, Nystrom & Cohen, 2013; Etzel, Zacks & Braver, 2013).

A major concern with previous memory decoding studies in the visual domain is the lack of control conditions in order to prevent such errors. We can only be certain whether a brain area stores specific content if another unrelated area doesn’t, which was addressed by the current study (e.g. the contents of auditory memory are stored in HG, but not in somatosensory cortex). However, previous studies decoding retinotopically represented visual features (Harrison & Tong, 2009, Xing et al., 2013) or non-retinotopically represented features (Christophel et al., 2012) as the contents of visual memory lack similar comparisons of interest, except for one study (Riggall & Postle, 2012) comparing classifier performance in different ROIs and periods of interest to performance at rest (ITI). Future research using MVPA might profit from ensuring the use of such controls in order to better identify the functional signatures of patterns revealed in neural activity.

6.6 Conclusion

This study was designed to investigate the frequency specific involvement of pure tones in memory processes. The results demonstrate that auditory regions (HG, PT and STG) can retain specific information about basic auditory features (frequency of pure tones) held in memory. This is in line with the idea that patterns of brain activity observed in auditory cortex reflect distributed activity at the population level, which is sustained over several seconds in the absence of an external stimulus. Such contents held in memory reflect *percepts* or *acoustic features* depending on the levels of processing involved. Future research might investigate the degree of continuity from *percepts* to *acoustic features* as a function of tonotopic specificity found in auditory cortex.

Chapter 7. General discussion

In this thesis, the mechanisms underlying WM across the senses, including vision, audition and touch were investigated. In the current chapter, I first discuss how these studies add to previous work in the field, before examining their limitations and making suggestions for future research.

7.1 The resource model applied to auditory and tactile WM

In vision, a novel approach to measure recall precision (instead of measuring whether an item was remembered or not) has stimulated reconsideration of models describing processes in WM (Wilken & Ma, 2004; Bays & Husain, 2008; Bays et al., 2009). Instead of regarding visual WM as an all-or-none system, measuring precision allows us to quantify the fidelity of a memory representation. Before the development of such novel measures, it has been widely held that visual WM is limited to a fixed number of items it can hold (about four: Luck & Vogel, 1997; Cowan, 2001). Similarly, with regards to other sensory modalities as only assessed by forced-choice or 'binary response' methods, storage has been considered to be capacity limited to a fixed number of items (one or two in audition, for complex sounds (timbre): Golubock & Janata, 2013) and verbal material: Fougnie & Marois, 2011; Saults & Cowan, 2007; one or two in touch, for vibrotactile flutter: Bancroft & Servos, 2011; Bancroft, Hockley & Servos, 2012).

Measuring precision has led to a fundamentally different conclusion: visual WM represents a limited resource which can be allocated flexibly across objects. As the resource is shared out across an increasing amount of information, there is a decline in the mnemonic resolution strength with which items are represented in WM and therefore in response fidelity. Thus, with an increased amount of information being held in WM, representations become noisier. However, previous research has only shown that the resource model accounts for WM processes of different types of stimulus material in the visual modality (Bays & Husain, 2008; Bays et al, 2009; Gorgoraptis et al., 2011; Zokaei et al., 2011) and has not been tested for material perceived through senses other than vision. Here, the aim was to develop continuous measures assessing the fidelity of WM representations for auditory (chapter 2 and 3) and tactile (chapter 4) modalities.

7.1.1 Auditory WM resource: Pitch

In chapter 2, a new pitch-matching task was introduced, designed to measure auditory WM precision for sequences of pure tones at different pitches. In this task a variable number of tones (1-4 items) were presented in succession on each trial and subjects had to match the pitch of one of these tones from memory. The results from this study were in line with predictions made by the resource model of WM: the more items held in WM, the less precisely each item was recalled. Precision clearly declined with an increase in memory load (number of tones presented within a sequence). Even when adding a single item to a previous one, a significant drop in precision was produced (Figure 2-3A).

The fixed capacity account (slot model) cannot explain this decline in precision, which predicts optimal performance until a capacity limit is reached (Cowan, 2001; Fougnie & Marois, 2011; Luck & Vogel, 1997; Saults & Cowan, 2007). Even at the highest memory load of four items, capacity limits were not yet reached as performance remained above chance. Again, this result cannot be explained on the basis of a fixed capacity or hybrid models (*discrete* resource, see Figure 1-4), which both predict memory performance to drop to chance once the capacity limit (here 1 -2 items; Golubock & Janata, 2013; Fougnie & Marois, 2011; Saults & Cowan, 2007) is exceeded.

The shared resource model developed for vision proposes that memory representations are noisy and that the amount of noise varies with memory load (Wilken & Ma, 2004; Bays & Husain, 2008; Bays et al., 2009). To confirm that the increase in variability (associated with noisy pitch representations) was not due to random guesses made by the subjects, a probabilistic mixture model (Bays & Husain, 2008; Zhang & Luck, 2008) was fitted to the results from the pitch-matching experiment (Figure 2-2). The model separates the variability due to random guesses from variability with which each item is maintained in WM. While the variability due to random guesses was constant across memory loads, the variability associated with representations of each item held in WM increased with memory load.

Thus, as the number of tones held in WM increased, the memory representation for each item became noisier, whilst the amount of random responses was equal across memory loads (even when the number of items exceeded two – previously estimated as the capacity limit for auditory WM). As there was no increase in guessing at such at capacity limits, the results from the mixture model further support the view that pitch WM, like visual WM, is best described by the resource model of WM.

As WM performance depends on the ability to encode and maintain selected information most relevant to the task (Awh et al., 2006; McNab & Klingberg, 2007), in another pitch-matching experiment, I manipulated the factor of task-relevance of different item-positions within a sequence by pre-cueing. Precision was highest when the tone was most likely to be probed and therefore more task-relevant compared to a non-cued tone or neutral condition (Figure 2-5). Enhancing the priority of a specific tone in the sequence thereby resulted in a clear benefit in precision, but importantly it came at a cost in precision for other tones. This is just what one might predict from a limited resource account, where devoting more resource to one item necessarily involves giving less to another.

Similar results have been obtained previously for visual WM, where a gain in precision has been demonstrated for the cued item (Gorgoraptis et al., 2011; Zokaei et al., 2011; Bays et al., 2011), as well as a cost for other stimuli recalled less precisely (Bays & Husain, 2008; Gorgoraptis et al., 2011; Bays et al., 2011). Thus, it has been shown that WM resources can be allocated flexibly according to priorities, which represents another important distinction between slot and resource models of WM.

Although the measure of recall precision is more informative than binary response methods, the way it was employed here cannot directly capture the amount of resource allocated to each item in the sequence. To be able to estimate the proportional amount of resource allocated to each item would be useful to determine which type of resource model explains the data best, e.g. the equal precision or variable precision model (see Figure 1-4; Mazyar et al., 2012; Van den Berg et al., 2012; Bays, Husain & Ma, 2014). While the *equal precision* model predicts that the resource is allocated evenly across all items, the *variable precision* model predicts that the amount of resource an item receives varies randomly across items. To draw a distinction between different resource accounts would be an important goal for future research. It would also be interesting to investigate, whether the effects observed (e.g. recency and prioritization effects) are due to properties of resource allocation only (proportion allocated), or whether they may also be due to temporal parameters, such as memory decay over time.

Although the continuous response method of pitch-matching employed here is a useful means to measure active reproduction of information from memory, it also comes with certain limitations. Pitch-matching (the adjustment of a continuous probe) takes several seconds. The duration of the continuous probe sound is therefore much longer compared to the duration of each individual sound presented at encoding. Such a perceptual difference may simply add noise to memory representations. Furthermore, as the continuous probe is played for several seconds until an adjustment is made (final response), the sound may cause interference with the internal memory representation of the target held in mind. Additionally, the target representation may be fragile and simply as matching takes several seconds it decays over time.

Both parameters, interference and decay over time, are not accounted for in the experiments presented in chapter 2. However a direct comparison between binary response measures (e.g. with different magnitudes of change between the target and probe) and continuous measures may be useful to disentangle factors (interference, decay etc.) influencing the quality of memory representations other than variability due to responses made at the target stimulus at a given memory load.

With regards to future research, the pitch-matching task might profitably be applied to measure recall precision of representations held in auditory WM in different populations: e.g. musicians or individuals with perfect pitch. Furthermore, it would be interesting to investigate the trade-off (interference) between multiple items within a single modality (different types of sounds). The pitch-matching task could therefore be combined with another adjustment task for a different auditory feature (e.g. speech sounds, see chapter 3; loudness, duration, timbre, harmonicity etc.).

Combining the auditory feature of pitch with another feature represented in another sensory modality (e.g. visual orientation) in a dual task would be another method to test for trading relationships, here: across modalities. Using dual tasks represent a means to test for inter- or cross-modal interference, which gives insights on whether information stores are modality specific or not (see Figure 1-3 for the multi-sensory models of WM). However, the challenge with the use of continuous measures (recall precision) is to match stimulus ranges in terms of the amount of values falling within this space as well as in terms of difficulty.

Although the resource model can explain results obtained behaviourally, the exact nature of this representational medium remains to be established and is an important goal for neurophysiology studies. Most electrophysiology (Fuster & Alexander, 1971; Goldman-Rakic, 1996; Romo et al., 1999) as well as computational studies (Wang, 2001) are limited to only testing WM performance for a single item. In order to distinguish between different models of WM, it would be important to test memory performance beyond a single item and to also include higher memory loads.

As recall variability is likely to reflect noise in memory representations, the resource model makes directly testable predictions towards the firing behaviour of neuronal populations. Firing patterns may reflect different mnemonic codes found at different memory loads. A number of visual behavioural animal studies in non-human primates have begun to use higher memory loads in order to distinguish among models (Buschman et al., 2011; Elmore et al., 2011; Heyselaar, Johnston & Pare, 2011; Lara & Wallis; 2012). Future research along those lines could also be extended to other sensory modalities.

7.1.2 Auditory WM resource: Phonemes

After showing that the resource model can be extended to audition, at least for the basic stimulus feature of pitch, it was of interested to test whether its predictions also hold for more complex auditory information. Traditionally, in the WM literature a wide body of research has used verbal stimulus material (e.g. spoken digits (Morgan, Chambers, & Morton, 1973), letters (Conrad, 1964; Conrad & Hull, 1964), syllables (Gupta, Lipinski, Abbs, & Lin, 2005) or words (Drewnowski, & Murdock, 1980; Haberlandt et al., 2005)). In order to test, whether the resource model of WM can also account for verbal information, a speech matching task was designed to measure precision (see chapter 3).

Participants were presented with sequences of speech sounds of variable length (1-4 items), where the precision with which listeners matched the vowel quality of a probe (composed of a particular F1 & F2 value, see Figure 3-1) to the vowel quality of the target was analysed. Importantly, as stimuli were sampled from a circular speech sound continuum, participants could also adjust vowel sounds along this dimension in order to make a response. The development of this analogue reporting procedure allowed us to estimate the fidelity of speech sound representations, as previously applied to measure WM for pitch (see chapter 2 or Kumar et al., 2013) and visual features (Wilken, & Ma, 2004; Bays, & Husain; 2008, Bays, Catalao, & Husain, 2009; Ma, Husain & Bays, 2014).

As a result of speech matching, precision decreased with an increase in memory load, resembling the pattern previously observed for pitch (chapter 2, experiment 1), where the addition of each item to be held in WM let to a further decline in precision. This decline was especially pronounced when comparing between load 1 and 2, where the addition of a single item to a previous one produced a sharp drop in precision. This overall pattern of decrease is in line with previous results on auditory WM for pitch and also supports predictions made by the resource model. Performance on speech matching (as for pitch matching) also remained above chance at the highest memory load of four items.

Thus, the results from the speech matching paradigm cannot be explained on the basis of a fixed capacity account and not on the basis of a hybrid model (*discrete* resource, see Figure 1-4), but can be captured best by the resource model. Further research should clarify, which type of resource model, equal vs. variable precision (figure Figure 1-4), describes verbal WM best.

However, in addition to effects of memory load on precision, unlike the findings for pitch, serial order effects seemed to be more pronounced for WM of speech sounds: for memory load of 4 items, there was a clear recency and primacy effect. Such U-shaped serial position curves (Figure 3-3B, memory load 4) demonstrating that the first and last items in the sequence are remembered best, has been reported in many verbal WM studies: for letters or digits (Baddeley, 1986; Conrad, & Hull, 1964; Levy, 1971), syllables (Gupta, Lipinski, Abbs, & Lin, 2005) and words (Haberlandt et al., 2005). Primacy and recency effects can also be described as 'edge effects' (Botvinick, & Plaut, 2006) or 'temporal distinctiveness effects' (Glenberg, & Swanson, 1986). As no item precedes the first or follows the last one in the sequence, those positions are most distinct, which also relates to the temporal order of events (sequential presentation of items; Acheson, & MacDonald, 2009).

Thus, in comparison to WM for pitch, although the resource model can capture the results from both studies, there also seem to be some fundamental differences in processing across types of information: pitch may be treated more acoustically as compared to speech. On the one hand, one might assume that speech sounds as presented and manipulated in this study could be considered as a purely acoustic dimension, as they represent different timbres. This is supported by the fact that vowels are perceived as continuous (Fry et al, 1962; Shouten, & van Hessen, 1992). However, vowels are also considered to belong to different *categories* (Harnad, 1987; Liberman et al., 1967; Holt & Lotto, 2010). Thus, on the other hand, speech sounds may not only be represented in WM as purely acoustic timbres, but as sounds that belong to *categories* at the same time.

In order to understand how the shift from continuous to categorical representations of speech sounds is affected by memory load, a MDS (multidimensional scaling) algorithm as well as a mixture model were fitted to the results. While MDS was useful in mapping responses made by participants back onto the continuous vowel space, the same mixture model as previously applied to pitch was adopted to account for information that is both continuous and categorical.

Using MDS, an overall shrinkage of the vowel space (see Figure 3-5) with an increase in memory load was observed, where memory representations become less continuous and noisier. Additionally, particular regions of the space indicate the formation of vowel clusters, which is especially pronounced in the region around the British vowel /u/. Thus, WM representations near /u/ may be considered categorical compared to other regions in the space. Interestingly, the mixture model (see Figure 3-6) further supported the finding from MDS that representations become noisier with an increase in memory load (increase in the proportion of continuous responses). Additionally, the proportion of guessing as well as proportion of categorical responses relative to continuous ones increased with memory load. Thus, as representations of the target vowel become less continuous, memory for the actual stimulus fades and participants begin to guess, whilst their memory for the stimulus category is retained. Therefore, *categorical memory* may be more resilient to interference from other items.

the

This issue is a fundamental question for many aspects of experimental psychology and cognitive neuroscience: to what extent is information is stored in a direct manner, e.g. in stimulus space, or in a more abstract or categorical format? The speech matching paradigm employed here offers a novel method of probing the interaction of continuous and categorical information storage in the brain, within a single task. Speech sounds are particular interesting to investigate, since categorical speech boundaries emerge early in life (Eimas et al., 1971) and are further modified based on one's native language (Kuhl et al., 1992). Thus, for different languages category boundaries are represented at different locations along the formant space, on which our task is based.

With regards to our particular speech sound continuum, if probed for speakers of different languages, categorical effects may represent a function of the particular vowel system underlying one's native language. For example, here native British English speakers were tested, while the stimulus space employed here was also designed around the corresponding vowel system. One of the reasons why clustering around /u/ was observed might be because only little variation of this particular speech sound occurs in British English.

However, in German there is comparably more variation along this region, where more /u/-sounds correspond to actual sounds found in German language (see Bohn & Fledge, 1992 for a study on how German speakers perceive English vowels). Thus native German speakers tested on the same task might show less clustering around the region of /u/, whereas more clustering may occur in other regions of the space.

It would also be interesting to test this for other languages with fundamentally different vowel systems: e.g. Danish (Uldall, 1933; Ejstrup & Hansen, 2004) or Japanese (Keating & Huffman, 1984). In this manner one could specify for speech under which circumstances we remember information by category as opposed to along a sensory continuum.

Although the idea that continuous physical signals can be mapped onto discrete mental categories was first demonstrated in speech perception (Lieberman et al., 1957), continuous vs. categorical perception has also been shown in other domains. Those include other types of auditory information and information represented in sensory modalities different from audition, implicating a more general phenomenon.

Another example of auditory stimuli is musical intervals (Burns & Ward, 1978; Zatorre & Halpern, 1979), while in vision this phenomenon applies to the encoding of colour (Bornstein & Korda, 1984; Kay & Kempton, 1984; Bird et al., 2014) and facial expressions (Bruce & Young, 2012; Harris, Young & Andrews, 2012). In the sensory modality of touch, judgement of continuous signals such as tactile distances on the body surface can be distorted by categories of anatomical landmarks such as body parts (Knight, Longo & Bremner, 2014). Thus future research could extend the interaction of continuous and categorical information established in the domain of perception to WM for various types of information. It would also be interesting to define neural correlates, particularly of the location in the brain where the shift from continuous to categorical representations occurs in the context of information storage. Computational models of WM (Pascanu & Jaeger, 2011; Compte et al. 2000; Bays 2014) might also be able to address this transition of representational states in the future.

7.1.3 Tactile WM resource: Mechanical vibrations

A third set of studies was designed to capture WM processes in the tactile modality (see chapter 4), where a frequency matching paradigm was used to measure the fidelity of vibrotactile WM representations for different memory loads and different frequency ranges associated with different channels of encoding. Importantly, this study assessed for the first time, storage of multiple items (beyond 2) in vibrotactile WM. The overall pattern of results demonstrated a monotonic decline in recall precision with an increase in memory load (Figure 4-2), in a similar way as observed for other sensory modalities (see Figure 7-1 for a comparison across sensory modalities).

In line with previous reports in vision (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Bays et al., 2009; Gorgoraptis et al., 2011; Wilken & Ma, 2004; Zokaei et al., 2011) and audition (chapter 2 or Kumar et al., 2013; chapter 3), the shared resource model captures the relationship between information load and mnemonic resolution strength at which information is represented. For tactile information, a strong fall in precision was observed when an item was added to a previous one (comparison of load 1 vs. load 2), where performance remained above chance for the highest memory load of 5 items. Thus the results suggest that there is not a fixed upper limit to the number of vibrations that can be stored in vibrotactile WM, but instead would be consistent with the shared resource model of WM.

Critically, the results demonstrate that vibrotactile WM can store multiple items (up to 5), whereas most previous evidence suggested that only a single (Bancroft et al., 2011; Bancroft & Servos, 2011) or up to two representations can be stored (Bancroft et al., 2012). Additionally, evidence on the mnemonic resolution strength (precision) at which tactile items are held in mind was not previously established. Not only the number of items (memory load), but also the order in which they appeared had an effect on recall precision. At the higher memory loads of 4 and 5 items, the most recent ones were remembered at a higher precision compared to items presented earlier in the sequence (Figure 4-3).

In addition to plotting mean precision across the group of participants at each serial position, the variability of memory recall around its true stimulus value was determined as raw response error histograms (Figure 4-4). The histograms not only show the response error for each target stimulus (in pink) at each memory load and serial position, but also for each non-target or distractor stimulus (in cyan). Although participants were not asked to reproduce the vibration of non-target stimuli, they were initially instructed to always remember the whole set of vibrations appearing within each sequence. Thus, they may sometimes inadvertently report the vibrational frequency of a non-target (unprobed) stimulus instead of reporting the target frequency, demonstrating a form of misbinding.

As visible from the histograms in Figure 4-4, different types of response were made: (1) those directed to the target or probed item, (2) responses directed to a non-target frequency and (3) sometimes, as observed in chapters 2 and 3, participants might simply guess. Therefore, in another analysis participants' responses were decomposed into three categories, each capturing a different type of response error. The proportion of responses directed at the target decreased with memory load (see Figure 4-5, purple line), in line with the overall pattern of decline in precision with an increase in memory load (see Figure 4-2). This result confirms that the increase in response variability with memory load is largely due to responses directed at the target (majority of responses) and not due to the other types of response.

The finding is consistent with previous findings in the visual (Bays et al., 2009, Gorgoraptis et al., 2011) and auditory modality (chapter 2 or Kumar et al., 2013, chapter 3). As more items are presented within a sequence, the larger the proportion of non-target responses (misbinding errors), (Figure 4-5, cyan line). Such a result has also previously been observed in vision when items are presented sequentially (Gorgoraptis et al., 2011), but is different from the findings for auditory WM. Finally, the proportion of guesses remained constant across all memory loads mirroring the finding for auditory pitch WM (chapter 2).

Since there are two types of receptors in the skin (Meissner and Pacinian corpsulses), distributed at different densities, forming RFs of different sizes and responding to different frequency ranges, the study on vibrotactile WM also addressed the question, whether precision is scaled to receptor density. To address this, the entire dataset was split into two frequency ranges (range 1: 10-50Hz corresponding to channel 1; range 2: 51-100Hz corresponding to channel 2) in order to differentiate between channels. Analysis revealed that there was indeed a significant difference in error scores between channels across all individual memory loads. Therefore the findings raise the possibility that mnemonic resolution is scaled to receptive field (RF) size/receptor density and their properties (e.g. slowly vs. rapidly adapting) within a given channel. Encoding a stimulus via Meissner corpsules is associated with a more precise stimulus representation in WM compared to a less precise representation of a stimulus frequency encoded via Pacinian corpsules.

Such findings make direct predictions towards information processing in other sensory modalities. For example in vision, precision may also be scaled to RF size, where a stimulus encoded via foveal vision might be encoded at a higher resolution strength compared to peripheral vision, since the receptor density in the centre of a RF is higher than in the surrounding regions (Hubel, 1963; Kandel, Schwartz & Jessell, 2000). In the auditory system, receptive fields can correspond to regions in auditory space or frequencies (Warren, 2008). Thus, future research could test how RF size constraints at *encoding* are associated with the formation of a stimulus representation for WM in the brain, as measured by recall precision.

The resolution at which a stimulus is represented in the brain may therefore not only be due to its context (e.g. memory load and amount of distractors), decay over time, familiarity etc., but first of all it is dependent on constraints of a given sensory system at the level of perception. Thus, the amount of receptors available to originally encode the stimulus before information reaches multiple stages of processing (e.g. filtering of information) in the brain initially determines mnemonic resolution at a later stage of cognition.

Although one of the interpretations of results from the vibrotactile study is that mnemonic resolution is scaled to RF size, its validity is debatable. The dataset collected on the basis of a single frequency range was split into two separate frequency ranges post hoc. While on a single trial, participants were instructed to memorize the frequencies (38Hz, 75Hz, 86Hz and 12Hz) two frequencies would have been encoded by channel 1 and the remaining ones by channel 2. Thus, encoding of frequencies was not studied for each channel in isolation. However, at test they had to reproduce only a single frequency of each set, which could belong to a frequency perceived through either channel 1 or channel 2.

It remains to be determined whether information perceived through either channel may be stored in a common or in separate memory stores. If separate stores are available to represent either type of information, it may offer a potential explanation why it appears that participants can store multiple items (up to 5). Future research might test more directly whether single channels can store fewer items and compare it to the potentially combined capacity of both. As sequentially presented stimuli are encoded in the context of distractors, there may also be specific patterns of interference between them. It would therefore be of interest to test how information associated with 'flutter' interferes with information associated with 'fusion'.

7.1.4 Comparison across sensory modalities

Measuring recall precision for visual information has led to a new conceptualization of WM as a limited resource, which can be flexibly allocated across objects. Here, I have demonstrated ways in which recall precision can also be measured for auditory and tactile WM. In sum, I show that the resource model can be extended to also account for information storage across the senses, where the common currency becomes the mnemonic resolution at which information is held in mind in contrast to regarding WM capacity as limited to a fixed number of items. To put the resource model in the joint context across sensory modalities, I will first discuss commonalities in multi-sensory information storage, before pointing out some key differences.

For all three sensory modalities – vision, audition and touch – results show that as the resource is shared out across an increasing amount of information, there is a decline in the mnemonic resolution strength and fidelity with which items are recalled from WM. There was a monotonous decline in memory precision with increases in memory load, with no discrete step to suggest an item capacity limit, consistent across sensory modalities (see Figure 7-1). Fidelity of memory representation (indexed by precision of recall) decreased when a single item was added to a previous one stored in WM, with additional decline in memory performance when further items were added. Performance remained above chance even at the highest memory load tested (up to 4 or 5 items depending on task and modality), where capacity limits were not reached. The resource model captures these findings, across sensory modalities. Recall precision may reflect neural noise in WM representations, providing a novel framework for the interpretation of neurophysiological data across modalities.

As discernible in Figure 7-1, a similar pattern of decline in precision was found across sensory modalities. Note that it is only the overall pattern which makes the results comparable, not the actual data points in relation to each other because each continuous measurement of e.g. precision for speech sounds vs. precision for pitch was obtained by non-equivalent response methods. Whereas WM for pitch was based on a linear scale (frequency range with lower and upper bound edges), WM for speech sounds was based on a circular stimulus space (no edges), where also the amount of values within a given stimulus space varied. Furthermore stimulus ranges underlie different perceptual thresholds, where continues responses along either stimulus dimension are not directly comparable with each other.

Further differences across results obtained from stimulus matching tasks across the senses concern serial order effects as well as different types of response error. For visual orientations presented sequentially in the study by Gorgoraptis et al. (2011), there was a clear recency effect for all memory loads beyond a single items. Similar order effects were observed for tactile stimuli here (mechanical vibrations, see chapter 4, Figure 4-3). However, for auditory stimuli (sound frequency, chapter 2, Figure 2-3B) there was only a trend towards order effects at the highest memory load of 4 items, while a U-shaped curve, indicating a primacy and recency effect was observed for the same memory load for speech sounds.

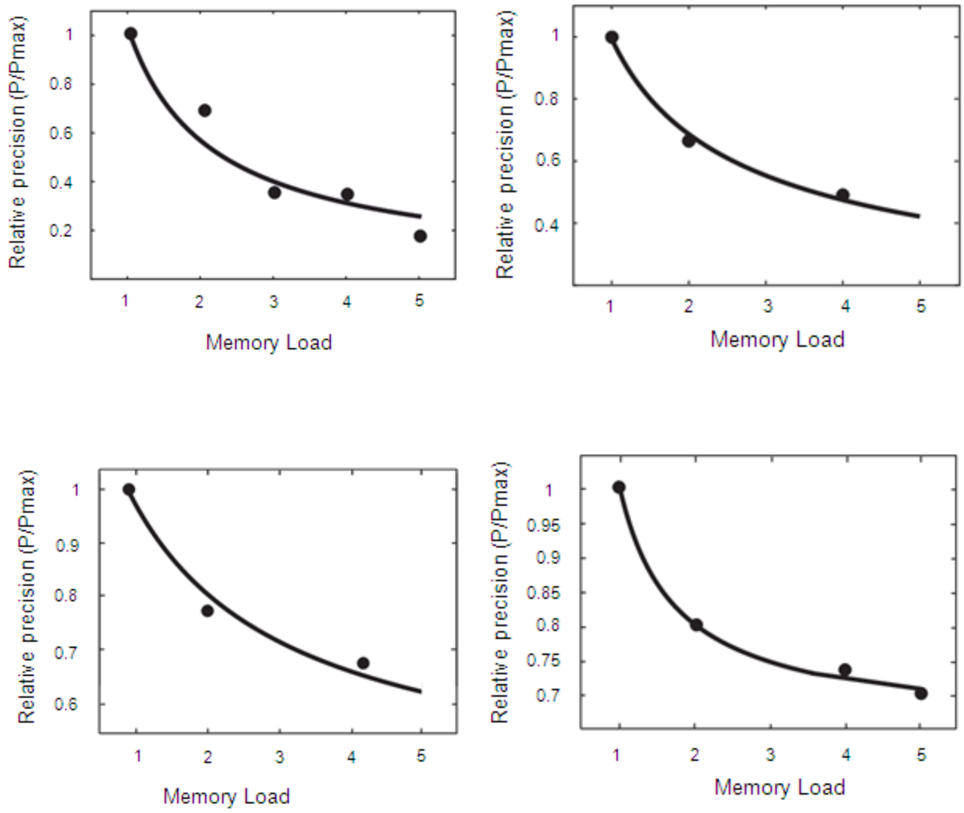


Figure 7-1 ***Relative mean precision by memory load across the senses***

There was a monotonous decline in memory precision with increases in load, with no discrete step to suggest an item capacity limit for (A) visual information (spatial orientation) based on data from Gorgoraptis et al. (2011), (B) auditory information (pitch), see Kumar et al. (2013) or chapter 2, (C) speech sounds), see chapter 3, as well as (D) vibrotactile information (see chapter 4).

Speech sounds are special. They can be perceived as both continuous and *categorical*, which was accounted for by two separate mixture model components (Figure 3-6). Thus when making comparisons across sensory modalities it is important to be aware that responses are biased in different ways according to how stimuli are perceived before being encoded in WM. Another type of response error, misbinding, commonly found in studies of visual WM (Bays & Husain, 2008; Bays et al, 2009; Gorgoraptis et al., 2011; Zokaei et al., 2011, Pertzov et al., 2012) was not observed for auditory WM (pitch and speech sounds), but did occur for vibrotactile WM (Figure 4-4 & Figure 4-5). Thus there might be differences with respect to feature bindings in WM too. In conclusion, recall precision provides an informative measure to assess the nature of WM representations across sensory modalities, which appear to have common underlying principles as well as important differences.

7.2 Representations of sound: objects vs. features

Findings with regards to the resource model presented in chapter 2-4 for different auditory features (pitch and speech sounds) and a tactile feature (vibrational frequency) suggest that memory representations become noisier with an increase in information load. These results have been interpreted on the basis of the theory that WM representations are coded in distributed patterns. However, it remains unknown what type of information such a pattern represents, especially for stimuli composed of multiple features.

In the visual system, most of the evidence suggests that a pattern – as detected by fMRI methods – codes for individual features (Harrison and Tong, 2009; Xing et al, 2013), but it may also code stimuli as objects (Christophel et al., 2012), where features are bound into higher-level object-representations. Such processes remain to be understood in the light of a processing hierarchy of a given sensory system, as well as across sensory systems. The study presented in chapter 4 aimed to address this behaviourally for the auditory system by asking whether sounds in auditory WM are represented as integrated objects or individual features.

Auditory objects composed of two fundamental dimensions of natural sounds (spectral content and temporal amplitude modulation rate) had to be held in WM. Sequences of auditory objects of variable length (1-4 items) were presented, where participants either maintained objects as a whole or their individual features until recall for one of the items was tested by change detection.

Memory recall was more accurate when the object had to be maintained as a whole compared to its individual features alone.

One interpretation of this findings is that, at some level in the auditory system, sounds may be stored as objects, as there is an extraction cost for single features stored in WM. This cost appeared to be higher when extracting temporal compared to spectral content, which may be due to stimulus spacing of features along each dimension. Whether this asymmetry in results for each feature dimension is of functional relevance (e.g. reflecting different WM capacities for each feature) might be addressed by future research. However, since performance was best when objects had to be maintained as a whole, binding features into coherent wholes might potentially serve as a mechanism effectively to increase WM capacity. The same resource might be devoted to more objects.

One limitation of this study is that recall precision was not assessed, as only the percentage of responding correctly was measured using standard tools of change detection. Although a distributed pattern of brain activity may correspond to a representation of an object or one of its features, the design of the study in chapter 5 did not measure the mnemonic resolution at which an object or a feature is represented. The results may therefore give the impression that auditory WM storage is object- instead of resource-based. One possibility is that at different levels of WM, there might actually be both object-based as well as resource-based systems, representing information in the brain. An alternative is that the WM network resource holds objects or features as a flexible medium, which can be represented at different resolution strength depending on how much information has to be held in mind simultaneously.

A shortcoming in the design of the study presented here is that it cannot distinguish which account is correct, an issue that could be addressed by future research. Additionally, it remains to be investigated whether there is a common resource for objects representing multiple features including their bindings, or whether there are separate resources (memory stores) representing features in separation. Both systems – a single common resource and separate stores – might exist, possibly at different levels of processing.

Measuring recall precision based on adjustment error can serve as a means to determine whether WM storage is object- or resource-based. A key prediction of an object-based account is that the error distributions associated with recall of individual features belonging to the same object should be correlated, where features of the same object are stored as bound entities. In contrast, if error distributions for feature recall are not correlated, this would suggest that each feature is stored in separation.

The study by Bays et al. (2011) on object-feature storage in WM tested this for visual information. The authors showed that not only the error distributions, but also the occurrence of guessing (uniform distribution) associated with each feature, were highly *independent* across features. Their results support the idea that different visual features are maintained in independent memory stores (Wheeler & Treisman, 2002). Such stores hold information about each feature in parallel with independent capacities. Additionally, object formation involves binding of features, which become reorganized to create more complex unified representations of previously distributed information (Treisman & Schmidt, 1982). According to feature integration theory in the context of WM, the information of this binding or conjunction is again maintained separately and independently from the features themselves.

Adapting the task I used from change detection to the method of adjustment in order to obtain the measure of recall precision, the results would be more informative, as different response distributions associated with different types of errors could be assessed. However, modifying the task in this way also brings a number of challenges with it. Considering spectral and temporal features employed here, only a limited number of stimulus values are contained within the stimulus range along each dimension (e.g. spectral: 250, 322.92, 417.12, 538.8, 695.98, 899, 1161, 1500 Hz; and temporal: 6, 7.62, 9.67, 12.29, 15.61, 19.83, 25.19, 32 Hz). Even if the stimulus space along each dimension is fully continuous (e.g. filling the gaps between each two neighbouring stimuli in each rage with further stimulus values), there would not be the same number of values along each dimension. However, a matching task would not only require an equal amount of possible stimulus values along each feature dimension, but they should also be matched in terms of difficulty. As response errors are distributed along the stimulus and response space, feature dimensions and their corresponding stimulus values act as a reference frame on which basis interpretations on the correlation or independence of WM stores are made. In a future experiment such design considerations would need to be taken into account to assess response error distributions, including binding errors across features in auditory WM.

In contrast to binding errors, where information about individual items is not lost but is confused with information from another dimension or item, information is also fragile, subject to *interference* (Nairne, 2002; Nairne, Neath & Serra, 1997; Oberauer & Kliegl, 2001; Wickelgren, 1965). A possible form of interference is feature *overwriting*, which is supposed to occur when objects held in WM share some of their features (Nairne, 1990; Oberauer & Kliegl, 2001; 2006). Simultaneously activated memory representations may impair each other limiting WM capacity limits. The cost in feature extraction observed in the study I performed might be a result of interference. However, this interference can occur at various memory stages: encoding, maintenance and recall, where limited representational space is available to hold information at each stage. Additionally, interference can occur within objects as well as across objects (Krumhansl & Iverson, 1992). Thus future research could explore when interference appears, the processing level (within or across objects) it is associated with and its underlying neural correlates.

Spectro-temporal stimuli used in the current study are perceived as objects because both features are presented as overlapping in time (common onsets and offsets) are perceptually grouped together forming coherent wholes. They seem to emanate from a single source (Alain & Arnott, 2000), while such features may also be perceived as belonging to a common auditory event or entity (Bregmann, 1990; Blauert, 1997; Hartmann, 1988). According to the interference model by Oberauer & Kliegl (2006), features are conceptualized as neuronal units that can be bound to only one specific object, where binding in WM occurs via temporal phase synchronization amongst units (Fell & Axmacher, 2011; Feldman, 2013). The mechanism of phase synchronization establishes associations between feature coding neurons of stimulus features and different brain regions.

In the model, neurons coding for features of one object fire in synchrony with each other, whereas neurons coding for features belonging to different objects are not synchronized to each other. As there may be a limited pool of resources (limited amount of neurons coding the same feature), features along a shared dimension are overwritten. This leads to the loss of entire features or results in the formation of low resolution memory representations. Such predictions have been tested for visual WM (Juenger, Kliegel & Oberauer, 2014), but could also be probed for auditory WM in the future.

Additionally, neuroimaging methods with a high temporal resolution, such as MEG or EEG could be employed to directly test the phase synchronization account as a mechanism of binding information in auditory WM. fMRI with MVPA could be used to identify information distributed across different areas in the brain (Haynes & Rees, 2006; Norman et al., 2006), that might discriminate between memory representations of features, their bindings, as well as objects. When decoding patterns of brain activity from different regions, this could foster our understanding on the levels of processing involved in representing such different types of information in memory.

7.3 Decoding the contents of auditory WM

The study presented in chapter 6 addressed the levels of information processing in the auditory system by decoding patterns of brain activity from different regions. The major goal was to resolve the debate on the role of early sensory cortices in cognition: are they primarily involved in the perception of low-level stimulus features or also in higher level cognitive processes such maintenance of the same features in memory? While in vision, evidence from fMRI studies with MVPA suggests that the same neural circuitry (i.e. early visual cortex) involved in perception of basic stimulus features also mediates maintenance of the same features in WM (Harrison & Tong, 2009; Xing et al., 2013; Albers et al., 2013; Pratte

& Tong, 2014). The study in chapter 6 tested whether similar principles apply to auditory memory.

Using event-related fMRI, activity patterns were decoded from time-periods associated with perceptual as well as memory processes to identify if the same brain regions are involved in both cognitive processes. Whether a low or a high tone was perceived as a probe sound, as well as memorized as a target sound, could be decoded successfully from Heschl's gyrus (HG). The classification accuracy associated with perception was much higher than the accuracy associated with memory maintenance. The results therefore confirm the primary involvement of tonotopically organized HG in perception and its recruitment for memory maintenance of tonal information. Thus, the same neural circuits involved in the perception of sound frequency are also involved in memory of the same feature.

In addition to HG, it could also be decoded from other auditory regions, such as PT and STG, which tone was held in mind, where classification accuracy was equal across auditory regions tested. What are the neural codes underlying pattern recognition with regards to such findings? Since low and high tones are represented in areas well separated on the tonotopic plane (Formisano et al., 2003; Humphries, Liebenthal & Binder, 2010; Da Costa et al., 2011), activity patterns found are likely to reflect tonotopic organization. While tonotopic maps correspond to the anatomical shape of HG, a less detailed level of tonotopic organization is preserved beyond HG (Da Costa et al., 2011). Thus, the activity pattern found may resemble a combination of at least different types of neural codes: (1) tonotopy (regions of STG covered by HG and to a certain extent of PT), (2) a more complex neural code (non-tonotopic parts of PT and remaining region of STG). The second code could represent information at the level of precepts (e.g. pitch) or objects, where information is in a format to be integrated with other type of auditory information or in coming from other sensory modalities.

In a control analysis, activity patterns were decoded from early sensory cortices other than auditory cortex during the maintenance period. While the analysis resulted in at chance classification for somatosensory cortex, it resulted in above chance classification for visual cortex during maintenance of pure tones, but not the perception of pure tones, indicating that visual mnemonic strategies (e.g. visual-spatial mapping scheme) might have been employed by participants to aid tone memory. The use of visual strategies to build auditory memories and their underlying neural codes could be addressed by future research.

It would also be interesting to understand better what the neural codes underlying pattern recognition actually resemble in terms of levels of processing (features, percepts or objects). This current paradigm could be adopted to include experimental conditions similar to the ones employed by the study in chapter 5, where one either holds an object as a whole or one of it's feature in mind. MVPA could then be used to predict, whether the object or its feature was held in mind, which could be tested for various regions and stimulus material. Such an experimental manipulation would be needed to identify whether representations coded as distributed patterns represent individual features (Harrison and Tong, 2009; Xing et al, 2013; Pratte & Tong, 2014) or objects (Christophel et al., 2012). Future research might also investigate the degree of continuity between different types of representations (from features to objects, e.g. within auditory cortex).

Further analysis could be applied to data obtained from the current paradigm. Firstly, it would be interesting to classify which tone was held in mind at individual fMRI time points. For example, in the study by Harrison & Tong (2009) on WM for visual orientation, such a time-series analysis of the maintenance period revealed a steep build up of classification accuracy within the first 2 seconds of this period, followed by an asymptotic level, consistent with the slow hemodynamic BOLD response. Interestingly, orientation-selective activity persisted throughout the maintenance period for 11 seconds until recall was tested. Applying this analysis to auditory WM, can one expect similar findings of persistent delay period activity? Memory for sound frequency is subject to rapid decay over time (in humans: Harris, 1952, and monkeys: Scott, Mishkin & Yin, 2012); and interference (Deutsch, 1970; 1972; Zatorre, Evans & Mayer, 1994), where auditory WM representations may be more fragile than visual ones. As the auditory decoding paradigm encompassed a lengthy maintenance period of 16 seconds, activity may fluctuate and only persist for a few seconds during this period.

Apart from not providing details on how classification evolves over time, another shortcoming of the MVPA as it currently stands is that all voxels contained in each ROI were included in classifier analysis. It means that apart from informative voxels, also voxels representing noise or activation unrelated to the events of interest were used as input to the classifier. In order to remove such non-informative voxels, one can determine the classification accuracy for each voxel individually and simply restrict below chance voxels to be entered into further analysis (e.g. as in Harrison & Tong, 2009).

Another promising multivariate technique exemplifies the use of searchlights to construct information-based maps to localize patterns in the brain, which finds the most effective voxels within a given volume (e.g. whole brain or defined ROIs), (Haynes et al., 2007; Kriegeskorte, Goebel & Bandettini, 2006). In this method, a 3D spherical volume is defined around each voxel, where voxels situated within the sphere are jointly analysed with MVPA to inform us about experimental conditions (here: which tone was held in mind). The searchlight analysis specifies an unbiased estimate of information contained in local patterns of activity around every voxel location within the volume (Haynes & Rees, 2006; Pereira et al., 2009).

Using fMRI with unbiased multivariate techniques, future studies could test whether more complex non-tonotopically organized auditory stimuli, such as multi-feature auditory objects or verbal material could also be decoded from early auditory cortex, when perceived or held in mind. For example for visual WM, even non-retinotopically stimuli (colour: Christophel et al., 2012; motion: Riggall & Postle, 2012; complex flow field patterns: Christophel & Haynes, 2014) could still be decoded from early visual cortex. It would also be interesting to decode the mnemonic resolution strength of memory representations across sensory modalities.

Emrich et al. (2013) decoded patterns of neural activity from motion sensitive visual areas (calcarine and extra-calcarine occipital cortex) as a function of memory load. This study revealed a decrease in classification performance with an increase in load (amount of motion stimuli), which is correlated with changes in mnemonic resolution on the single-subject level. Since there appears to be a relationship between memory load, classification accuracy and recall precision, it would be interesting to test, whether it also applies to other sensory modalities or whether precision can be decoded directly. For example the perceptual decoding study by Pratte and Tong (2014) used different contrasts, where one could argue that low contrast stimuli may be encoded at a lower resolution relative one presented at a higher contrast.

Another recent visual study using fMRI with MVPA revealed that neural activity patterns associated with stimulus identity could be decoded from early visual cortex (V1-V3) during perception, working memory (oriented gratings) and mental imagery (rotated gratings) (Albers et al., 2013). Thus, the study could directly compare multiple cognitive processes, where similar neural activity as observed when participants either held visual gratings in mind as they appeared in the task, as well as when they transformed gratings by internally rotating them. Whether such analogous neural codes can be found for auditory WM and auditory imagery remains subject to future research. In addition to decoding across cognitive processes, another interesting topic for future research is to use decoding to distinguish between different levels of processing at which sensory information can be represented, such as features, bindings, percepts or objects.

Alternatively, sensory information can also be represented as continuous or categorical (for the example of speech sounds, see chapter 3). For example the study by Klein & Zatorre (2014) compared classification of categorically perceived musical intervals with continuously perceived pitch height in musicians. In contrast to our findings, only the categorical, but not the continuous stimuli could be decoded from auditory right superior temporal sulcus and left intra-parietal sulcus.

7.4 Conclusion

The experiments included in this thesis were directed toward understanding the nature of sensory representations, as well as the capacity limits underlying information storage across the senses. Using psychophysics, I devised analogue reporting procedures to characterize auditory and tactile memory representations in terms of their fidelity. Results from such experiments (auditory: chapter 2 &

chapter 3, tactile: chapter 4) show how WM resources can be allocated in order to store information perceived through different senses. Next, I identified that there might be an advantage to binding auditory features into coherent wholes, suggesting that at some level in the auditory system complex sounds are stored as objects (chapter 5). Finally, using fMRI with MVPA, auditory memory traces represented as distributed patterns of brain activity could be decoded from early auditory cortex, showing that perception and memory share common neural substrates (chapter 6).

The research presented here shows that memory representations can be measured in different ways, where recall precision appears as particularly informative. The thesis provided converging evidence from different sensory modalities, supporting the idea that WM is best characterized as a shared resource. However, the levels of processing, their neural signatures and in particular what type of information is represented across the senses – features, bindings or objects – clearly needs further work. Although there have been a number of studies decoding memory content from patterns of brain activity using multi-variate techniques, we are just at the beginning of understanding what type of information such patterns actually represent.

Bibliography

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50–68. doi:10.1037/a0014411

Adriani, M., Maeder, P., Meuli, R., Thiran, A. B., Frischknecht, R., Villemure, J.-G., … Clarke, S. (2003). Sound recognition and localization in man: specialized cortical networks and effects of acute circumscribed lesions. *Experimental Brain Research*, *153*(4), 591–604. doi:10.1007/s00221-003-1616-0

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience: A Journal and Virtual Library*, *5*, D202–212.

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology: CB*, *23*(15), 1427–1431. doi:10.1016/j.cub.2013.05.065

Alloway, T. P., & Gathercole, S. E. (2006). *Working Memory and Neurodevelopmental Disorders* (1 edition.). Hove, East Sussex, England ; New York, NY: Psychology Press.

ALLUISI, E. A., MORGAN, B. B., Jr, & HAWKES, G. R. (1965). MASKING OF CUTANEOUS SENSATIONS IN MULTIPLE STIMULUS PRESENTATIONS. *Perceptual and Motor Skills*, *20*, 39–45.

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. *Psychological Science*, *15*(2), 106–111.

Amit, D. J., & Brunel, N. (1995). Learning Internal Representations in an Attractor Neural Network With Analogue Neurons. *NETWORK*, *6*, 359–388.

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(3), 1128–1138. doi:10.1523/JNEUROSCI.4125-10.2011

Anourova, I., Rämä, P., Alho, K., Koivusalo, S., Kahnari, J., & Carlson, S. (1999). Selective interference reveals dissociation between auditory memory for location and pitch. *Neuroreport*, *10*(17), 3543–3547.

ANSI. (1973). *Psychoacoustical Terminology*. New York: American National Standards Institute.

Auvray, M., Gallace, A., Hartcher-O'Brien, J., Tan, H. Z., & Spence, C. (2008). Tactile and visual distractors induce change blindness for tactile stimuli presented on the fingertips. *Brain Research*, *1213*, 111–119. doi:10.1016/j.brainres.2008.03.015

Auvray, M., Gallace, A., & Spence, C. (2011). Tactile short-term memory for stimuli presented on the fingertips and across the rest of the body surface. *Attention, Perception & Psychophysics*, *73*(4), 1227–1241. doi:10.3758/s13414-011-0098-6

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. doi:10.1111/j.1467-9280.2007.01949.x

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208. doi:10.1016/j.neuroscience.2005.08.023

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 3228–3233. doi:10.1073/pnas.0911531107

Axmacher, N., Schmitz, D. P., Wagner, T., Elger, C. E., & Fell, J. (2008). Interactions between medial temporal lobe, prefrontal cortex, and inferior temporal regions during visual working memory: a combined intracranial EEG and functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(29), 7304–7312. doi:10.1523/JNEUROSCI.1778-08.2008

Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Vecchio, F., Cola, B., … Rossini, P. M. (2004). Functional frontoparietal connectivity during short-term memory as revealed by high-resolution EEG coherence analysis. *Behavioral Neuroscience*, *118*(4), 687–697. doi:10.1037/0735-7044.118.4.687

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839. doi:10.1038/nrn1201

Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A., & Liberman, K. (1980). Spatial working memory. In R. Nickerson (Ed.), *Attention and performance* (Vol. 3, pp. 521–539). Hillsdale: Lawrence Erlbaum Associates.

Baddeley, A., & Speelman, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. NORMAN (Ed.), *Models of Human Memory*. New York: Academic Press.

Baddeley, A. D. (1987). *Working Memory*. Clarendon Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D., Kopelman, M., & Wilson, B. A. (2004). *The Essential Handbook of Memory Disorders for Clinicians*. John Wiley & Sons.

Baddeley, A. (2010). Working memory. *Scholarpedia*, *5*(2), 3015. doi:10.4249/scholarpedia.3015

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*(6), 907–918. doi:10.1016/j.neuron.2005.07.023

Bancroft, T. D., Hockley, W. E., & Servos, P. (2011). Vibrotactile working memory as a model paradigm for psychology, neuroscience, and computational modeling. *Frontiers in Human Neuroscience*, *5*, 162. doi:10.3389/fnhum.2011.00162

Bancroft, T. D., Hockley, W. E., & Servos, P. (2012). Can vibrotactile working memory store multiple items? *Neuroscience Letters*, *514*(1), 31–34. doi:10.1016/j.neulet.2012.02.044

Bancroft, T. D., Hockley, W. E., & Servos, P. (2013). Irrelevant sensory stimuli interfere with working memory storage: evidence from a computational model of prefrontal neurons. *Cognitive, Affective & Behavioral Neuroscience*, *13*(1), 23–34. doi:10.3758/s13415-012-0131-9

Bancroft, T. D., Servos, P., & Hockley, W. E. (2011). Mechanisms of interference in vibrotactile working memory. *PloS One*, *6*(7), e22518. doi:10.1371/journal.pone.0022518

Bancroft, T., & Servos, P. (2011). Distractor frequency influences performance in vibrotactile working memory. *Experimental Brain Research*, *208*(4), 529–532. doi:10.1007/s00221-010-2501-2

Baseler, H. A., Morland, A. B., & Wandell, B. A. (1999). Topographic organization of human visual areas in the absence of input from primary cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *19*(7), 2619–2627.

Baumann, S., Petkov, C. I., & Griffiths, T. D. (2013). A unified framework for the organization of the primate auditory cortex. *Frontiers in Systems Neuroscience*, *7*, 11. doi:10.3389/fnsys.2013.00011

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(10), 3632–3645. doi:10.1523/JNEUROSCI.3204-13.2014

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

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science (New York, N.Y.)*, *321*(5890), 851–854. doi: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. doi:10.1016/j.neuropsychologia.2010.12.023

Belger, A., Puce, A., Krystal, J. H., Gore, J. C., Goldman-Rakic, P., & McCarthy, G. (1998). Dissociation of mnemonic and perceptual processes during spatial and nonspatial working memory using fMRI. *Human Brain Mapping*, *6*(1), 14–32.

Bird, C. M., Berens, S. C., Horner, A. J., & Franklin, A. (2014). Categorical encoding of color in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4590–4595. doi:10.1073/pnas.1315275111

Bisley, J. W., Zaksas, D., Droll, J. A., & Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *Journal of Neurophysiology*, *91*(1), 286–300. doi:10.1152/jn.00870.2003

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology. Human Perception and Performance*, *23*(2), 353–369.

Blauert, J. (1997a). *Räumliches Hören 2. Nachschrift 2. Nachschrift*. Stuttgart: Hirzel.

Blauert, J. (1997b). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press.

Bohn, O., & Fledge, J. E. (1992). The Production of New and Similar Vowels by Adult German Learners of English. *Studies in Second Language Acquisition*, *14*(2), 131–158.

Bonda, E., Petrides, M., & Evans, A. (1996). Neural systems for tactual memories. *Journal of Neurophysiology*, *75*(4), 1730–1737.

Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychological Research*, *46*(3), 207–222.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological Review*, *113*(2), 201–233. doi:10.1037/0033-295X.113.2.201

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4. doi:10.1167/11.5.4

Brain, W. R. B., Donaghy, Michael. (2009). *Brain's diseases of the nervous system.* Oxford; New York: Oxford University Press.

Brandt, T., Caplan, L. R., Dichgans, J., Diener, H. C., & Kennard, C. (Eds.). (2003). Contributors. In *Neurological Disorders (Second Edition)* (pp. xi–xvi). San Diego: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780121258313501933

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, Mass.: MIT Press.

Brener, R. (1940). An experimental investigation of memory span. . *Journal of Experimental Psychology*, *26*, 467–482.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*, 349–368.

Bruce, V., Georgeson, M. A., & Green, P. R. (2003). *Visual Perception: Physiology, Psychology and Ecology* (4 edition.). Hove ; New York: Psychology Press.

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *The Journal of the Acoustical Society of America*, *112*(6), 2985–2995.

Burak, Y., & Fiete, I. R. (2012). Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17645–17650. doi:10.1073/pnas.1117386109

Burnett Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science*, *15*(4), 528–539. doi:10.1111/j.1467-7687.2012.01148.x

Burns, E. M., & Ward, W. D. (1978). Categorical perception--phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, *63*(2), 456–468.

Burton, H., & Sinclair, R. J. (2000). Attending to and remembering tactile stimuli: a review of brain imaging data and single-neuron responses. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, *17*(6), 575–591.

Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11252–11255. doi:10.1073/pnas.1104666108

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, *18*(12), 1959–1972. doi:10.1162/jocn.2006.18.12.1959

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, *118*(1), 471–482.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*(6768), 428–430. doi:10.1038/35000206

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: a reconciliation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(6), 1235–1249. doi:10.1037/0278-7393.31.6.1235

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: the limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology (2006)*, *62*(7), 1420–1429. doi:10.1080/17470210802453977

Christophel, T. B., & Haynes, J.-D. (2014). Decoding complex flow-field patterns in visual working memory. *NeuroImage*, *91*, 43–51. doi:10.1016/j.neuroimage.2014.01.025

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *The Journal of*

Neuroscience: The Official Journal of the Society for Neuroscience, *32*(38), 12983–12989. doi:10.1523/JNEUROSCI.0184-12.2012

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. doi:10.1016/j.conb.2007.03.005

Clarke, S., Adriani, M., & Bellmann, A. (1998). Distinct short-term memory systems for sound content and sound localization. *Neuroreport*, *9*(15), 3433–3437.

Clarke, S., Bellmann Thiran, A., Maeder, P., Adriani, M., Vernet, O., Regli, L., … Thiran, J.-P. (2002). What and where in human audition: selective deficits following focal hemispheric lesions. *Experimental Brain Research*, *147*(1), 8–15. doi:10.1007/s00221-002-1203-9

Clarke, S., & Thiran, A. B. (2004). Auditory neglect: what and where in auditory space. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *40*(2), 291–300.

Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *The Journal of the Acoustical Society of America*, *106*(5), 2805–2811.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086–1095.

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex (New York, N.Y.: 1991)*, *10*(9), 910–923.

CONRAD, R., & HULL, A. J. (1964). INFORMATION, ACOUSTIC CONFUSION AND MEMORY SPAN. *British Journal of Psychology (London, Engla*

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Brain Research. Cognitive Brain Research*, *7*(1), 1–13.

Deco, G., Rolls, E. T., & Romo, R. (2009). Stochastic dynamics as a principle of brain function. *Progress in Neurobiology*, *88*(1), 1–16. doi:10.1016/j.pneurobio.2009.01.006

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*(10), 1189–1199.

Deutsch, D. (1970a). Dislocation of tones in a musical sequence: a memory illusion. *Nature*, *226*(5242), 286.

Deutsch, D. (1970b). Tones and numbers: specificity of interference in immediate memory. *Science (New York, N.Y.)*, *168*(3939), 1604–1605.

Deutsch, D. (1972a). Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology*, *93*(1), 156–162.

Deutsch, D. (1972b). Mapping of interactions in the pitch memory store. *Science (New York, N.Y.)*, *175*(4025), 1020–1022.

Deutsch, D. (1973). Interference in memory between tones adjacent in the musical scale. *Journal of Experimental Psychology*, *100*(2), 228–231.

Deutsch, D. (1974). Generality of interference by tonal stimuli in recognition memory for pitch. *The Quarterly Journal of Experimental Psychology*, *26*(2), 229–234. doi:10.1080/14640747408400408

Drewnowski, A., & Murdock, B. B., Jr. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(3), 319–332.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. doi:10.1016/j.neuroimage.2007.09.034

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. doi:10.1016/j.neuroimage.2004.12.034

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science (New York, N.Y.)*, *171*(3968), 303–306.

Ejstrup, M., & Hansen, G. F. (2004). Vowels in regional variants of Danish. In P. Branderud & H. Traunmueller (Eds.), *Proceedings of FONETIK* (pp. 88–91). Department of Linguistics, Stockholm University.

Ellis, K. A., Silberstein, R. B., & Nathan, P. J. (2006). Exploring the temporal dynamics of the spatial working memory n-back task using steady state visual evoked potentials (SSVEP). *NeuroImage*, *31*(4), 1741–1751. doi:10.1016/j.neuroimage.2006.02.014

Elmore, L. C., Ma, W. J., Magnotti, J. F., Leising, K. J., Passaro, A. D., Katz, J. S., & Wright, A. A. (2011). Visual short-term memory compared in rhesus monkeys and humans. *Current Biology: CB*, *21*(11), 975–979. doi:10.1016/j.cub.2011.04.031

Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in

Visual Short-Term Memory. *Journal of Neuroscience*, *33*(15), 6516–6523. doi:10.1523/JNEUROSCI.5732-12.2013

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, *369*(6481), 525. doi:10.1038/369525a0

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269. doi:10.1016/j.neuroimage.2013.03.041

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: an investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America*, *115*(1), 352–361.

Feldman, J. (2013). The neural binding problem(s). *Cognitive Neurodynamics*, *7*(1), 1–11. doi:10.1007/s11571-012-9219-8

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews. Neuroscience*, *12*(2), 105–118. doi:10.1038/nrn2979

Ferrera, V. P., Rudolph, K. K., & Maunsell, J. H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(10), 6171–6186.

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge [England]; New York: Cambridge University Press.

Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*(4), 859–869.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12). doi:10.1167/11.12.3

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12), 27. doi:10.1167/10.12.27

Fougnie, D., & Marois, R. (2011). What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(6), 1329–1341. doi:10.1037/a0024834

Frey, S., & Petrides, M. (2000). Orbitofrontal cortex: A key prefrontal region for encoding information. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(15), 8723–8727. doi:10.1073/pnas.140543497

Frick, R. W. (1988). Issues of representation and limited capacity in the auditory short-term store. *British Journal of Psychology (London, England: 1953)*, *79 (Pt 2)*, 213–240.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189.

Fujisaki, H., & Kawashima, T. (1971). A model of the mechanisms for speech perception: Quantitative analyses of categorical effects in discrimination. *Annual Report of the Engineering Research Institute, Faculaty of Engineering, University of Tokyo*, 59–68.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–349.

Fuster, J. M. (1990a). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology*, *64*(3), 681–697.

Fuster, J. M. (1990b). Neuronal discrimination and short-term memory in association cortex. In A. B. Scheibel & A. Wechsler (Eds.), *The Neurobiological Foundations of Cognitive Functions* (pp. 85–98). New York: Guilford.

Fuster, J. M. (1999). Prefrontal Cortex. In G. Adelman & B. H. Smith (Eds.), *Encyclopedia of Neuroscience* (pp. 1690–1692).

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science (New York, N.Y.)*, *173*(3997), 652–654.

Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Research*, *330*(2), 299–307.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*(6784), 347–351. doi:10.1038/35012613

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. *NeuroImage*, *19*(4), 1417–1426.

nd: 1953), *55*, 429–432.

Conrad, R. (1964). Acoustic Confusions in Immediate Memory. *British Journal of Psychology*, *55*(1), 75–84. doi:10.1111/j.2044-8295.1964.tb00899.x

Constantinidis, C., Franowicz, M. N., & Goldman-Rakic, P. S. (2001). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature Neuroscience*, *4*(3), 311–316. doi:10.1038/85179

Constantinidis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology*, *76*(2), 1352–1355.

Cornoldi, C., Rigoni, F., Venneri, A., & Vecchi, T. (2000). Passive and active processes in visuo-spatial memory: double dissociation in developmental learning disabilities. *Brain and Cognition*, *43*(1-3), 117–120.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *6*(1), 39–49.

Cowan, N. (1995). Sensory memory and its role in information processing. *Electroencephalography and Clinical Neurophysiology. Supplement*, *44*, 21–31.

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

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. doi:10.1016/S0079-6123(07)00020-9

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(3), 731–747. doi:10.1037/a0029687

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: a logical sequel to Miller (1956). *Psychological Science*, *15*(9), 634–640. doi:10.1111/j.0956-7976.2004.00732.x

Crowder, R. G. (1981). General forgetting theory and the locus of amnesia. In L. S. Cermak (Ed.), *Human memory and amnesia*. Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 472–478. doi:10.1037/0096-1523.15.3.472

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl’s gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(40), 14067–14075. doi:10.1523/JNEUROSCI.2000-11.2011

Darwin, C. J., & Carlyon, R. P. (1995). Auditory Grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). San Diego: Academic.

Gale, R. M. (2005). *The Philosophy of William James : an introduction*. Cambridge University Press.

Gallace, A., & Spence, C. (2008). The cognitive and neural correlates of ‘tactile consciousness’: a multisensory perspective. *Consciousness and Cognition*, *17*(1), 370–407. doi:10.1016/j.concog.2007.01.005

Gallace, A., & Spence, C. (2009). The cognitive and neural correlates of tactile memory. *Psychological Bulletin*, *135*(3), 380–406. doi:10.1037/a0015325

Gallace, A., Tan, H. Z., Haggard, P., & Spence, C. (2008). Short term memory for tactile stimuli. *Brain Research*, *1190*, 132–142. doi:10.1016/j.brainres.2007.11.014

Gallace, A., Tan, H. Z., & Spence, C. (2006). Numerosity judgments for tactile stimuli distributed over the body surface. *Perception*, *35*(2), 247–266.

Gallace, A., Tan, H. Z., & Spence, C. (2007). Do ‘mudsplashes’ induce tactile change blindness? *Perception & Psychophysics*, *69*(4), 477–486.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working Memory and Language*. Psychology Press.

GELDARD, F. A., & SHERRICK, C. E. (1965). MULTIPLE CUTANEOUS STIMULATION: THE DISCRIMINATION OF VIBRATORY PATTERNS. *The Journal of the Acoustical Society of America*, *37*, 797–801.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376.

Gescheider, G., & Verillo, R. T. (1979). Vibrotactile frequency characteristics as determined by adaptation and masking procedures. In D. Kenshalo (Ed.), *Sensory functions of the skin of humans*. New York: Plenum Press.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex (New York, N.Y.: 1991)*, *7*(4), 374–385.

Giard, M. H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, *7*(2), 133–143. doi:10.1162/jocn.1995.7.2.133

Gilson, E. Q., & Baddeley, A. D. (1969). Tactile short-term memory. *The Quarterly Journal of Experimental Psychology*, *21*(2), 180–184. doi:10.1080/14640746908400211

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *12*(1), 3–15.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *351*(1346), 1445–1453. doi:10.1098/rstb.1996.0129

Golubock, J. L., & Janata, P. (2013). Keeping timbre in mind: working memory for complex sounds that can't be verbalized. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(2), 399–412. doi:10.1037/a0029720

Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, *67*(1-2), 181–207.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Gorden, G. (Ed.). (1978). *Active Touch: the mechanism of recognition of objects by manipulation: a multi-disciplinary approach*. Oxford: Pergamon Press.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(23), 8502–8511. doi:10.1523/JNEUROSCI.0208-11.2011

Gosselin, N., Jolicoeur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 270–272. doi:10.1111/j.1749-6632.2009.04762.x

Gottlieb, Y., Vaadia, E., & Abeles, M. (1989). Single unit activity in the auditory cortex of a monkey performing a short term memory task. *Experimental Brain Research*, *74*(1), 139–148.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–328.

Greenlee, M. W., Magnussen, S., & Reinvang, I. (2000). Brain regions involved in spatial frequency discrimination: evidence from fMRI. *Experimental Brain Research*, *132*(3), 399–403.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270–1277.

Grey,, J.M., & Grey. (1977). Multidimensional scaling of musical timbres. *J. Acoust. Soc. Am.*, *61*, 1270–1277.

Griffiths, T. D., Rees, G., Rees, A., Green, G. G., Witton, C., Rowe, D., … Frackowiak, R. S. (1998). Right parietal cortex is involved in the perception of sound movement in humans. *Nature Neuroscience*, *1*(1), 74–79. doi:10.1038/276

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews. Neuroscience*, *5*(11), 887–892. doi:10.1038/nrn1538

Grimault, S., Lefebvre, C., Vachon, F., Peretz, I., Zatorre, R., Robitaille, N., & Jolicoeur, P. (2009). Load-dependent brain activity related to acoustic short-term memory for pitch: magnetoencephalography and fMRI. *Annals of the New York Academy of Sciences*, *1169*, 273–277. doi:10.1111/j.1749-6632.2009.04844.x

Grimault, S., Lefebvre, C., Vachon, F., Peretz, I., Zatorre, R., Robitaille, N., & Jolicoeur, P. (2010). Load-dependent brain activity related to acoustic short-term memory for pitch. *The Neurosciences and Music III: Disorders and Plasticity*, *1169*, 273–277.

Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, E., Ludden, D., … Newman, R. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 599–603. doi:10.3758/BF03206540

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: the roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*(6), 1001–1016.

Haberlandt, K., Thomas, J. G., Lawrence, H., & Krohn, T. (2005). Transposition asymmetry in immediate serial recall. *Memory (Hove, England)*, *13*(3-4), 274–282.

Haegens, S., Osipova, D., Oostenveld, R., & Jensen, O. (2010). Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. *Human Brain Mapping*, *31*(1), 26–35. doi:10.1002/hbm.20842

Hall, M. D., & Wieberg, K. (2003). Illusory conjunctions of musical pitch and timbre. *Acoustics Research Letters Online*, *4*(3), 65–70. doi:10.1121/1.1578951

Harnad, S. (1987). *Categorical perception: the groundwork of cognition*. New York: Cambridge University Press.

Harnad, S. R. (1990). *Categorical Perception: The Groundwork of Cognition*. Cambridge; New York: Cambridge University Press.

Harris, J. A., Miniussi, C., Harris, I. M., & Diamond, M. E. (2002). Transient storage of a tactile memory trace in primary somatosensory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(19), 8720–8725.

Harris, J. D. (1952). The decline of pitch discrimination with time. *Journal of Experimental Psychology*, *43*(2), 96–99. doi:10.1037/h0057373

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi:10.1038/nature07832

Hartmann, W. M. (1988). Pitch perception and the segregation and integration of auditory entities. In G. W. Edelman & W. M. Cowan (Eds.), *Auditory Function – Neurobiological Bases of Hearing* (pp. 623–645). John Wiley and Sons.

Haynes, J.-D., & Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Current Biology: CB*, *15*(14), 1301–1307. doi:10.1016/j.cub.2005.06.026

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, *7*(7), 523–534. doi:10.1038/nrn1931

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology: CB*, *17*(4), 323–328. doi:10.1016/j.cub.2006.11.072

Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *The Journal of the Acoustical Society of America*, *125*(1), 447–456. doi:10.1121/1.3035837

Hernández, A., Salinas, E., García, R., & Romo, R. (1997). Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(16), 6391–6400.

Hernández, A., Zainos, A., & Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(11), 6191–6196. doi:10.1073/pnas.120018597

Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, *137*(3-4), 346–353.

Herrmann, C. S., Munk, M. H. J., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: memory match and utilization. *Trends in Cognitive Sciences*, *8*(8), 347–355. doi:10.1016/j.tics.2004.06.006

Heyselaar, E., Johnston, K., & Paré, M. (2011). A change detection approach to study visual working memory of the macaque monkey. *Journal of Vision*, *11*(3). doi:10.1167/11.3.11

Hickok, G., & Buchsbaum, B. (2003). Temporal lobe speech perception systems are part of the verbal working memory circuit: Evidence from two recent fMRI studies. *Behavioral and Brain Sciences*, *26*(06), 740–741. doi:10.1017/S0140525X03340166

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception & Psychophysics*, *72*(5), 1218–1227. doi:10.3758/APP.72.5.1218

Holz, E. M., Glennon, M., Prendergast, K., & Sauseng, P. (2010). Theta-gamma phase synchronization during memory matching in visual working memory. *NeuroImage*, *52*(1), 326–335. doi:10.1016/j.neuroimage.2010.04.003

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., … Kahana, M. J. (2003). Gamma oscillations correlate with working memory load in humans. *Cerebral Cortex (New York, N.Y.: 1991)*, *13*(12), 1369–1374.

Hsieh, T.-H., Chen, J.-J. J., Chen, L.-H., Chiang, P.-T., & Lee, H.-Y. (2011). Time-course gait analysis of hemiparkinsonian rats following 6-hydroxydopamine lesion. *Behavioural Brain Research*, *222*(1), 1–9. doi:10.1016/j.bbr.2011.03.031

Hubel, D. H. (1963). The visual cortex of the brain. *Scientific American, 209*(5), 54-62.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685–701. doi:10.1016/0749-596X(91)90032-F

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211. doi:10.1016/j.neuroimage.2010.01.046

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: perception of formant targets, formant movement, and duration.

The Journal of the Acoustical Society of America, *122*(5), 2842–2854. doi:10.1121/1.2783198

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, *97*(1), 553–562.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? *Perception & Psychophysics*, *62*(4), 874–886.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87 (1) B47 - B57.*

Iverson, P., Smith, C. A., & Evans, B. G. (2006). Vowel recognition via cochlear implants and noise vocoders: effects of formant movement and duration. *The Journal of the Acoustical Society of America*, *120*(6), 3998–4006.

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A. & Evans, B.G., (2008). Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *J EXP PSYCHOL HUMAN, 34(5), 1305 - 1316.*

Jensen, O., & Lisman, J. E. (1998). An oscillatory short-term memory buffer model can account for data on the Sternberg task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(24), 10688–10699.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *The European Journal of Neuroscience*, *15*(8), 1395–1399.

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, *50*(4), 544–552. doi:10.1016/j.neuropsychologia.2011.11.007

Jokisch, D., & Jensen, O. (2007). Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(12), 3244–3251. doi:10.1523/JNEUROSCI.5399-06.2007

Jump, R. L., & Ries, D. T. (2008). Effect of interaural level and phase cues on intervening interference in auditory working memory for loudness. *Hearing Research*, *240*(1-2), 112–115. doi:10.1016/j.heares.2008.04.004

Jünger, E., Kliegl, R., & Oberauer, K. (2014). No evidence for feature overwriting in visual working memory. *Memory (Hove, England)*, *22*(4), 374–389. doi:10.1080/09658211.2013.794243

Kaas, A. L., van Mier, H., Visser, M., & Goebel, R. (2013). The neural substrate for working memory of tactile surface texture. *Human Brain Mapping*, *34*(5), 1148–1162. doi:10.1002/hbm.21500

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Research Bulletin*, *44*(2), 107–112.

Kaas, J. H. (2000). Organizing principles of sensory representations. *Novartis Foundation Symposium*, *228*, 188–198; discussion 198–205.

Kaiser, J., Heidegger, T., & Lutzenberger, W. (2008). Behavioral relevance of gamma-band activity for short-term memory-based auditory decision-making. *The European Journal of Neuroscience*, *27*(12), 3322–3328. doi:10.1111/j.1460-9568.2008.06290.x

Kaiser, J., Leiberg, S., Rust, H., & Lutzenberger, W. (2007). Prefrontal gamma-band activity distinguishes between sound durations. *Brain Research*, *1139*, 153–162. doi:10.1016/j.brainres.2006.12.085

Kaiser, J., Ripper, B., Birbaumer, N., & Lutzenberger, W. (2003). Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *NeuroImage*, *20*(2), 816–827. doi:10.1016/S1053-8119(03)00350-1

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. doi:10.1038/nn1444

Kay, P., & Kempton, W. M. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, (86), 65–79.

Kandel E.R., Schwartz, J.H., Jessell, T.M. (2000). *Principles of Neural Science*, 4th ed., pp. 515–520. McGraw-Hill, New York.

Keating, P. A., & Huffman, M. K. (1984). Vowel Variation in Japanese. *Phonetica*, *41*(4), 191–207. doi:10.1159/000261726

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology*, *9*(2), e1002927. doi:10.1371/journal.pcbi.1002927

Killackey, H. P., Rhoades, R. W., & Bennett-Clarke, C. A. (1995). The formation of a cortical somatotopic map. *Trends in Neurosciences*, *18*(9), 402–407.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87*(2), 820–857.

Klein, M. E., & Zatorre, R. J. (2014). Representations of Invariant Musical Categories Are Decodable by Pattern Analysis of Locally Distributed BOLD Responses in Superior Temporal and Intraparietal Sulci. *Cerebral Cortex (New York, N.Y.: 1991)*. doi:10.1093/cercor/bhu003

Klingberg, T. (2009). *The Overflowing Brain: Information Overload and the Limits of Working Memory* (1 edition.). Oxford ; New York: OUP USA.

Klingberg, T., Kawashima, R., & Roland, P. E. (1996). Activation of multi-modal cortical areas underlies short-term memory. *The European Journal of Neuroscience*, *8*(9), 1965–1971.

Knight, F. L. C., Longo, M. R., & Bremner, A. J. (2014). Categorical perception of tactile distance. *Cognition*, *131*(2), 254–262. doi:10.1016/j.cognition.2014.01.005

Koch, K. W., & Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Experimental Brain Research*, *76*(2), 292–306.

Kopp, F., Schröger, E., & Lipka, S. (2006). Synchronized brain activity during rehearsal and short-term memory disruption by irrelevant speech is affected by recall mode. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *61*(2), 188–203. doi:10.1016/j.ijpsycho.2005.10.001

Kosslyn, S. M., & Ochsner, K. N. (1994). In search of occipital activation during visual mental imagery. *Trends in Neurosciences*, *17*(7), 290–292.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. doi:10.1073/pnas.0600244103

Krumhansl, C. L. (1989). *Why is musical timbre so hard to understand? In: Structure and perception of electroacoustic sound and music*. New York: Excerpta Medica.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(3), 739–751.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science (New York, N.Y.)*, *255*(5044), 606–608.

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. *Cognitive Neuroscience*, *4*(1), 12–20. doi:10.1080/17588928.2012.716416

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, *62*(7), 1426–1439.

Lalonde, J., & Chaudhuri, A. (2002). Task-dependent transfer of perceptual to memory representations during delayed spatial frequency discrimination. *Vision Research*, *42*(14), 1759–1769.

LaMotte, R. H., & Mountcastle, V. B. (1975a). Capacities of humans and monkeys to discriminate between vibratory stimuli of different frequency and amplitude: A correlation between neural events and psychophysical measurements. *Journal of Neurophysiology*, (38), 539–599.

LaMotte, R. H., & Mountcastle, V. B. (1975b). Capacities of humans and monkeys to discriminate vibratory stimuli of different frequency and amplitude: a correlation between neural events and psychological measurements. *Journal of Neurophysiology*, *38*(3), 539–559.

Langers, D. R. M., & van Dijk, P. (2012). Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(9), 2024–2038. doi:10.1093/cercor/bhr282

Lara, A. H., & Wallis, J. D. (2012). Capacity and precision in an animal model of visual short-term memory. *Journal of Vision*, *12*(3). doi:10.1167/12.3.13

Lehnert, G., & Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Memory & Cognition*, *34*(5), 1080–1090.

Lehnert, G., & Zimmer, H. D. (2008a). Common coding of auditory and visual spatial information in working memory. *Brain Research*, *1230*, 158–167. doi:10.1016/j.brainres.2008.07.005

Lehnert, G., & Zimmer, H. D. (2008b). Modality and domain specific components in auditory and visual working memory tasks. *Cognitive Processing*, *9*(1), 53–61. doi:10.1007/s10339-007-0187-6

Leiberg, S., Lutzenberger, W., & Kaiser, J. (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Research*, *1120*(1), 131–140. doi:10.1016/j.brainres.2006.08.066

Levy, B. (1971). Role of Articulation in Auditory and Visual Short-Term Memory. *Journal of Verbal Learning and Verbal Behaviour*, *10*, 123–132.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79. doi:10.1162/jocn_a_00140

Li, D., Cowan, N., & Saults, J. S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception & Psychophysics*, *75*(1), 145–160. doi:10.3758/s13414-012-0383-z

Li Hegner, Y., Lee, Y., Grodd, W., & Braun, C. (2010). Comparing tactile pattern and vibrotactile frequency discrimination: a human FMRI study. *Journal of Neurophysiology*, *103*(6), 3115–3122. doi:10.1152/jn.00940.2009

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Linden, D. E. J. (2007). The working memory networks of the human brain. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *13*(3), 257–267. doi:10.1177/1073858406298480

Linden, D. E. J., Oosterhof, N. N., Klein, C., & Downing, P. E. (2012). Mapping brain activation and information during category-specific visual working memory. *Journal of Neurophysiology*, *107*(2), 628–639. doi:10.1152/jn.00105.2011

Linke, A. C., Vicente-Grabovetsky, A., & Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(31), 12961–12966. doi:10.1073/pnas.1102118108

Lisman, J. E., & Jensen, O. (2013). The θ-γ neural code. *Neuron*, *77*(6), 1002–1016. doi:10.1016/j.neuron.2013.03.007

Lu, Z. L., Williamson, S. J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science (New York, N.Y.)*, *258*(5088), 1668–1670.

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

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639. doi:10.1016/j.neuropsychologia.2010.11.031

Lutzenberger, W., Ripper, B., Busse, L., Birbaumer, N., & Kaiser, J. (2002). Dynamics of gamma-band activity during an audiospatial working memory task in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(13), 5630–5638. doi:20026570

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. doi:10.1038/nn.3655

Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science (New York, N.Y.)*, *307*(5712), 1121–1124. doi:10.1126/science.1104171

MacLeod, C. M. (2007). Cognitive inhibition: Elusive or illusion? In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 301–305). New York: Oxford University Press.

Macmillan, N. A., Goldberg, R. F., & Braida, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *The Journal of the Acoustical Society of America*, *84*(4), 1262–1280. doi:10.1121/1.396626

Macmillan, N.A. & Creelman, C.D. (1991). Detection Theory: A User's Guide. Cambridge University Press.

Maddox, R. K., & Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *Journal of the Association for Research in Otolaryngology: JARO*, *13*(1), 119–129. doi:10.1007/s10162-011-0299-7

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., … Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *NeuroImage*, *14*(4), 802–816. doi:10.1006/nimg.2001.0888

Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, *23*(6), 247–251.

Magnussen, S., & Greenlee, M. W. (1992). Retention and disruption of motion information in visual short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(1), 151–156.

Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, *31*(7-8), 1213–1219.

Makar, A. B., McMartin, K. E., Palese, M., & Tephly, T. R. (1975). Formate assay in body fluids: application in methanol poisoning. *Biochemical Medicine*, *13*(2), 117–126.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, *114*(5), 2946–2957.

Massaro, D. W. (1970). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, *83*(1), 32–39.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*(2), 124–145.

Mathiak, K., Hertrich, I., Grodd, W., & Ackermann, H. (2004). Discrimination of temporal information at the cerebellum: functional magnetic resonance imaging of nonverbal auditory memory. *NeuroImage*, *21*(1), 154–162.

Mathias, S. R., & von Kriegstein, K. (2014). Percepts, not acoustic properties, are the units of auditory short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 445–450. doi:10.1037/a0034890

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision*, *12*(6), 10. doi:10.1167/12.6.10

McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *336*(1278), 383–389. doi:10.1098/rstb.1992.0072

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192. doi:10.1007/BF00419633

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. doi:10.1038/nn2024

Medendorp, W. P., Kramer, G. F. I., Jensen, O., Oostenveld, R., Schoffelen, J.-M., & Fries, P. (2007). Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(10), 2364–2374. doi:10.1093/cercor/bhl145

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: timbre, pitch, and loudness. *Perception & Psychophysics*, *48*(2), 169–178.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews. Neuroscience*, *1*(1), 59–65. doi:10.1038/35036228

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *13*(4), 1460–1478.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:10.1037/h0043158

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural Mechanisms of Visual Working Memory in Prefrontal Cortex of the Macaque. *Journal of Neuroscience*, *16*, 5154–5167.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*(1), 57–77.

Miyake, A., & Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press.

Mohr, H. M., & Linden, D. E. J. (2005). Separation of the systems for color and spatial manipulation in working memory revealed by a dual-task procedure. *Journal of Cognitive Neuroscience*, *17*(2), 355–366. doi:10.1162/0898929053124929

Molter, C., Salihoglu, U., & Bersini, H. (2007). The road to chaos by time-asymmetric Hebbian learning in recurrent neural networks. *Neural Computation*, *19*(1), 80–110. doi:10.1162/neco.2007.19.1.80

Moran, R. J., Campo, P., Maestu, F., Reilly, R. B., Dolan, R. J., & Strange, B. A. (2010). Peak frequency in the theta and alpha bands correlates with human working memory capacity. *Frontiers in Human Neuroscience*, *4*, 200. doi:10.3389/fnhum.2010.00200

Morgan, B., & Chambers, S. M. (1973). Acoustic confusion of digits in memory and recognition, *14*(2), 375–383. doi:10.3758/BF03212408

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, *11*(2), 296–301.

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(4), 703–713. doi:10.1037/0278-7393.31.4.703

Morgan, B. J. T., Chambers, S. M., & Morton, J. (1973). Acoustic confusion of digits in memory and recognition. *Perception and Psychophysics*, *14*, 375–383.

Mottaghy, F. M., Döring, T., Müller-Gärtner, H.-W., Töpper, R., & Krause, B. J. (2002). Bilateral parieto-frontal network for verbal working memory: an interference approach using repetitive transcranial magnetic stimulation (rTMS). *The European Journal of Neuroscience*, *16*(8), 1627–1632.

MOUNTCASTLE, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, *20*(4), 408–434.

MOUNTCASTLE, V. B. (1984). Central nervous system mechanisms in mechanoreceptive sensibility. In S. R. Geiger, I. Darian-Smith, & J. M. Brookhart (Eds.), *Handbook of Physiology* (pp. 789–878). Baltimore: Waverly.

Mountcastle, V. B., Talbot, W. H., Darian-Smith, I., & Kornhuber, H. H. (1967). Neural basis of the sense of flutter-vibration. *Science (New York, N.Y.)*, *155*(3762), 597–600.

Mukari, S. Z., Umat, C., & Othman, N. I. (2010). Effects of age and working memory capacity on pitch pattern sequence test and dichotic listening. *Audiology & Neuro-Otology*, *15*(5), 303–310. doi:10.1159/000283007

Myung, I. J. (2003). Tutorial on Maximum Likelihood Estimation. *J. Math. Psychol.*, *47*(1), 90–100. doi:10.1016/S0022-2496(02)00028-7

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269.

Nairne, J. S., Neath, I., & Serra, M. (1997). Proactive interference plays a role in the word-length effect. *Psychonomic Bulletin & Review*, *4*(4), 541–545. doi:10.3758/BF03214346

Ng, C.-W., Plakke, B., & Poremba, A. (2009). Primate auditory recognition memory performance varies with sound type. *Hearing Research*, *256*(1-2), 64–74. doi:10.1016/j.heares.2009.06.014

Niki, H. (1974). Differential activity of prefrontal units during right and left delayed response trials. *Brain Research*, *70*(2), 346–349.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. doi:10.1016/j.tics.2006.07.005

Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, 38(5), 415-432.

O'Sullivan, B. T., Roland, P. E., & Kawashima, R. (1994). A PET study of somatosensory discrimination in man. microgeometry versus macrogeometry. *The European Journal of Neuroscience*, *6*(1), 137–148.

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(3), 411–421.

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

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology*, *13*(1-2), 187–215. doi:10.1080/09541440042000278

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. doi:10.1016/j.jml.2006.08.009

Oliveri, M., Turriziani, P., Carlesimo, G. A., Koch, G., Tomaiuolo, F., Panella, M., & Caltagirone, C. (2001). Parieto-frontal interactions in visual-object and visual-spatial working memory: evidence from transcranial magnetic stimulation. *Cerebral Cortex (New York, N.Y.: 1991)*, *11*(7), 606–618.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. doi:10.1016/j.tics.2011.05.004

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., … Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biology*, *5*(11), e288. doi:10.1371/journal.pbio.0050288

Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7580–7585. doi:10.1073/pnas.0913113107

Palva, S., Kulashekhar, S., Hämäläinen, M., & Palva, J. M. (2011). Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(13), 5013–5025. doi:10.1523/JNEUROSCI.5592-10.2011

Pascanu, R., & Jaeger, H. (2011). A neurodynamical model for working memory. *Neural Networks: The Official Journal of the International Neural Network Society*, *24*(2), 199–207. doi:10.1016/j.neunet.2010.10.003

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378.

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews. Neuroscience*, *6*(2), 97–107. doi:10.1038/nrn1603

Payne, L., & Kounios, J. (2009). Coherent oscillatory networks supporting short-term memory retention. *Brain Research*, *1247*, 126–132. doi:10.1016/j.brainres.2008.09.095

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: implications for a working-memory model. *Memory & Cognition*, *20*(3), 314–320.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, *45*(1), S199–S209. doi:10.1016/j.neuroimage.2008.11.007

Perlstein, W. M., Cole, M. A., Larson, M., Kelly, K., Seignourel, P., & Keil, A. (2003). Steady-state visual evoked potentials reveal frontally-mediated working memory activity in humans. *Neuroscience Letters*, *342*(3), 191–195.

Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron*, *35*(5), 975–987.

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid Forgetting Prevented by Retrospective Attention Cues. *Journal of Experimental Psychology: Human Perception and Performance,* 39(5): 1224–31. doi:10.1037/a0030947.

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*, *136*(8), 2474–2485. doi:10.1093/brain/awt129

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290. doi:10.3758/BF03203943

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253–260. doi:10.3758/BF03214136

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*(1), 7–18. doi:10.3758/BF03198202

Polanía, R., Paulus, W., & Nitsche, M. A. (2012). Noninvasively decoding the contents of visual working memory in the human prefrontal cortex within high-gamma oscillatory patterns. *Journal of Cognitive Neuroscience*, *24*(2), 304–314. doi:10.1162/jocn_a_00151

Posner, M. I. (1967). Retention of visual and name codes of single letters. In A. F. Sanders (Ed.), *Attention and Performance* (pp. 267–284). Soesterberg: Acta Psychologia.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. doi:10.1016/j.neuroscience.2005.06.005

Pratte, M. S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *Journal of Vision*, *14*(3), 22. doi:10.1167/14.3.22

Preuschhof, C., Heekeren, H. R., Taskin, B., Schubert, T., & Villringer, A. (2006). Neural correlates of vibrotactile working memory in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(51), 13231–13239. doi:10.1523/JNEUROSCI.2767-06.2006

Rager, G., & Singer, W. (1998). The response of cat visual cortex to flicker stimuli of variable frequency. *The European Journal of Neuroscience*, *10*(5), 1856–1877.

Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., … Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *21*(9), 3175–3183.

Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15008–15013.

Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *The Journal of the Acoustical Society of America*, *119*(3), 1597–1605.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science (New York, N.Y.)*, *276*(5313), 821–824.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11800–11806. doi:10.1073/pnas.97.22.11800

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science (New York, N.Y.)*, *268*(5207), 111–114.

Rauschecker, J. P., Tian, B., Pons, T., & Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *The Journal of Comparative Neurology*, *382*(1), 89–103.

Ravizza, S. M., Hazeltine, E., Ruiz, S., & Zhu, D. C. (2011). Left TPJ activity in verbal working memory: implications for storage- and sensory-specific models of short term memory. *NeuroImage*, *55*(4), 1836–1846. doi:10.1016/j.neuroimage.2010.12.021

Reinhart, R. M. G., Heitz, R. P., Purcell, B. A., Weigand, P. K., Schall, J. D., & Woodman, G. F. (2012). Homologous mechanisms of visuospatial working memory maintenance in macaque and human: properties and sources. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(22), 7711–7722. doi:10.1523/JNEUROSCI.0215-12.2012

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology. Human Perception and Performance*, *5*(1), 129–145.

Ricciardi, E., Bonino, D., Gentili, C., Sani, L., Pietrini, P., & Vecchi, T. (2006). Neural correlates of spatial working memory in humans: a functional magnetic resonance imaging study comparing visual and tactile processes. *Neuroscience*, *139*(1), 339–349. doi:10.1016/j.neuroscience.2005.08.045

Ries, D. T., & DiGiovanni, J. J. (2007). Release from interference in auditory working memory for pitch. *Hearing Research*, *230*(1-2), 64–72. doi:10.1016/j.heares.2007.04.003

Ries, D. T., & DiGiovanni, J. J. (2009). Effects of recurrent tonal information on auditory working memory for pitch. *Hearing Research*, *255*(1-2), 14–21. doi:10.1016/j.heares.2009.05.002

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(38), 12990–12998. doi:10.1523/JNEUROSCI.1892-12.2012

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *NeuroImage*, *6*(4), 288–304. doi:10.1006/nimg.1997.0304

Roberts, B. M., Hsieh, L.-T., & Ranganath, C. (2013). Oscillatory activity during maintenance of spatial and temporal information in working memory. *Neuropsychologia*, *51*(2), 349–357. doi:10.1016/j.neuropsychologia.2012.10.009

Robertson, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nature Reviews. Neuroscience*, *4*(2), 93–102. doi:10.1038/nrn1030

Roland, P. E. (1987). Somatosensory detection of microgeometry, macrogeometry and kinesthesia after localized lesions of the cerebral hemispheres in man. *Brain Research*, *434*(1), 43–94.

Roland, P. E., O'Sullivan, B., & Kawashima, R. (1998). Shape and roughness activate different somatosensory areas in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(6), 3295–3300.

Romo, R., Brody, C. D., Hernández, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, *399*(6735), 470–473. doi:10.1038/20939

Romo, R., Hernández, A., Zainos, A., Brody, C., & Salinas, E. (2002). Exploring the cortical evidence of a sensory-discrimination process. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *357*(1424), 1039–1051. doi:10.1098/rstb.2002.1100

Romo, R., & Salinas, E. (2001). Touch and go: decision-making mechanisms in somatosensation. *Annual Review of Neuroscience*, *24*, 107–137. doi:10.1146/annurev.neuro.24.1.107

Romo, R., & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nature Reviews. Neuroscience*, *4*(3), 203–218. doi:10.1038/nrn1058

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. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 5975–5979. doi:10.1073/pnas.0711295105

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, *18*(1), 16–25. doi:10.1016/j.tics.2013.10.010

Roux, F., Wibral, M., Mohr, H. M., Singer, W., & Uhlhaas, P. J. (2012). Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(36), 12411–12420. doi:10.1523/JNEUROSCI.0421-12.2012

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113–129. doi:10.1016/j.cognition.2005.01.004

Ruusuvirta, T., Wikgren, J., & Astikainen, P. (2008). Proactive interference in a two-tone pitch-comparison task without additional interfering tones. *Psychological Research*, *72*(1), 74–78. doi:10.1007/s00426-006-0094-y

Salinas, E., Hernandez, A., Zainos, A., & Romo, R. (2000). Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(14), 5503–5515.

Salmela, V. R., Lähde, M., & Saarinen, J. (2012). Visual working memory for amplitude-modulated shapes. *Journal of Vision*, *12*(6), 2. doi:10.1167/12.6.2

Salmela, V. R., Mäkelä, T., & Saarinen, J. (2010). Human working memory for shapes of radial frequency patterns. *Vision Research*, *50*(6), 623–629. doi:10.1016/j.visres.2010.01.014

Samson, S., Zatorre, R. J., & Ramsay, J. O. (1997). Multidimensional scaling of synthetic musical timbre: perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *51*(4), 307–315.

Sanchez-Panchuelo, R. M., Francis, S., Bowtell, R., & Schluppeck, D. (2010). Mapping human somatosensory cortex in individual subjects with 7T functional MRI. *Journal of Neurophysiology*, *103*(5), 2544–2556. doi:10.1152/jn.01017.2009

Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G. L., & von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(12), 7092–7096.

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology. General*, *136*(4), 663–684. doi:10.1037/0096-3445.136.4.663

Sauseng, P., Klimesch, W., Doppelmayr, M., Hanslmayr, S., Schabus, M., & Gruber, W. R. (2004). Theta coupling in the human electroencephalogram during a working memory task. *Neuroscience Letters*, *354*(2), 123–126.

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., … Hummel, F. C. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Current Biology: CB*, *19*(21), 1846–1852. doi:10.1016/j.cub.2009.08.062

Schiff, W., & Foulke, E. (1982). *Tactual Perception: A Sourcebook*. Cambridge; New York: Cambridge University Press.

Scholl, B. J., & Xu, Y. (2001). The magical number 4 in vision. *Behavioral & Brain Sciences*, *24*, 145–146.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. doi:10.1016/S0167-6393(02)00094-8

Schouten, M. E., & van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *The Journal of the Acoustical Society of America*, *92*(4 Pt 1), 1841–1855.

Schwabbauer, M. L. (1975). Use of the latent image technique to develop and evaluate problem-solving skills. *The American Journal of Medical Technology*, *41*(12), 457–462.

Scott, B. H., Mishkin, M., & Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(30), 12237–12241. doi:10.1073/pnas.1209685109

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *The Journal of the Acoustical Society of America*, *89*(5), 2404–2410.

Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *The Journal of the Acoustical Society of America*, *94*(3 Pt 1), 1315–1322.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214. doi:10.1111/j.1467-9280.2009.02276.x

Sereno, M. I., McDonald, C. T., & Allman, J. M. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *4*(6), 601–620.

Serrien, D. J., Pogosyan, A. H., & Brown, P. (2004). Cortico-cortical coupling patterns during dual task performance. *Experimental Brain Research*, *157*(1), 79–84. doi:10.1007/s00221-003-1822-9

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. doi:10.1016/j.tics.2008.02.003

Silberstein, R. B. (1995). Steady-state visually evoked potentials, brain resonances, and cognitive processes. In P. L. Nunez (Ed.), *Neucortical Dynamics and Human EEG Rhythms* (pp. 272–303). New York: Oxford University.

Silberstein, R. B., Nunez, P. L., Pipingas, A., Harris, P., & Danieli, F. (2001). Steady state visually evoked potential (SSVEP) topography in a graded working memory task. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *42*(2), 219–232.

Sinclair, R. J., & Burton, H. (1996). Discrimination of vibrotactile frequencies in a delayed pair comparison task. *Perception & Psychophysics*, *58*(5), 680–692.

Smith, E. E., & Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42. doi:10.1006/cogp.1997.0658

Snijders, T. M., Petersson, K. M., & Hagoort, P. (2010). Effective connectivity of cortical and subcortical regions during unification of sentence structure. *NeuroImage*, *52*(4), 1633–1644. doi:10.1016/j.neuroimage.2010.05.035

Sohal, V. S., Zhang, F., Yizhar, O., & Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature*, *459*(7247), 698–702. doi:10.1038/nature07991

Sörös, P., Marmurek, J., Tam, F., Baker, N., Staines, W. R., & Graham, S. J. (2007). Functional MRI of working memory and selective attention in vibrotactile frequency discrimination. *BMC Neuroscience*, *8*, 48. doi:10.1186/1471-2202-8-48

Sörqvist, P., Stenfelt, S., & Rönnberg, J. (2012). Working memory capacity and visual-verbal cognitive load modulate auditory-sensory gating in the brainstem: toward a unified view of attention. *Journal of Cognitive Neuroscience*, *24*(11), 2147–2154. doi:10.1162/jocn_a_00275

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. doi:10.1037/h0093759

Spitzer, B., & Blankenburg, F. (2011). Stimulus-dependent EEG activity reflects internal updating of tactile working memory in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(20), 8444–8449. doi:10.1073/pnas.1104189108

Spitzer, B., & Blankenburg, F. (2012). Supramodal parametric working memory processing in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(10), 3287–3295. doi:10.1523/JNEUROSCI.5280-11.2012

Spitzer, B., Wacker, E., & Blankenburg, F. (2010). Oscillatory correlates of vibrotactile frequency processing in human working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(12), 4496–4502. doi:10.1523/JNEUROSCI.6041-09.2010

Starr, G. E., & Pitt, M. A. (1997). Interference effects in short-term memory for timbre. *The Journal of the Acoustical Society of America*, *102*(1), 486–494.

Sternberg, R. J. (1999). *Cognitive psychology* (2nd ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Stevanovski, B., & Jolicœur, P. (2011). Consolidation of multifeature items in visual working memory: central capacity requirements for visual consolidation. *Attention, Perception & Psychophysics*, *73*(4), 1108–1119. doi:10.3758/s13414-011-0099-5

Stoeckel, M. C., Weder, B., Binkofski, F., Buccino, G., Shah, N. J., & Seitz, R. J. (2003). A fronto-parietal circuit for tactile object discrimination: an event-related fMRI study. *NeuroImage*, *19*(3), 1103–1114.

Strenberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

Striem-Amit, E., Hertz, U., & Amedi, A. (2011). Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI. *PloS One*, *6*(3), e17832. doi:10.1371/journal.pone.0017832

Sullivan, E. V., & Turvey, M. T. (1972). Short-term retention of tactile stimulation. *The Quarterly Journal of Experimental Psychology*, *24*(3), 253–261. doi:10.1080/14640747208400278

Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science (New York, N.Y.)*, *293*(5527), 120–124. doi:10.1126/science.1060496

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, *91*(3), 1282–1296. doi:10.1152/jn.01125.2002

Talbot, W. H., Darian-Smith, I., Kornhuber, H. H., & Mountcastle, V. B. (1968). The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *Journal of Neurophysiology*, *31*(2), 301–334.

Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(11), 4244–4254.

Tarentino, A. L., & Maley, F. (1975). A comparison of the substrate specificities of endo-beta-N-acetylglucosaminidases from Streptomyces griseus and Diplococcus Pneumoniae. *Biochemical and Biophysical Research Communications*, *67*(1), 455–462.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science (New York, N.Y.)*, *292*(5515), 290–293. doi:10.1126/science.1058911

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

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage*, *77*, 157–165. doi:10.1016/j.neuroimage.2013.03.039

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., … Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(18), 7060–7078.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141.

Uldall, H. J. (1933). *A Danish phonetic reader*. London: University Press London.

Ungerleider, L. G., & Pasternak, T. (2004). Ventral and dorsal cortical processing streams. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (pp. 541–562). Cambridge: MIT Press.

Vallbo, A. B., & Johansson, R. S. (1984). Properties of cutaneous mechanoreceptors in the human hand related to touch sensation. *Human Neurobiology*, *3*(1), 3–14.

Van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. doi:10.1073/pnas.1117465109

Van Vugt, M. K., Schulze-Bonhage, A., Litt, B., Brandt, A., & Kahana, M. J. (2010). Hippocampal gamma oscillations increase with memory load. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(7), 2694–2699. doi:10.1523/JNEUROSCI.0567-09.2010

Ventre-Dominey, J., Bailly, A., Lavenne, F., Lebars, D., Mollion, H., Costes, N., & Dominey, P. F. (2005). Double dissociation in neural correlates of visual working memory: a

PET study. *Brain Research. Cognitive Brain Research*, *25*(3), 747–759. doi:10.1016/j.cogbrainres.2005.09.004

Verillo, R. T. (1968). A duplex mechanism of mechanoreception. In R. D. Kenshalo (Ed.), *The skin senses* (pp. 139–159). Springfield: Thomas, C.C.

Verrillo, R. T., & Gescheider, G. A. (1979). Backward enhancement and suppression of vibrotactile sensation. *Sensory Processes*, *3*(3), 249–260.

Vialatte, F.-B., Maurice, M., Dauwels, J., & Cichocki, A. (2010). Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in Neurobiology*, *90*(4), 418–438. doi:10.1016/j.pneurobio.2009.11.005

Vogels, R., & Orban, G. A. (1986). Decision processes in visual discrimination of line orientation. *Journal of Experimental Psychology. Human Perception and Performance*, *12*(2), 115–132.

Vuontela, V., Rämä, P., Raninen, A., Aronen, H. J., & Carlson, S. (1999). Selective interference reveals dissociation between memory for location and colour. *Neuroreport*, *10*(11), 2235–2240.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *3*(4), 255–274.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453. doi:10.1016/j.tics.2005.07.001

Walsh, V., Le Mare, C., Blaimire, A., & Cowey, A. (2000). Normal discrimination performance accompanied by priming deficits in monkeys with V4 or TEO lesions. *Neuroreport*, *11*(7), 1459–1462.

Wang, L., Bodner, M., & Zhou, Y.-D. (2013). Distributed neural networks of tactile working memory. *Journal of Physiology, Paris*, *107*(6), 452–458. doi:10.1016/j.jphysparis.2013.06.001

Wang, L., Li, X., Hsiao, S. S., Bodner, M., Lenz, F., & Zhou, Y.-D. (2012). Persistent neuronal firing in primary somatosensory cortex in the absence of working memory of trial-specific features of the sample stimuli in a haptic working memory task. *Journal of Cognitive Neuroscience*, *24*(3), 664–676. doi:10.1162/jocn_a_00169

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 455–463.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(13), 5799–5804.

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, *24*(4), 1052–1057. doi:10.1016/j.neuroimage.2004.10.031

Warren, R. M., (2008). *Auditory Perception: An Analysis and Synthesis*. 3rd ed. Cambridge, UK ; New York: Cambridge University Press.

WAUGH, N. C., & NORMAN, D. A. (1965). PRIMARY MEMORY. *Psychological Review*, *72*, 89–104.

Weber, L., & Experimental, P. S. (1978). *The sense of touch*. London: Academic Press for the Experimental Psychology Society.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology. General*, *131*(1), 48–64.

Wickelgren, W. A. (1965). Distinctive features and errors in short-term memory for English vowels. *The Journal of the Acoustical Society of America*, *38*(4), 583–588.

Wickelgren, W. A. (1966). Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, *71*(3), 396–404.

Wickelgren, W. A. (1969a). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, *6*, 13–61.

Wickelgren, W. A. (1969b). Auditory or articulatory coding in verbal short-term memory. *Psychological Review*, *76*(2), 232–235.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. doi:10:1167/4.12.11

Williams, P. E., Mechler, F., Gordon, J., Shapley, R., & Hawken, M. J. (2004). Entrainment to video displays in primary visual cortex of macaque and humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(38), 8278–8288. doi:10.1523/JNEUROSCI.2716-04.2004

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory (Hove, England)*, *18*(6), 657–669. doi:10.1080/09658211.2010.501339

Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., … Kang, X. (2009). Functional maps of human auditory cortex: effects of acoustic features and attention. *PloS One*, *4*(4), e5183. doi:10.1371/journal.pone.0005183

Xing, Y., Ledgeway, T., McGraw, P. V., & Schluppeck, D. (2013). Decoding working memory of stimulus contrast in early visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(25), 10301–10311. doi:10.1523/JNEUROSCI.3754-12.2013

Yost, W. A. (2007). *Auditory Perception of Sound Sources*. Springer.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(4), 1908–1919.

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, *26*(5), 384–395.

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

Zhou, Y. D., & Fuster, J. M. (1996). Mnemonic neuronal activity in somatosensory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(19), 10533–10537.

Zhou, Y. D., & Fuster, J. M. (2000). Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(17), 9777–9782.

Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P. M., & Husain, M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal of Vision*, *11*(14). doi:10.1167/11.14.2