TEMPORAL PROPERTIES OF REHEARSAL IN AUDITORY-VERBAL SHORT-TERM MEMORY

Rebecca Anne Gilbert

PhD

University of York

Psychology

September 2014

ABSTRACT

Subvocal rehearsal, the use of inner speech for the maintenance of phonological material, is thought to play an important role verbal short-term memory (STM). The importance of rehearsal is based largely on indirect measures, as it is difficult to detect and quantify. To address this issue and investigate rehearsal timing, a novel 'rehearsal-probe' task was developed. Individuals silently rehearsed an auditory-verbal sequence, responding after an unpredictable probe (tone) by indicating the item currently being rehearsed. The presentation of probes after variable and repeated delays provides item response proportions over time. The data were analysed using a theory-neutral measure of temporal precision; the circular standard deviations of response distributions.

The methods were established across seven experiments designed to explore whether timing precision is fixed or resource-limited. Experiment 3 showed that timing precision decreases with increased in memory load. Temporal precision was negatively correlated with auditory-verbal STM span in six experiments, including one designed specifically to examine individual differences. Experiments 6 and 7 investigated timing in developmental language disorders, which are characterized by serial ordering deficits. Adults with dyslexia and children with language impairments showed more temporal imprecision compared to matched controls. These results suggest that temporal precision is limited by shared resources and may play a role in language development.

A computational model was also developed to describe the data with four separable temporal properties. The model captured the main characteristics of the data and provided quantitative estimates of each property. In an EEG experiment, event-related responses to item probes were modulated by the contents of rehearsal, and there was increased spectral power at the item rate during sequence presentation and rehearsal, but not baseline, periods. The findings suggest an important role for fine-grained timing information in serial order STM and have broader implications for debates about models of serial order.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This thesis grew out of my unwavering determination to study short-term memory, despite a number of personal and financial obstacles. The journey has been a difficult but worthwhile one and I owe thanks to all who made this possible.

First and foremost I want to express my deepest appreciation for my two supervisors, Dr Tom Hartley and Professor Graham Hitch. This thesis only came to be because Tom was kind enough to respond to my rambling emails about research project ideas, and I'm so grateful that he took a chance on me. Tom, your enthusiasm is wonderfully contagious, and your patience and sense of humour kept me motivated. I'm truly inspired by your dedication to science and to your students. Graham, thank you for your thoughtful feedback, and for showing me that it's possible to achieve scientific greatness without sacrificing humility and kindness.

I am also very grateful for support and guidance from Dr Pádraig Kitterick and Professor Alan Baddeley, who provided valuable advice as members of my research committee.

I would like to thank the Department of Psychology for providing an annual stipend to support this project, and for creating an environment conducive to scientific dissemination, collaboration and personal development. I am also thankful for financial support in the form of an early career researcher award from the Royal Society to attend the Language in Developmental and Acquired Disorders conference and satellite meeting, and to the Experimental Psychological Society for a grant to attend the Computational Modelling Summer School in 2012.

Thank you to Dr Joni Holmes and Professor Sue Gathercole for allowing me the opportunity to visit the MRC CBU in Cambridge. I greatly appreciate the time and effort Joni put forth in organising my visit. Many thanks go to Sally Butterfield, who was instrumental in facilitating the data collection at the schools. I'd also like to thank Tom's two daughters, Cathy and Isobel Hartley, for helping me pilot the task used with the children in Cambridge.

I met a few wonderful people in York who supported me in different ways throughout this project. I'm grateful for having had the opportunity to work with Justyna Sobczak during the EEG pilot study. She was an essential part of the early EEG work and has been a wonderful friend ever since. I'd like to thank Tori Brattan for being such a kind person and considerate friend. I also owe thanks to Taliah Farnsworth, an absolutely delightful visiting undergraduate student who generously donated her time to help with EEG and behavioural pilot work.

To my family, I can't thank you enough for your love, patience and support. My sister and brother-in-law, Sarah and Pete Carlson, thank you for being so understanding and supportive, for having faith in me, and for helping me keep things in perspective. Thank you to Jane Gilbert, my mother-in-law, and Rachel Ward, my sister-in-law, for your generosity and kindness. All of our ridiculous adventures helped to break up the monotony of research.

To my husband, Dr Mat Gilbert – no one has been more directly affected by the ups and downs of this journey than you have. Thank you for all the sacrifices you've made so that I could do this. Thank you for having faith in me, for the practical support, for doing more than your share of the washing up, and for always making me laugh. Thank you for the 58 (and counting) 150+ mile journeys to see me on Friday evenings, even though it meant driving all the way back again on Monday morning. But most of all, thank you for being my inspiration.

Finally, to my mom, Deb Barnes, and my step-dad, Gordon Barnes, you have been the biggest influences on my life and it goes without saying that I couldn't have done this without you. You were with me every step of the way, through IB tests, college visits, financial aid applications, multiple moves and even hay deliveries. Mom, thank you for the packages full of crossword puzzles and the never-ending supply of encouragement. I wouldn't have been brave enough to move to England and pursue this degree without you.

DECLARATION

This thesis comprises the candidate's own original work and has not been submitted to this or any other University for a degree. All experiments were designed and conducted by the candidate under the supervision of Dr Tom Hartley and Professor Graham Hitch. Some parts of the work presented in this thesis have been presented elsewhere as posters, conference talks and informal lab group presentations.

Experiment 7 (Chapter 5) was done in collaboration with Dr Joni Holmes at the Medical Research Council Cognition and Brain Sciences Unit (MRC CBSU). The candidate was solely responsible for the rehearsal-probe task administration and data analysis. Dr Holmes and colleagues recruited the participants and collected all other data as part of an ongoing working memory training study.

The development of the computational model (Chapter 6) was done in collaboration with Dr Tom Hartley. Tom wrote the original version of the model, and later versions were done by the candidate or in collaboration.

The EEG experiment (Chapter 7) was based on a previous EEG study, which was carried out by the candidate and Justyna Sobczak under the supervision of Dr Tom Hartley. Justyna Sobczak and Taliah Farnsworth assisted with data collection in the EEG experiment presented in Chapter 7.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The topic of this thesis is the temporal properties of subvocal rehearsal in auditory-verbal short-term memory (STM). In this thesis I will present a novel behavioural method for probing the contents of rehearsal over time, referred to as the rehearsal-probe task, and analysis methods for quantifying the temporal properties of subvocal rehearsal (Chapter 2). After introducing these methods, I will investigate relationships between the rehearsal timing properties, auditory-verbal STM load (Chapter 3), and individual differences in immediate serial recall (Chapter 4). I will then explore the possible application of this method to studying developmental language disorders (Chapter 5). In addition to the analysis method described in Chapter 2, I will also present a computational model designed to quantify the separable temporal properties of rehearsal (Chapter 6). Finally, as a distinct approach toward the goal of detecting and measuring the temporal properties of subvocal speech, I will investigate the potential for detecting neural markers for subvocal rehearsal using the rehearsal-probe task with electroencephalography (EEG).

Before introducing the rehearsal-probe task, I will provide historical context about the debate surrounding the role of rehearsal in STM, and I will review the literature that has been instrumental in motivating this thesis. The rationale for the hypotheses and methods in this thesis draws from a diverse body of literature; therefore, a broad range of topics will be covered in the review of the literature that follows. In the review I will first evaluate the view that rehearsal plays an important role in STM performance. I will further show that many of the historical and contemporary techniques used for investigating subvocal rehearsal, while instrumental in laying the foundation for understanding STM and rehearsal, are insufficient for conclusively detecting and quantifying subvocal speech in STM tasks. As a result of methodological limitations, there are ongoing debates regarding the interpretations of many STM phenomena. This problem has motivated one aim of this thesis, which is to develop a new method of detecting and measuring rehearsal that has the potential to test some of the theoretical assumptions that have thus far provided the basis for our interpretations of STM performance.

As I will demonstrate in the literature review, there is evidence to suggest that fine-grained temporal properties of internal speech representations may be important for STM for serial order. Such temporal properties of rehearsal include rate and variability. However, the limitations of the methods that have been previously used to measure or manipulate rehearsal preclude their ability to address this issue. The review will demonstrate that sequence presentation timing affects auditory-verbal STM performance, and it will be argued that an ability to accurately encode and maintain these temporal attributes is related to superior STM for serial order. Therefore another aim of this thesis is to quantify the timing of rehearsal, and to investigate relationships between timing and auditory-verbal STM for serial order via rehearsal.

After considering the literature related to rehearsal and timing in STM, I will focus on the importance of temporal precision in the neural basis of speech perception. I will evaluate the evidence for dynamic time-based mechanisms in speech perception, such as the role of cortical oscillatory entrainment in temporal prediction and the segmentation of continuous speech into meaningful units. A review of the research on these topics will illustrate that temporal precision in auditory-verbal perception is critical for higher level speech encoding and processing. These mechanisms provide a potential link from low-level auditory temporal perception to the segmentation and sequencing of higher-level, abstract units of language such as phonemes, syllables and words. The effects of temporal variability may carry through from speech encoding to maintenance, and thus may be reflected in the subvocal rehearsal of a sequence of spoken material during the rehearsal-probe task. Some of the concepts discussed in this section will provide a basis for the EEG experiments presented in Chapter 7.

Next, STM for serial order, rhythm and timing will be discussed vis-à-vis language and literacy disorders. The aforementioned literature on temporal precision and speech perception also relates to the rationale for hypothesised relationships between the timing, STM and phonological deficits observed in developmental language disorders. That is, one possible implication of poor temporal precision in the neural encoding and maintenance of speech is that it could result in some of the difficulties typically reported in individuals with developmental language disorders. Groups with developmental language disorders such as dyslexia and SLI have been shown to have abnormal entrainment to speech rhythms, as well as deficits in reproducing auditory rhythmic sequences and maintaining spoken sequences in the correct serial order. I will address this topic in the literature review by evaluating the evidence for specific deficits in timing and STM for serial order in developmental language disorders, as well as discussing the possible causal relationships between these impairments and phonological deficits. Here I will argue that, like speech perception and new word learning, reading development depends critically on the ability to perceive and retain multi-level temporal properties of auditory-verbal stimuli. I will also argue that a fine-grained measure of the timing of STM representations would shed light on this topic. This will provide rationale for a preliminary investigation into rehearsal timing in developmental disorders, which is presented in Chapter 5.

1.1 Perspectives on Subvocal Rehearsal in Short-Term Memory

Subvocal rehearsal is the use of silent speech for the short-term maintenance of phonological material. It has long been thought to play a role in memory, but despite a history of interest in the topic, the precise role and relative importance of the rehearsal process in STM has remained elusive. Many early memory researchers, including Waugh and Norman (1965) and Atkinson and Shiffrin (1968), adopted the view that rehearsal serves the dual purpose of maintaining items in a transient short-term store (or 'primary memory') and transferring this information into a more stable, longer-term store ('secondary memory'). Waugh and Norman (1965) observed that, when rehearsal was restricted to only the item that immediately preceded each inter-item interval, the typical primacy effect was eliminated. The primacy effect refers to the recall advantage for the earliest sequence items. The authors interpreted this result as indicating that cumulative rehearsal aids in the transfer of items from primary to secondary memory, with the earlier items benefiting more from cumulative rehearsal and thus are able to be recalled from secondary memory. This view was supported by evidence for a positive linear relationship between the number of overt item rehearsals and immediate recall (Rundus & Atkinson, 1970; Rundus, 1971, 1974; Weist, 1972). Rundus and Atkinson (1970) argued that this view of rehearsal explained both the primacy and recency effects typical of immediate recall; the probability of correct recall for each item being a function of both the total number of rehearsals and the duration between rehearsal and recall. Other studies have also shown that there is a specific primacy benefit in immediate recall when participants have a greater opportunity or are instructed to rehearse (Brodie & Prytulak, 1975; Fischler, Rundus, & Atkinson, 1970; Glanzer & Meinzer, 1967; P. H. Marshall & Werder, 1972). This result was interpreted by most authors as showing that earlier items benefitted from an increased number of rehearsals.

In spite of the evidence that increased rehearsal in the short-term store leads to improved immediate recall, this effect did not transfer to improved recall over delays as many theorists had supposed. Craik and Watkins (1973) were among the researchers who showed that LTM performance does not benefit from an increase in the length of time spent rehearsing, nor with the number of overt rehearsals. This result was in conflict with the predictions of Atkinson and Shiffrin's Modal Model. Craik and Watkins' findings added to the set of conflicting evidence regarding whether or not time spent in the short-term store, which is increased via rehearsal, aids in the transfer of information into a longer-term store (R. A. Bjork & Whitten, 1974; Glanzer & Meinzer, 1967; Jacoby & Bartz, 1972; Jacoby, 1973; Maki & Schuler, 1980; Rundus & Atkinson, 1970; Rundus, 1971; Tulving, 1966; Tzeng, 1973; Weist, 1972). In addition, some studies showed that rehearsal improves long-term recognition but not necessarily recall (Glenberg & Adams, 1978; Glenberg, Smith, & Green, 1977).

The contradictory results from multiple studies on this topic were especially difficult to reconcile due to the diverse and indirect methods of measuring rehearsal activity. While some of the researchers attempted to directly control the pace and number of item rehearsals, others used intervening tasks to disrupt or prevent rehearsal, or assumed that subvocal rehearsal would occur for only for particular words and durations due to the nature of the task. Craik and Watkins (1973) accounted for the inconsistent pattern of results by distinguishing between maintenance and elaborative rehearsal types[1]. The authors explained that time spent rehearsing will benefit later recall if that time is spent on cognitive and semantic analysis of the material, which Craik and Lockhart (1972) refer to as deep processing in their Levels of Processing framework. Deeper level encoding of verbal material may involve semantic (rather than, say, orthographic or phonological) processing of items, and the formation of connections between the new verbal material with existing knowledge in LTM. In the case of rote repetition of material, such as that used in overt rehearsal paradigms, the Levels of Processing framework predicts that there will be no benefit at delayed recall. However, Rundus (1980) later showed that recency effects at delayed recall could be observed with an overt rehearsal paradigm over longer durations (up to 60 seconds), which is not predicted by the Levels of Processing framework. Nonetheless, the distinction between maintenance and elaborative rehearsal clarified some of the empirical discrepancies, and perhaps more importantly, it highlighted the need for direct, controlled and atheoretical techniques for manipulating and measuring rehearsal.

The introduction of the Baddeley and Hitch Working Memory model (Baddeley & Hitch, 1974; see also Baddeley, 1986, 2007) was instrumental in providing a common theoretical framework that has since driven a wealth of empirical investigation into rehearsal. The Working Memory model arose in part as a response to the problems faced by unitary accounts of short-term memory. One such problem is that the unitary accounts could not account for the finding that the maintenance of digit sequences of near span-length produced a less than expected deficit in performance on a concurrent cognitive task. According to a unitary store account of short-term memory, near-span digit sequences should have loaded the singular short-term store capacity and caused a massive performance deficit in the concurrent reasoning or learning task. Baddeley and Hitch (1974) addressed this and other modality effects inconsistent with a unitary store model by proposing the existence of a

---

[1] While useful at the time, the explicit distinction between maintenance (Type I) and elaborative (Type II) rehearsal is not usually needed in the context of verbal STM as the term ‘rehearsal’ now typically refers to simple rote repetition (i.e. ‘maintenance rehearsal’). The term ‘rehearsal’ in this thesis refers to the cyclic, forward-ordered repetition of verbalisable material for the purpose of maintenance in STM. References to rehearsal throughout this thesis should also be assumed to be silent (i.e. subvocal, covert) unless specifically labelled as vocal or overt.

multi-component system, comprised of a main STM executive controller aided by separate 'slave systems' dedicated to memory for auditory and visual information.

The current instantiation of the Working Memory model (Baddeley, 2000, 2007) consists of the *central executive* controller aided by the *visuo-spatial sketchpad* and the *phonological loop*, which are linked by the *episodic buffer* (see Figure 1). The slave systems were hypothesized to be linked via the central controller but operate independently in other respects, drawing from primarily separate limited resources. The central executive is responsible for the coordination of the subsystems, task switching, selective attention and the binding of information from multiple sources. The visuo-spatial sketchpad is specialized for processing visuo-spatial material, while the phonological loop is specifically concerned with STM for language. The episodic buffer is responsible for the linking of information across visuo-spatial, verbal and temporal domains to form integrated units. Finally, these three dynamic short-term storage subsystems have direct access to long-term knowledge and semantics.

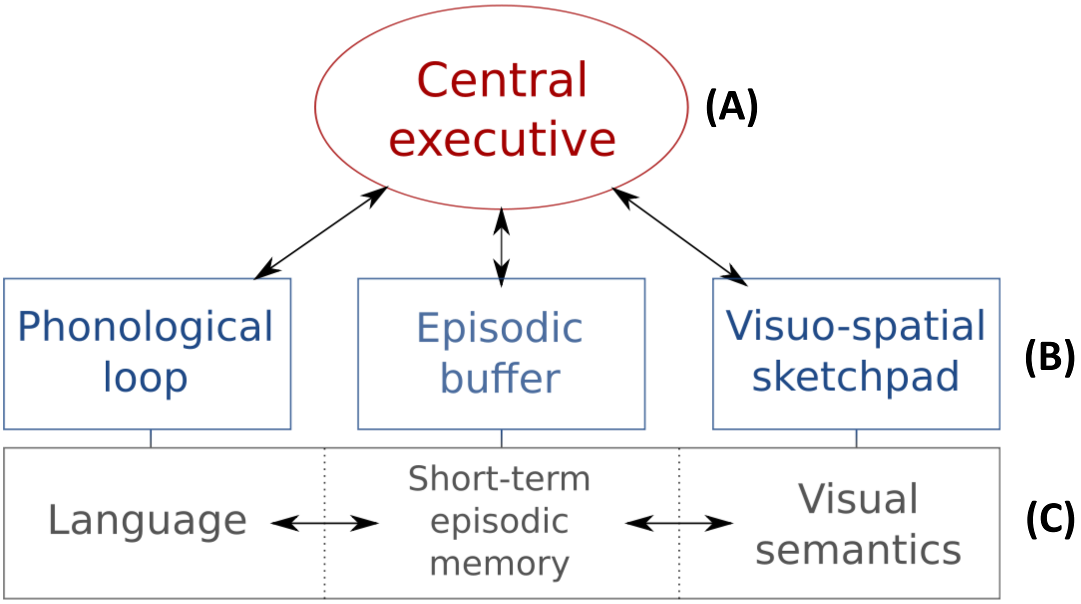

Figure 1. The multiple component Working Memory model, consisting of the central executive (A), the phonological loop, episodic buffer and visuo-spatial subsystems (B), and the links to long-term knowledge (C). The central executive and subsystems (A and B) are fluid, while the long-term stores (C) are crystallised (adapted from Baddeley, 2007).

1.1.1 Rehearsal in the Phonological Loop

The phonological loop component is of particular interest to this thesis as it deals exclusively with the short-term storage of auditory-verbal information. The phonological loop controls speech-based material in a temporally-limited short-term store, and is capable of maintaining this material long enough for recall or use in other cognitive processes. In terms of structure, the phonological loop consists of two separate components; the *phonological store* and the *articulatory loop* (see Figure 2). Speech stimuli enter the phonological store (Figure 2 A) and are represented in terms of its phonemic elements. A phoneme is the most basic unit of speech that provides the smallest level of semantic differentiation between two words (e.g. “cat” and “hat”, “ran” and “rant”). The phonological store also serves as a buffer for speech, preparing any to-be-articulated verbal material for rapid access to the articulatory loop. The contents of the phonological store decay rapidly but can be refreshed via rehearsal in the articulatory loop (Figure 2 B). As shown in Figure 2, speech inputs have direct access to the phonological store, and any speech input is automatically coded phonologically. In contrast, nonspeech inputs (e.g. written words, nameable pictures) do not have immediate access to the phonological loop system. In order to gain access, this material must be re-coded in terms of its phonemic-level representation.

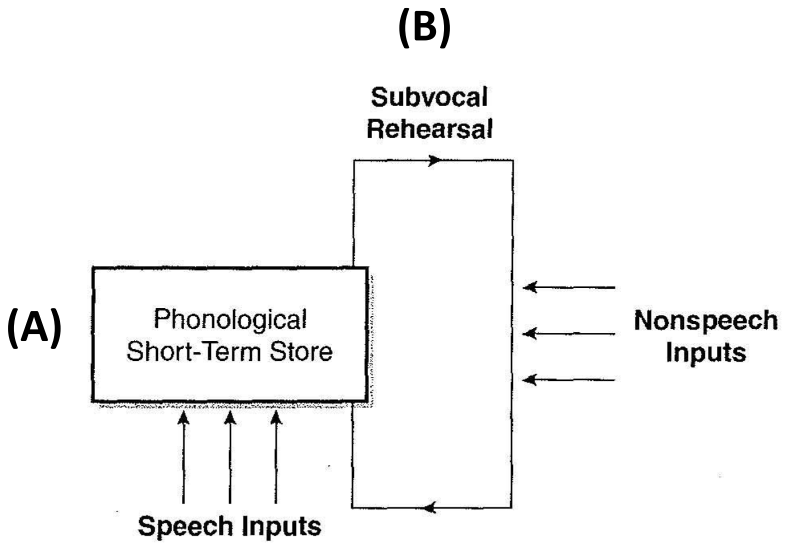

Figure 2. Representation of the phonological loop component of the Baddeley and Hitch Working Memory model. Included are the two components of the phonological loop; the phonological short-term store (A), and the articulatory loop (B). Source: Baddeley, 1986.

The phonological loop is occupied by self-produced speech which, importantly, is not necessarily spoken aloud. Within the phonological loop, vocal and subvocal speech types are presumed to be identically represented. The phonological loop model places the differentiation between these two modes of speech in the very last stages involved in the overt speech process, that is, in the transmission of motor commands and efference copies to peripheral articulatory musculature. This last point is very relevant to the EEG experiments presented in this thesis, and thus it will be discussed in more detail in Chapter 7.

One strength of the phonological loop model is that it comes with a set of experimental methods for testing its theoretical mechanisms. The importance of the subvocal rehearsal process in particular has been demonstrated by extensive testing using the experimental methods proposed by Baddeley and others to isolate the individual components. These methods are reviewed in the next section. The empirical support for rehearsal, as it is defined in the phonological loop model, has established a wealth of reliable findings whose implications must be considered when investigating relationships between rehearsal timing and STM performance.

Before reviewing the evidence regarding the set of factors known to affect verbal STM performance, definitions of the most common dependent measures are necessary. STM has been tested extensively using tasks which require the subject to recall the sequence in order either immediately after presentation, termed an *Immediate Serial Recall* (ISR) task, or after a given delay, referred to as a *Delayed Serial Recall* task. Digit span is perhaps the most widely-used ISR task as it has become a standard measure of general STM capacity. The memory span method, wherein the subject is tested for ISR of lists of increasing length until a certain threshold of errors is reached, is frequently used for other verbal material such as non-digit words and letters. In addition to tasks of (forward) serial recall, subjects can also be asked to recall the list items in any order, as in a *Free Recall* task, or in some other specified order, such as backwards. In order to test memory of serial order while controlling for the effects of output interference, some research has employed the *Probed Recall* method in which subjects must respond with the presented item in the probed serial position, or respond with the item that immediately preceded/followed a probe item from the list. Two other common methods for testing verbal STM are *Serial Reconstruction*, in which the items are re-presented in random order and subjects must arrange the items in the correct order, and *Serial Recognition*, in which subjects must make a same/different judgment about the order of the re-presented items. These two methods allow the specific testing of STM for serial order information, as all of the item information is available in the recall phase. Of course, these tasks tap into different aspects of STM, so the outcome STM measure may be an important consideration when interpreting the effects of a given manipulation. Most of the investigations which provide evidence for rehearsal use recall performance on an auditory-verbal ISR (e.g. digit span, word span) task as the dependent measure. Recall performance can be quantified in terms of maximum list length where performance is above some threshold (i.e. span), or in terms of the percentage of correct or erroneous responses.

1.2 Methods for Investigating Rehearsal in the Phonological Loop

Support for a separate speech-based memory storage system, such as the phonological loop model, comes from investigation into a set of factors which are known to affect performance in STM tasks. These factors are articulatory suppression, irrelevant speech, speech rate and word length, phonological similarity, and neuropsychological impairments. Each of these factors will be discussed in turn, with the exception of the irrelevant speech effect as this phenomena relates more to the phonological store than to the articulatory loop.

1.2.1 Articulatory Suppression

The articulatory suppression technique was first used by Murray (1967, 1968) as a strategy of preventing rehearsal during STM tasks. In the standard use of the articulatory suppression (AS) method, subjects engage in a STM task while repeating irrelevant words (e.g. "the the the...") throughout list presentation and/or retention intervals. In Murray's (1967, 1968) studies, as in many others (Hanley & Bakopoulou, 2003; Hanley, 1997; Larsen & Baddeley, 2003; Salamé & Baddeley, 1982; Surprenant, Lecompte, & Neath, 2000), subjects show a performance deficit under suppression as compared to non-suppression conditions. Baddeley and colleagues have reasoned that self-produced speech, such as that which occurs during AS, necessarily occupies the articulatory loop and thereby prevents concurrent articulatory rehearsal of other verbal material (Baddeley & Hitch, 1974; Baddeley, Thomson, & Buchanan, 1975; Baddeley, 1986). The fact that STM performance suffers but is not completely impaired with AS demonstrates that rehearsal appears to be beneficial, yet non-essential, process for the maintenance of verbal material in STM.

1.2.2 Speech Rate

Convincing evidence in support of a specific speech-based subsystem involved in STM, as well as the role of rehearsal within this subsystem, comes from the positive linear relationship observed between speech rate and memory span (Baddeley et al., 1975). The standard interpretation of this relationship is that speech rate is an index of rehearsal rate, and the rate of rehearsal is the sole or main determinant of STM span when the items are encoded phonologically (Baddeley, 1986). According to the phonological loop account, rehearsal prevents the temporal decay of phonological memory traces in a limited-capacity short-term store. Within this view, faster rehearsal results in more frequent refreshing of memory traces, which in turn leads to less decay of the item representations during the maintenance period.

If the mechanisms driving subvocal speech are equivalent or closely related to those governing overt speech, apart from absent or suppressed articulatory motor commands, then it should be the case that overt speech rate is highly correlated with covert speech rate (Baddeley et al., 1975; Baddeley, 1986). Many researchers have worked under this assumption and, using speech rate is an index of rehearsal rate, found high correlations between measures of rehearsal speed and STM span (Baddeley, 1986; Gathercole, Adams, & Hitch, 1994; Hulme, Thomson, Muir, & Lawrence, 1984; Nicolson, 1981; Standing, Bond, Smith, & Isely, 1980). Articulation rate has been shown to account for a portion of the variance in adult span scores (Baddeley et al., 1975; Gathercole, Adams, et al., 1994; Standing et al., 1980) as well as developmental changes in span (Hulme et al., 1984; Nicolson, 1981). For instance, Baddeley et al. (1975) found that adult memory span was highly correlated with reading rate ($r = .685$, $p < .005$), where reading rate was measured by requiring subjects to read printed words as rapidly as possible. These results suggest that subvocal rehearsal speed has a particular benefit for immediate serial recall performance.

Studies have found a relationship between children's articulation speed and STM spans over a wide range of ages (Cowan et al., 1994; Hulme et al., 1984; Hulme & Tordoff, 1989; Kail & Park, 1994; Kail, 1997; McDougall, Hulme, Ellis, & Monk, 1994), which supports the view that the increase in verbal STM span throughout childhood is attributable to corresponding increases in children's subvocal rehearsal speeds. Furthermore, these and other studies have shown that the relationship between speech rate and span in children cannot be fully explained by other factors such as age, phonological ability, processing speed, or item identification time (L. A. Henry & Millar, 1991; Hitch, Halliday, & Littler, 1989; Kail & Park, 1994; Kail, 1997; Tam, Jarrold, Baddeley, & Sabatos-DeVito, 2010). Some studies on the relationship between speech rate and memory span in younger children have failed to find this correlation, and this has been interpreted as evidence that young children do not use a subvocal rehearsal strategy (Gathercole, Adams, et al., 1994; Gathercole & Adams, 1994). Gathercole, Adams and Hitch (1994) did not find any relationship between memory span and articulation rates in 4 year olds. Gathercole and Adams (1994) found that, in 5 year olds, speech rate was significantly correlated with digit span but not word span. Because a significant correlation between speech rate and memory span would not be expected in the absence of rehearsal, based on these results it was concluded that children do not begin to rehearse until around age 6. Thus it seems from this evidence that, once young children begin to use a rehearsal strategy, subsequent increases in subvocal rehearsal speed could explain some of the age-related increase in ISR performance.

1.2.3 Word Length Effect

While the relationship between speech rate and STM span is examined at the level of individual differences, the relationship between word length and STM span is a closely related phenomenon that can be more easily manipulated within individuals. The word length effect (WLE) is the observation that, given a certain list length (number of words), lists of longer words are less likely to be correctly recalled than shorter words in ISR tests (Baddeley et al., 1975). The WLE is thought to occur because longer words, which take longer to articulate, are refreshed at a slower rate via rehearsal in the articulatory loop, making the traces more susceptible to degradation (Baddeley, 1986; Jacquemot, Dupoux, & Bachoud-Lévi, 2011). To test the rehearsal rate hypothesis against item-based storage capacity and other possibly confounding factors, Baddeley et al. (1975) tested immediate serial recall words from sets differing in terms of spoken duration but which were matched for frequency, number of phonemes, number of syllables, and within-word phoneme repetition. The authors found a significant performance deficit in recall for longer words across multiple experiments using words with 1-5 syllables and sequences of 4 to 8 words. The WLE reported by Baddeley et al. has been widely replicated in adults and children as young as 4 years old, and occurs with both visual and auditory presentation modalities (Coltheart & Langdon, 1998; Hitch, Halliday, Dodd, & Littler, 1989; Hitch, Halliday, & Littler, 1989; Hitch, Halliday, Schaafstal, & Heffernan, 1991; Hulme, Silvester, Smith, & Muir, 1986; Hulme et al., 1984; Hulme & Tordoff, 1989; Nicolson, 1981; Standing et al., 1980). The fact that the WLE is observed in adults and children has given support to the view that subvocal rehearsal speed may account for some individual and developmental variability in STM span.

That the WLE is also observed with visual stimuli presentation is further evidence for the importance of subvocal rehearsal in STM, as it suggests that re-coding nameable visual stimuli into phonological form has a beneficial effect on recall performance. Investigations into the WLE in STM for nameable pictures with children have produced inconsistent results; some authors finding that this effect emerges in children as young as 4 (Hulme et al., 1986), and others failing to find the effect before age 8 (Hitch, Halliday, Dodd, et al., 1989; Hitch et al., 1991). The discrepancy between the presence of the WLE for pictorial and auditory-verbal material during development could be due to the fact that auditory stimuli have direct access to the phonological loop, while subvocal rehearsal for pictures requires phonological recoding. Young children may differ in their age of acquisition of a phonological recoding strategy, thus resulting in discrepant findings with regard to the youngest onset of the WLE in the visual modality. The notion that young children may use visuo-spatial rather than phonological encoding strategies during visual ISR tasks is supported by evidence of higher susceptibility to visual similarity and retroactive interference in 5-year-old compared to 10-year-old children (Hitch, Halliday, Schaafstal, & Schraagen, 1988). Because STM recall for pictures seems to benefit from overt labelling in young children who do not yet spontaneously rehearse (Hitch et al., 1991), it can be concluded that subvocal rehearsal is advantageous for recall even when the material must be re-coded in order to gain access to the phonological loop. Taken with the evidence regarding the effects of speech rate on recall, the robust effects of word length on STM recall provide evidence for the view that the speed of articulatory rehearsal plays a critical role in immediate serial recall.

Further evidence for the relationship between the speed of subvocal rehearsal and immediate serial recall performance comes from cross-linguistic studies comparing STM spans for words in different languages, where the same words vary in spoken duration. Analyses of the memory spans and spoken word durations in adult (Naveh-Benjamin & Ayres, 1986; Standing et al., 1980; Stigler, Lee, & Stevenson, 1986) and child (Chen & Stevenson, 1988; Stigler et al., 1986) speakers of English, Welsh, Chinese and other languages have shown that articulatory duration of the words used in span lists can account for differences in span, both between native speaker groups and within bilingual subjects. In an influential paper by Ellis and Hennelly (1980), a cross-linguistic WLE was identified in Welsh versus English bilinguals. This research was prompted by the observations that Welsh children have on average a lower digit span than American children, and that Welsh digits take longer to pronounce than their English equivalents. The authors reported that spoken word duration could account for span performance differences between Welsh and English digits in Welsh-English bilinguals. Like the monolingual WLE, the bilingual effect is eliminated with the use of AS, which further points to subvocal rehearsal as the source of the effect (Chincotta & Hoosain, 1995). Also, the effects of cross-language differences in word length on span are reduced or eliminated for backward span, which suggests that backward serial recall is not as dependent upon subvocal rehearsal speed, presumably because it involves a different strategy (Chen & Stevenson, 1988). These results reinforce the findings related to the within-language speech rate and WLE on immediate serial recall, which together provide convincing support for the view that subvocal rehearsal speed is positively related to ISR performance.

1.2.4 Phonological Similarity Effect

In contrast to the effects of AS, speech rate and word length, the phonological similarity effect (PSE) does not depend on the hypothesized articulatory loop with auditory stimuli and is therefore used to define the circumstances when rehearsal is not a factor in recall performance. The PSE is the observation that the probability of correct recall for items decreases as a function of increasing phonemic similarity. This finding supports the supposition that verbal material is coded phonologically in STM (Baddeley, 1968; Conrad, 1963, 1965; Wickelgren, 1966). The finding that the PSE is dissociable from rehearsal was a key contributor to the dual-component structure of the phonological loop system. The PSE does not depend on active rehearsal in the articulatory loop ‘inner voice’ system, but rather it is thought to occur as a result of the automatic phonological coding of auditory verbal input in the ‘inner ear’ (Baddeley, 1986). While the PSE does not originate in the rehearsal process, there has been at least one study showing that there may still be a link between speech rate and the PSE. Hulme and Tordoff (1989) found that the magnitude of the PSE increases with age, and this is also closely tied to speech rates. The authors provide the following explanation for these findings. As age and speech rate increase, so does the amount of information read and rehearsed in storage over a given time interval. The error rates for phonologically similar and dissimilar lists are not the same, so although more items are being rehearsed in the phonologically similar condition with higher speech rates in older children, the error rate also increases proportionately. Therefore, the PSE is greater with faster rehearsal rates. This result shows that, contrary to the belief that STM processes can be entirely isolated from one another, the rehearsal process can affect performance beyond the principal causal factors which are known not to originate with rehearsal. It therefore seems that, because rehearsal involves the repeated activation of encoded material, rehearsal may amplify STM effects that originate during the initial encoding in the phonological loop. This is one example of a situation where the use of a rehearsal strategy has a detrimental effect on recall.

1.2.5 Neuropsychological Evidence

Additional evidence for rehearsal in the phonological loop comes from case studies of individuals with specific neuropsychological deficits which affect one or more theoretical component of the STM system. For example, evidence supporting the structure of the phonological loop system comes from case studies of phonological STM in individuals who have specific deficits which render them aphasic, that is, unable to produce overt speech. Aphasic patients with dyspraxia are unable to set up the speech codes necessary for articulation, a deficit in speech planning that would occur at the level of the phonological loop. In verbal STM tests, dyspraxic patients have reduced memory spans, they do not show the normal PSE or WLE, and they have more difficulty with written rhyme than with homophone judgments (Waters, Rochon, & Caplan, 1992). In short, dyspraxic patients show performance patterns similar to those found with normal subjects under AS. Because AS is thought to block rehearsal, the similar pattern found in dyspraxic patients supports the view that their speech-planning deficit occurs in the articulatory loop and prevents rehearsal. On the other hand, aphasic patients with dysarthria have secondary speech problems without the loss of language. In this case, the phonological loop model would predict that the verbal STM of these aphasic patients would remain intact because the deficit lies not at the level of the phonological loop, but rather at the later stage of peripheral motor functioning. Indeed, dysarthric patients are affected by phonological similarity and word length in STM tasks, indicating that they are capable of normal phonological storage and rehearsal in the articulatory loop (Baddeley & Wilson, 1985). The results of studies using aphasic patients show that the subvocal rehearsal process does not rely on the ability to operate peripheral musculature involved in speech production, nor is it reliant on feedback from motor activity. Instead, rehearsal appears to be equivalent to overt speech only up to the earlier planning stages of articulation.

Patients with neuropsychological impairments have also provided evidence that the WLE emerges as a consequence of rehearsal. An investigation by Jacquemot, Dupoux and Bachaoud-Levi (2011) was made in response to the alternative proposal that the reduction of the WLE under AS results from interference due to self-generated speech rather than from an inability to rehearse. The authors studied two patients with STM deficits, only one of whom also showed a pseudoword repetition deficit which provided an ecological case in which rehearsal was disrupted without AS. The patient with the pseudoword repetition deficit showed a PSE but not a WLE, a pattern that resembles the effects of AS on PSE and WLE in healthy subjects. The other patient with STM deficits displayed the typical the PSE and WLE, demonstrating that the pattern found in the rehearsal-disrupted patient could not be attributed to the general STM deficit. Based on these results, the authors concluded that the WLE emerges due to rehearsal and is abolished when rehearsal is disrupted.

Taken together, the empirical evidence for the effects of articulatory suppression, speech rate, word length, and phonological similarity on immediate serial recall are strongly suggestive of the articulation-based subvocal rehearsal process such as that described in the Working Memory model. Patterns of ISR performance in neuropsychological studies have supported the view that subvocal rehearsal relies on early articulatory planning systems, and not on articulatory-motor execution or feedback processes. Neuropsychological case studies have further confirmed the predictions of the phonological loop model by demonstrating that the effects of rehearsal-based (WLE, AS) and non-rehearsal based (PSE) factors dissociate as expected in patients with specific STM deficits. Two relevant conclusions are clear from these findings. The first is that the use of a subvocal rehearsal strategy, whether in ISR for auditory stimuli or re-coded visual stimuli, has a generally beneficial effect on performance compared to situations where rehearsal is prevented. Second, increases in the speed of subvocal rehearsal have a further benefit on ISR performance.

It should be noted that, while the Baddeley and Hitch Working Memory model uses a time-based decay mechanism to explain the benefits of rehearsal and the positive correlation between rehearsal speed and span, this relationship does not depend on the acceptance of the trace decay hypothesis. An alternative interpretation is that the speed of refreshing in the articulatory loop counteracts the detrimental effects of interference, as opposed to time-based decay. For the sake of this thesis, the precise source of forgetting in the phonological loop is not pertinent. Regardless of the mechanism of forgetting, these results present a convincing argument in favour of a reliable, beneficial effect of rehearsal speed in ISR performance.

1.3 Other Methods of Investigating Rehearsal

The conclusions regarding subvocal rehearsal discussed thus far are mostly based on experimental paradigms that rely on the phonological loop model, however there have been important findings obtained using other methods that provide a separate source of information about the nature of inner speech and rehearsal in STM. These methods, while reliant on various other assumptions, are less dependent upon the phonological loop model for interpretation. Like the experimental paradigms that are based on the phonological loop, some of these other methods are indirect, such as overt rehearsal, observations of speech-related activity, and self-report. Others are more direct, such as physiological, psychophysiological and neuroimaging measurements.

1.3.1 Overt Rehearsal

One very straightforward way to study the effects of rehearsal on STM is to have participants rehearse aloud. This allows researchers to control or monitor the exact time course of the rehearsal activity, and to avoid making assumptions about the occurrence, pace or contents of rehearsal. Investigations using the overt rehearsal method have shown that immediate and delayed free recall performance can be affected by manipulation of overt rehearsal factors (Fischler et al., 1970; Palmer & Ornstein, 1971; Rundus, 1971; Ward, Woodward, Stevens, & Stinson, 2003). Fischler, Rundus and Atkinson (1970) found that there was a reduced primacy effect during an overt rehearsal condition where only the current item was rehearsed during presentation, compared to an unconstrained overt rehearsal condition. The authors interpret this result as showing that the primacy effect is a function of the increased number of rehearsals for earlier items under typical silent rehearsal conditions. Support for this was also found by Rundus (1971), who showed that the recall advantage for repeated items in a list as a function of spacing could be explained by rehearsal frequency. Rundus demonstrated that the number of overt rehearsals of items increased when the item was repeated in a list, and that it also increased with the duration of the lag between successive presentations of the repeated items. The increase in number of rehearsals as a function of item presentation spacing was attributable to the increased delay between presentations, which allowed more time for rehearsal of the repeated item between presentations. Rundus also showed that the number of rehearsals for an item that was repeated immediately after its first occurrence was actually lower than that for the first presentation, which he suggested was due to decreased attention paid to the item when it is presented twice in succession. These results provide strong evidence in favour of the role of both rehearsal frequency and spacing in determining immediate recall probabilities.

Similar evidence comes from Palmer and Ornstein (1971), who showed that cumulative rehearsal improved immediate free recall relative to paired-associate rehearsal, and that this recall performance advantage is attributable to the increased primacy effect in the cumulative rehearsal condition. In cumulative rehearsal, the preceding items are rehearsed in forward serial order as they are presented (e.g. A, AB, ABC...), while in the paired-associate rehearsal, the current item is rehearsed with the immediately preceding item (e.g. A, AB, BC...). In the Palmer and Ornstein study, only the current item was overtly rehearsed in both conditions, and any preceding items were rehearsed silently. Recall performance was similar for the items at the end of the list, but there was a primacy effect for the beginning items (positions 1-4) in the cumulative rehearsal condition only. This enhanced primacy effect persisted in a delayed recall condition, but the effect was reduced relative to that in the immediate recall condition. Although both silent and overt rehearsal methods were used in the same conditions, the frequency of overt rehearsals was matched across the conditions which allowed the comparison of silent rehearsal factors. This study showed that silent cumulative rehearsal benefits recall for earlier items compared to paired-associate rehearsal, presumably because of the increased number of rehearsal for earlier items in the cumulative condition. This result is consistent with the view that rehearsal frequency is a critical determinant of recall performance in that greater rehearsal frequency will more effectively counteract the forces of forgetting (i.e. decay, interference) for a given sequence presentation and time interval.

There is also some evidence from overt rehearsal studies to support the notion that age-related increases in the use of subvocal rehearsal can explain corresponding developmental increases in STM span. Lehmann and Hasselhorn (2010, 2012) observed the overt rehearsal of 8 to 10 year olds in a free recall task and found that children tended to rehearse the items in their original serial order, or in groups of successive items from nearby serial positions. These authors also found that the serial order of rehearsal was predictive of the serial order at recall, replicating this finding from other studies (e.g. Rundus, 1971). In terms of the qualitative changes with development, the authors reported that older children tended to engage more in cumulative rehearsal, and that cumulative rehearsal was related to the successful recall for serial order. The authors concluded that rehearsal facilitates recall by strengthening item and order associations, and that this strategy improves with age. This result is consistent with other reports of age-related increases in a forward-order cumulative rehearsal strategy (Cuvo, 1975; Ornstein, Naus, & Liberty, 1975).

These findings reinforce the view that rehearsal factors such as number of repetitions and serial order are influential in verbal STM performance, including developmental changes and serial order errors. The developmental change in STM performance may reflect a strategy change, where the strategy for immediate free recall becomes more similar to that for ISR with increasing age. The finding that the order of item rehearsal relates to recall serial order is of particular importance to the present research, as this thesis will explore the relationships between subvocal rehearsal timing and ISR, where rehearsal timing variability is hypothesized to be a sensitive predictor of serial order errors at recall. However, these conclusions rest on the assumption that the effects of overt rehearsal on performance can be extended to covert rehearsal, a suggestion that will be critically evaluated in section 1.6.4.

1.3.2 Observation of Speech-Related Activity

Young children appear to have a greater tendency to audibly whisper or make inaudible lip movements when studying memory lists, a tendency that seems to diminish with age (Flavell, Beach, & Chinsky, 1966; Keeney, Cannizzo, & Flavell, 1967). Researchers have used overt signs of speech-related activity as a measure of rehearsal, and at least one study has shown that children who exhibit more spontaneous visible signs of speech activity during memory list presentation perform better at recall (Keeney et al., 1967). Like the results of studies using overt rehearsal, the possible relationship between overt signs of speech activity and STM performance has been used as evidence to support the view that rehearsal has a beneficial effect on recall in STM. These findings also support the assumption that subvocal rehearsal is closely related to overt speech, and that subvocal rehearsal develops when children learn to suppress speech-related motor activity.

1.3.3 Self-Report

To further investigate the issue of whether overt speech accurately represents inner speech, self-report techniques have been used to make comparisons between the representations of items in the two speech modes. Oppenheim and Dell (2008) investigated subvocal speech by comparing the patterns of errors in overt and covert speech. The authors had participants recite tongue-twisters overtly and covertly; overt errors were recorded by the experimenter and covert errors were self-reported by participants. The authors found a lexical bias effect in both overt and covert conditions, meaning that there was a higher tendency for errors to create words (e.g. “hit” to “sit”) than nonwords (e.g. “hit” to “rit”). However, they also found that, in the overt condition only, phonemic-level errors were more likely to involve similar sounding phonemes. Oppenheim and Dell (2010) also compared silent rehearsal to an articulated but soundless (i.e. ‘mouthing’) condition, and again found that the lexical bias effect occurred in both conditions but the phonemic similarity effect only occurred in the articulated condition. Based on these results, these authors argue that verbal representations in covert speech are impoverished at the subphonemic (feature) level, and that subvocal speech is more flexible and abstract than articulated speech (Oppenheim & Dell, 2010; Oppenheim, 2012). This result is particularly interesting as it is one of only a few behavioural findings which suggest that overt and covert speech may be qualitatively different during the earlier speech planning stages, as opposed to diverging during the motor execution stage as previously thought.

However, the evidence from self-report studies on rehearsal is not consistent. Corley, Brocklehurst and Moat (2011) responded to the Oppenheim and Dell (2008) study with two main criticisms. First, the authors point out that Oppenheim and Dell's approach rests on the assumption that participants can attend to their own inner speech, and can accurately detect and report errors. Corley and colleagues argue that the reason for a decreased phonemic similarity effect for self-reported covert speech may have occurred because people are better at perceiving feature-level errors in the presence of acoustic feedback. Second, these authors note that the effects of lexicality and phonemic similarity may be related to the adaptive editing of speech plans rather than to activation feedback, as Oppenheim and Dell propose. Corley, Brocklehurst and Moat used an auditory masking technique to test whether the reduction in self-reported phonemic similarity errors were due to the absence of auditory feedback. No effect of masking was found, supporting Oppenheim and Dell's assumption that there is no difference in error detection abilities between overt and covert speech. However, the authors failed to replicate the main findings by Oppenheim and Dell (2008, 2010); Corley et al. reported a phonemic similarity effect in both the overt and covert speech conditions.

The results of research using self-report of errors in covert speech appear to remain open to interpretation. In the context of the current review, this research brings up two important considerations. The first is that the representations of verbal material during subvocal rehearsal may be qualitatively different from those during articulated speech (Oppenheim & Dell, 2008, 2010; Oppenheim, 2012). Specifically, verbal representations in covert speech may be impoverished at the feature level. A second consideration is that, as Corley, Brocklehurst and Moat (2011) point out, there may be problems with the assumption that people can attend to and perceive errors in their own subvocal speech. These authors claim that error detection during covert speech may differ as a result of the lack of acoustic feedback or the adaptive and continuous editing of speech plans. Thus, acoustic feedback from self-generated speech and the inability to dynamically edit vocalized speech content may mean that overt rehearsal is qualitatively different from subvocal rehearsal, at least in certain aspects. This presents a consideration not only for the empirical findings based on an overt rehearsal method, but also for interpretations that assume a qualitative equivalence between covert and overt speech, such as that regarding speech rates and the WLE. Finally, the conflicting results of these studies highlight the need for further research, ideally using an objective measure of inner speech that does not rely on self-report.

1.3.4 Physiological Measures of Speech-Motor Activity

The need for a quantifiable and objective measure of inner speech prompted the investigation into physiological measures of speech-motor activity as possible markers of this internal activity. Patterns of electrical potentials generated by muscle activity in the lips, chin, jaw and throat have been used to detect the presence of subvocal speech, and to discriminate the subvocalisation of different words (Betts & Jorgensen, 2006; Garrity & Donoghue, 1977; Garrity, 1975; Jorgensen & Binsted, 2005; Jorgensen & Dusan, 2010; Locke & Fehr, 1970a, 1970b; Meltzner et al., 2008). The detection of subvocal speech using electrical signals from muscle activity, or electromyography (EMG), relates to the assumption that subvocal rehearsal is closely related to overt speech, and it exploits the notion that peripheral articulatory musculature can respond to articulatory commands without the production of speech. The EMG method differs from vocalised and whispering speech conditions in that EMG does not require the production of audible speech, thus eliminating confounding issues such as auditory feedback. Also, EMG is sensitive to very small movements, and thus differs from an inaudible articulatory (i.e. mouthing) condition in that, in theory, it is more similar to typical overt speech in terms of the degree of motor execution required. Finally, a key advantage of EMG is that it is quantifiable, allowing the analysis of both the continuous and categorical features of muscular activity.

In practice, EMG has been shown to be capable of making broad distinctions among dissimilar subvocal speech patterns. Locke and Fehr (1970a) differentiated the rehearsal of visually-presented disyllabic words containing labial or nonlabial phonemes, with the EMG recordings for labial words showing a significantly greater mean peak amplitude relative to non-labial words. These differences were observed during both the stimuli presentation and rehearsal intervals, suggesting that subvocalisation co-occurs with the sequence encoding. The authors used these results to support the hypothesis that subvocal rehearsal is a form of speech, and thus is of theoretical and practical significance for verbal learning, reading and related topics. Locke and Fehr (1970b) also used this technique with 4 and 5 year old children in a picture recall task, and reported significant differences in EMG in for trials with familiar pictures containing labial or nonlabial phonemes. Interestingly, the EMG differences between labial and nonlabial trials were observed during the presentation phase but not during the rehearsal interval. This suggests that the children were covertly articulating the picture names during their presentation, but not continuing to do so during the delay period, which is consistent with other evidence suggesting that young children do not rehearse until after about age 6.

However, there is conflicting evidence from EMG studies regarding spontaneous subvocal rehearsal in young children. In contrast to Locke and Fehr's (1970b) results, studies by Garrity (1975) and Garrity and Donoghue (1977) detected EMG differences for labial and nonlabial material during both the presentation and delay periods in 4 and 5 year old children. These authors also found that the magnitude of the difference between labial and nonlabial EMG activity during the delay period was positively related to both age and IQ, suggesting that only the older children and children with higher IQ scores tended to subvocally rehearse during the delay periods. Furthermore, the magnitude of the EMG difference was also related to overall recall performance, but not to any serial position effects such as primacy or recency. These results provide evidence that, not only do young children subvocally rehearse, but that 1) rehearsing has a beneficial effect on recall in young children, and 2) some young children rehearse spontaneously during delay periods in addition to rehearsing during item presentation. The absence of serial position effects suggests that the children may have not have been engaging in cumulative rehearsal. Garrity (1977) suggests the discrepancy between these results and those of Locke and Fehr (1970b) is due to differences in the EMG signal scoring. Specifically, Locke and Fehr relied on a simple maximum amplitude measure, which may not have been sensitive to lower-amplitude or frequency related differences, and which may have been overly sensitive to non-articulatory movement artefacts.

However, an important question that remains is whether or not the speech-motor responses are an obligatory effect of subvocal speech, both in children and in adults. Although participants were not given explicit instructions to produce articulatory motor gestures during these experiments, the presence of the electrodes on the lip and chin may have alerted them to the purpose of the experiment and perhaps caused a change in their subvocal rehearsal behaviour. To combat this possibility, Locke and Fehr (1970a, 1970b) used dummy recording leads on the body to distract the participants' attention from the lip-chin EMG site. Furthermore, the authors noted that while the young children's subvocal speech was inaudible, their EMG recordings during subvocal speech were nearly as great in amplitude as those during vocalised speech. Given that young children may tend to shift from early voiced articulation to later unvoiced articulation and finally covert speech during memory tasks, this finding suggests that kinaesthetic involvement during subvocal speech lies on a continuum. This issue leaves open the possibility that EMG correlates of subvocal speech are optional and effortful, rather than automatic and ubiquitous responses, particularly in older children and adults who have mastered the internalisation of their speech.

Recent technological advances allow for more sophisticated and sensitive analyses of EMG activity, which has prompted a re-investigation into the potential for detecting the contents of inner speech using EMG. A NASA research group investigated the use of EMG subvocal speech decoding for the purposes of recognizing or synthesizing speech commands, particularly for use in situations in which the acoustic signal is likely to be distorted by environmental noise or physical obstruction. Multiple studies by this group report the training of classifier algorithms on individuals' EMG responses during subvocalisations of limited vocabulary sets, and then testing the accuracy of these algorithms on EMG recordings after training (Betts & Jorgensen, 2006; Jorgensen & Dusan, 2010; Jorgensen & Binsted, 2005). Because the computational classifier modelling of these subvocalised EMG signals requires a high number of training trials, the researchers limited the training set to a small number of short and dissimilar-sounding words and letters. The highest classification accuracy rates in these experiments ranged from about 74 to 92% for sets of 5 to 15 words, leading the authors to conclude that content classification of subvocal speech using EMG is possible with sufficient classifier training and a small vocabulary set.

The possibility of decoding subvocal speech from EMG signals is revealing, as it shows that rehearsal of different sounds may produce small but reliable physiological differences which can be recognised using sensitive physiological measures. This type of measure would present an opportunity to observe the actual time course and contents of rehearsal, rather than retrospectively infer its occurrence due to the presence or absence of an effect (e.g. PSE, WLE). However, the results in this area have been inconsistent, and a study by Meltzner et al. (2008) provides convincing evidence that articulatory-motor activity is not an obligatory artefact of subvocal speech. While Meltzner and colleagues found that high classification accuracy (up to ~87-97%) was achievable using EMG responses during a voiceless articulation condition, they also reported that EMG responses during a mental rehearsal condition could not be analysed due to a consistent lack of discernable EMG activity. These results are in surprising contrast to those obtained by Jorgensen and colleagues, suggesting that the detection of subvocal speech via EMG is only possible when participants make a conscious effort to produce articulatory gestures. This again suggests that participants in previous studies reporting positive results were perhaps aware of the purpose of the experiments and may have been motivated to exaggerate their articulatory muscle activity. Thus investigations into subvocal speech using EMG have not provided a conclusive answer to the questions about the nature of the similarities between articulated and unarticulated voiceless speech.

1.3.5 Psychophysiological and Neuroimaging Methods

As mentioned previously, one of the downsides of using EMG for detecting rehearsal is that it relates to nonessential rather than obligatory processes. An obvious solution to this problem is to attempt to measure subvocal speech as it occurs in the brain. Such attempts have been made by researchers using EEG (Hwang et al., 2005), MEG (Curio, Neuloh, Numminen, Jousmäki, & Hari, 2000; Kauramäki et al., 2010; Numminen & Curio, 1999; Tian & Poeppel, 2013; Ylinen et al., 2014), PET (Mcguire, Murray, David, Frackowiak, & Frith, 1996) and fMRI methods (Baciu, Rubin, Décorps, & Segebarth, 1999; Rauschecker, Pringle, & Watkins, 2008; Shergill et al., 2002; Shergill, Tracy, Seal, Rubia, & McGuire, 2006). One of the most consistent findings regarding the brain basis of inner speech comes from electromagnetic (EEG, MEG) studies showing modulation of the temporal lobe in response to auditory stimuli during covert speech. Internal forward models propose that the sensory consequences of motor actions are predicted via internal simulations or 'efference copies' (Sperry, 1950; von Holst & Mittelstaedt, 1950). In the case of normal speech, planned articulation results in a forward prediction of the acoustic consequences, which changes the activity of auditory cortex. More precisely, the responses to auditory stimuli are attenuated in auditory cortex during subvocal speech. Numminen and Curio (1999) demonstrated this in an experiment where participants subvocalised vowel sounds under magnetoencephalograpy (MEG) recording, and during the subvocalisation period the experimenters presented recordings of the participant's overt vocalizations of these sounds. On some trials the replayed sounds matched the contents of the participants' subvocalised speech, and on other trials the replayed sounds did not match. The authors reported a greater M100 response on trials in which the recording of participants' overt speech did not match the contents of the participants' internal speech. Although this experiment lacked a memory component, it demonstrated that the temporal cortex is primed for the contents of subvocal speech, and so the contents of rehearsal may be accessed through these more direct means.

Functional magnetic resonance imaging (fMRI) has high spatial resolution and thus is specifically useful for determining the functional anatomy of cognitive processes. Although the question of where subvocal speech occurs in the brain may seem lateral to the purposes of the current research, evidence from fMRI studies has supported assumptions about subvocal speech as well as evidence from studies using other methods. Baciu et al. (1999) computed lateralisation indices for overt and covert speech, and found that these indices were highly correlated, indicating left hemispheric dominance for both vocal and subvocal speech. This result supports the notion that subvocal speech is similar to vocal speech, and that many of the neurological processes known to be involved in speech production may also occur, at least to some degree, in the production of covert speech. Shergill and colleagues (2002) varied the rate of inner speech under fMRI and found that rate was significantly associated with the magnitude of temporal cortical activation. The authors argue that these results are consistent with the results reported by Numminen and Curio (1999), as an increased hemodynamic response in the temporal cortex with faster rehearsal rates would be expected if modulation of the areas involved in auditory perception (located in the temporal cortex) occurred in response to internally-generated speech. Again, while these studies do not address inner speech in the context of STM tasks, they do offer insight into what assumptions might be made about rehearsal in STM and how the rehearsal process may be measured more directly.

1.4 Serial Order and Timing in Auditory-Verbal Short-Term Memory

There is a limit to the amount of information that an individual can immediately recall in the correct serial order. STM span tasks were developed as a method of quantifying the limits of this ability, which is thought to reflect a theoretical STM capacity. The STM span task was first used in experimental work by Jacobs (1887) and Galton (1887), who immediately highlighted individual and developmental differences in the immediate serial recall task, as well as a relationship between STM span and academic rank. In an early review, Blankenship defined memory span as

> the ability of an individual to reproduce immediately, after one presentation, a series of discrete stimuli in their original order. Practically any sort of material may be presented, such as digits, letters, words, and sounds, and almost any sense organ or combination of sense organs may be used to receive the impressions. (Blankenship, 1938 p. 2)

As Blankenship notes, the span task paradigm is flexible, and it has been used extensively under different modalities and experimental conditions. Here the focus is on auditory-verbal STM span; the limit of immediate forward serial recall for lists of spoken words.

Auditory-verbal ISR tasks, such as digit span, have become a fundamental tool in research and clinical settings. The widespread use of span tasks to assess STM is likely because the tests are simple to administer and performance is closely linked to outcomes of interest. For instance, auditory-verbal STM span has been shown to relate to phonological processing (Brady, 1986), which is known to be a crucial precursor to literacy (Snowling, 2000; Wagner & Torgesen, 1987). Verbal STM has also been shown to predict a variety of more complex cognitive skills, such as word learning (Service, Maury, & Luotoniemi, 2007), verbal IQ

(Cantor, Engle, & Hamilton, 1991), and reading abilities (Anvari, Trainor, Woodside, & Levy, 2002).

While the correlation between auditory-verbal STM span and other cognitive measures does not necessarily indicate a causal relationship, many researchers would agree that auditory-verbal STM capacity is likely to be one of many factors that determine an individual's performance in cognitive tasks which rely upon, or are aided by, STM for verbal material. The main reason for this assumption is that STM span for spoken words appears to reflect a higher degree of bottom-up (i.e. stimulus-driven, attention independent) processing, and thus might reflect a low-level bottleneck that constrains the upper limits of high-level cognitive skills. The notion that auditory-verbal STM span reflects a relatively large contribution from bottom-up processes than other is supported by the fact that digit span performance becomes stable after middle childhood and, compared to memory tasks that are thought to rely on central executive function, does not benefit as much from training (Holmes, Gathercole, & Dunning, 2009). Unlike working memory (WM) and complex span tasks, which contain a combination of STM, processing and task switching elements, the simpler STM span task is assumed to provide a less contaminated measure of domain-specific STM capacity. In a comparison of the relative contributions of visual and verbal STM and WM tasks to reasoning abilities, Kane et al. (2004) reported that verbal and visuospatial WM tasks largely reflected a single domain-general factor that was strongly related to general fluid intelligence but more weakly predictive of domain-specific reasoning. Kane and colleagues found that verbal and visuospatial STM span tasks were stronger predictors of verbal and visuospatial reasoning, respectively, than were WM tasks of the same modalities. The absence of correlations between auditory and visual STM span tasks is itself evidence that the simple span paradigm is a relatively pure measure of domain-specific capacity, because performance would be expected to correlate across modalities to the extent that the constructs rely on domain-general factors. Similarly, STM span performance is not highly correlated with tasks that tap executive function and processing speed, which suggests that these factors do not play a determining role in span performance.

Individual and group differences in STM capacity seem to persist after controlling for top-down, attention-dependent processing, which supports the notion that a portion of the variance in STM capacity is attributable to bottom-up, stimulus driven factors. STM span performance generally benefits from circumstances that provide the opportunity for subvocal rehearsal, but span performance is not disastrously impaired and individual differences persist when the opportunity to rehearse is limited (Cowan, Nugent, Elliott, Ponomarev, & Saults, 1999; Lyon, 1977). For instance, Cowan et al. (1999) reported that, when attention was directed away from incoming auditory digit stimuli with a concurrent visual task, there are still observable individual[2] and age-related differences in the limits of STM capacity. Based on this finding, Cowan et al. concluded that there is a central capacity limit, and this capacity limit is supplemented by attention-dependent mnemonic strategies during typical STM span tasks. Similarly, experiments using fast presentation rates and AS to reduce the opportunity for rehearsal have shown that performance suffers but the effect is not catastrophic. This provides further support to the conclusions of Dempster (1981), Tehan and Lalor (2000), Lyon (1977) and others that the use of rehearsal and other mnemonic strategies cannot fully account for individual or developmental differences in auditory-verbal STM span performance. This relates to the points made in the section about the role of rehearsal in STM, where the consensus seems to be that a subvocal rehearsal strategy aids STM in many instances, but is not required.

One explanation for the persistence of differences in auditory-verbal STM span is that these are due in part to differences in the bottom-up processing of the temporal structure of auditory stimuli, and that precision in the perception and maintenance of the temporal structure supports STM for serial order. This hypothesis depends on empirical support for a dissociation between STM for item and serial order information, where serial order STM abilities are disproportionately tapped by familiar word span tasks. In the next section, I review the evidence that auditory-verbal serial order memory is separable from phonological (item) information, and that the serial order component is critical for STM span performance. Following that, I will discuss the stimulus-driven processing of temporal structure in auditory stimuli, and argue for its relevance to auditory-verbal ISR tasks.

1.4.1 Dissociations between Short-Term Memory for Item and Order Information

Many STM models posit that auditory-verbal STM involves at least a partial distinction between the encoding of item and order information (Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Gupta, 2003; Henson, 1998). The independence of STM for item and order information was proposed after experiments revealed that recall patterns differ depending on whether the task isolates STM for either item or order information (Bjork & Healy, 1974; Healy, 1974). For instance, memory for item and order information is differentially affected by phonological and semantic similarity, lexicality, word frequency and AS (Lian, Karlse, & Eriksen, 2004; Nairne & Kelley, 2004; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999a, 1999b). These factors are known to affect performance on STM tasks using standard scoring techniques. Phonological similarity of words has a detrimental effect on memory for

---

[2] Cowan and colleagues (1999) do not discuss individual differences in this report, but the figures listed in Table 2 (p. 1088) clearly show that the individual differences within age groups, indicated by the standard deviations of the mean number of items correctly recalled, are not reduced or eliminated in the unattended condition. In fact, the standard deviations increase from the attended to unattended conditions within all age groups.

serial order, resulting in increased transpositional order errors (Fallon, Groves, & Tehan, 1999; Gathercole, Pickering, Hall, & Peaker, 2001). However, when manipulated within recognition tasks which emphasize memory for item information and minimize the need for order retention, phonological similarity has a beneficial effect on performance (Fallon et al., 1999).

LTM also has a separable influence on item and order STM. In word span tasks, LTM influences the short-term representations of familiar words at the phonological, lexical and/or semantic levels, and is capable of contributing to memory for serial order if the item-order associations are familiar (e.g. recalling days of the week in order, reciting digits in ascending numerical order). In nonword repetition tasks, LTM can influence memory for item information as evidenced by superior performance associated with familiar consonant-vowel structures and 'wordlikeness' of the nonwords (Gathercole, Willis, Emslie, & Baddeley, 1991; Gathercole, 1995; Thorn & Gathercole, 1999; Treiman, Goswami, & Bruck, 1990). Thus, the relative emphasis on either item or order retention in a verbal STM task will interact with LTM factors, as LTM mainly aids in recall for item information. The contribution of LTM to item information is further revealed by investigations of item and order memory in serial recall and serial recognition tasks. One such study has shown that the advantage for words over nonwords (i.e. the lexicality effect) is diminished in recognition tasks, presumably because the reconstruction of item information from LTM occurs only during recall (Gathercole et al., 2001). Consequently, item STM is disproportionately measured over order STM in recall versus recognition tasks, and STM measures of item information are confounded by contributions from LTM during recall for familiar material. Therefore, the independent (and sometimes even opposing) effects of multiple factors on STM for item and order information may help to explain the seemingly contradictory evidence in support of several competing theoretical processes thought to play causal roles in auditory-verbal STM performance. Furthermore, the decomposition of auditory-verbal STM into separate item and order STM capacities may reveal that the relationships observed between verbal STM performance and other measures is specific to either item or order STM capacity.

The traditional tests of auditory-verbal STM capacity, digit/word span and nonword repetition, each require the coding and maintenance of both item and order information. Word span tasks require the storage of phonological, lexical and semantic content of the words, which can be considered item information, as well as the order in which the words are presented, which is the serial order information (Majerus, Poncelet, Elsen, & van der Linden, 2006). However, ISR tasks using an arbitrary sequence of familiar words from a single category, such as digit span, are more likely to tap into serial order STM than item STM. This is because the item information is strongly supported by over-learned phonological representations and a highly constrained lexical subset in LTM. That is, the phonological and lexical information (the digits from zero to nine) are readily available, so the difficult aspect of the task is mainly attributable to the arbitrary ordering. That digit span is mainly an indicator of STM for serial order is supported by the fact that most errors that occur at the limit of digit span are serial order errors as opposed to item errors (Aaronson, 1968).

Like the word span task, nonword repetition requires both item and serial order memory, but in this case, item information is likely to be only sublexical as there are no existing word form representations or semantic information associated with novel words. Indeed, some consider nonword repetition to be mainly a measure of phonological memory (Gathercole, Willis, Baddeley, & Emslie, 1994). What is often neglected in interpreting the nonword repetition task is the role of serial order memory. In the case of nonword repetition tasks, serial order memory is needed for the maintenance of a novel sequence of phonemes. However, in the case of pseudowords and 'wordlike' non-words, the ordering of phonemes is constrained by linguistic knowledge from LTM, such as language-specific phonotactic rules (Majerus, Linden, Mulder, Meulemans, & Peters, 2004). Therefore, nonword repetition tasks are likely to load on serial order STM, but errors in nonword repetition tasks are likely to be constrained by knowledge of the conditional probabilities of sublexical phonemic sequencing.

Because memory for item and order information have been so often confounded in these two standard verbal STM tasks, it is difficult determine whether the correlations that have been previously established between verbal STM and other skills can be more precisely attributed to STM for either item or order. Furthermore, the conflicting evidence and ongoing debates regarding the relative importance of phonological skills, rehearsal and other factors to verbal STM performance may be resolved by subtleties in the methods for measuring verbal STM. If the specific verbal STM task and scoring method disproportionately emphasizes the retention of either item or order information, then the results of the experiment are likely to reflect relationships with one of these two separable cognitive processes. By making an effort to address STM processes for item and order separately, researchers can begin to narrow down the specific determinants of factors that have found to relate to verbal STM such as reading skills and vocabulary acquisition.

Given the dissociation between STM for item and order information, as well as the evidence that different task designs seem to disproportionately tap into either item or order STM, one relevant question is whether associations between verbal STM and language abilities can be best explained by either item or order components of auditory-verbal STM. Verbal STM capacity, as measured by word span or nonword repetition tasks, is strongly associated with vocabulary knowledge and new word learning in both children and adults

(Baddeley, Gathercole, & Papagno, 1998; Gathercole, Hitch, Service, & Martin, 1997; Gathercole, Service, Hitch, Adams, & Martin, 1999; Gathercole, Willis, & Baddeley, 1991; Gupta, 2003; Jarrold, Thorn, & Stephens, 2009; Papagno & Vallar, 1995). Recently, several investigations have sought to clarify these relationships using STM measurements which maximize retention of either item or order memory. Majerus and colleagues have shown that STM for order information in particular is a unique predictor of language-related outcomes previously linked to verbal STM span; namely, vocabulary development in children (Majerus, Poncelet, Greffe, & Van der Linden, 2006) and lexical learning in adults (Majerus, Poncelet, Van der Linden, & Weekes, 2008). The findings from these and other investigations point to the conclusion that order STM plays a critical and causal role in vocabulary development and new word learning (Leclercq & Majerus, 2010; Majerus, Poncelet, Elsen, et al., 2006). Consistent with empirical evidence for the link between serial order STM and new word learning abilities, some models propose a STM system dedicated for the maintenance of order which plays a specific and crucial role in vocabulary acquisition (Burgess & Hitch, 1999; Gupta, 2003; Hartley & Houghton, 1996). It therefore seems possible that many previously-reported associations between performance on verbal ISR tasks, such as digit span, and higher-level language skills are largely due to individual differences in the serial ordering system.

1.4.2 Bottom-Up Temporal Processing and Serial Order in Auditory-Verbal Short-Term Memory

STM for order is no doubt influenced by both top-down (intentional/strategic, attention-dependent) and bottom-up (automatic, stimulus-driven) processes. Previous research has addressed aspects of STM for order that are driven by top-down control (e.g. strategy use), but the bottom-up aspects of serial order for speech have been relatively neglected in the context of auditory-verbal STM span tasks. In particular, the temporal features of auditory sequences, such as between-item duration and hierarchical metrical structure, may have a bottom-up effect on the encoding and maintenance of serial order in auditory-verbal sequences. Serial order necessarily relies on timing information to some degree, as the ordering of events requires coarse knowledge of relative timing. This is especially true in the auditory-verbal domain, where the speech stimuli have intrinsic and complex hierarchical temporal properties.

The relative importance of processing across different time scales in serial order recall is likely to depend on the context. For example, the ability to perceive a phoneme-level timing structure would be expected to be more important for novel words, where the phonological sequence information cannot be reconstructed from LTM, or for phonologically-similar words in the absence of semantic cues. In the context of highly familiar, phonologically-dissimilar words from a small categorical set, the critical timing for entrainment is that of the perceived 'beat' or centre of the word itself (i.e. the perceptual center or "p-centre", Morton, Marcus, & Frankish, 1976). This is likely the case for sequences of digits, where the sublexical features (e.g. syllables, phonemes) are readily available from long-term memory. That word-level timing is critical for accurate serial ordering in the digit span task is supported by the finding that the most common errors are transpositions of neighbouring words (Bjork & Healy, 1974; Healy, 1974). This is notable given all other possible types of incorrect recall patterns, including myriad combinations of non-adjacent items. The relative frequency of transposition errors in digit span can be readily explained by a bottom-up timing mechanism with variable accuracy that directly impacts the serial position coding.

In order for a bottom-up timing variability hypothesis to be capable of explaining a portion of variance in familiar word ISR scores, there are a few necessary (but not sufficient) empirical points that must be true. First, individuals must differ in their ability to accurately perceive and maintain the temporal structure of auditory sequences. Second, these individual differences must be associated with the same cognitive and language measures as the span task (to the extent that the relationship between STM span and these abilities are mediated by a common temporal regulatory mechanism). Third, the presentation timing of auditory-verbal stimuli in ISR tasks should have an automatic (attention-independent) effect on performance. That is, auditory timing should affect serial order recall even when there is no explicit instruction to maintain and recall the more precise temporal information.

With regard to the first and second points, individuals have indeed been shown to vary in the ability to track and reproduce simple auditory rhythms, and there is emerging evidence that this variation has associations with relevant cognitive and language measures. Temporal variability has typically been tested using simple motor-timing tasks, where participants are asked to synchronize their finger-tapping to an auditory rhythm and then continue tapping the rhythm in the absence of the external stimuli. Using this method, researchers have developed scoring formulas and models to distinguish between multiple sources of motor-timing errors that become integrated in motor-timing tasks (Wing & Kristofferson, 1973a). When separated from the overall pace of tapping and absolute timing accuracy, the inter-tap (or 'local') variability of finger-tapping has been found to relate to both intelligence and language-related skills. In a summary of the studies on the relationship between isochronous serial interval production (i.e. tapping to a steady beat) and intelligence, Ullén and colleagues (2012) reported consistent results across three studies and five samples; there was a significant negative correlation between the local variability measure and non-verbal intelligence. Furthermore, Ullén et al. manipulated the participants' motivation using monetary rewards and found that, while the high motivation condition resulted in improved performance overall, the motivation manipulation did not affect the correlation between tapping variability and intelligence scores. This finding suggests that although this timing measure can be influenced by top-down control, there is a separable bottom-up mechanism that contributes to variability in motor synchronisation tasks, and it is this aspect of the task that relates to intelligence.

In terms of the established relationships between auditory-verbal STM span and language abilities, measures of auditory timing and rhythm perception have been shown to also relate to phonological skills and auditory-verbal serial order STM. Tierney and Kraus (2013a) found that the standard-deviation of inter-tap intervals relates to reading ability in typically-developing adolescents. In studies of children and adults with and without developmental language disorders, finger-tapping variability has been shown to relate to individual differences in digit span, phonological skills, and reading and spelling abilities after controlling for IQ (Thomson, Fryer, Maltby, & Goswami, 2006; Thomson & Goswami, 2008). One key limitation of these motor synchronisation studies, however, is that they cannot separate internal representations of timing and rhythm from the precision of motor timing and sensorimotor feedback mechanisms that are also involved in the task.

Support for the notion that these relationships are attributable to variability in temporal processing of auditory sequences rather than motor timing variability comes from auditory timing and rhythm change detection tasks. In these tasks, the discrepancy between target and reference sequences of tones is adaptively adjusted to test the perceptual thresholds for auditory rhythm and timing discrimination. Grube, Cooper and Griffiths (2013) found that thresholds for detecting regularity and deviations in metrical patterns was related to phonological skills tasks such as irregular word reading, nonword reading and poem reading[3]. It is therefore evident that individual differences exist in bottom-up timing variability as measured by motor synchronisation and auditory change detection tasks. Further, timing variability measures relate to ISR performance directly as well as to the high-level cognitive and language-based skills that have been previously shown to associate with auditory-verbal STM span scores. While these results do not necessarily indicate a specific, causal relationship between variability in a bottom-up temporal processing mechanism and serial order STM, they are consistent with such a hypothesis.

Perhaps the most compelling support for the notion that bottom-up temporal variability is a determining factor in auditory-verbal STM span comes from a study that aimed to

---

[3] These two rhythmic sequence processing tasks were also correlated with digit span scores in the expected direction, however the significance of the correlations did not survive Bonferroni correction for multiple comparisons.

directly test individual differences in these abilities. Saito (2001) showed that individual differences in rhythmic tapping variability was related to digit span in typical adults, and that this relationship was not mediated by articulation speed, reading speed, or a mental rotation task. To assess each individual's ability to maintain the relative timing of a sequence of beats, Saito calculated the number of tapped sequences in which the deviation from the timing of inter-tap intervals relative to the total duration of the tapped sequence was below a threshold of 15%. Simple regression analysis revealed that the rhythmic tapping measure accounted for unique variance in both auditory and visual digit span scores even after controlling for the contributions to variance from other-modality digit span, reading speed, and the mental rotation task. The fact that individual differences in articulation and reading speeds did not contribute to the relationship between digit span and memory for rhythm suggests that the relationship between span and timing cannot be explained by articulatory control or attention-dependent processes. As discussed in previous sections, the use of a subvocal rehearsal strategy and other attention-dependent factors are likely to account for some of the variance in span scores, but individual and group differences in span performance persist after considering these variables. Based on the results of Saito (2001), it seems that a timing control process that affects the precision of rhythm memory has potential for explaining some of the variance in STM span performance.

Regarding the third empirical point, that auditory timing should be automatically encoded and affect ISR performance, the evidence here is more varied and open to interpretation. The temporal grouping of items is known to affect ISR; temporal groups can have beneficial or detrimental effects on performance, depending on the grouping pattern. Regular temporal grouping patterns both in sequence presentation (Henson, Hartley, Burgess, Hitch, & Flude, 2003; Hitch, Burgess, Towse, & Culpin, 1996; Ryan, 1969a) and via rehearsal instructions (Wickelgren, 1964, 1967) are known to benefit serial order recall. More precisely, temporal grouping improves recall for middle-list items through primacy and recency effects in each subgroup. By contrast, the use of irregular auditory temporal grouping patterns has been shown to be detrimental to recall (Ryan, 1969a). Temporally grouping items in a regular pattern produces a specific benefit for serial order memory over item memory (Henson et al., 2003; Ryan, 1969a), and changes the pattern of order errors whereby the middle items in each subgroup are more likely to be transposed (Ryan, 1969b), despite the fact that they are in non-adjacent serial positions and thus unlikely to be transposed in ungrouped conditions.

The view that auditory temporal grouping effects are not entirely due to intentional or attention-dependent processes is supported by the rather surprising finding that these effects are not dependent upon rehearsal or the phonological store. The advantage for regular temporal groups occurs during articulatory suppression with auditory but not visual presentation, and this effect is not sensitive to phonological similarity or word length (Hitch et al., 1996). In tasks measuring the variability in the reproduction of irregular temporal sequence structures, the variability is equivalent for items presented in auditory and visual modalities, except under articulatory suppression, in which case there is less variability in reproduction of auditory timing relative to visual timing (Watkins, LeCompte, Elliott, & Fish, 1992). One interpretation of this result is that auditory sequence timing is automatically encoded and processed separately from the phonological store and articulatory loop, whereas visual sequence timing is perhaps recoded into phonological form or otherwise supported by subvocal speech. This seems plausible given that the auditory system is more sensitive to temporal information.

The results of the studies discussed thus far are consistent with the view that bottom-up temporal factors can account for some of the variance in auditory-verbal serial order recall, and that temporal variability in particular may be predictive of poorer auditory-verbal span and phonological skills (Grube et al., 2013; Henson et al., 2003; Hitch et al., 1996; Saito, 2001; Tierney & Kraus, 2013; Ullén et al., 2012). As Saito (2001) explains, one explanation for the relationship between memory for rhythm and digit span, given the absence of a relationship between rhythm memory and articulatory and reading speeds, is that there is timing control system independent of phonological store and articulatory loop. While the proposed timing system is independent of these phonological loop components, it is necessarily engaged during paced subvocal rehearsal in the articulatory loop. AS has been shown to disrupt memory for rhythms (Saito & Ishio, 1998), and this disruption may be due to the involvement of a timing regulation system rather than due to the articulatory loop per se. That the timing control system is separable from the articulatory loop is supported by evidence that the phonological similarity effect, which is normally eliminated during intermittent AS (e.g. "ah, ah, ah"), is not eliminated when the concurrent articulation is continuous (e.g. "ahhhh...") (Saito, 1997). Thus, the effects of AS on memory for timing information may be due to the regulation of concurrent articulation rate via a timing control system.

The proposal of a mechanism involved in the processing of temporal structure that is also required for the maintenance of serial order in STM is supported by selective interference between tasks. Timing estimation experiments show that concurrent STM tasks have an adverse affect on estimation of short (2-3 s) intervals, while equally-difficult and attentionally-demanding tasks without a STM component do not affect timing estimation (Fortin & Massé, 1999; Fortin, Rousseau, Bourque, & Kirouac, 1993). One of these experiments further showed that time estimation error varies as a function of concurrent memory load, for example when processing of order information is required in addition to item information (Fortin & Massé, 1999). Similarly, Henson et al. (2003) found that paced tapping during ISR tasks had a greater detrimental effect on the order recall than on item recall. In addition to the impairment on STM recall, these authors reported an increase in tapping variability from baseline (tapping only) to dual-task (tapping plus memory) conditions. That paced tapping and time estimation tasks are affected by concurrent STM load, and specifically by serial order STM processing, is consistent with the view that a timing control mechanism is required both for serial order recall and temporal control and processing.

Despite these points, there is ongoing debate surrounding the nature of the serial order coding system in STM. This is due in part to disagreement as to whether the sequential order of list items are likely to be encoded in terms of a positional/ordinal system as opposed to a time- or oscillation-based mechanism. One source of the lack of consensus among models of serial order is the conflicting evidence regarding the impact of item timing on recall. As discussed earlier in this section, temporal grouping reliably affects ISR performance, but there is evidence that grouping effects result from item distinctiveness and hierarchical factors in a positional coding system rather than precise temporal coding. For instance, the input temporal structure of an auditory sequence tends to be reflected in recall timing structure via increased latencies between subgroups, but the recall timing does not reflect the temporal ratios of within- and between-subgroup timing as would be expected if STM for serial order were driven in large part by automatic encoding of the temporal structure of the stimuli (Maybery, Parmentier, & Jones, 2002). Also, non-temporal grouping cues such as pitch have shown to have similar effects on ISR performance (Frankish, 1995; Parmentier & Maybery, 2008), which is not explained solely by a bottom-up timing hypothesis for encoding auditory-verbal serial order. Issues related to the positional versus temporal nature of encoding in serial order STM are discussed in more depth in Chapter 8.

To conclude, variability in bottom-up temporal processing may play an important in auditory-verbal STM for serial order. While the exact mechanisms required for the processing of auditory temporal structure are not fully understood, the empirical findings suggest limited but significant associations between temporal precision and auditory-verbal STM for serial order. These abilities are also related to both phonological skills and non-verbal IQ, so one possible explanation is that automatic, attention-independent temporal processing factors mediate the established relationships between auditory-verbal ISR and higher-level skills that support language and cognition. If the ability to recall auditory-verbal sequences in the correct serial order is related to bottom-up variability over intervals corresponding to word-level timing, then this relationship has the potential to explain the individual and developmental variation remaining in digit span performance after considering attention-dependent and LTM factors. Therefore an investigation into the temporal properties of sequence representations in auditory-verbal STM may clarify the uncertain role of temporal processing in serial order capacity.

1.4.3 Auditory-Verbal Short-Term Memory Timing, Capacity and Load

One possible interpretation of the relationship between temporal processing and auditory-verbal STM for serial order is that the temporal accuracy of item representations may be constrained by the same resources that limit other aspects of STM. The issue of whether STM resources are limited by a fixed number of discrete items or groups, or whether these resources are dynamically redistributed has been examined in the context of visuo-spatial WM (Awh, Barton, & Vogel, 2007; Bays, Catalao, & Husain, 2009; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Bays & Husain, 2008; Ester, Fukuda, May, Vogel, & Awh, 2014; Fukuda, Awh, & Vogel, 2010; Gorgoraptis, Catalao, Bays, & Husain, 2011; Huang, 2010; Luck & Zhang, 2004; Zhang & Luck, 2011, 2008). According to the flexible resource sharing view presented by Bays and Husain (2008), a dynamically updating trade-off exists between the number (load) and quality (precision/resolution) of representations in visuo-spatial WM. In the variable precision model, not only are resources continuous, but encoding precision is also variable across items and trials (van den Berg, Shin, Chou, George, & Ma, 2012). Alternatively, Zhang and Luck (2008, 2011) and others have argued that the resource limitations in visuo-spatial WM are best characterized as discrete, fixed-resolution representations of a limited number of items. Of the results supporting the view that resources can be flexibly allocated during visual WM tasks, precision of item representations was shown to decline with increases in memory load in within-subject designs (Gorgoraptis et al., 2011). While there has not been any empirical evidence showing a link between individual differences in visuospatial WM capacity and precision of representations, some research suggests that differences in capacity reflect attentional processes that serve to filter out irrelevant information during encoding (Luck & Vogel, 2013). According to this view, it is not the capacity per se that differs between individuals, but rather the efficient direction of resource allocation to the most relevant aspects of the input.

The issue of discrete versus dynamic resource allocation in WM has been approached mainly in the context of visual WM, but little attention has been paid to this conceptual framework as it may apply to auditory WM. Recently, two experiments have provided preliminary support for a dynamic resource sharing account in the auditory domain. These experiments used an item probe paradigm, similar to those used in visual WM experiments, to test precision in memory for pitch. Kumar and colleagues (2013) presented sequences of pure tones and asked participants to match the pitch of the tone in the probed position. Participants matched the pitch by adjusting a dial, which changed the pitch of a random-frequency starting tone within a set range. These authors found that the precision in memory for pitch was better than chance in all conditions, but decreased with increasing memory load (number of tones in the sequence). Moreover, precision in pitch-matching was far superior for tones at the end of the sequences, and there was a slight advantage in precision for the first sequence tones. It is notable that the serial position curves for item precision in these experiments reflect the same bowed shape as those for recall probability in ISR tasks.

An analogous relationship may exist between temporal accuracy of sequence representations in STM, where the precision of temporal structure representations could be either fixed within individuals or could vary dynamically with the STM load of the material. Also, given that individuals vary in their motor synchronization precision and rhythm change detection thresholds, differences in temporal accuracy of item representations may explain some of the individual and group variance in STM capacity for serial order. A major focus of the current research is to investigate these relationships by measuring the precise temporal accuracy of item representations under different load conditions, and across individuals and groups. In the present research, temporal precision is measured during a subvocal rehearsal task, with the aim of revealing possible relationships with STM load and individual differences in STM capacity for serial order.

As discussed by Luck and Vogel (2013), the evidence from visuospatial WM points toward differences in the allocation of attention during encoding as a key source of individual variability in tests of capacity. It also seems probable that, while a subvocal rehearsal task may reveal differences in temporal precision of in auditory-verbal sequence representations, the source of imprecision would occur during perception and encoding stages. Therefore it is worth examining the role of auditory timing and rhythm in driving bottom-up auditory perception. In particular, it is important to consider the role of rhythm and timing in speech perception in order to understand how timing precision may relate specifically to auditory-verbal STM. As the next section will show, variable precision in the perception of auditory timing and rhythm has the potential to explain variance in auditory-verbal STM for serial order as well as phonological awareness.

1.5 Rhythm and Timing in Speech Perception and Language Development

Auditory information is continuous and necessarily perceived along a temporal dimension. This presents an important distinction between auditory and visual memory, where visual stimuli can be presented simultaneously as well as sequentially. Due to the necessarily sequential nature of auditory-verbal stimuli, STM of this type is unique in that the initial perception stage is not clearly distinct from the stages that normally defines memory, that is, the maintenance of an internal representation of external stimuli when the stimuli is no longer perceptually available. Just as perceptual accuracy has an obvious effect on memory for speech, the ability to accurately remember time-linked auditory information also affects the perception of speech. It is therefore possible that the ability to encode and maintain the temporal information in speech signals has an effect on performance on auditory-verbal STM tasks. In this section I will discuss the role of timing and rhythm in speech perception, and then review the evidence that speech and language impairments are related to specific deficits or abnormalities in neural entrainment to external auditory rhythms. Finally, I will argue that temporal accuracy may play an important role in auditory-verbal STM, and that fine-grained measures of rehearsal timing can inform our current understanding of the links between these topics.

The necessary representation of speech as changing signals over time gives auditory-verbal stimuli a rich set of temporal information at multiple levels. The smallest levels of temporal analyses begin with perception at the acoustic level. In speech perception, *acoustic*-level temporal accuracy in perception is associated with detection of signal amplitude (volume) and fundamental frequency (F0), which is one of criteria used for differentiating speech from non-speech sounds. At the *phoneme*-level, accuracy in the temporal domain is associated with the awareness and detection of the most basic sound unit capable of distinguishing word meanings. Phoneme awareness is a skill that is measured in rhyming, phoneme elision and Spoonerism tasks. *Syllable*-level temporal accuracy is required for the segmentation of continuous speech into vowel/consonant chunks, and for the perception of rhythm and stress patterns of these chunks (prosody). Awareness of syllables is revealed by performance on rise time (change in volume) detection, syllable counting and syllable-stress pattern judgment tasks. *Word*-level temporal accuracy is associated with serial ordering of multiple words, as commonly tested in ISR auditory-verbal memory tasks. Finally, temporal information can also be processed at the *list*-level, when the serial ordering of multiple word sets may be important. For example, list-level timing may be involved in n-back tasks, where subjects are required to retain information from specific sequences presented in past trials. List-level timing may also be important for the chunking of items within a single list, which is known to benefit STM performance. A chunking strategy might involve creating smaller lists within a single list and maintaining the serial order of the smaller (sub)lists, in effect shifting the burden from word-level only to a more hierarchical structure of word- and list-level timing.

As highlighted in previous sections, auditory-verbal STM tasks often test coarse temporal processing at the item-level, namely, the serial order of familiar words. Beyond memory for item-level serial order, memory for the within- and between-item temporal information is not explicitly tested in auditory-verbal STM span tasks. These relatively untested levels of temporal accuracy in auditory-verbal memory may contribute to explaining effects of memory load on STM for serial order, as well as individual and group differences in STM performance.

While largely unexamined in terms of speech, the role of hierarchical structures in fine-grained temporal precision has been extensively studied in non-verbal auditory tasks. In particular, the strength and coherence of auditory metrical structure has been shown to have a beneficial effect on precision of temporal processing. Grube and Griffiths (2009) found lower temporal change detection thresholds for sequences of tones with a stronger sense of meter (more accented tones on downbeats) and with metrical closure (final tones falling on downbeats) compared to sequences with the same number of tones and overall duration, but with a weaker sense of meter and with open endings. Moreover, metrical structure was found to impact the subjective perception of sequences, as the participants in this study rated the temporal change detection for metrically weak and open sequences as being more difficult. This finding suggests that precision of temporal processing in auditory perception is not solely a function of stimuli quantity and complexity, but also depends on the hierarchical rhythmic structure over multiple time scales. The role of hierarchical, entrained rhythms in auditory perception is likely to be particularly important for speech perception. Evidence for neural entrainment to hierarchical speech rhythms, and its role in speech segmentation and temporal prediction, will be reviewed in the next section.

1.5.1 Neural Oscillatory Entrainment and Temporal Prediction in Speech Perception

The speech perception literature highlights the importance of automatic neural entrainment to multiple speech-relevant rhythms in order to aid the segmentation of the continuous signal. On a physical level, speech is characterized by continuous changes in the frequency and power of an acoustic signal over time, rather than a series of discrete units. Because speech is inherently temporal, when speech is segmented into meaningful units, the discrete events are serially ordered (Kotz & Schwartze, 2010). The frequencies of neural oscillations that are critical for speech processing correspond to the timing of phonemes, syllables and stressed-syllable patterns (Luo & Poeppel, 2012; Peelle & Davis, 2012; Peelle, Gross, & Davis, 2013). In auditory perception, the automatic perception of the underlying temporal structure of a series of acoustic events is thought to aid in processing by facilitating the prediction of the timing of future events, which then directs attention to the salient features of incoming stimuli (Schwartze, Rothermich, Schmidt-Kassow, & Kotz, 2011). This is particularly important for speech processing, where semantically-relevant information is conveyed over multiple time intervals embedded within complex, continuous acoustic signals.

The multi-level processing of rhythm and timing information for speech is more complex than that for sequences of simple beats and tones. One way to conceptualize the multiple levels of temporal processing involved in speech is to imagine a set of endogenous neural oscillations which entrain to external stimuli and fluctuate in intensity (amplitude) over time (Goswami, 2011). Some of these frequencies correspond to the relative sizes of each level of speech segmentation and can be categorized within the synchronous neural oscillatory activity known to be associated with cognition (e.g. Delta, 0.5-4 Hz; Theta, 4-7.5 Hz; Alpha, 7.5-14 Hz; Beta, 14-30 Hz; Gamma, 30+ Hz). The fine-grained temporal analysis required for the integration of acoustic-level features in single phonemes would relate to higher frequency oscillations, while lower frequency oscillations relate to the coarser temporal segmentation required for entrainment of syllables (occurring at about 5 Hz, within the Theta band) and prosodic rhythms (2 Hz, Delta). Retaining the serial order of a word sequence would also relate to a lower (1-2 Hz, Delta) frequency oscillation. The accuracy in speech perception at each level could be dependent on stimulus-induced oscillatory modulation (phase locking) at the specific frequencies important for speech perception (Goswami, 2011).

Temporal regularity and coherence is thought to aid in the segregation of complex acoustic information (Andreou, Kashino, & Chait, 2011; Gross et al., 2013; Teki, Chait, Kumar, Shamma, & Griffiths, 2013). It has been suggested that temporal structure provides auditory segmentation cues, particularly for faster rates and when frequency cues are insufficient (Andreou et al., 2011). Broadly in agreement with Goswami (2011), Giraud and Poeppel (2012) and Ghitza (2011) argue that the frequencies specifically used in speech processing are Delta, Theta and Gamma, and that these oscillations segment speech signals into temporally-based units to track ongoing speech rhythms. Both the phase and amplitude modulations of speech have been shown to entrain cortical oscillations at these frequencies (Ghitza, Giraud, & Poeppel, 2013; Ghitza & Greenberg, 2009; Giraud et al., 2007; Hickok & Poeppel, 2007; Peelle & Davis, 2012; Peelle et al., 2013; Remez et al., 2013). Using MEG to record phase patterns from the human auditory cortex, Luo and Poeppel (2007) found that 4-8 Hz Theta band responses discriminated spoken sentences and were correlated with speech intelligibility. The authors suggested that the Theta oscillation period served as a temporal window for incoming speech, segmenting and tracking ongoing speech dynamics that correspond to syllabic rhythms. Similarly, Ghitza (2012) suggested that Theta-level entrainment is critical for speech processing. Ghitza manipulated the temporal envelopes of speech spectra and found that, while speech intelligibility is very poor when the temporal envelopes of critical frequency bands are flattened, the insertion of Theta band rhythms results in a significant improvement. Based on this and similar findings, Ghitza (2011, 2012, 2013) proposes that the neural Theta band serves as a 'master' frequency, which tracks syllabic rhythms in speech and sets the hierarchical structure of cascading temporal windows corresponding to multiple neural oscillators. Similar perspectives have been put forward by other researchers, who also suggest that phase resetting to low-frequency cues in the speech envelope entrains a hierarchy of cortical oscillations (Gross et al., 2013; M. J. Henry & Obleser, 2012). This perspective is also in line with the bottom-up syllabic phase model proposed by Hartley (2001). Hartley describes a process whereby representations of syllabic structure can be developed by neural tracking of the position (phase) within syllables, without a need to assume top-down knowledge about syllable structure.

Neural oscillatory entrainment to speech rhythms is thought to facilitate perception and processing by providing a predictive model of the temporal structure of incoming stimuli, which in turn modulates limited resources to the most salient aspects of the stimuli. This perspective of speech processing is in line with the more general Dynamic Attending Theory (Jones, 1976; Large & Jones, 1999), which states that internal oscillators entrain to the dynamics of external events in order to direct attentional energy to particular points in time. In support of this theory, Schwartze and colleagues (2011) found that attention-dependent ERPs to deviant pitch tones were facilitated by temporal regularity.

In the current research, the timing of subvocal rehearsal is used as an index of the temporal precision of internal auditory-verbal sequence representations. If perception of temporal structure and phase-locking to critical low-frequency speech rhythms facilitate the processing and accurate perception of speech stimuli, then individual differences in this process during encoding might be apparent when sequence timing is reproduced during rehearsal. Consequently, rehearsal timing precision may be expected to relate to the recall of broad, relative item timing, and therefore to auditory-verbal STM span. Furthermore, if temporal processing factors such as the precision of auditory hierarchical metrical structure and neural entrainment to critical-band speech rhythms relate to speech segmentation and attention, then these factors may also relate to individual differences and group differences in syllable- and phonological-level processing.

1.5.2 Auditory Timing in Speech and Language Development

In considering the possibility that speech is encoded using multiple-frequency tracking and modulation to external speech rhythms, one can begin to contemplate how certain types of speech and language disorders may arise from deficient phase-locking or inaccurate temporal sampling at certain frequencies. Goswami (2011) and Goswami and colleagues (2008) suggest that the phonological deficits characteristic of dyslexia and SLI arise specifically from deficiencies in perceiving the auditory cues to temporal organization at the levels of syllable and syllable stress rhythms. These authors argue that low-level auditory features (e.g. amplitude rise times) serve as signals for the tracking and entrainment to syllable-stress rhythms, and an inability to perceive these cues results in problems with both phonological skill development and in the perception of syllable structure and syllable-stress patterns.

Difficulties in entraining the rhythm of stressed syllables would explain the finding that the tapping rhythm deficits observed in dyslexics appear to be specific to a rate of about 2 Hz (Goswami, 2011; Thomson et al., 2006; Thomson & Goswami, 2008). This theory is also consistent with Wolff's (2002) finding that, in contrast to age-matched normal readers, dyslexic students showed more difficulties with reproducing simple finger tapping rhythms as well as speech rhythms of nonword (linguistically-neutral) syllables. One may reasonably speculate that inaccuracies in phase locking at word-level frequencies may relate to a higher proportion of serial order errors. Likewise, list-level timing might relate to interposition serial order errors, which are between-list transpositions in which the within-list position of the intruding item is preserved.

Some researchers have indeed suggested that the phonological and syllable awareness deficits characteristic of individuals with developmental dyslexia can be attributed to ineffective phase-locking to specific frequencies (Goswami, 2011; Hämäläinen, Rupp, Soltész, Szücs, & Goswami, 2012; Leong & Goswami, 2014a; Luo & Poeppel, 2007; Poelmans et al., 2011; Poeppel, Idsardi, & van Wassenhove, 2008; Power, Mead, Barnes, & Goswami, 2013). According to multi-time resolution models (MTRM) of speech processing (Ghitza & Greenberg, 2009; Giraud & Poeppel, 2012; Hickok & Poeppel, 2007; Poeppel, 2003), impaired syllable-level processing occurs as a result of inefficient phase locking to Theta rhythms. Theta rhythm deficits are thought to be accompanied by normal Gamma (acoustic-level) phase locking in developmental dyslexia, which accounts for dyslexic individuals' over-sensitivity to acoustic features (Luo & Poeppel, 2007). The normal Gamma and impaired Theta phase locking explains dyslexics' unimpaired performance on acoustic-level discrimination tasks, such as intensity (volume) discrimination between tones, and impairment on tasks which include a temporally-changing component, such as the discrimination based on intensity changes (rise times) between two tones. The temporal sampling framework (TSF) proposed by Goswami (2011) accounts for the symptoms of dyslexia by extending the MTRM to include a phase locking deficit in the slower Delta (as well as Theta) rhythms.

Many studies have shown that poorer performance on paced motor tapping tasks is related to speech, language and reading deficits. Two common tasks include *paced tapping*, in which subjects tap to a metronome beat, and *unpaced tapping*, in which subjects continue tapping to an internally-generated rhythm after the external rhythmic stimulus has been removed. Performance on these tapping tasks is measured using the inter-tap interval (ITI) variability, and by anticipation time (AT) (see Figure 3). ITI variability is a measure of the timing accuracy at the target frequency (rate), while AT captures any static temporal shift in tapping (leading ahead or lagging behind the metronome tone). Children with dyslexia (Goswami et al., 2002; Thomson & Goswami, 2008) and specific language impairment (SLI) (Corriveau & Goswami, 2009) show higher ITI variability AT measures in paced tapping tasks, and this relationship cannot be attributed to differences in age, non-verbal IQ or motor dexterity (Thomson & Goswami, 2008). Studies with adult subjects have also shown that dyslexics perform significantly worse in paced tapping ITI variability (Thomson et al., 2006) and AT (Wolff, 2002) measures. Wolff (2002) reported that adult dyslexics showed marked difficulty in reproducing the speech rhythms of nonsense syllables, and that the dyslexic group took significantly longer to recalibrate their tapping rates to changes in the metronome pace than did age- and reading-matched controls. Similar relationships between dyslexia and beat perception deficits have been found in cross-linguistic studies, demonstrating that this link is not specific to the rhythmic properties of the English language (Muneaux, Ziegler, Truc, Thomson, & Goswami, 2004). These findings are consistent with the previously-mentioned results of Martinez Perez et al. (2012) and Beneventi et al. (2009) showing that dyslexics have a specific deficit for the retention of item-order information. Taken together, these results support the theory that speech and language disorders are linked to a fundamental deficit in language-independent auditory rhythm perception.

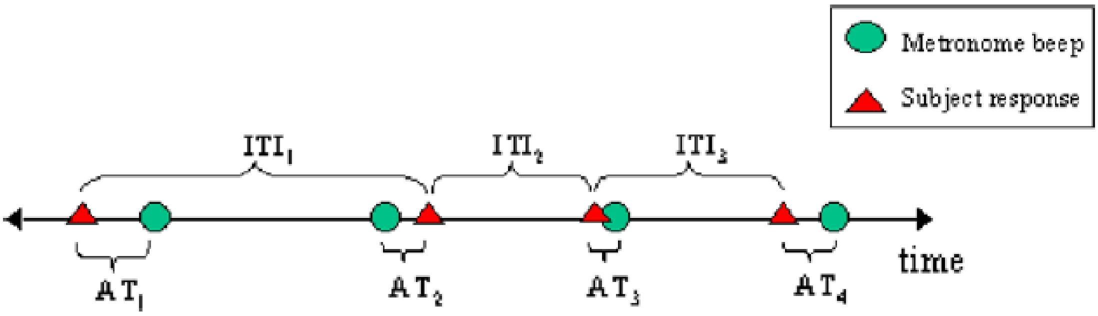

Figure 3. Inter-tap interval (ITI) and anticipation time (AT) performance measures in a paced tapping task (Source: Thomson & Goswami, 2008).

The theory that STM for speech is directly related to timing accuracy is supported by findings that children with dyslexia perform worse on STM tasks when recall for order is tested separately from recall for item information. These findings include studies showing that dyslexic children perform significantly worse than chronological- and reading-age matched controls on measures of serial order reconstruction, a STM task that maximizes order over item retention (Martinez Perez, Majerus, Mahot, et al., 2012). Furthermore, a one-year developmental study showed that unique variance in first graders' reading decoding abilities was predicted by their order STM capacity, but not item STM capacity, measured during the previous year (Martinez Perez, Majerus, & Poncelet, 2012). In an fMRI study of the neural networks involved in short-term retention of item and order information, dyslexics showed reduced bilateral activation in the superior parietal cortex (BA7) relative to controls during a sequence probe task, indicating a distinct STM deficit with temporal processing that could not be caused by a deficit in phonological processing (Beneventi et al., 2009). Evidence from these studies suggests that speech and language disorders are linked to a fundamental impairment with timing and rhythm, and these disorders cannot be explained purely by difficulties with phonological processing.

Difficulties with memory for temporal information may not be categorical and specific to individuals with speech and language disorders, but are rather more likely to vary on a continuum in which individuals with these disorders are highly represented at one end of the spectrum. Not only does AT and ITI performance measures on paced tapping tasks differentiate clinical groups (e.g. dyslexic versus controls), these measures have also been shown to account for unique variance in individual differences in literacy and language abilities (Corriveau & Goswami, 2009; Thomson & Goswami, 2008). A study of children who varied in reading ability revealed that significant differences in amplitude rise time detection were found between dyslexic and normally reading children, and between normally reading children and young early readers (Goswami et al., 2002). This study showed that amplitude rise time detection accounted for substantial unexplained variance in individual reading and spelling acquisition after controlling for age, nonverbal IQ, and vocabulary. Studies using similar methods have also found that impairments in the detection of rise times, known to be critical for the rhythmic timing of speech, predict individual differences in language and literacy outcomes (Goswami, Fosker, Huss, Mead, & Szucs, 2011; Hämäläinen, Leppänen, Torppa, Müller, & Lyytinen, 2005; Huss, Verney, Fosker, Mead, & Goswami, 2011; Leong, Hämäläinen, Soltész, & Goswami, 2011; Pasquini, Corriveau, & Goswami, 2007). Finally, a study of typical adolescents revealed a limited but significant relationship between timing and rhythm tasks and phonological skills (Grube, Kumar, Cooper, Turton, & Griffiths, 2012). Phonological skills were specifically related to change detection tasks that required the processing of short, isochronous sequences of roving tempo. Moreover, a general auditory processing deficit was ruled out as an explanatory factor because there were no correlations between phonological skills and non-temporal aspects of auditory perception (e.g. key violation). The authors concluded that the ability to detect rhythmic deviations may help in extraction of important cues for segmentation, and aid in the parsing of an incoming speech stream into smaller units.

In conclusion, both neural and behavioural measures of the precision of auditory temporal processing have been linked to speech and language abilities. This relationship is substantiated by the fact that individuals with speech and language deficiencies also show impairments with entraining to auditory rhythms, and that timing and rhythm deficiencies predict unique variance in speech and language outcome measures after controlling for non-verbal IQ, age and vocabulary. Because specific serial order deficits have been linked to speech and language deficits independent of memory for item information, it is unlikely that these findings can explained by a common impairment in phonological abilities. Instead, some researchers have proposed that impaired phonological abilities can be explained by temporal accuracy deficits and abnormal phase-locking (Goswami, 2011). Fine-grained measures of temporal accuracy in motor entrainment tasks using simple tones and rhythms have revealed frequency-specific (Delta and Theta band) impairments that are related to dyslexia and SLI. Measures of temporal accuracy are not usually obtained at this level in STM tasks using speech stimuli, which contain rich temporal information at multiple rhythmic levels. If applied to auditory-verbal STM tasks, temporal accuracy measures at multiple frequencies may be useful in accounting for performance differences.

1.6 Limitations of Current Methods for Measuring Rehearsal and its Temporal Properties

As Brown and Hulme point out, “if rehearsal does exert an important causal influence on memory span, it is important to know exactly when such rehearsal is taking place” (Brown & Hulme, 1995, p. 599). A direct and temporally-precise measure of rehearsal activity has the potential to inform the ongoing debate surrounding the determinants of STM performance. Such a measure may also elucidate the specific role of rehearsal in STM. That is, a direct measure of rehearsal has the potential to address prior assumptions regarding when rehearsal does and does not occur, as well as under what circumstances rehearsal is or is not a crucial determinant of STM performance.

The first few sections of this Chapter have provided an outline of the reasons why rehearsal is thought to have an important role in verbal STM. The rehearsal process has traditionally been considered to be beneficial, though not required, for the short-term retention of verbal information. The phonological loop model describes rehearsal as a method of reactivating the rapidly decaying phonological representations within the articulatory system. Both word length and speech rate, which are considered to be indices of rehearsal speed, have been shown to have a linear relationship verbal STM span. Further, performance on verbal STM tasks suffers when subjects engage in concurrent speech during the list presentation, and this detrimental effect on performance is thought to occur because use of the articulatory loop blocks the rehearsal process. Psychophysiological, electromyographical and neuroimaging methods have been used to support the premise that subvocal rehearsal can be detected and that it appears to be similar to vocalized speech. Rehearsal is part of a system whose main purpose, it has been suggested, is for language learning and vocabulary acquisition (Baddeley et al., 1998). A more complete and unambiguous understanding of rehearsal would contribute to the pursuit of a full comprehension of the verbal STM system and may illuminate the roles of STM for words, timing and rhythm in the development of speech, language and literacy.

If it can be agreed that a measure of rehearsal would make a valuable contribution toward a complete understanding of these issues, the next question is how best to measure rehearsal. As described in a previous section, there have been a variety of approaches to the examination of rehearsal. The bulk of the current evidence surrounding the role of rehearsal in STM rests on indirect manipulations of rehearsal using recall as the dependent measure; these methods rely critically on assumptions about the underlying processes in STM. Because rehearsal is not usually directly measured it is not always clear whether the performance effects thought to be attributable to rehearsal in fact occur at the level of encoding, maintenance, recall, or some interaction of effects at multiple stages. These methods use speech rates and spoken word durations as indicators of rehearsal rate, and they use concurrent articulation as a way of interrupting or blocking rehearsal.

Overt speech and observations of speech-related behaviour have also been used as indirect methods of examining rehearsal. In addition, researchers have attempted to detect and measure rehearsal directly using imaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), psychophysiological measures such as electroencephalography (EEG) and magnetoencephalography (MEG), and physiological methods such as electromyography (EMG). Case studies using patients with specific neuropsychological and sensorimotor impairments have provided ecological tests of alterations within theoretical STM components, and these tests have been used to support and enlighten the current views of rehearsal in STM.

While these methods have established the foundations for the current understanding of rehearsal in STM, some are insufficient for the purposes of the current research. In this section I will evaluate each of these methods and, where applicable, describe why the method 1) is limited in the amount of information that it can provide about subvocal rehearsal, and 2) does not adequately meet the task criteria of the present research, which aims to directly measure rehearsal with high temporal resolution. Following a critical evaluation of the historical and contemporary rehearsal measurement techniques, a novel task paradigm will be presented which, it will be argued, can be used to measure the fine-grained temporal properties of rehearsal.

1.6.1 Articulatory suppression

The articulatory suppression (AS) method has been frequently used to examine STM when subvocal rehearsal is disrupted or prevented. Given the premise that AS selectively blocks rehearsal, the suppression technique can be used with a variety of other task manipulations in order to determine whether the rehearsal process contributes to a given performance effect in STM. If the effect is abolished under suppression, then this is interpreted as evidence that the effects of the experimental manipulation on recall performance rely on rehearsal. If, on the other hand, a given effect is equally present under AS and non-AS conditions, then this is seen as evidence that the effect does not originate in the rehearsal process (Baddeley, 1986).

However, there are reasons why the classic interpretations of the outcomes of experiments employing this technique should be questioned. First, there are at least two possible reasons for the persistence of an effect when rehearsal is blocked with AS; one is that rehearsal does not contribute to the effect observed in non-AS (i.e. silent) conditions, and another is that subjects tend not to engage in rehearsal during the non-AS condition. If subjects do not rehearse in the non-AS condition, this could occur either out of strategy choice or because subjects are unable to do so, for example because the AS blocks the phonological recoding of visual STM material. In the latter scenario, the elimination of the WLE and PSE in visual STM would not be due to the blocking of subvocal rehearsal but rather to the prevention of phonological encoding. If this were the case then the effects of AS would have implications for the importance of phonological encoding, rather than for the subvocal rehearsal process per se. Without a direct measure of rehearsal, it is not possible to make conclusive distinctions among these possibilities.

Second, it is possible that, apart from the opportunity for rehearsal, there are differences between AS and non-AS conditions that can explain the observed differences in STM recall performance. Such differences could be related to an increase in general processing demands or some categorical difference between overt and covert speech. While there is evidence to suggest that overt and covert speech involve highly similar processes, the mechanisms of covert speech are not fully understood and thus the possibility that there are unknown pre-motor cognitive differences between these two modes of speech must be considered. Also, it is possible that AS merely adds noise/disruption to the subvocal rehearsal process, rather than abolishing it completely.

1.6.2 Speech Rate and Word Length

Like the articulatory suppression technique, one problem with measures of speech rate and word length as indicators of rehearsal speed is that they make certain assumptions about underlying cognition. One such assumption is that the representational quality and temporal characteristics of subvocal speech are equivalent to those of vocalized speech. Subvocal speech may be qualitatively different from vocal speech or may differ in terms of temporal characteristics, which are of particular interest for the purposes of the current research. A more detailed comparison between articulated and unarticulated speech is presented in a later subsection (1.6.4) on the use of overt rehearsal techniques to investigate rehearsal. However it should be noted here that the evidence for qualitative and quantitative differences between overt and covert speech challenges the premise that speech rates and word lengths are indices of subvocal rehearsal speed, and consequently that slower overt speech rates and longer word pronunciation durations are causally related to poorer ISR performance.

The notion that relationships between speech rate, word length and STM span support the role of subvocal rehearsal speed in determining ISR performance is also undermined by alternative explanations of these relationships that do not invoke rehearsal. One criticism of the rehearsal speed explanation is that it minimises long-term memory factors. Da Costa Pinto (1991) found a native language advantage that affected both STM span and digit reading rates in bilinguals, even if the mean number of syllables was higher in the native language. As da Costa Pinto suggests, it is possible that internal representations of highly familiar long words are abbreviated during rehearsal, as the decayed and incomplete representations of highly familiar words may be more easily recovered from LTM during redintegration (i.e. reconstruction of degraded traces, see Brown & Hulme, 1995). If highly familiar long words are abbreviated during rehearsal, then the rehearsal speed explanation for the WLE would not hold in cases where word familiarity was equated across the long and short word lists. The results of da Costa Pinto's experiment lend support to the notion that there may not be a one-to-one correspondence between overt and covert speech, and that in fact covert speech may be underspecified, allowing for faster rehearsal of highly familiar words. This question cannot be resolved by relying on recall performance in tasks that manipulate speech rate and word length; rather, only a direct measure of rehearsal timing can confirm whether rehearsal for longer words is indeed slower than rehearsal for the same number of shorter words.

Another potential explanation for the relationship between speech rate and STM span is that phonological skills mediate these variables. This point is related to the more general criticism that the phonological loop account of the speech rate and span relationship does not take into account LTM factors, because the concept of 'phonological skills' includes the quality of phonological representations in LTM. A number of studies have shown that phonological skills are correlated with verbal STM span (Brady, 1986; Gathercole, Willis, & Baddeley, 1991; Kail, 1997; Nithart et al., 2011; Rapala & Brady, 1990; Wagner & Torgesen, 1987) as well as speech rate (Kail, 1997; McDougall et al., 1994; Muter & Snowling, 1998). Research has also shown that phonological skills and speech rate share a substantial amount of the variance when entered into a regression equation as predictors of verbal STM (Kail, 1997). Furthermore, phonological skills and verbal STM also have shared variance as predictors of reading abilities (Hansen & Bowey, 1994; McDougall et al., 1994). Based on relationships between phonological processing and verbal STM span, Brady (1986) concluded that the efficiency of phonological processes is a key determinant of span, and that phonological efficiency accounts for both individual and developmental differences. Consistent with this view, Rapala and Brady (1990) interpreted these results as indicating that superior phonological skills reduce the processing requirements during phonological encoding of words, thereby increasing the availability of resources that can be dedicated to the storage of phonological codes.

Like the evidence for rehearsal based on the speech rate and span correlation, many researchers have raised similar non-rehearsal alternative explanations of the WLE since its original observation. One key argument that has been raised in opposition to the notion that the WLE is caused by increased decay due to slower subvocal rehearsal for longer words is that longer words take longer to articulate during spoken recall. The increased delay at output for longer words could increase forgetting (due to either more interference or time-based decay) during overt recall (Cowan et al., 1992, 1994; Nairne, 2002). Thus it is not clear from studies on the WLE whether rehearsal speed during item presentation and delay intervals is affected by word length, or whether poorer recall for lists of longer words can be explained entirely by output delays.

In addition, studies showing correlations between speech rates and ISR performance to buttress arguments about developmental changes in the use of rehearsal do not preclude explanations that do not necessitate the rehearsal process. Gathercole and Adams (1994) and Gathercole, Adams and Hitch (1994) did not find a correlation between speech rate and STM span in young children, and claim that this is evidence that young children do not rehearse. However, another possibility is that the motor execution aspect of speech is slower to develop than the rest of the internal articulatory system, and as a result, articulation rate is not a valid index of subvocal rehearsal until the motor skills involved in overt speech 'catch up' to the comparatively faster development of the internal articulation. This interpretation goes against the idea that overt speech develops prior to internal speech (Flavell, Green, Flavell, & Grossman, 1997; Vygotsky, 1962). However, it would explain discrepancies between the absence of a correlation between speech rate and digit span in four year olds, who may be capable of rehearsing faster than they can speak. It would also explain the emergence of a modest correlation in five year olds, as the motor skills involved in articulation continue to develop, and finally the near-perfect correlations found between the spans and speech rates in adult subjects whose articulatory skills are fully developed. This interpretation undermines the notion that subvocal rehearsal is equivalent to speech in all ways apart from the execution of certain motor commands in the latest stage of speech production.

In summary, the relationships observed between speech rate, word length and STM span can be explained by several alternatives that do not appeal to the speed of subvocal rehearsal, and there is evidence to suggest that some of these alternatives can provide as good or better accounts of the empirical data. Many of the methods used to investigate these issues rely on indirect manipulations of rehearsal, where there is no guarantee that 1) subvocal rehearsal is used in these tasks, 2) the speed of subvocal rehearsal is affected by variations in overt articulation rate, and 3) there are no qualitative differences between subvocal and overt speech that result in non-uniform discrepancies in the correspondence between speech and rehearsal rates. Thus, the ongoing debate about the nature of the relationships between speech rate, word length and STM span is difficult to resolve without a direct measure of rehearsal and its timing.

1.6.3 Neuropsychological Impairments

The evidence for the role of rehearsal that comes from case studies of neurological impairments suffers mainly from the limitations that apply to this methodology in general. That is, patients with neurological impairments may not have a 'pure' or isolated deficiency, both in terms in brain structure and cognitive function. In fact, in most cases it is highly likely that a patient's deficit is contaminated by other factors, some of which cannot easily be assumed, measured or otherwise controlled for. In addition, case studies of neurological patients are nearly impossible to replicate, as the type and magnitude of brain damage will rarely be perfectly matched across individuals. Geva et al.'s investigation into relationships between brain lesions and inner and overt speech abilities in post-stroke aphasia patients drew attention to the wide distributions of scores across both measures, and the authors noted that "The anatomical results presented here suggest that the group of patients tested is a mixed one and likely to present all sources of impairments to inner speech" (2011, p. 3080). Finally, while evidence from neurological impairments has been useful for clarifying the nature of rehearsal, including its reliance on motor planning systems and its independence from motor execution functionality, this evidence seems limited to defining the conditions for the presence or absence of subvocal rehearsal. To this point, there is no indication that studies of neuropsychological patients can provide more specific information about rehearsal, such as its contents and timing, to address questions about the relationships among rehearsal rate, precision of temporal representations, and STM span.

1.6.4 Overt Rehearsal

With regard to overt speech, the results of Oppenheim and Dell (2008, 2010; Oppenheim, 2012) and Corley, Brocklehurst and Moat (2011) raise issues regarding the validity of comparisons between overt and covert speech. Oppenheim and Dell suggest that inner speech may differ categorically from overt speech due to an underrepresentation of the contents of subvocal speech at the subphonemic (feature) level. Corley and colleagues disagree with the subphonemic attenuation hypothesis of inner speech, arguing that inner speech is fully specified at the feature level. Instead these authors suggest that covert speech primarily differs from overt speech in that there is no ability to dynamically monitor and edit the contents of subvocal representations, and they highlight the importance of the lack of acoustic and motor feedback during the production of internal speech.

The possibility that vocal and subvocal rehearsal may differ categorically is also supported by the somewhat surprising finding that recall performance can worsen during overt rehearsal conditions compared to silent conditions. Fischler, Rundus and Atkinson (1970) reported superior recall in silent compared to two overt rehearsal conditions; free rehearsal and rehearsal restricted to the currently-presented item. Einstein, Pellegrino, Mondani and Battig (1974) also reported that immediate free recall performance was superior in a silent rehearsal condition compared to free overt rehearsal and forward serial order overt rehearsal conditions. These results suggest that rehearsal may be qualitatively or quantitatively different in the silent compared to overt conditions. For instance, participants may have been able to rehearse more quickly or use a different strategy during the silent condition. Jacoby (1973) suggested the latter case was a factor in his experiment, which showed poorer free recall for the earliest items in overt rehearsal compared to both a silent condition and a distracter condition. This experiment involved the presentation of four sets of five words, where each set was followed by a 15 second delay filled with either silent rehearsal, overt rehearsal of the words from the current five word set, or mental subtraction. Participants then had to recall all twenty items after the presentation of all four sets. Jacoby proposed that participants were engaging in the cumulative rehearsal of items from prior sets in the silent rehearsal condition only, which selectively improved recall for the words from the earliest sets.

However, there is another possible explanation for poorer recall performance during overt compared to covert rehearsal, which is that overt rehearsal produces interference due to the additional auditory feedback and/or motor execution. This explanation is somewhat counterintuitive since one might expect that overt action and perceptual feedback related to the memoranda would strengthen the representations and provide more avenues for re-encoding the material. It is not clear whether this account could explain the selective reduction in the primacy effect reported by Jacoby (1973), as opposed to a uniform impairment across serial positions. Still, it is not possible to distinguish between these explanations for superior recall in silent versus overt conditions because there was no measure of, or control over, the silent rehearsal conditions in these experiments.

A separate line of evidence that underscores a dissociation between overt and covert speech processes comes from studies of post-stroke aphasia patients. Geva and colleagues (Geva, Bennett, Warburton, & Patterson, 2011) used rhyme judgment, familiar word homophone and non-word homophone tasks to tap the patients' inner and overt speech functioning, where the patients were required to read the word pairs either silently or aloud before making a response. As expected, these authors found significant correlations between overt and covert speech performance which were driven by clusters of participants whose performance was either normal or severely impaired for both speech types. However, they also observed patients with preserved inner speech and impaired overt speech, and surprisingly, patients with the opposite pattern of results; impaired inner speech and normal overt speech. In the former case, this can be explained by speech motor difficulties (i.e. oral apraxia) or difficulties translating the phonological to articulatory code. However, the latter case is unexpected given the premise that inner speech is simply overt speech without articulation. The authors suggest that the aphasia patients in this category have a specific deficit in the transfer of information from the output to the input phonological stores (i.e. from production to comprehension systems). The conclusion that inner speech relies on processes outside of those implicated in overt speech was supported by a lesion mapping study, where Geva et al. (2011) showed that inner speech abilities were significantly related to lesions in the dorsal language route after controlling for overt speech production scores.

In summary, unexplained differences in recall performance based on rehearsal speech type, as well as conflicting results regarding the correspondence of representations in overt and covert speech, cast doubt on the validity of the overt rehearsal method as a tool for understanding subvocal rehearsal. Behavioural and brain lesion studies of post-stroke aphasia patients also strongly support the claim that inner speech is not equivalent to overt speech without the motor execution component. Therefore, investigations into subvocal rehearsal should be cautious when interpreting relationships between overt rehearsal and STM performance.

1.6.5 Behavioural Indicators of Speech-Related Activity

As previously mentioned, a key limitation of using behavioural indicators of speech-related activity to detect or measure subvocal speech is that subvocal speech does not rely on the functionality of articulatory musculature (Baddeley & Wilson, 1985), so articulatory-motor systems are peripheral rather than central to subvocal rehearsal. Behavioural signs of speech such as whispering or lip and chin movements may incidental rather than reliable markers of rehearsal, and these signs may vary with factors such as developmental stages, individual differences and memory list material. The notion that articulatory-motor movements are not obligatory is supported by Meltzner et al.'s (2008) inability to conduct subvocalised speech recognition due to insufficient EMG activity. This result suggests that even very subtle articulatory movements can be suppressed during covert speech, at least in adults. While behavioural indicators of speech-related activity have revealed some relationship to recall in young children (Keeney et al., 1967), the validity of the measures should be questioned due to uncertainty about whether the behaviours are obligatory in young children and when they can be voluntarily suppressed between and within individuals throughout development.

1.6.6 Self-Report

The use of self-report in studies of rehearsal necessitates the assumption that people can attend to their own inner speech, and then accurately report this information in retrospect. This assumption may be problematic because it is possible that individuals, and especially children, do not have much awareness of their own inner speech (Flavell, Green, & Flavell, 2000; Flavell et al., 1997). As previously mentioned, Oppenheim and Dell (2008) conducted an experiment in which participants were required to monitor their internal speech errors while subvocalising tongue twisters, and they found that participants reported more phonemic errors in overt speech compared to inner speech. In a critique of Oppenheim and Dell's (2008) study, Corley, Brocklehurst and Moat (2011) pointed out a few reasons Oppenheim and Dell's conclusion about the underspecification of inner speech may be flawed. First, these authors proposed that inner speech may be fully specified, yet errors in inner speech could be more frequently undetected. They suggest that this could occur because of the lack of auditory and motor feedback during covert speech, and that the smallest (single phoneme) errors may be especially susceptible to underreporting in this case. Second, Corley et al. state that adaptive editing of speech plans may occur to repair speech production errors while they are still in the planning stage, and that this process may be specific to situations where there is an intention to articulate. Both of these issues would cause individuals to overlook errors in their own inner speech, which calls into question the validity of the self-report method for examining subvocal rehearsal.

In an attempt to replicate and clarify Oppenheim and Dell's (2008) findings, Corley and colleagues (2011) used similar methods and found that the frequencies and types of errors reported during the recitation of tongue twisters did not differ between overt speech and silent conditions. The correspondence between error types in overt and covert speech was not due to an absence of feature-level reported errors in the covert condition, which shows that the participants were indeed capable of detecting these errors in their internal speech. Corley and colleagues authors also reported that auditory masking in an overt speech condition did not affect the frequency or type of reported errors, which showed that error reporting in overt speech was not dependent upon auditory feedback. Moreover, independent ratings of overt speech errors revealed that participants were fairly accurate at reporting their own speech errors. Although there was an overall tendency for participants to underreport their own overt speech errors, the magnitude of this reduction was consistent across all experimental conditions and for all error types.

The results of phoneme monitoring tasks also suggest that participants are able to accurately monitor their own inner speech. For instance, in an experiment requiring participants to respond to the presence of target phonemes in their subvocal Dutch translations of words presented in English, reaction times to the target phonemes were a function of both the onset time of the target sound and the stress pattern within the translated word (Wheeldon & Levelt, 1995). That is, the target detection times were faster when target phonemes were located earlier in the word and when stress was on the initial syllable, and the effect of target phoneme position was greater for initial stress compared to final stress words. Given that the participants were required to monitor their own subvocal translations, and that the reaction times provided a sensitive measure, this result suggests that individuals are capable of accurately monitoring their own inner speech.

Based on the studies employing self-report methods to investigate inner speech, it seems that self-report methods have the potential to be useful for studying the content and timing of rehearsal. Evidence from error reporting and phoneme detection times suggest that individuals can accurately monitor their own unarticulated speech, although it remains uncertain as to whether subvocal speech can be monitored at the phonemic or subphonemic level. Research employing a self-monitoring technique to investigate inner speech should therefore be aware that errors due to phonemic similarity and lexical bias may be underreported, and that the representations in inner speech may be more flexible and abstract compared to overt speech. The surprising finding that auditory masking did not affect error reporting in overt speech suggests that the lack of auditory feedback during subvocal speech may not be problematic in terms of awareness and self-monitoring. Finally, it should be noted that the utility of the self-report method is likely to be unique to situations where participants provide quantifiable data that is fully dependent upon the use of a subvocal rehearsal strategy. In contrast to the error reporting and reaction time measures discussed here, self-reporting about strategy type alone cannot provide a continuous measure of rehearsal activity, but is perhaps only useful for determining whether or not a rehearsal strategy was used.

1.6.7 Physiological Measures of Speech-Motor Activity

The detection of the electrical signals emitted as a result of muscular activity (EMG) associated with speech is more objective and quantifiable than observations of speech-motor behaviour, although these methods are subject to similar criticisms. First, internal speech does not rely on functional articulatory musculature (Baddeley & Wilson, 1985), so articulatory-motor systems are peripheral rather than central to subvocal rehearsal. This method may not reliably detect motor signals associated with subvocal speech in all people, and it does not decouple the subvocal rehearsal system involved in the internal representations of language from the articulatory system which is involved in the production of speech.

Second, there are issues with the practicality of this method. Jorgensen and Dusan (2010) concluded that subvocal speech detection using EMG is possible with a small vocabulary set, but it is limited because 1) the signals are energy-modulated in time (as opposed to acoustic signals, which are also spectrally-modulated in frequency and thus easier to decode), and 2) the averaged and combined of signals recorded by surface EMG electrodes makes it difficult to extract the activity of individual articulatory muscles. In the most successful instances, the training of a small word set was performed with multiple EMG sensors on the neck, lip, chin and jaw muscles, and the analysis of the EMG signals employed complex signal processing techniques where classification was based on multi-dimensional patterns of signal data from different electrodes. Jorgensen et al. (2010) concludes that this method falls short due to "an inherent lack of information present in the EMG signal envelope or due to inadequate sensing" (p. 364).

Lastly, some evidenced suggests that subvocal speech can be detected with EMG but only when the participant consciously exaggerates the articulatory-motor gestures. Meltzner et al. (2008) reported an inability to conduct subvocalised speech recognition due to insufficient EMG activity. Meltzner et al.'s finding shows that even very subtle articulatory motor gestures can be inhibited, and the level of voluntary inhibition is likely to vary across development. This suggests that the EMG method may not be ideal for studying subvocal speech in the context of STM experiments, where a focus on speech-motor activity during rehearsal may affect memory performance.

1.6.8 Psychophysiological and Neuroimaging Methods

Investigations into rehearsal using psychophysiological (M/EEG) and imaging (fMRI) provide a more direct and atheoretical view of the rehearsal process as it occurs, and have the critical advantage of measuring differences in central processes rather than peripheral artefacts. Unlike many behavioural measures, these methods do not rely on STM recall performance as a dependent measure, so some confounding processes that occur during recall can be avoided. These measures can be used in conjunction with performance data in order to more specifically identify the behaviourally-relevant neural signatures of subvocal rehearsal.

A major advantage of recording the brain's electrical (EEG) and magnetic (MEG) signals is that they provide measures of ongoing neural activity with excellent temporal resolution. The high frequency sampling of these methods allows the detection of transient changes in brain activity over relatively fast (millisecond range) timescales. An MEG study by Numminen and Curio (1999) has shown that the auditory cortex is selectively attenuated to auditory stimuli during covert speech, which is thought to be a result of the internal forward models that predict the auditory stimuli that result from one's own vocalized speech. These authors asked participants to covertly articulate one of two vowel sounds, during which time they presented a recording of the participants' overt production of either a matching or mismatching sound. Numminen and Curio reported that left hemisphere M100 responses, which occur 100-200 ms after the stimulus onset, to matching vowel sounds during covert speech were attenuated relative to responses to the same stimuli without concurrent covert speech. The authors also found that the left hemisphere M100 peak latency was slightly delayed in the match condition relative to the mismatch condition. The effects of covert speech production on the auditory event-related potential (ERP) amplitudes and latencies have been replicated in many other MEG and EEG studies (Behroozmand & Larson, 2011;

Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Kauramäki et al., 2010; Niziolek, Nagarajan, & Houde, 2013; Parkinson, Korzyukov, Larson, Litvak, & Robin, 2013; Sato, Troille, Ménard, Cathiard, & Gracco, 2013; Shergill et al., 2002; Tian & Poeppel, 2013; Ylinen et al., 2014), which together provide convincing evidence for the reliability of the effects. Timm and colleagues (2014) further confirmed that the attenuation of self-produced sounds is result of motor planning, rather than motor execution, using TMS of the motor cortex to produce unplanned movements and MEG to measure response attenuation. While this content-specific covert speech attenuation effect has yet to be demonstrated in the context of ISR tasks where multiple items are held in memory, the selective priming of the auditory cortex during subvocal speech presents a promising opportunity for detecting the contents of rehearsal.

As mentioned in section 1.3.5, the high sampling rates in electromagnetic recordings of brain activity are useful for providing insight into changes in neural oscillatory dynamics during various cognitive tasks. EEG and MEG have been used to investigate the changes in neural frequency dynamics during subvocal speech, and during rehearsal in particular. One method of localising the neural dynamics that relate specifically to rehearsal is to compare rehearsal intervals for words or nameable visual stimuli to control intervals, where the control interval is either an attention-demanding non-rehearsal task or a task requiring the maintenance of non-nameable visual items (Harmony et al., 1999; Hwang et al., 2005). However, two experiments using this method of manipulating rehearsal have yielded inconsistent results, with one attributing a change in left parieto-temporal Theta band power to rehearsal (Harmony et al., 1999), and the other reporting enhanced frontal and occipital Beta during rehearsal (Hwang et al., 2005). These studies have provided a starting point toward identifying the frequency-modulated cortical networks involved in subvocal speech, but the conflicting results may point to problems with the lack of control over rehearsal in this task paradigm.

Kopp, Schröger and Lipka (2004, 2006) took a different approach to this question, using the irrelevant speech effect to disrupt rehearsal during a verbal delayed serial recall EEG task. These authors found a reduction in the duration of Gamma band coherence in the irrelevant speech condition compared to silent and white noise rehearsal intervals, and they reported that participants who did not show a behavioural irrelevant speech effect also did not show a reduction in Gamma coherence in the irrelevant speech condition (Kopp et al., 2004). In a second study, the reduction in Gamma coherence in the irrelevant speech condition was replicated for spoken recall, but not for written recall (Kopp et al., 2006). Given that most accounts of the effect of recall mode suggest that the differences occur at the time of recall, this result is particularly surprising. Kopp et al. proposed that, because the participants were aware of the recall mode in each trial, they were perhaps adjusting their rehearsal strategy away from using only the presupposed phonological rehearsal to using a visual code that is more resistant to irrelevant speech input. These findings underscore both the potential for identifying neural markers for rehearsal using EEG, as well as the importance of using a behavioural task that demands a particular type of subvocal rehearsal in order to minimise unintentional and undetected differences in encoding and maintenance strategies. Specifically, a suitable task is one that demands and verifies rehearsal in the absence of ongoing auditory, motor or visual stimuli that might be confounded with it, while also exerting control over rehearsal timing to allow for signal averaging during a silent maintenance interval.

Functional magnetic resonance imaging (fMRI) has provided evidence that the neural structures involved in overt and covert speech production have substantial overlap (Herman, Houde, Vinogradov, & Nagarajan, 2013; Paulesu, Frith, & Frackowiak, 1993; see also Tian & Poeppel, 2012), which supports the assumption that subvocal rehearsal is closely related to overt speech production. Consistent with the results of EEG and MEG studies, attenuation of the dorsolateral temporal cortex in response to spoken sentences during silent articulated speech has also been demonstrated using fMRI (Agnew, McGettigan, Banks, & Scott, 2013). An fMRI investigation revealed that a right hemisphere network consisting of dorsolateral prefrontal cortex, parietal and temporal lobe areas control the self-pacing of subvocal speech (Shergill et al., 2006). This one of a minority of studies that attempted to control the pacing of subvocal speech during the task, and the authors assessed the participants' abilities to keep the correct rehearsal pace by asking them to tap the rate with a finger before the task. However, Shergill and colleagues acknowledged that their self-paced subvocal speech condition may have differed from the visually-cued subvocal speech control condition not only in terms of temporal regulatory mechanisms, but also in demands on attention and concentration. Nonetheless, these and other fMRI results that pinpoint the structures and networks implicated in specific aspects of subvocal speech are useful for identifying the structural and functional commonalities and dissociations among these cognitive processes, which can then also be used to guide interpretations of the rich topographical and spectral patterns in EEG data. In spite of its more general utility, the low temporal resolution in fMRI means that it is not an ideal tool for determining the relationship between rehearsal timing properties and STM performance.

1.6.9 Conclusion

Subvocal rehearsal has traditionally been studied indirectly via behavioural tasks that manipulate the hypothesised articulatory loop and require interpretation of corresponding changes in STM recall. These methods are limited in that they do not rule out unobserved and unintentional differences in encoding and maintenances strategies, such as the use of elaborative rehearsal (Hanley & Bakopoulou, 2003) and visual coding (Kopp et al., 2006), as well as confounding differences at the overt recall stage. Many discrepant results in the literature may therefore be due to the lack of control over rehearsal, where serial ordered subvocal rehearsal is not required by the task but rather is encouraged and assumed. Furthermore, the use of overt speech measures as indices of subvocal speech appears to be problematic given a number of differences between covert and overt speech that may differentially affect STM performance. The lack of control over or measure of subvocal rehearsal timing in particular means that many questions remain about how the precise temporal properties of internal auditory-verbal representations affect STM performance. Finally, there are conflicting outcomes regarding the specific neural markers for inner speech, such as M/EEG oscillatory dynamics and fMRI spatial activations, which could also be a result of problems with uncontrolled strategic and timing differences.

In order to address these issues, a rehearsal task is needed that 1) demands the use of a single, well-defined rehearsal strategy, 2) controls or measures the precise timing of the covert speech activity, 3) requires only the minimally necessary elements of internal speech, without any involvement of overt sound production or motor execution, and 4) attempts to access rehearsal as it occurs, rather than relying on recall as a dependent measure. The high temporal resolution of EEG and MEG makes these methods potentially useful in conjunction with the behavioural task for revealing neural markers of both the temporal properties and contents of subvocal rehearsal.

1.7 A Novel Measure of Rehearsal Timing

In this thesis I will present a novel measure of the temporal properties of rehearsal. The rehearsal-probe task is described in detail in Chapter 2. In brief, participants are presented with a list of auditory-verbal material. The participants are instructed to rehearse the list exactly as it is presented, that is, silently reproducing the sequence items in the same order and at the exact same rate. After an unpredictable rehearsal delay, subjects are probed for a response with a brief tone. Upon hearing the tone, participants must indicate the list item that was currently being rehearsed at the moment that the probe was presented. The participants complete several trials with probes presented after variable delays, and with repeated trials at the same delay, which together provides distributions of responses within and across probe times.

This task produces a rich data set consisting of multiple probability distributions for the activations of each list item over time, and these distributions can be summarised by a single variability measure (circular standard deviation, see Chapter 2) or described in more detail using model parameters. The model, which is presented in Chapter 6, estimates the values of four parameters that we expect to correspond to theoretical temporal properties of rehearsal; a temporal offset (i.e. shift), rate, constant imprecision, and cumulative imprecision. Where timing precision is an important factor in auditory-verbal STM, it is expected that subvocal rehearsal timing properties will relate to individual and group differences in STM span, or to STM task manipulations that are known to affect performance.

Finally, the measures of rehearsal timing obtained from this task can be used with high temporal resolution signals obtained from electromagnetic psychophysiological recordings in order to reveal possible neural correlates of subvocal rehearsal. Chapter 7 describes the use of this task with EEG, where I compare ERPs to matching and mismatching rehearsal list items, based on the known sequence presentation timing. I also compare oscillatory activity during sequence presentation, rehearsal delays, and baseline intervals in an effort to detect neural markers that correspond to the rate of subvocal rehearsal.

1.8 Conclusions

Many researchers have argued that subvocal rehearsal, the active and silent repetition of speech-based material, is a key contributor to STM performance. Rehearsal has been specifically implicated in verbal ISR tasks, as rehearsal typically involves the forward-ordered cyclic rehearsal of verbalisable material. The importance placed on rehearsal in these tasks is based on a foundation of empirical results from behavioural, physiological and neuropsychological that consistently demonstrates a variety of separate phenomena that can be parsimoniously explained by subvocal phonological articulation during STM tasks. However, others have argued that too much emphasis is placed on the importance of rehearsal in STM, and that the empirical evidence typically used to support a role for rehearsal can be equally or better explained by other factors.

A key reason for the lack of resolution concerning these issues is that the majority of the evidence for the role of rehearsal in verbal STM relies on indirect methods of detecting rehearsal, and on methods of manipulating the rehearsal process which require theoretical assumptions about the underlying mechanisms involved. In addition to behavioural methods, subvocal speech has been studied using EMG, fMRI, EEG and MEG. While these more direct methods have shown that subvocal speech is measureable and suggests that it is related to overt speech, the majority of the research using these methods has not focused on the role and properties of subvocal speech in the context of STM. Thus, a direct measure of subvocal rehearsal is still needed in order to elucidate the role and relative importance of the rehearsal process in auditory-verbal STM.

Not only would a direct measure of rehearsal contribute to a better understanding of the role of this process, it would also provide the opportunity to examine the nature of the internal temporal structure of auditory-verbal sequence representations in greater detail. Recent models of STM have proposed that memory for serial order information is separate from memory for item information, and a number of these models assume that temporal information is required for the representation of serial order (discussed in Chapter 8). Some empirical results have also shown that the ability to reproduce an auditory rhythm correlates with the ability to retain verbal material in the correct serial order. It is therefore possible that the accuracy and precision of internal timing representations of subspan auditory-verbal sequences varies dynamically with changes in memory load, which are known to affect within-subject serial order recall, and is predictive of individual differences in STM capacity for serial order.

Furthermore, there is increasing evidence that temporal STM and STM for serial order are uniquely linked to speech, literacy and word learning outcomes. The observation that individuals with disorders of speech and literacy tend to show poorer performance on tests of timing and rhythm have supported the notion that these disorders may be related to a more central deficit in the encoding and storage of auditory temporal information. EEG and MEG studies have provided further support for this notion, showing that individuals with dyslexia tend to have abnormal neural phase-locking to speech rhythms. An underlying deficit in the perception and memory for temporal information may be a contributing factor to impaired phonological development during early childhood, due to difficulties in segmenting ongoing speech into meaningful units and predicting the temporal onsets of these units. This would explain the fact that individuals with disorders such as developmental dyslexia and specific language impairment (SLI) perform below average in tasks requiring an awareness or manipulation of phonemes. This theory also accounts for the higher occurrence of serial order errors reported in the disordered groups, as the retention of serial order information reflects the accurate coding of the timing of items relative to one another. Together, these lines of evidence suggest that the temporal accuracy of verbal representations may be the critical factor linking auditory-verbal STM, and STM for serial order in particular, to speech, reading and vocabulary skills.

In order to shed light on these issues, the rehearsal-probe task was developed as a novel method of directly measuring the temporal properties of subvocal rehearsal activity. The aim of this method is to probe the contents of paced inner speech in order to obtain a reflection of the temporal properties of the participants' auditory sequence representations. If the accurate and precise maintenance of auditory-verbal sequence timing is a limited resource in STM, then it is expected that task manipulations known to affect ISR by loading on shared

STM resources will result in changes to rehearsal timing variability. Moreover, we may also expect to find that specific temporal properties of rehearsal are related to individual differences in STM performance, and possibly to speech and reading impairments. A computational model will be presented with the aim of examining whether estimates of multiple independent timing properties can be identified from the rehearsal data, and if so, whether any of these properties have any additional predictive power beyond a simpler measure of timing precision. Finally, it is predicted that the measures of rehearsal timing obtained from this task can be used with high temporal resolution signals obtained from EEG recordings in order to reveal possible neural correlates of subvocal rehearsal.

CHAPTER 2

METHODS

There are two main tasks used in the experiments presented in this thesis; a novel rehearsal-probe task and a standard auditory digit span task. In this chapter I will describe the audio stimuli used in both tasks. I will then outline the general procedures and data analysis methods for each task in turn. With regard to the rehearsal-probe task, some experiments used variations of this task, and these variations are described in the appropriate methods sections in the experimental chapters. There were additional tasks used in some experiments, and these tasks are described in the appropriate methods sections of experimental chapters as well.

2.1 Audio Stimuli

The spoken digits from 0 to 9 were digitally recorded in a sound-attenuated booth by a female native English speaker at 44100 Hz sampling rate as a single channel (mono). Post-recording processing on the stimuli was done using Audacity (Mazzoni & Dannenburg, 2000) and Praat (Boersma & Weenink, 1992) audio software. All digit recordings were adjusted to equal duration (400 ms) and maximum amplitude using the 'tempo' and 'max amp' features in Audacity. The 'tempo' function extends or compresses the audio stimuli to a specified duration without altering the pitch. The 'max amp' function adjusts the amplitude (volume) of the audio file to the maximum level without clipping the sound wave.

The digit stimuli underwent further processing in order to reduce pitch fluctuations in order to control the effect of pitch changes on the perception of digit sequence timing. The grand mean pitch of all digits was obtained using the mean pitch output data in Praat. For each digit, the pitch contour was extracted and shifted such that the mean pitch for the digit matched the grand mean pitch. The distance from the grand mean pitch to each pitch contour point was reduced by 50% for all digits. This 50% reduction in the pitch contour deviations was made to reduce the magnitude of the natural and variable pitch fluctuations produced by the speaker's voice while simultaneously limiting the unnatural resonance sounds that result from flattening the pitch to a greater degree. The probe stimulus used in the rehearsal-probe task was created by adding 15 ms onset and offset ramps to a 150 ms 440 Hz steady-state tone.

2.2 Perceptual-Centre Adjustment of Digit Stimuli

In order to create digits that could be concatenated into randomly-ordered, perceptually isochronous sequences, it was necessary to estimate the perceptual centre (p-centre) of each digit. The p-centre is the perceived 'beat' of a word, i.e. its "psychological moment of occurrence" (Morton et al., 1976, p. 405). The p-centres of words are dependent upon their initial acoustic properties, although this relationship is not a straightforward one (Marcus, 1981; Morton et al., 1976). Because the spoken digits from zero to nine have different initial acoustic and phonemic properties, they are not perceptually isochronous when evenly spaced in time according to acoustic onset time.

The p-centres of each digit were estimated using a rhythmic adjustment task (Morton et al., 1976). In this task, participants heard alternating presentations of one of the digits from zero to nine (the target) and the spoken digit "one" (the reference) in a continuous loop. The participant's task was to adjust the timing of the target digit relative to the steady "one" sound until s/he judged the two sounds to be occurring at equal intervals. The final intervals for multiple trials with the same target digit were averaged within individuals, and a grand average was obtained across individuals for each digit. The grand average adjusted offsets for each digit provided the basis for the temporal shifts in acoustic onsets for each digit in the rehearsal-probe and digit span tasks.

2.2.1 Methods

2.2.1.1 Participants

Eight postgraduate students from the University of York took part in exchange for a small cash reward. All participants were native English speakers, and none had any self-reported hearing impairments or learning disabilities.

2.2.1.2 Design

A within-subject design was used; all participants completed adjustment trials for all ten target digits and all starting interval durations. The dependent measure for each trial was the offset time (in ms) of the target digit acoustic onset relative to the midpoint of the reference digit onset-to-onset interval. A positive offset time indicated that the target digit acoustic onset was adjusted to occur later than the midpoint of the reference digit onset-to-onset interval, while a negative offset time meant that the target digit acoustic onset was adjusted to occur earlier than the midpoint.

2.2.1.3 Materials

The stimuli used were the spoken digits from zero to nine described in an earlier section in this Chapter.

2.2.1.4 Procedure

In each trial, the target and reference digits were presented binaurally through over-ear headphones via a custom E-Prime program. Participants were instructed to listen to the reference-target sequences and adjust the target sound until it sounded as though the two sounds were occurring at equal intervals, like a metronome. The timing of the target digit was controlled with four keyboard keys, corresponding to smaller or larger shifts forward or backward in time (see Figure 4). The experimenter explained that if it sounded as though the target digit was occurring too early in the loop then the participant should use one of the forward/plus (+) buttons, whereas if it sounded like the target was occurring too late then the participant should use one of the backward/minus (-) buttons. Participants were told that pressing a single-symbol button (+ or -) would produce a smaller shift in the target timing, while pressing a double-symbol button (++ or --) would produce a larger shift. The single-symbol buttons corresponded to 10 ms shifts forward or backward, while the double-symbol buttons resulted in +/- 50 ms changes. Because the timing shifts were subtle, visual feedback would also appear on the screen in response to a shift, indicating that the button press had registered. The visual feedback was in the form of arrow(s) indicating the direction and magnitude of the shift. The symbols ‘>’, ‘>>’, ‘<’, or ‘<<’ appeared on the screen to indicate small or large forward or backward temporal shifts. The experimenter then demonstrated the use of these buttons to adjust the timing of two different target digits, and the participant had the opportunity to practice using these buttons to adjust a target digit before starting the task.

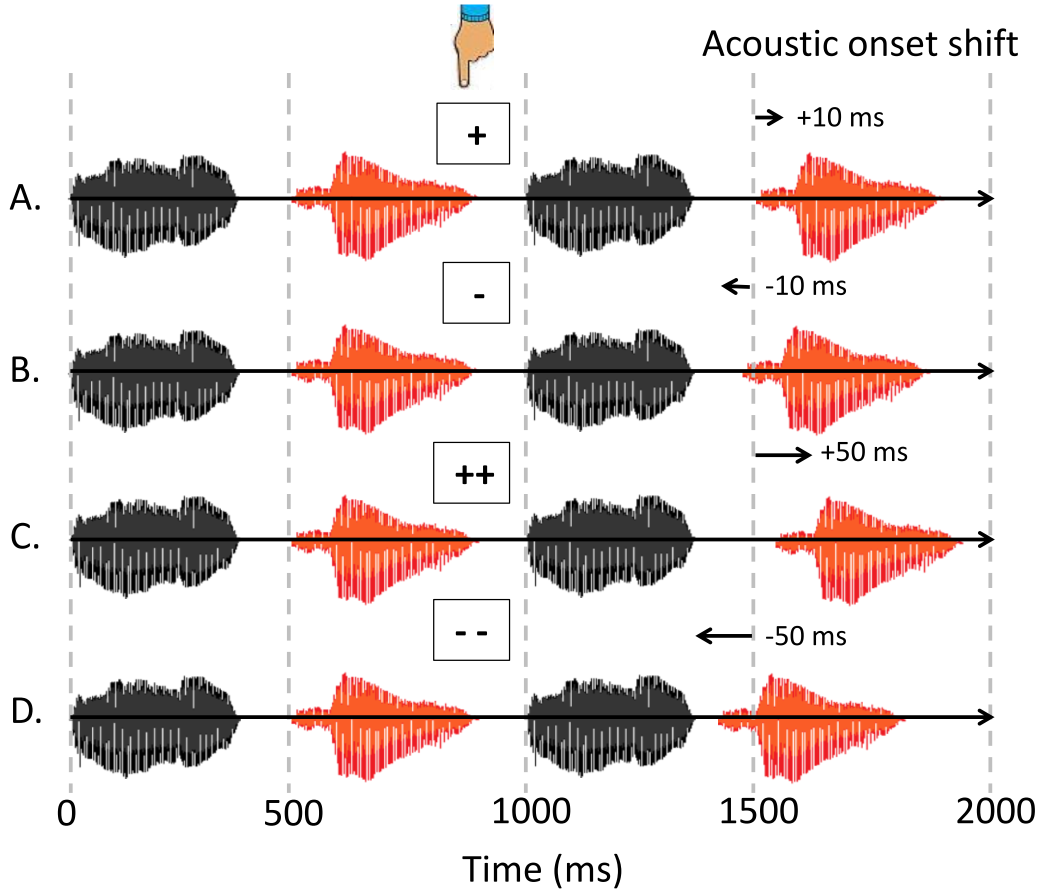

Figure 4. Schematic representation of the timing of reference and target audio files in the p-centre adjustment task. Each row A-D represents a possible audio sequence occurring within a single trial, where the reference sound (black) and target digit (red) are presented in a continuously alternating fashion. When the participant presses an adjustment button (represented in the figure by the finger and boxes with + and – symbols), the onset of the target digit is adjusted accordingly in the next presentation of the target digit. Rows A-D show the four possible adjustments that the participant can make to the timing of the target digit during the task. Here the target adjustments are shown relative to the midpoint of the reference onset-to-onset interval, but note that the target acoustic onset shifts are always relative to the target acoustic onset times on the previous trial.

There was no time limit for the trials; rather, participants were encouraged to continue listening to the reference-target loop and adjusting the target timing until they were satisfied that the two sounds were occurring at equal intervals. Once satisfied, the participant pressed the Enter key to end that trial and start the next. There were blocks of five trials for each target digit, with self-paced breaks between blocks. Participants were told that the starting intervals would vary randomly for each trial.

2.2.2 Results

The reference acoustic offset to target acoustic onset times (in ms) were recorded at the end of each trial. These durations were averaged over the five trials per target digit, giving each participant a single mean adjustment value per digit. These adjustment values were then averaged over the eight participants. The grand means and between-subject standard deviations are shown in Table 1.

Table 1. Means and standard deviations of the adjusted reference offset to target onset intervals for each digit. A mean of 100 ms indicates that the target digit was adjusted to perfect onset-to-onset isochrony relative to the reference digit, whereas values below and above 100 indicate that the target digit was adjusted to occur earlier or later, respectively, than the midpoint between the reference onset-to-onset interval.

Digit	Mean (ms)	Standard Deviation (ms)	Pre-acoustic silence (ms)
0	97	37	6
1	100	38	9
2	101	36	10
3	106	30	15
4	93	30	2
5	93	31	2
6	102	49	11
7	91	36	0
8	123	35	32
9	98	34	7

To determine the values of the silent offsets at the beginning of each sound file, the digit with the smallest adjusted mean interval from reference offset to target onset was identified. This was the digit 7, with a mean reference offset to target onset interval duration of 91 ms. The pre-acoustic silent durations for the other digits were then calculated by subtracting 91 from the corresponding mean for each digit. The resulting pre-acoustic silent intervals, listed in Table 1, were added to the start of the audio files.

2.2.3 Discussion

The results of this experiment were roughly consistent with those reported by Morton, Marcus and Frankish (1976). These authors measured the p-centres to the spoken digits from one to nine and found that the digits with the latest p-centres were those with initial sibilant fricative phonemes (“seven”, “six”), followed by non-sibilant fricatives and nasals (“four”, “three”, “five”, “two”, “nine”). The digits with the earliest p-centres were those starting with vowels (“eight”, “one”). In the results of the present experiment, the latest p-centres were found in some of the digits beginning with fricatives (“seven”, “four”, “five”, “zero”) and, consistent with Morton et al., the spoken digit “eight” had the earliest p-centre. However, the p-centres for some fricatives (“six”, “two”, “three”) were slightly later than expected, and the p-centre for the spoken digit “one” was earlier than expected. Overall, there was also less variation in the p-centre offsets among the spoken digits in the present experiment relative to those reported by Morton et al. The slight discrepancies between the present results and those of Morton et al., as well as the reduced variation in p-centre timing across digits, is very likely due to the processing done on the digit recordings used in the present study. The audio processing performed on the digit stimuli prior to the p-centre adjustment task was intended to normalise the stimuli in terms of duration, amplitude and pitch fluctuation. This normalisation affected some of the acoustic determinants of p-centre latency, thereby differentially affecting the individual digit latencies and reducing the latency differences across digits.

2.3 Rehearsal-Probe Task

2.3.1 General Procedure

In this task, random digit sequences are presented binaurally over headphones via E-Prime software at a rate of 2 Hz. The actual acoustic onset times of the 400 ms digits are offset by the silent padding (see Perceptual Centre Adjustment section in this chapter). The offset of the digit is followed by up to 100 ms silence, depending on the initial acoustic onset, which results in one digit per 500 ms. The random digit sequence is presented twice at a steady rate in order to provide rhythmic continuity between the end and beginning digits in the sequence. The variable parameters in this task include the number of digits in the sequence, the timing of the digit sequence presentation, the rehearsal delay durations (i.e. the probe presentation times), and the number of trials per probe time. These parameters differ across experiments and will be stated in the methods sections of each experiment. Figure 5 shows a representation of a single trial from the rehearsal-probe task using a 4-digit sequence.

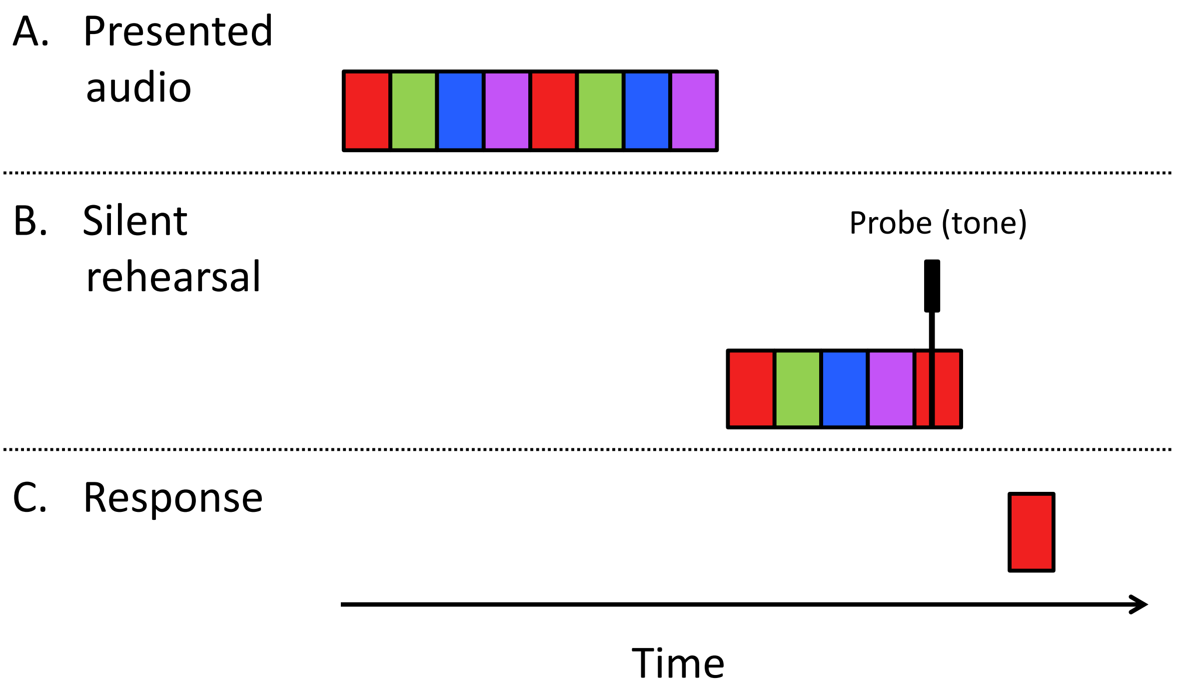

Figure 5. Schematic representation of a single trial in the rehearsal-probe task. Each rectangle represents a spoken digit and inter-item interval in a random 4-digit sequence, where the four colours denote the serial positions 1-4. Row A shows the audio presentation of two consecutive isochronous repetitions of the 4-digit sequence. At the end of the second sequence presentation, the participant immediately begins silently rehearsing the sequence exactly as it was presented (row B). Upon hearing the probe (150 ms tone) during the rehearsal interval, the participant responds with the item that s/he was rehearsing at the same time that the tone was presented (row C). The tone also signals the end of that trial, and there is a brief inter-trial interval before the next audio sequence is presented.

Participants are told that in each trial they will hear a sequence of digits repeated twice, and that their task is to silently rehearse the digits exactly as they hear them. They are told to continue silently rehearsing the list in this manner until they hear a tone, at which point they should respond with the digit that they were rehearsing at the time of the tone. The task instructions specifically emphasise that participants should rehearse the digits in the same order and at the exact same pace. The experimenter also directs participants not to speak aloud or move any part of his/her body during the trials, as any overt speech or repetitive movements during the rehearsal period may reduce the load on STM resources. The participants' movement is monitored by the experimenter present in the room during the testing session.

Before starting the task, participants observe the experimenter demonstrate three trials, including responses, with the rehearsal portion of the task spoken aloud. After hearing the tone, the experimenter explains "I was saying the number X when I heard the tone, so I'm responding with number X" before pressing the corresponding key on the computer keyboard. Participants then complete six practice trials; three trials with the rehearsal portion spoken aloud and the other three trials with silent rehearsal. The overt rehearsal practice trials allow the experimenter to verify that the participant understands the task and determine whether the participant is able to correctly identify the articulated digit at the time of the probe presentation. If the participant responds incorrectly during the overt practice trials then the experimenter reviews the instructions and the practice trials are repeated. No feedback is given about the participant's overt rehearsal timing, except in the case of serious deviations from the task instructions, such as rehearsing the digits as quickly as possible, or stopping or pausing after the first rehearsal list cycle. In these cases, the experimenter clarifies the instructions and the participant then repeats the overt practice trials.

2.3.2 Response Probability Plots

The participant's response to the probes on a single trial is coded according to the serial position of the digit response. For example, if the digit sequence on a given trial is "2 7 9 0" and the participant responds the probe by pressing the "7" key, then that response is coded as serial position 2. Data from one participant and condition consists of a set of serial positions obtained at each probe time (time from the end of the presented sequence to the onset of the probe). This data can be visualised by plotting the proportions of responses for each serial position obtained at each probe time, as shown in Figure 6. Probe times (i.e. rehearsal interval durations) are shown along the x-axis, reflecting the progression of time during the probed section of the silent rehearsal interval. The points indicate the marginal proportions (plotted on the y-axis) of a given serial position response at each probe time. The points are colour-coded according to the serial position of the response probabilities, and the points for each serial position and probe time are connected by a line to depict the interpolated response probabilities for each serial position between probe times. At the top of the plot, a set of boxes are shown that correspond to the 'correct' item timing, i.e. the timing of the items if they were rehearsed exactly as presented during the silent interval.

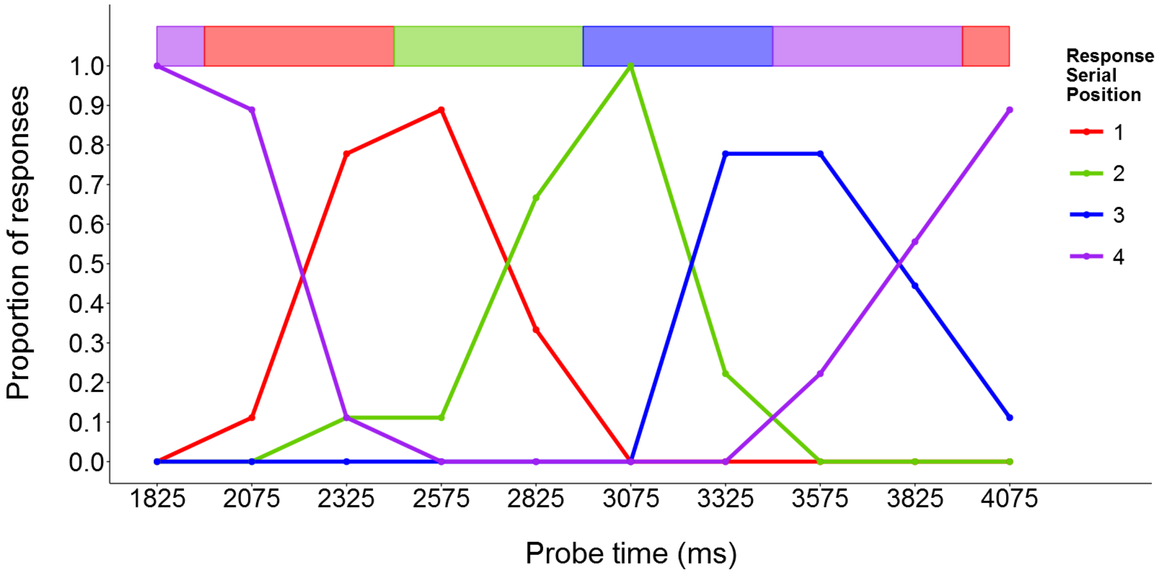

Figure 6. Example of a response proportion plot from a single participant in Experiment 6.

There are two key features of the response probability plots that reveal relevant characteristics about the participant's rehearsal timing. The first feature to note is the order of the centres or peaks of each serial position curve, and correspondence between these peaks and the boxes indicating the presentation sequence timing at the top of the plot. In Figure 6, the maximum proportions of the serial position occur in the expected order, which provides convincing evidence that the participant rehearsed the items in the correct forward serial order throughout the task. The frequency of these peaks gives some indication of the rehearsal rate, where a contraction in the peak frequency suggests increased rate relative to the presentation rate and an expansion in this frequency suggests a slower rate. Discrepancies between the peaks of the response distributions and the boxes at the top of the plot may also indicate a constant shift in the participant's rehearsal timing, for example if they begin rehearsing the sequence too early or too late at the end of the sequence presentation. The peaks of the serial position response distributions in Figure 6 lag slightly behind the correct timing, which indicates that this participant may have paused before beginning to rehearse or rehearsed at a slower rate.

The second feature of these plots, which is the more relevant aspect for the purposes of the current research questions, relates to the shapes of the serial position curves. Temporal precision during rehearsal is reflected by the heights and widths of the peaks, as well as the spread of the tails of the distributions over probe times. In the task used in Experiment 6, which produced the data plotted in Figure 6, there were two probe times per digit during the probed rehearsal interval (apart from the first and last probe times, of which there were only one per digit). If the digit timing is represented with high precision during rehearsal, then the response proportions for each serial position would be very high (close to 1) for two consecutive probe times, and very low (close to 0) for all other probe times. Therefore, high temporal precision in the rehearsal-probe task will produce serial position response proportion distributions that resemble a narrow, uniform distribution. As precision of the digit timing decreases, the tails of the distributions widen, and there is greater overlap in the distributions of response proportions across serial positions. Another consequence of greater variability or uncertainty in the digit timing is that the peak response proportions for each serial position will decrease due to the increase in responses for neighbouring items. Low temporal precision in this task is thus characterised by wide serial response proportion distributions with low peaks, and as timing precision decreases, the response proportion distributions will approach a wide uniform distribution across all probe times at the level of chance response proportion (one divided by the number of serial positions).

The data from each participant/condition produces a single response proportion plot. The individual sets of response proportions for each serial position and probe time can then be averaged over participants, resulting in mean serial position response proportion values with some measure of between-participant variance (see Figure 7). The between-subject variance (standard error) surrounding these values is represented by the width of ribbons plotted around the mean proportion values, and extended linearly between probe times.

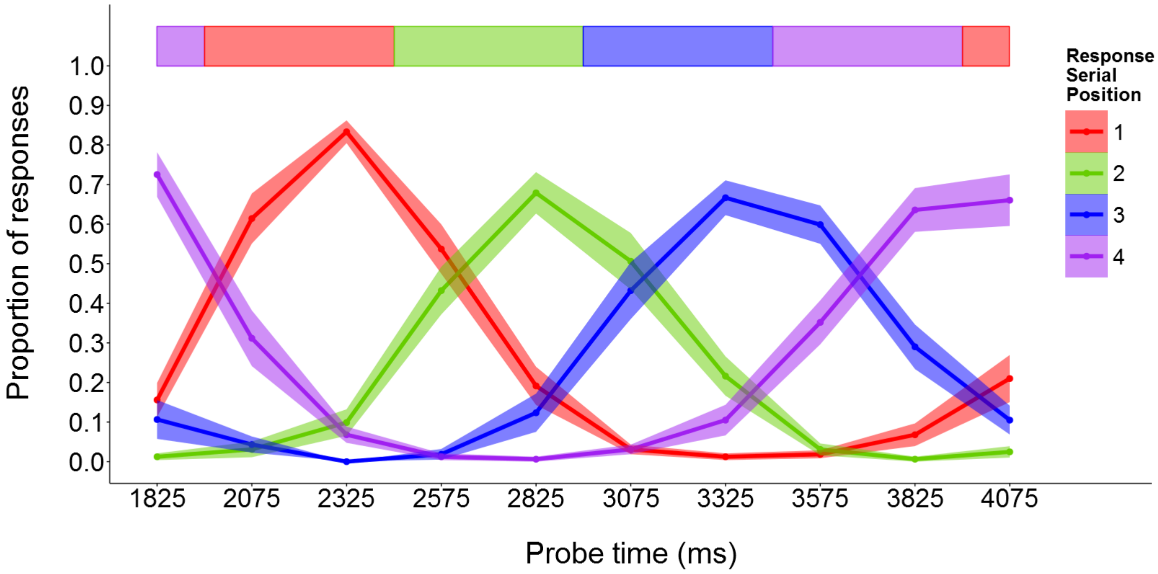

Figure 7. Changes in mean response proportions for each serial position across probe times. Response proportions in this example are averaged over the control group participants from Experiment 6. The width of the ribbon at each probe time reflects the between-subject standard error at that point.

2.3.3 Circular Standard Deviation Measure

As the focus of this thesis is on temporal variability, the dependent measure was chosen to reflect the variability of responses across probe times. There are two methods of looking at the variability of responses within each rehearsal-probe data set. The data can be viewed in terms of A) the sets of response serial positions at each probe time (within-probe variability), or B) the sets of probe times that resulted in each response serial position (within-position variability). Thus, variability can be expressed in terms of either the consistency of serial position responses within each probe time, or the consistency of probe times that led to each serial position response. These two alternatives can be visualised in Figure 7, where within-probe variability reflects the mean variability in response proportions at any single probe time along the x-axis, and within-position variability reflects the spread of the response proportion distribution over time within each serial position. For example, at the 2325 ms probe time, there is relatively low mean within-subject response variability, with participants responding with the digit in serial position 1 on most trials. In contrast, at the 3075 ms probe time there is greater mean response variability, with a near-even proportion of responses for serial positions 2 and 3. In terms of the variability of the probe times associated with each serial position response, the shape of the serial position 1 response distribution (red) appears slightly less variable over probe times than the other three serial position response distributions.

The experiments presented in this thesis present at least two probes during the duration of a single digit, which means that there are at least twice as many probe times as there are possible serial response positions. Given the relationship between item durations and probe frequency, it was expected that participants may respond with perfect consistency at a single probe time (i.e. within-probe variability, standard deviation of zero), but it would be very unlikely for a serial position response not to vary across more than one probe time (i.e. within-position variability, standard deviation of zero). Therefore the distributions of probe times were analysed within each response serial position in order to obtain a more continuous and less biased measure of variability.

Circular standard deviation (CSD) is used as a measure of variability. The CSD measure is similar to a typical linear standard deviation measure in that it captures the variation or dispersion around the average. In the case of the rehearsal-probe data, the average refers to the mean probe time across all responses that correspond to a particular serial position. A circular analysis method was applied because of the cyclical nature of these response distributions over time. The cyclic distributions make linear descriptive statistics inappropriate, a point that is illustrated in Figure 8. If the peak proportion of responses for a given serial position falls squarely within the probed time interval, then the responses tend to take a normal Gaussian distribution (Figure 8, A) and thus can be described by linear statistics; the means and standard deviations of this distribution are what we would expect. However, if the probed time interval captures partial distributions from responses which do not fall entirely within the boundaries of the interval (Figure 8, B), then linear statistics will provide very a misleading description of the rehearsal data.

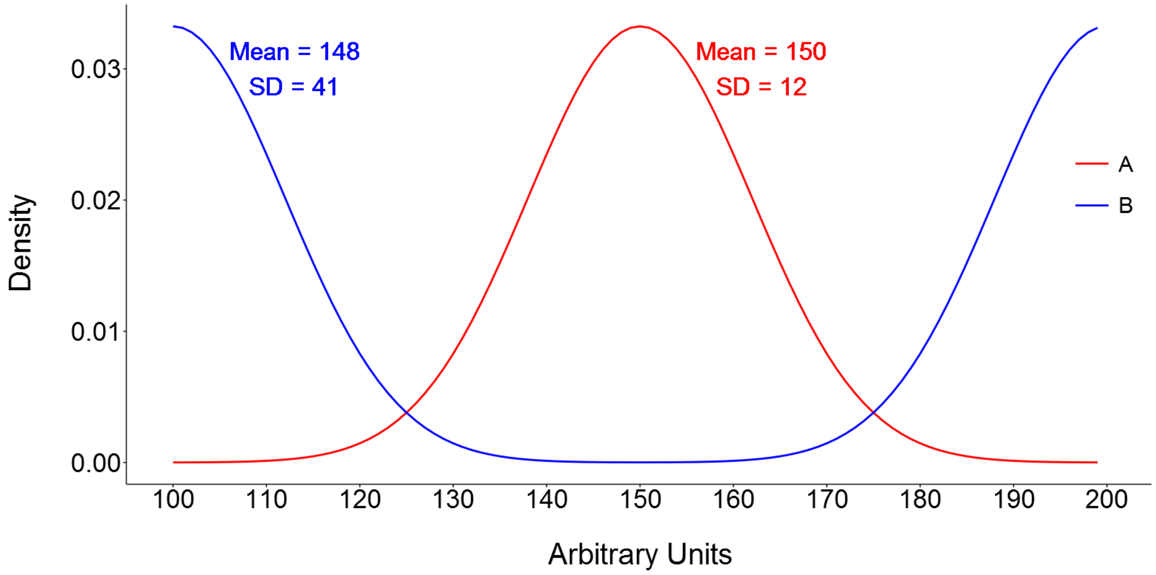

Figure 8. Linear means and standard deviations of two distributions, A and B. Note that the standard deviation of the distribution B (blue) is much greater than the standard deviation of the distribution A (red).

This issue was addressed by representing the duration of the rehearsal cycle as cyclical rather than linear, where the distribution of probe times for each serial position response are represented in terms of positions around a circle. To calculate the CSDs for each response serial position, the probe times are first converted into radians such that the beginning and end times of the rehearsal cycle are represented at the same point around the circle (0 or 2π radians). The distribution of probe times in radians corresponds to their distribution in real time with respect to the rehearsal cycle start and end times. The CSDs are calculated for the distributions of the probe times (in radians) that correspond to each response serial position. CSD calculation is performed using MATLAB (The MathWorks Inc., 2010) and the MATLAB CircStats toolbox (Berens, 2009), and will be described briefly here (see Appendix for a full description of the CSD calculation). First, the mean resultant vector is computed by averaging the unit vectors that correspond to each data point (i.e. probe time, in radians). The mean resultant vector has a direction and length, and the length of the resultant vector is related to the dispersion of vector angles in the sample. The resultant vector length falls between 0 and 1, where a value closer to one indicates more concentration of the sample around the mean direction (see Figure 9).

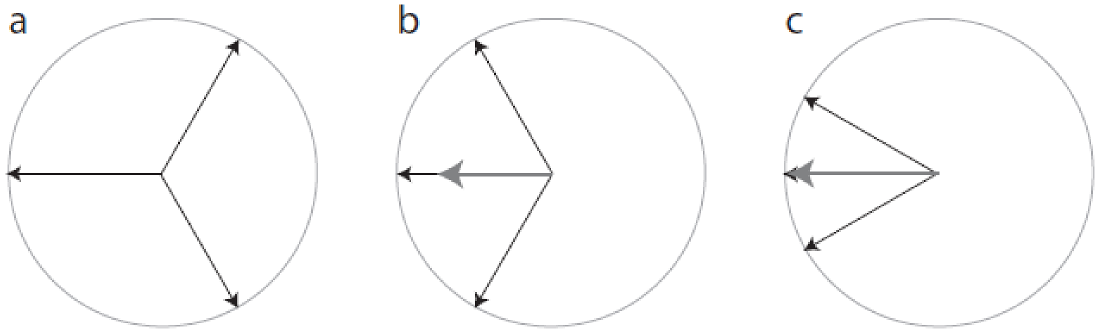

Figure 9. Examples of resultant vectors for three unit vectors. The resultant vectors are obtained by vector addition. In (a), the three unit vectors are evenly spaced around the circle, so the resultant vector has a length of zero. This case represents the largest possible CSD value of √2 (see equations in text). In (b), the resultant vector (grey) is formed by vectors at 120, 180 and 240 degrees, resulting in a mean direction of 180 degrees and a length of 2/3. In (c), there is less dispersion in the unit vector directions (150, 180, and 210 degrees) relative to (b), which results a longer vector length of 0.91 with the same mean direction. Source: Berens (2009, p. 5).

CSD is based on the circular variance. The circular variance S is computed using the formula

$$S = 1 - R$$

where R is the resultant vector length. Recall that R lies in the interval [0, 1], where a value of 1 is perfect consistency, so the inverse relationship is true for S; values of S close to 0 indicate low variability, and values of S close to 1 indicate high variability. The CSD (also termed angular deviation) is defined as

$$s = \sqrt{2 * \mathrm{S}}$$

where the resulting CSD quantity s lies within the interval [0, √2]. It is important to note that the maximum CSD value, √2, can result either from a uniform distribution of points around a circle, or from an even distribution of points that are separated by half the circle length (π radians or 180 degrees). It is therefore important that the length of the circle represents a single rehearsal cycle duration, where probe times that correspond to the same within-item timing on different rehearsal cycles are represented by the same angular values.

When circular statistics are applied to the two distributions plotted linearly in Figure 8, the results are very different. To demonstrate this, the x-axis values in Figure 8 were mapped to angular values around the entirety of a circle (0 to 2π radians). The radian values were then sampled according to the probability densities shown in Figure 8 to form a large sample of radian values (N = 10000) for distributions A and B. The circular histograms of these two distributions are shown in Figure 10, along with the circular mean and standard deviation values. The circular means and standard deviations more accurately describe the B distribution than the linear values, provided that the distribution is cyclical, and that the sampled interval reflects one period of the cycle. In the case of the rehearsal-probe task, serial position responses are indeed cyclical, and the probed interval mapped to the angular values of a circle is the length of a list cycle.

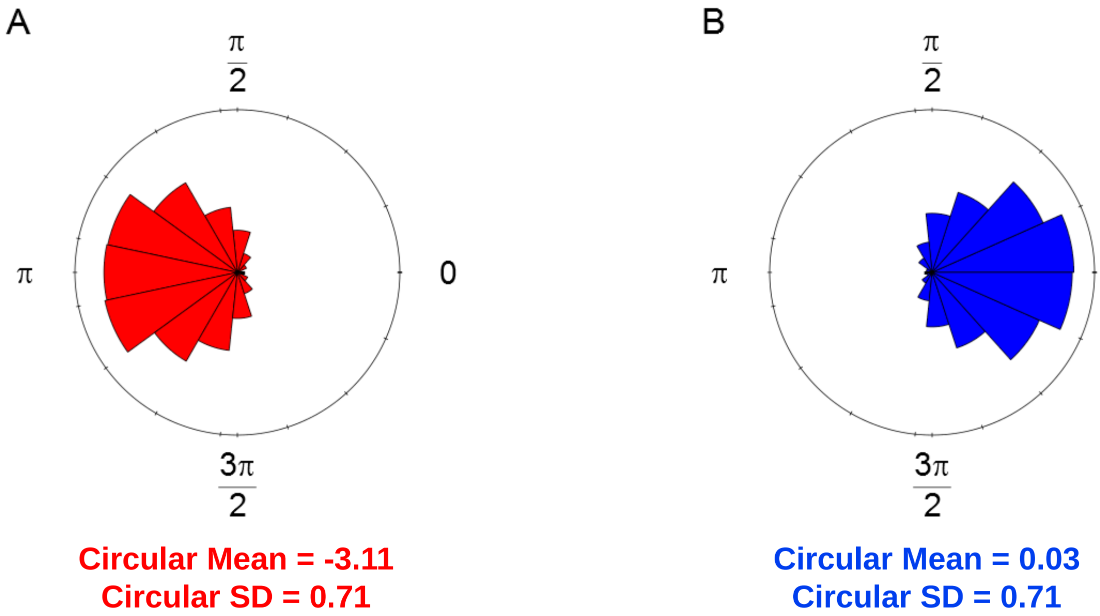

Figure 10. Circular histograms, means and standard deviations of the two distributions A and B that are shown as linear distributions in Figure 8. Means and standard deviations are given in radians. Note that the circular standard deviations of these two distributions are the same, and the circular means (A ~= π, B ~= 0) are different.

CSD values are calculated for the distributions of each serial position across probe times within each data set (participant/condition). In some cases it is interesting to investigate response variability as a function of serial position. However, for most of the experiments presented in this thesis, there were no specific hypothesis with regard to temporal variability of specific serial positions, and the aim was to determine whether any relationships exist between a general measure of temporal variability and STM capacity and span. Therefore, to obtain a single measure of response variability per data set, the CSD values are averaged over serial positions. This dependent measure is termed 'mean CSD', where 'mean' refers to the averaging of CSDs across serial positions within a participant/condition.

2.4 Digit Span Task

2.4.1 Procedure

The digits were presented via an E-Prime program to ensure consistency in stimulus presentation. Participants heard three different digit sequences of equal length. The three sequence sets increased in length by one digit, starting with a list length of three digits. The audio files used in this task were the same as those used in the rehearsal task, and were paced at one digit per second. For each sequence, the participants' spoken immediate serial recall was recorded by the experimenter. The task continued until the participant completed all three trials at the longest list length (12 digits) or responded incorrectly to all three sequences of a given length. Thus the continuation criteria is at least one correct response out of three trials at a given list length.

2.4.2 Scoring

Digit span scores were calculated as the average list length of the last three correct trials. For example, if a participant correctly recalled one of three 5-digit, 6-digit and 7-digit sequences but did not correctly recall any 8-digit sequences, then their score would be 6.00 (calculated as (5 + 6 + 7)/3). If a participant correctly recalled two 6-digit and one 7-digit sequences but did not correctly recall any 8-digit sequences, then their score would be 6.33 (calculated as (6 + 6 + 7)/3). The maximum possible score is 12.00.

CHAPTER 3

INITIAL TESTS OF THE REHEARSAL-PROBE TASK PARADIGM AND THE EFFECTS OF MEMORY LOAD ON TEMPORAL PRECISION

The general aims of the set of experiments presented in this chapter were twofold. The first aim was to establish the rehearsal-probe task paradigm and determine its potential as a method for examining the temporal properties of verbal sequence representations in STM. Toward this end, the goals were to examine the characteristics of the response data, test the mean CSD measure as a dependent variable, and refine the experimental design. The second general aim was to investigate the relationship between temporal precision in STM and memory load.

Performance on auditory-verbal ISR tasks varies as a function of various task characteristics, including properties of the to-be-remembered words, sequence timing, and concurrent demands on attention and processing (Dempster, 1981; Hall & Gathercole, 2011; Larsen & Baddeley, 2003; Ryan, 1969b). There is some degree of both item- and order-related memory demands in typical ISR tasks but in the case of sequences of short, familiar words without syntactic structure, such as random digit lists, there is relatively little cognitive demand placed on the maintenance of the item information, and greater demand on STM processes related to retaining the serial order (Hulme, Roodenrys, Brown, & Mercer, 1995; Jefferies, Frankish, & Lambon Ralph, 2006; Saint-Aubin & Poirier, 1999b). Memory for order information may therefore be particularly important in explaining performance on auditory ISR tasks such as digit span. However, the extent to which the recall of serial order depends on accurate STM for sequence timing is not well understood. In particular, it is unclear whether the fine-grained precision of the representation of sequence timing impacts the ability to recall items in the correct order. Conversely, it may be that the ability to maintain auditory-verbal sequence timing beyond only a coarse representation of relative item timing (i.e. only enough to discriminate the serial order) has no effect on ISR performance. The aim in the current set of experiments was to determine whether the temporal precision of auditory-verbal sequence representations in STM varies as a function of the task difficulty.

Current models of visual STM capacity limitations can be categorised as fixed/discrete slot models or dynamic resource allocation models. A purely ‘fixed slot’ model would predict that a number of items can be remembered with a given precision, and that no information is retained for additional items beyond the number of discrete slots (see e.g. Awh et al., 2007;

Ester et al., 2014; Fukuda et al., 2010). By contrast, a dynamic resource account proposes that a shared pool of resources can be used to represent items of any number or complexity, where the precision of the item representations decreases as the overall cognitive load increases[4] (see e.g. Bays et al., 2009; Bays & Husain, 2008; Ma, Husain, & Bays, 2014). These different characterisations of memory limitations can be tested by manipulating cognitive load, for example by increasing the number of items to be remembered or their complexity, and then examining the effect on precise recall for the features of items that vary on a continuum.

While capacity limitations have been widely researched in the visual modality, little work has been done to determine whether there is an analogous issue in auditory STM. One exception is a recent study by Kumar et al. (2013), where participants' recall precision for the pitch of auditory tones was measured with a continuous pitch-matching paradigm. Kumar et al. showed that the precision of pitch recall decreased with increasing memory load (number of tones in the sequence). In the experiments presented in this chapter, the aim was to test whether a similar trade-off exists between memory load and the temporal precision of auditory-verbal sequence representations.

In the four experiments presented here, the general aims were to 1) establish and refine the rehearsal-probe task method, and 2) determine whether STM load affects the temporal precision of auditory sequence representations. The expectation was that the response proportions for each serial position would be characterised by a cyclic increase and decrease pattern over probe times. It was also predicted that the spread of the response proportions for each item over time would vary meaningfully between individuals and/or within individuals under different conditions. Memory load was manipulated using the number of items in the rehearsal sequence (Experiment 1), the predictability of item-order associations (Experiment 2), a pre-load paradigm (Experiment 3), or temporal grouping (Experiment 4). The prediction was that temporal precision of sequence representations during rehearsal would not be fixed, as might be predicted by a discrete-slot interpretation of the relationship between temporal precision and load, as well as by positional coding models of STM for serial order. Instead, it was expected that there would be an inverse relationship between temporal precision of sequence representations and memory load.

---

[4] While a dichotomy between 'fixed/discrete' and 'dynamic resource' models is used to discuss these issues in the visual WM literature, many models of visual WM capacity are more nuanced than this, often involving aspects of both the fixed slot and dynamic resource allocation systems. See Suchow et al. (2014) for a less polarized framing of this debate.

3.1 Experiment 1

As a first test of the rehearsal-probe task, the aims were to examine the patterns of responses derived from this task, assess the mean CSD measure as a dependent variable, and determine whether this measure would be sensitive to a substantial change in memory load. A considerable difference between the memory load conditions was used in this experiment in order to determine whether the task was at all sensitive to memory load manipulations under the most favourable circumstances (i.e. a large difference in load) before testing if the paradigm is sensitive to smaller, more subtle load manipulations.

Perhaps the most straightforward way of manipulating the STM load is by changing the number of items to be remembered. A number of studies have been conducted with visual stimuli where the number of sequence items (in the case of sequential presentation) or set size (in the case of simultaneous presentation) is used to manipulate load (e.g. Barton, Ester, & Awh, 2009; Bays & Husain, 2008; Lavie, Hirst, de Fockert, & Viding, 2004).

In the present experiment, memory load of the rehearsal-probe trials was changed via the number of digits in the sequence; there were 3 digits in the low load trials and 6 digits in the high load trials. Importantly, both of these list lengths were below the typical digit span for typical undergraduate students (Tehan & Lalor, 2000). This was important to ensure that the sequences could be remembered accurately during the rehearsal period.

One issue that became apparent when designing a rehearsal-probe experiment using two different list lengths is that it was not possible to match the two conditions for both sequence duration and presentation rate. In the current experiment, the two conditions were matched for sequence duration, but not presentation rate. Matching the two conditions for sequence duration had the advantage that the conditions would also be matched for the number of item rehearsals during the rehearsal delay, and temporal precision could be measured over the same time scale. However, in order to match the total sequence durations, the 3-digit lists were presented at half the speed of the 6-digit lists. The implications of this design on the results are evaluated in the Discussion section.

The first set of hypotheses relates to the general patterns of responses that were expected to result from the rehearsal-probe task. It was expected that the average proportions of responses corresponding to each serial position would be characterised by a rise and fall over time, where the modal probe time for each item response would indicate the time of greatest subvocal activation for that item. Thus the modal probe times for each serial position should occur in the correct serial order, so long as the sequences are rehearsed without order errors. Also, it was expected that average performance would not be at floor (i.e. no regularity in responses, highest possible mean CSD) or at ceiling (i.e. perfectly consistent responses, lowest possible mean CSD), and that there would be some between-participant variability in the dependent measure.

The second hypothesis was that temporal variability (mean CSD) would be higher in the high load condition than in the low load condition. The rationale for this view is that more cognitive resources are required to maintain the item and order information for 6-digit sequences than for 3-digit sequences, and the increased demands on shared, limited resources in the high load condition would result in fewer available resources dedicated to the maintenance of precise sequence timing. This hypothesis is consistent with a dynamic resource sharing account of timing precision in auditory-verbal STM.

3.1.1 Methods

3.1.1.1 Participants

Twenty-four undergraduate students (16 women) from the University of York participated in this experiment for cash (£4/hour) or psychology course credit hours. Participants were deemed eligible if they reported English as a first language, had no self-reported diagnosis of hearing problems and no self-reported diagnosis of a developmental language disorder (e.g. dyslexia, SLI). The experiment received ethical approval from the University of York Departmental Ethics Committee. All participants were informed about the nature of the tasks and their right to withdraw from the study at any time without penalty. Participants provided written consent before the experiment and were debriefed at the end of the session.

3.1.1.2 Design

The experiment took a between-participant design; participants were assigned to either the high or low memory load conditions in an alternating order for a total of 12 participants in each experimental group. The dependent variable was mean CSD (described in Chapter 2).

3.1.1.3 Materials

The digit stimuli used in this experiment[5] were the spoken names of digits 0-9. These digits were recorded by a male native English speaker at 44100 Hz and 32-bit mono. The

---

[5] The materials used in this experiment differ from those described in Chapter 2 and used in all subsequent experiments. This is because Experiment 1 was conducted as an initial test of the paradigm, before the digit stimuli were matched for duration and pitch differences, and adjusted for p-center alignment.

durations of the digits ranged from 294 to 574 ms (mean = 400 ms, SD = 78 ms). The digit recordings were matched for maximum amplitude.

3.1.1.4 Procedure

The task procedure was generally the same as that described in Chapter 2, but with some differences noted here. Random 3- or 6-digit lists were presented twice, followed by a silent rehearsal interval lasting until a probe (tone) was presented. Upon hearing the probe, the participant was instructed to respond with the digit s/he was rehearsing at that moment. There was a two-second response window after the probe, which was used to manage the duration of the testing session and encourage participants to respond quickly. The inter-trial interval was two seconds, which began at either the time of the response or after the two-second response window had elapsed.

The rehearsal lists consisted of three digits presented at a slower rate (1 item per 900 ms) or six digits presented at a faster rate (1 item per 450 ms, see Figure 11). Two digit recordings ("five" and "six") were longer than 450 ms, which meant that there was slight overlap between the ends of these digits and the beginnings of the next digits in the 6-digit fast-rate group. These sequences were tested and found to be clear and intelligible.

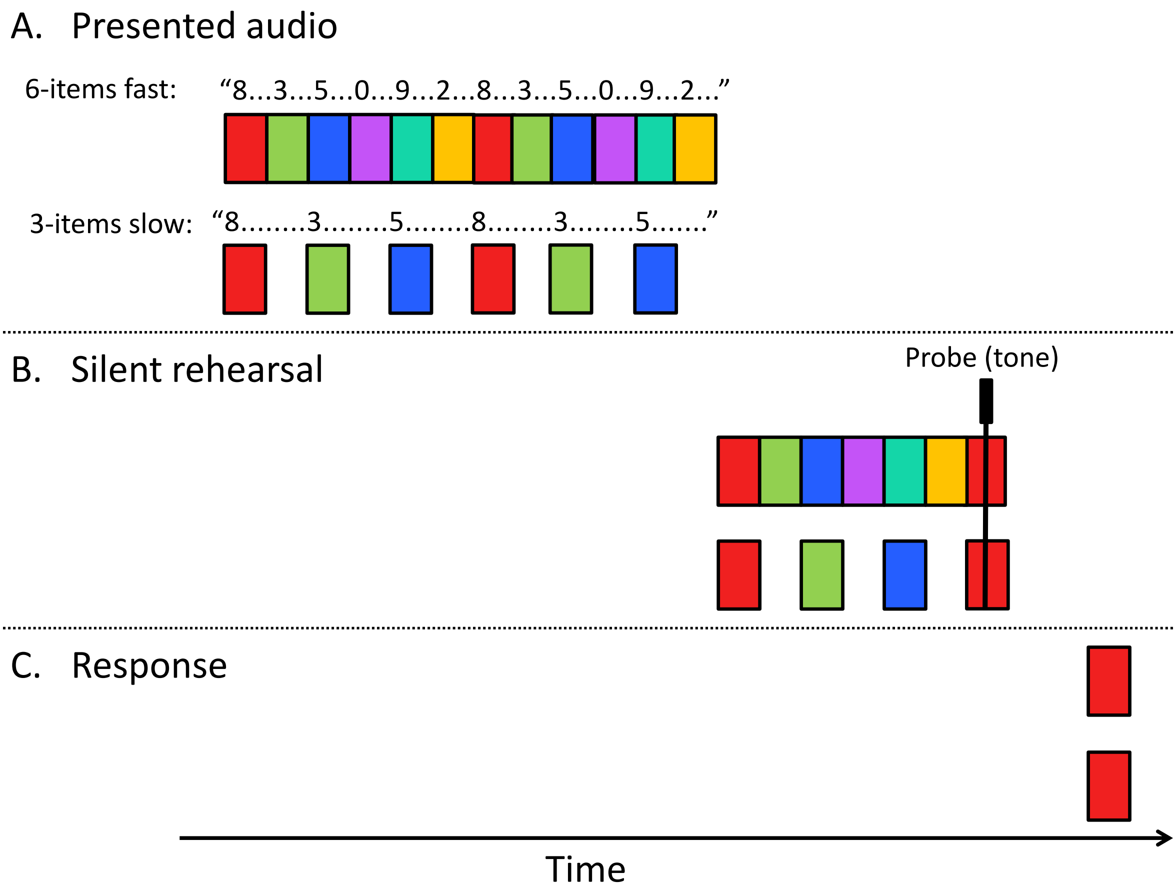

Figure 11. Schematic of the Experiment 1 trial design. The boxes represent the ~400 ms digits, where the colour indicates the serial position within the sequence. Isochronous random digit sequences were either 6-items long, presented at one digit per 450 ms (A, top row) or 3-items long, presented at one digit per 900 ms (A, bottom row). After hearing two repetitions of the sequence, the participant began subvocally rehearsing the sequence exactly as it was presented (B) until hearing a tone (B, black vertical line). Upon hearing the tone, the participant was instructed to stop rehearsing and respond with the digit that was actively being rehearsed when the tone was presented (C). This example shows the first probe time, which corresponds to the beginning of the first item in the second rehearsal cycle.

In both the high and low load groups, there were three probe times per item (digit and inter-digit silent interval). These probe times were aligned with the beginning, middle and end of each of item during the second rehearsal cycle, if the list were rehearsed exactly as it was presented (see Figure 12 for a graphical representation of probe timing). In the low load group, there were 9 probe times in total, which corresponded to 3 probes aligned within each of the 3 digits during the second rehearsal cycle. In the high load group, there were 18 different probe times, corresponding to 3 probes within each of the 6 digits during the second rehearsal cycle. In both groups, the probes were presented within the same range of delays after the end of the sequence presentation period, which was between 2700 and 5400 ms.

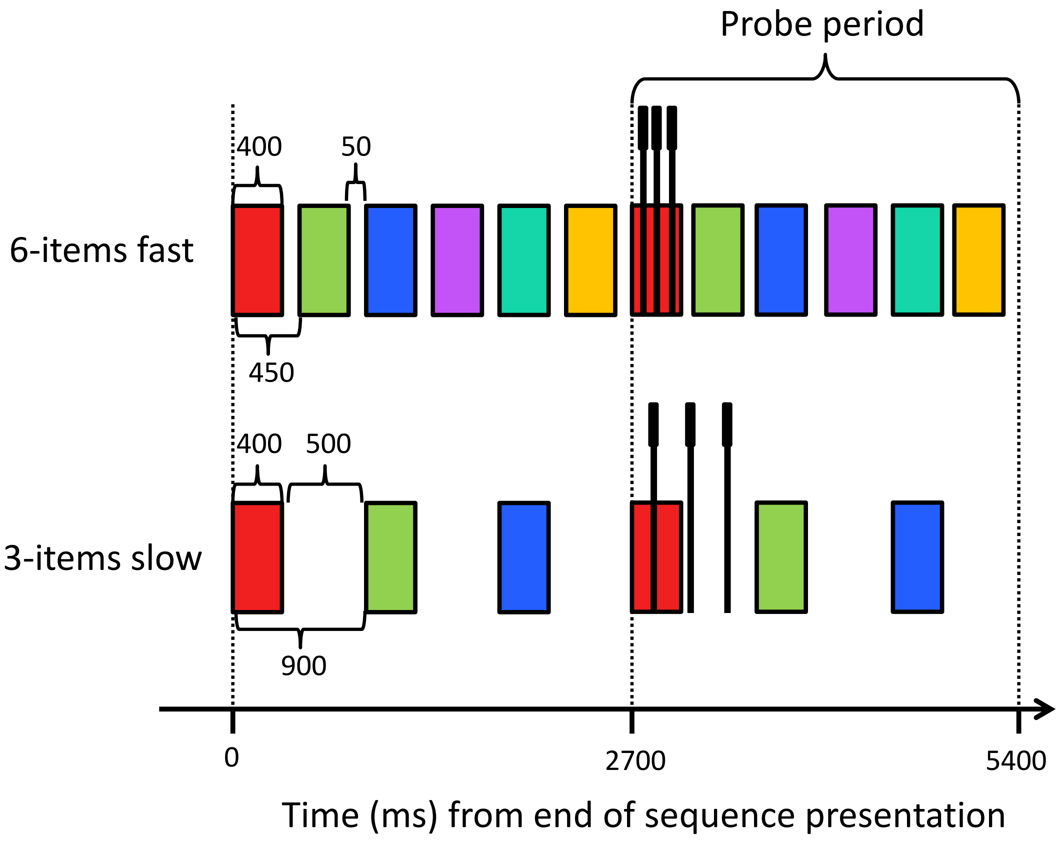

Figure 12. Representation of the item durations (coloured boxes), silent inter-item durations and probe timing in the high load (top row) and low load (bottom row) groups. Time is shown on the x-axis and starts from the end of the sequence presentation period, i.e. the beginning of the silent rehearsal period. The item timing shown represents the time course of item rehearsal if the sequence was rehearsed exactly as presented. Only three probe times are shown here (black vertical lines) for ease of visual display. However, there were three probe times per sequence item that occurred only during the second rehearsal cycle, as indicated by the time segment labelled ‘probe period’. The probe times shown indicate the alignment of the probe times relative to the items within each experimental group.

The participants were tested individually in a quiet room by a single experimenter. After the participant read the instructions and asked any questions, the experimenter demonstrated 3 trials by rehearsing aloud, and then the participant completed 3 practice trials with both overt and silent rehearsal. Participants then completed 3 blocks of trials, where each block consisted of three trials per probe time. In the low load group, there were 27 trials per block (3 trials x 9 probe times) and 81 trials in total (9 trials per probe time x 9 probe times). In the high load group, there were 54 trials per block (3 trials x 18 probe times), for a total of 162 trials (9 trials per probe time x 18 probe times). Because there were twice as many probe times in the high load group compared to the low load group and the same number of trials per probe time, the high load testing sessions lasted approximately twice as long as the low load sessions (~50 minutes vs. ~25 minutes).

3.1.2 Results

The responses were filtered according to three possible response types; valid, invalid or no-response. Valid trials were defined as a digit response that was present in the rehearsal sequence for that trial. Invalid responses were digits that were not present in the presented sequence (i.e. extra-list responses). A no-response occurred when the participant did not respond within two-seconds after the probe. The frequencies and percentages of each response type according to group are shown in Table 2, which shows that there was a small percentage of no-response trials and an even smaller percentage of invalid responses. A Fishers Exact Test on the response type by group frequency table showed that there was a significant relationship between the two factors (p = .001), and the standardised residuals (z > +/- 1.96) revealed that this difference was driven by the unequal distribution of no-responses. The number of no-responses was higher than expected in the 6-item group (z = 2.0) and lower than expected in the 3-item group (z = -2.8).

Table 2. Response type frequencies and percentages (in parentheses) within the two experimental groups and combined. Percentages are marginal by row. Bottom row shows the frequencies and percentages for the total responses (two groups combined). Rightmost column shows the total number of responses (trials) in each group and in total.

	Response type			
Group	Valid	Invalid	No response	Total
3-items	903 (92.9%)	2 (0.2%)	67 (6.9%)	972
6-items	1720 (88.5%)	8 (0.4%)	216 (11.1%)	1944
Total	2623 (90.0%)	10 (0.3%)	283 (9.7%)	2916

After removing the invalid and no-response trials (10.0% of total trials), the remaining valid trials for each participant were grouped according to the serial positions of the responses (positions 1-3 in the low load group and positions 1-6 in the high load group). For each participant, the proportions of responses corresponding to items in each serial position were calculated over the total responses made at each probe time. These serial position response proportions were then averaged over participants in each group, resulting in mean proportions of responses for each serial position at each probe time. The averaged response proportions are shown as points in Figure 13, where the widths of the ribbons around the points show the standard error.

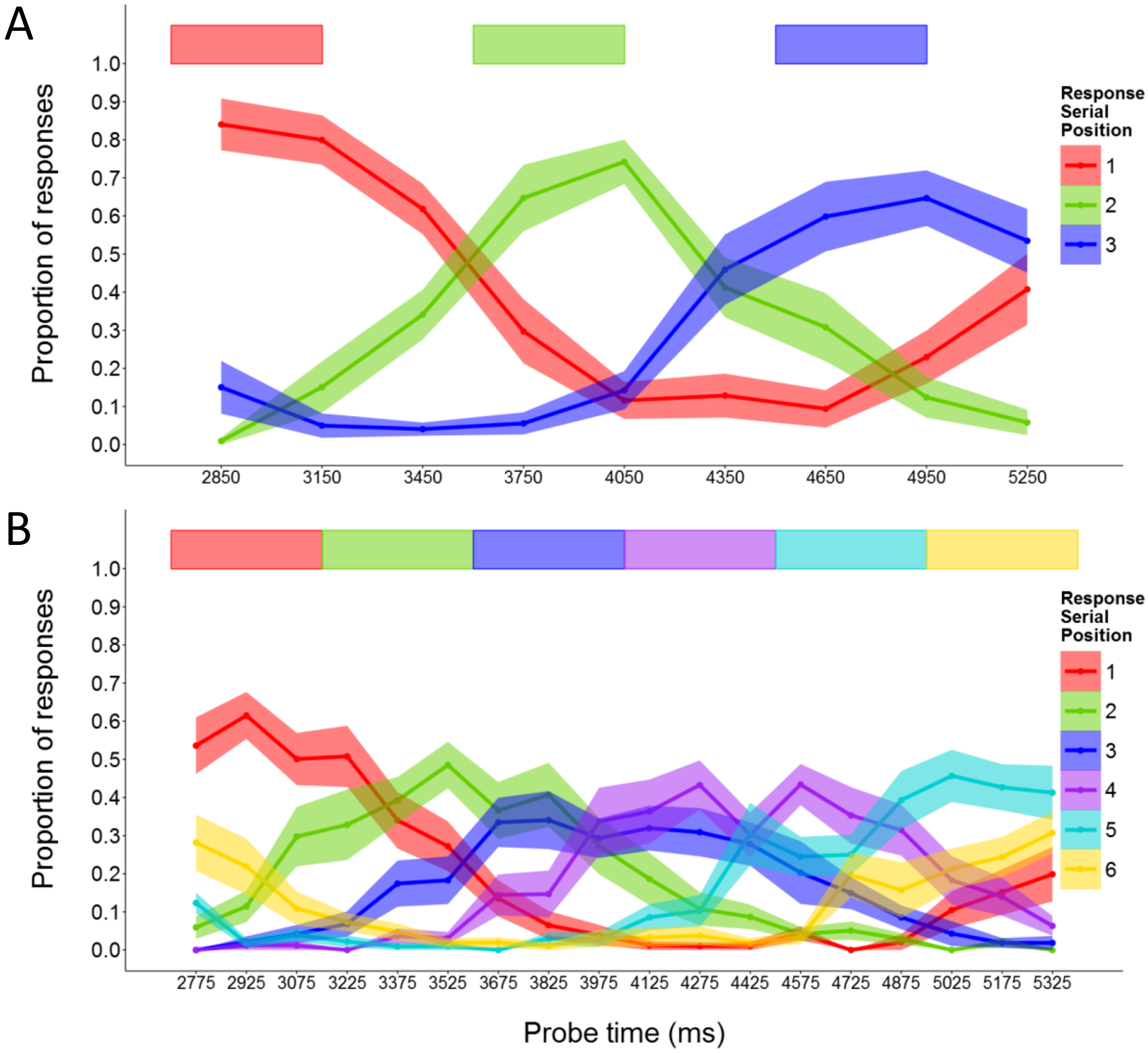

Figure 13. Mean response proportions for 3-digit (A) and 6-digit (B) sequences during the rehearsal-probe task. Colours represent the serial positions of the responses, points show the mean proportion of responses for each item within each probe time, and width of the ribbon shows the SEM of the response proportions. Boxes at the top of the plot reflect the 'correct' item timing, i.e. assuming the sequence was rehearsed exactly as it was presented. The x-axes in A and B are matched for duration (2700 to 5400 ms after the end of sequence presentation), and the ticks show the specific probe times used for that experimental group.

CSDs of the responses for the two load conditions were calculated according to the procedure outlined in Chapter 2. The CSDs of the serial position response distributions were averaged across participants in each group, resulting in a mean CSD value for each serial position within each group. The serial position mean CSD values were then averaged within groups to produce a grand mean CSD for the two memory load groups. Table 3 shows the mean CSDs by group and serial position.

Table 3. Mean CSDs (SEM in parentheses) by serial positions in the two memory load conditions; 3-items (low load) and 6-items (high load). The maximum possible CSD value is 1.41. The minimum possible CSD value is 0.56 in the 3-item group and 0.28 in the 6-item group. The minimum and maximum possible CSD values apply to serial position CSDs as well as the mean CSD within each group.

	Serial Position						
Group	1	2	3	4	5	6	Mean
	0.88	0.82	0.80				0.83
3-items	(0.06)	(0.05)	(0.04)	-	-	-	(0.04)
	0.67	0.76	0.78	0.76	0.73	0.76	0.75
6-items	(0.04)	(0.03)	(0.06)	(0.03)	(0.04)	(0.06)	(0.04)

The maximum possible CSD value was 1.41. The lowest possible CSD value differs according to the number of items and probe times, and therefore differed between the two groups. In the low load group, the minimum possible CSD score was 0.56, and mean CSD scores in this group ranged from 0.64 to 1.05 (skewness = 0.24, kurtosis = 1.64). In the high load group, the minimum possible CSD score was 0.28. Mean CSD scores in the high load group ranged from 0.55 to 0.97 (skewness = 0.25, kurtosis = 2.41). Shapiro-Wilks tests confirmed that the distributions of mean CSD scores in both groups were normally distributed (low load: *p* = .210; high load: *p* = .996). Figure 14 shows the mean CSD scores for each group with 95% confidence intervals, and with dashed horizontal lines showing the lowest possible CSD value that could be obtained in each task.

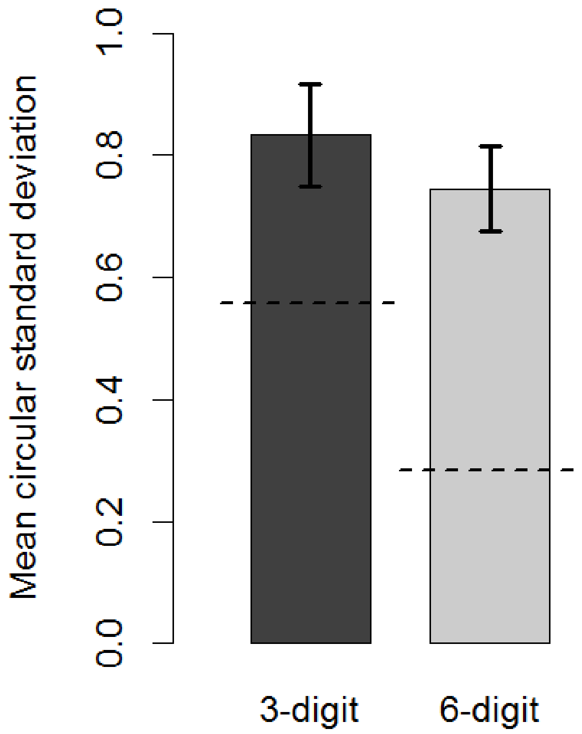

Figure 14. Bar plots of mean CSD scores in the low load (3-item) group and high load (6-item) group. Error bars show 95% confidence intervals. The dashed lines reflect the minimum possible mean CSD scores in the two tasks.

Because the minimum possible CSD score differed between the two groups, it was not possible to directly compare the raw mean CSD values between groups. It was possible, however, to investigate whether temporal variability was related to serial position of the response. One-way repeated-measures ANOVAs were conducted on the data within each group, where the serial position factor had either 3 or 6 levels. In the low load group, there was no significant effect of serial position on CSD, $F(2,22) = 2.08$, $p = .149$. In the high load group, Mauchly's test of sphericity indicated that the assumption had been violated ($p = .045$), therefore the degrees of freedom were corrected the Greenhouse-Geisser estimate ($\varepsilon = 0.52$). There was also no effect of serial position in the high load group, $F(2.59,28.53) = 1.98$, $p = .146$.

In order to compare the mean CSD values between the two groups in this experiment, the scores were adjusted to the account for the between-group difference in the range of possible mean CSD scores. The adjustment formula was *(CSD – minimum)/range*, where *CSD* is the observed CSD score, *minimum* is the lowest possible score in the appropriate group, and *range* is the range of possible CSD values in that group. The adjusted scores can thus be interpreted as a proportion out of the range of possible scores, where an adjusted score of 0 reflects the lowest possible mean CSD within that group and 1 reflects the highest possible response variability. Group means and 95% confidence intervals of the adjusted mean CSD scores are shown in Figure 15.

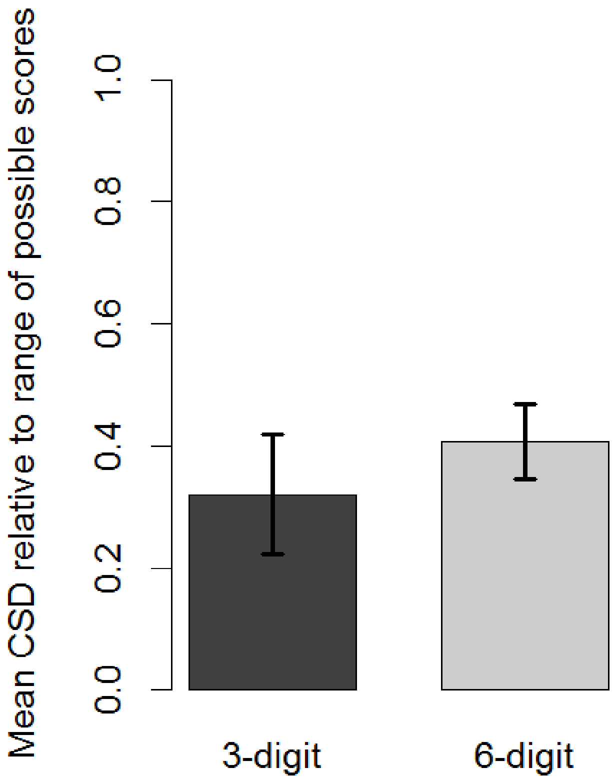

Figure 15. Group averages of the adjusted mean CSD scores (error bars show the 95% confidence intervals). Adjusted scores reflect the proportion within the full range of possible scores within each group, where 0 is the lowest possible CSD score and 1 is the highest possible score. See text for details about the computation.

The data were tested and found to be normally distributed. A Levene's Test revealed that variances were not homogenous ($p = .030$), therefore an independent samples *t*-test was used with the Welch modification to degrees of freedom. The results of this test showed that there was no significant effect of load group on the adjusted mean CSD scores, $t(18.584) = -1.49$, $p = .154$.

3.1.3 Discussion

The first hypothesis dealt with the expected qualitative patterns of responses derived from the rehearsal-probe task, and this hypothesis was confirmed. In both the 3-digit and 6-digit groups, the average response proportions for each serial position showed the expected rise and fall over probe times. These distributions reflect the assumed increase and decrease in the activation of the words during subvocal rehearsal, where the modal probe time for each serial position corresponds to the time when the item is most clearly being subvocalised. In the 3-digit group, the modal (peak) probe times for the serial position response distributions were in the correct serial order and roughly in time relative to the expected/correct timing (indicated by the boxes above the response proportion functions in Figure 13). In the 6-digit group, the general pattern was similar but more difficult to discern. The responses in this group tended to lag behind the expected timing, which is likely due to the fact that these digits were presented at a faster rate. Some of the serial position response distributions in the 6-digit group appeared to have multiple peaks, which may have resulted from either serial order errors or as an artifact of averaging over data sets with markedly different rehearsal timings.

With regard to the mean CSD measure, it was expected that there would be some between-participant variability and that average performance would not be at floor or ceiling. These predictions were also confirmed. Participants varied in terms of their response variability in these tasks, and there was significantly more between-participant variability in the 3-item group than in the 6-item group. The increased variance among participants in the low load group might have been due to the slower rate and/or longer silent intervals within the 3-digit sequences, which may have allowed for increased opportunity for between-subject deviation in rehearsal rate and response ambiguity for probes occurring during the silent intervals between digits. The mean CSD scores in both groups were normally distributed and comfortably located within the range of possible mean CSD values (see Figure 14).

The second main hypothesis was that temporal variability (mean CSD scores) would be higher in the high load condition than in the low load condition. This hypothesis was not supported; there were no significant differences in the mean CSD scores after adjusting for differences in the range of possible scores. However, the results were also difficult to interpret because of two major issues with the experimental design and dependent measure that were brought to light.

The first limitation is that the conditions were not matched for presentation rate, number of items, or the number probes per time interval (i.e. sampling frequency). The properties of the CSD measure differ according to the number of possible responses and probe frequency, which means that it is difficult to compare this measure between two conditions that differ in either factor. Differences in presentation rate will also bias the CSD measure because of the increased amount of inter-item silence in the slower rate condition. The inter-item silence may create additional ambiguity in responses to the probes, and participants may further differ in their strategy to responding to a probe that occurs between two items, e.g. tending to respond with either the preceding or succeeding item, or determining which item was closer in time to the probe on a trial-by-trial basis. The effect of inter-item silence in this experiment may have contributed to the counterintuitive result, that temporal variability was not lower in the high load group, because the reduced inter-item silence in the high load condition may have actually reduced this particular source of response variability.

A second issue is that the groups were not matched for number of trials or duration of the testing session. The high load task took roughly twice as long to complete as the low load task. This difference may have introduced fatigue effects, which could account for the greater proportion of no-response trials in high load group. The increased task difficulty in the high load condition may also explain the difference in the distributions of response types between the two groups, where participants in the high load group may have experienced more item and order errors during the task, as well as failures to encode the sequences due to lapses in attention. It cannot be determined from this experiment whether the larger proportion of no-responses in the 6-digit group was due to fatigue effects, task difficulty or a combination of the two.

These issues have provided opportunities for refinement in future experiments. Ideally, the effect of memory load on timing should be tested with a memory load manipulation that allows the rehearsal-probe conditions to be matched for number of sequence items, presentation rate and number of probe times. This solution will eliminate the problems encountered with the mean CSD measure by equating the minimum possible CSD values, as well as the amount of inter-item silence. Rehearsal-probe conditions should also be matched for overall duration in order to eliminate possible fatigue effects. Given the between-subject variability in performance on these tasks, future experiments would benefit from within-subject manipulations. Finally, because this task essentially requires the immediate recall of auditory-verbal sequences, albeit subvocal recall, performance on tasks requiring serial order recall of auditory-verbal sequences might be relevant. In future studies, individual scores on an ISR task, such as digit span, should also be collected. Individual differences in digit span scores might explain some of the between-subject variability in response patterns. Also, digit span would be useful for ensuring that the participants can recall digit sequences at the length required in the rehearsal-probe task without making order errors.

In conclusion, this experiment was generally successful in establishing the rehearsal-probe method and confirming the expected qualitative patterns of responses. The results of this experiment have shown that the rehearsal-probe task produces systematic patterns of response data, with periodic increases and decreases in response probabilities for each item over time, and with peak response probabilities for each item occurring in the correct serial order. The mean CSD scores have the properties of a useful dependent measure, although this experiment has highlighted its limitations. Thus the rehearsal-probe paradigm looks promising at this stage but is clearly in need of a bit of refinement to overcome the problems identified. These problems arose when trying to compare the mean CSD scores between groups, and they made it difficult to draw firm conclusions about the effects of memory load on timing variability. However, these issues did prompt experimental design revisions that will allow the relationship between memory load and temporal precision to be more successfully investigated with the rehearsal-probe task in future experiments.

3.2 Experiment 2

The aims of this experiment were again to establish and refine the new rehearsal-probe task paradigm, and to determine whether temporal variability of sequences representations are sensitive to changes in memory load. Based on the previous experiment, there were a few changes made to the task design. In this experiment, the number of items, presentation rate, number of probe times and number of trials (i.e. task duration) were matched between the two load conditions in order to avoid biasing the dependent variable and to prevent differences in attention and fatigue. Also, a within-participant manipulation of load was used here to control for the between-participant variance in mean CSD scores, and digit span data were collected to examine relationships with ISR performance.

In this experiment, the memory load of 4-digit rehearsal sequences was manipulated via sequence predictability. In the low memory load condition, the digits occurred in a predictable, ascending numerical order (e.g. “2 3 4 5”). The high load trials consisted of 4-digit sequences where the digits were selected at random without replacement from the set 0-9 (e.g. “5 0 2 7”). In the low load condition, the item-order information was supported by LTM, whereas in the high load condition all of the item-order associations cannot be retrieved from LTM and thus must be actively maintained. Critically, the two load conditions were matched for number of items, sequence duration, presentation rate and probe times. Also of interest was whether individual differences in STM capacity for serial order mediates the effects of the load manipulation on temporal variability, so digit span data was also collected. The hypothesis was that temporal variability during rehearsal, as indexed by the mean CSD score, would increase for high load (random order) sequences compared to low load (familiar order) sequences. It was also predicted that the response proportion data would show the same systematic patterns as those in Experiment 1, and that the mean CSD measure would show normally distributed variation across individuals.

3.2.1 Methods

3.2.1.1 Participants

Participants were 18 undergraduate volunteers (15 women) recruited from the University of York with a mean age of 20.17 years (range 19-21 years). One participant was excluded due to severe difficulties concentrating throughout the rehearsal-probe task, as observed by the experimenter and reported by the participant during debriefing. Therefore there were 17 participants included in the analysis. The ethics approval, compensation and informed consent details were the same as those in Experiment 1. In addition to the eligibility criteria listed in Experiment 1, participants must have been able to correctly recall at least one 4-digit sequence in the digit span task in order to continue with the rehearsal-probe task.

3.2.1.2 Design

The experiment took a within-subjects design; all participants completed blocks of both high and low memory load trials. The order of the first block condition was counterbalanced, and blocks alternated between the two memory load conditions after the first block. The dependent variables were the mean CSDs across serial positions (for each condition and averaged across conditions), and the digit span score.

3.2.1.3 Materials

The digit stimuli used were those described in Chapter 2.

3.2.1.4 Procedure

The rehearsal-probe task was administered according to the procedures outlined in Chapter 2. Participants completed trial blocks of alternating high and low memory load conditions with the first block condition assigned in counterbalanced order. The blocks consisted of 36 trials of 4-digit sequences in either random order (high memory load) or ascending numerical order (low memory load; see Figure 16). In the high memory load trials, digits were selected at random from the set 0-9 without replacement (e.g. 8-0-2-5). In the low memory load trials, the digit in the first serial position was selected at random from the set 0-6, and the three digits in the subsequent serial positions followed from the first digit in numerically increasing order (e.g. 4-5-6-7). Probe times were selected randomly during each block, with each of the 12 probe times occurring three times per block. In total, there were 108 trials (9 trials per probe time) in each condition, resulting in 216 total trials. Participants were given optional self-paced breaks between the six blocks.

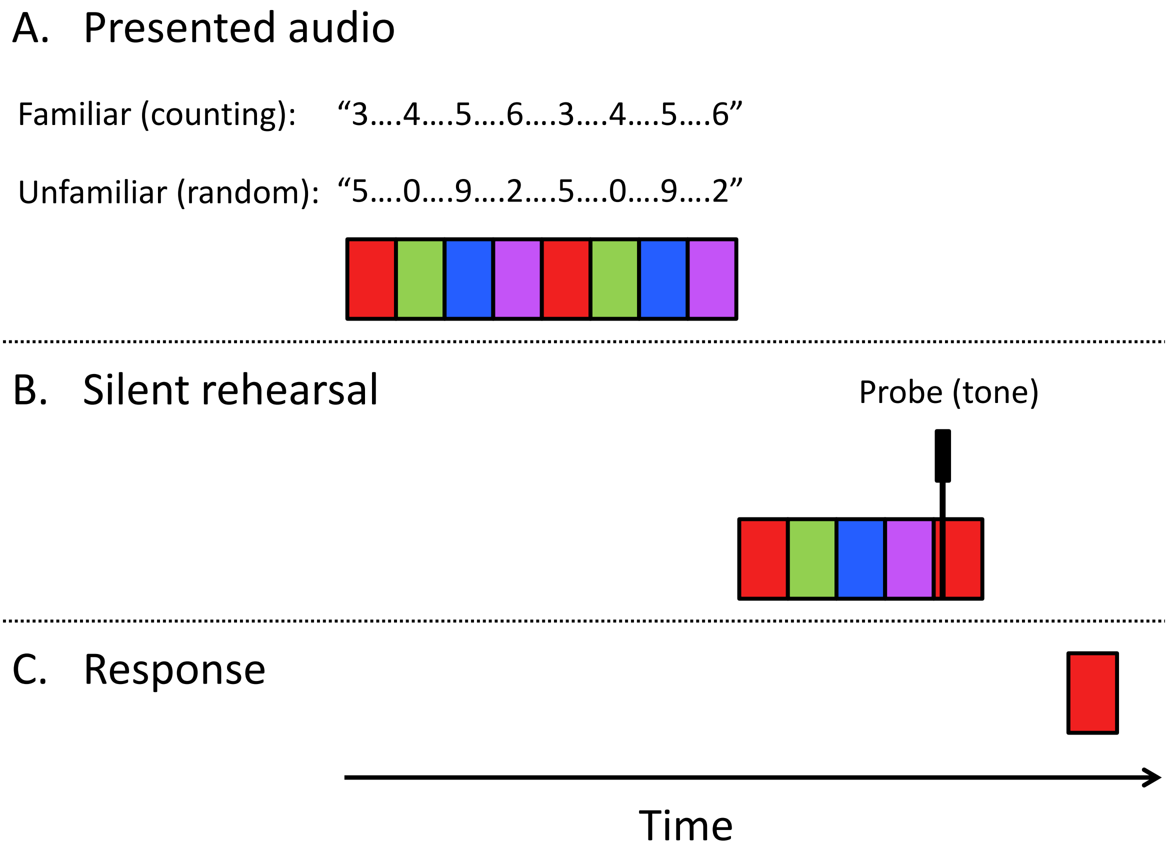

Figure 16. Representation of the Experiment 2 rehearsal-probe task design. The four colours represent the serial positions of the items. Row A refers to the presented audio sequence, where the participants heard two repetitions of a 4-digit list. The digits were either in numerically ascending order (familiar condition) or were not in numerical order (random). The participants silently rehearsed the list exactly as it was presented until hearing a probe (B). Upon hearing the probe, participants responded with the item that s/he was rehearsing when the probe was presented (C). The probe shown represents the first probe time, which corresponds to the beginning of the first item during the second rehearsal cycle.

The probe times were set such that the probes would occur during the beginning, middle or end of each of the four items during the second list cycle if the sequence were perfectly replicated during the rehearsal delay. For each of the 400 ms digits, the first probe occurred 67 ms after the digit onset, the middle probe occurred 200 ms after the digit onset, and the last probe occurred 333 ms after the digit onset. No probes were aligned with the first rehearsal list cycle (0 to 2000 ms after stimulus offset) and no probes were set to occur during the 100 ms of silence between digits. The probe times, measured from the final digit stimulus offset time to the probe onset time, are listed along the x-axis in Figure 17.

All participants were tested individually in quiet room by a single experimenter and in a single session. Digit span data was collected prior to the rehearsal-probe task. The digit span task was administered according to the procedures outlined in Chapter 2. Because the digit stimuli used in the digit span task was the same as those used in the rehearsal task, the task order provided an added benefit of familiarizing the participants with the digit stimuli before it was presented at a faster rate during the rehearsal task. The instructions for the rehearsal-probe task alerted participants to the fact that in some blocks the trials would consist of predictable (ascending numerical order) digit sequences, and in other blocks the trials would consist of sequences of randomly ordered, unpredictable digits. The experimenter instructed participants not to respond to the probe if they forgot the sequence during the rehearsal delay.

3.2.2 Results

The digit responses were first categorised according to response type: valid, invalid, or no-response. These categories are the same as those used in Experiment 1. The frequencies and percentages of each type of response were highly similar between conditions (Table 4).

Table 4. Response type frequencies and percentages within the two experimental groups; low (familiar numerical order) and high (random order). Percentages are marginal by row. Bottom row shows the frequencies and percentages for the total responses (two groups combined). Rightmost column shows the total number of responses (trials) in each group and in total.

	Response type			
Load condition	Valid	Invalid	No response	Total
Low	1771 (96.5%)	2 (0.1%)	63 (3.4%)	1836
High	1773 (96.6%)	2 (0.1%)	61 (3.3%)	1836
Total	3544 (96.5%)	4 (0.1%)	124 (3.4%)	3672

After removing the no-response and invalid response trials from the data (3.5% of the total trials), the proportions of serial position responses were calculated at each probe time for each participant, then these proportions were averaged over participants within each condition. Figure 17 shows the mean proportions of responses for all probe times in the low load (numerical order, A) and high load (random order, B) conditions.

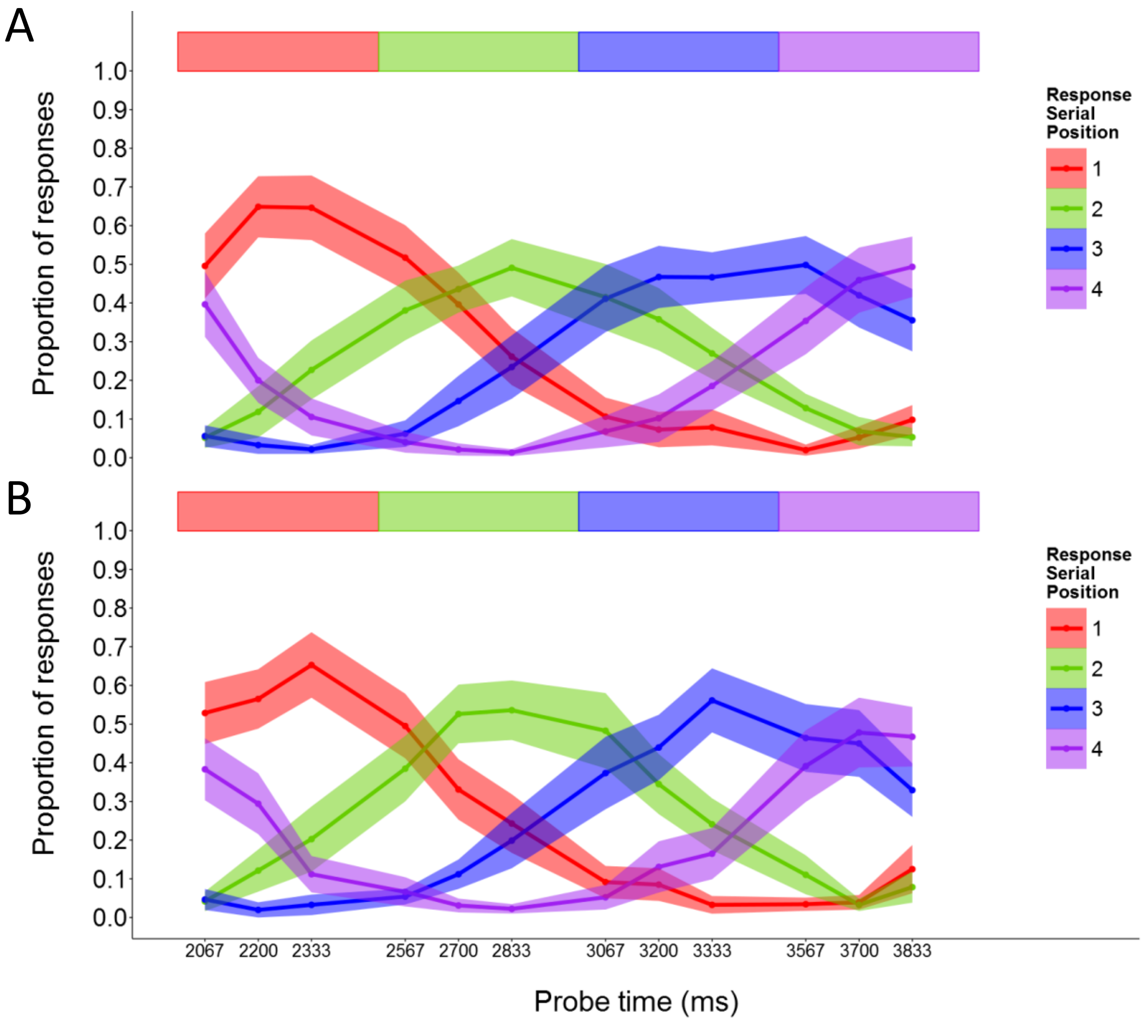

Figure 17. Mean proportion of responses for each serial position across probe times in the numerically ordered (A) and randomly ordered (B) sequence conditions. The widths of the ribbons at each probe time reflect the between-subject SEM.

CSDs of the responses for the two load conditions and four response serial positions were calculated according to the procedure outlined in Chapter 2. Overall, the grand mean CSD was higher in the low load condition than in the high load condition (see Table 5), meaning that participants showed more variation in their responses to probes during rehearsal of numerically ordered digit sequences than to randomly ordered sequences. The grand mean CSDs across the serial positions of responses were similar with the exception of the last digit in the sequence (4th serial position), which was lower than the other positions.

Table 5. Mean CSDs (SEM in parentheses) for the 4 serial positions and the two memory load conditions. The maximum possible CSD value is 1.41, and the minimum possible CSD value is 0.42 in this experiment.

	Serial Position				
Load condition	1	2	3	4	Mean
Low	0.73 (0.05)	0.77 (0.04)	0.74 (0.04)	0.67 (0.04)	0.73 (0.04)
High	0.70 (0.05)	0.69 (0.03)	0.73 (0.04)	0.63 (0.04)	0.69 (0.04)
Mean	0.71 (0.05)	0.73 (0.04)	0.74 (0.04)	0.65 (0.04)	

The load (2 levels) x serial position (4 levels) grand mean CSD data were tested and found to have met the assumptions for a two-way repeated-measures ANOVA. There was a significant effect of load condition on grand mean CSD, *F*(1,15) = 8.19, *p* = .012 (see Figure 18). The grand mean CSD of the high load (randomly ordered) trials was lower than the grand mean CSD of the low load (numerically ordered) trials (see Table 5). There was no significant main effect of serial position, *F*(3,45) = 2.28, *p* = .093, nor was there a significant interaction between load and position, *F*(3,45) = 0.66, *p* = .581.

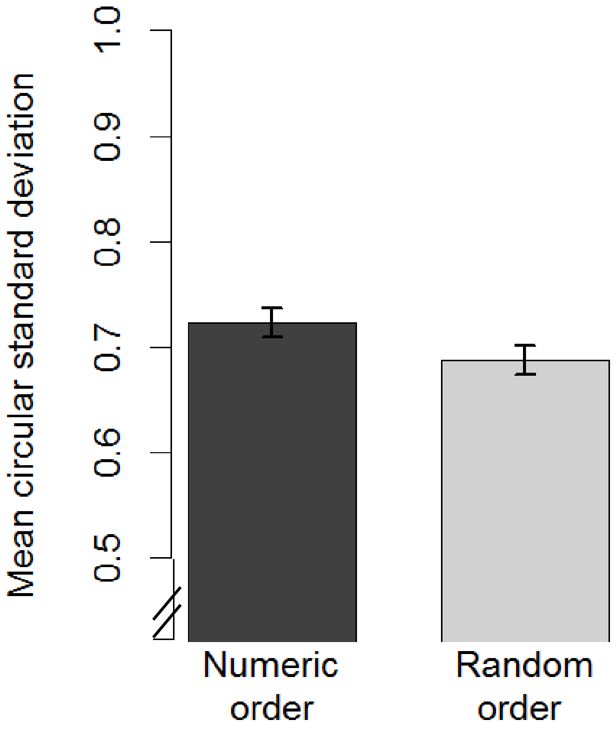

Figure 18. Mean CSDs in the low load (familiar numeric order, dark gray) and high load (random order, light gray) 4-digit rehearsal conditions. Error bars show 95% confidence intervals, adjusted for within-subject error using the method proposed by Cousineau (2005). The y-axis minimum is adjusted to the lowest possible mean CSD score in this experiment.

The relationship between mean CSD and digit span scores was also examined. The digit span task was scored using the method described in Chapter 2. Digit span was significantly negatively correlated with individual mean CSD scores averaged over the two load conditions, *r* = -.49, *p* = .045. Correlations between digit span and the mean CSD scores for each condition revealed that there was a significant negative correlation between digit span and the mean CSD in the high load condition, *r* = -.58, *p* = .015. The correlation coefficient between digit span and mean CSD in the low load condition was also in the negative direction but did not reach significance, *r* = -.38, *p* = .129. Figure 19 shows the relationship between digit span scores and mean CSDs, averaged over the two load conditions.

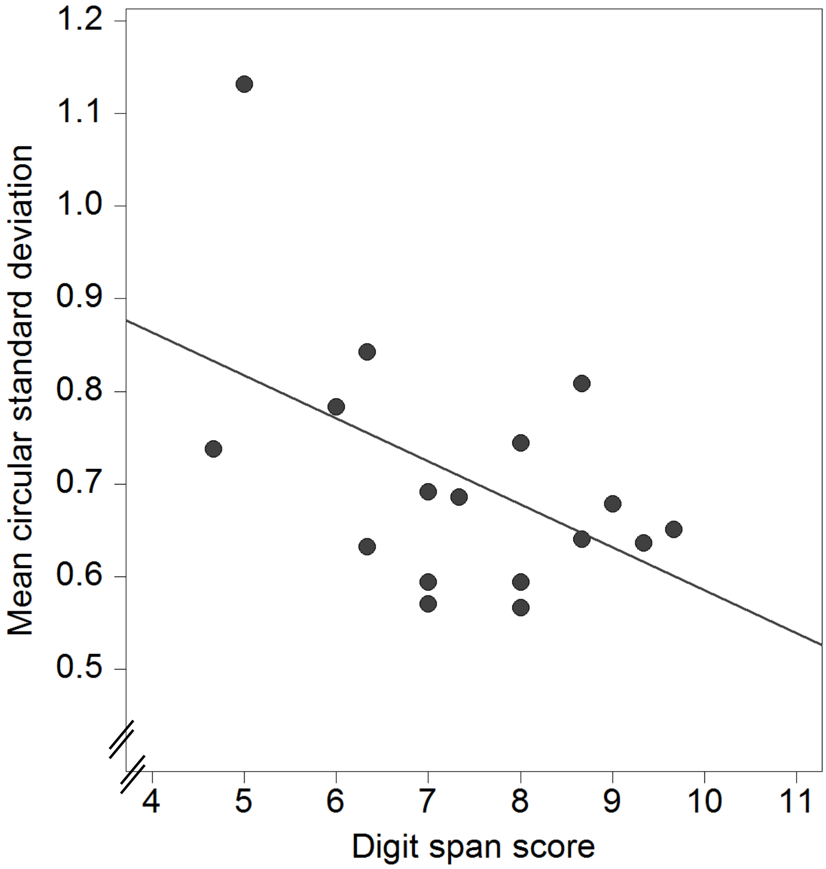

Figure 19. Scatterplot of the relationship between digit span and mean CSD averaged over the two load conditions. Solid black line represents the linear regression model for mean CSD as predicted by digit span. The y-axis minimum is the lowest possible mean CSD score obtainable in this experiment.

Finally, a correlational analysis of digit span and the within-subject difference between mean CSD in the high and low load conditions revealed a significant positive relationship, *r* = .49, *p* = .040. Participants with higher digit spans tended to have a higher mean CSD in the low load compared to the high load condition, as reflected by more positive mean CSDs differences (low load mean CSD minus high load mean CSD, see Figure 20). In contrast, participants with lower digit spans tended to have higher mean CSDs in the high load condition compared to the low load condition, as reflected by negative difference values in Figure 20.

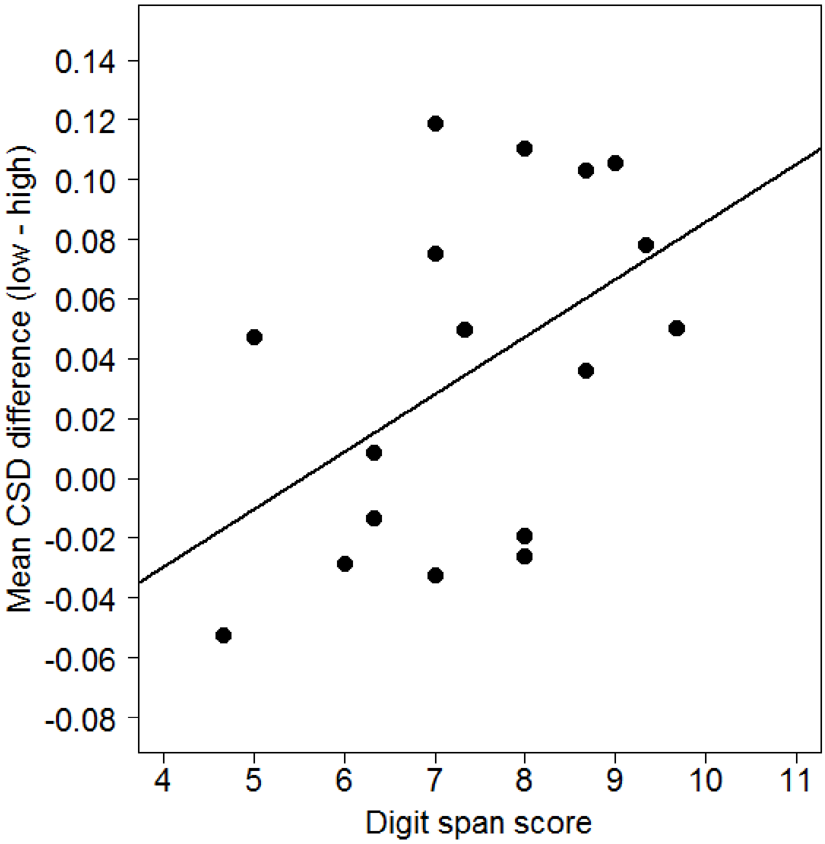

Figure 20. Scatterplot of individual digit span scores and differences between mean CSDs in the two rehearsal conditions (low load mean CSD minus high load mean CSD). Solid line shows the least squares regression linear model for the mean CSD difference predicted by digit span.

3.2.3 Discussion

In this experiment, the aim was to further establish and refine the rehearsal-probe task paradigm, and to determine whether the temporal variability of sub-span sequence representations is responsive to changes in memory load. The experiment was a successful replication of the lawful pattern of responses over time that was observed in Experiment 1. The response probabilities for each item again demonstrated a characteristic rise and fall over the probe times, where the modal probe times for each response item occurred in the expected serial order and roughly in line with the expected timing. These results suggest that the participants were performing the task correctly. Also, as in Experiment 1, this experiment showed considerable variability in these response position patterns between participants, as well as an absence of floor or ceiling effects. Because the number of items, presentation rate and probe times were matched between conditions in this experiment, the comparison of the mean CSD variable between conditions was straightforward.

With regard to the effect of memory load on timing, the results of this experiment showed that knowledge about item-order associations affected within-participant timing variability during rehearsal, but the effect of load on timing variability did not occur in the hypothesized direction. Rather, significantly greater temporal variability was observed in the low memory load condition compared to the high memory load condition. While this result was not expected, it also is not predicted under the null hypothesis, i.e. that temporal precision in STM is invariant to memory load manipulations. The effect of memory load on temporal variability in this experiment was most likely due to unpredicted effects of prior knowledge about the item-order associations on the precision of isochronous sequence timing during the STM maintenance interval.

Upon reflection, the comparison between random and familiar digit sequences appears not to be a valid manipulation of memory load. The rationale for using familiar digit sequences in the low load condition was that the task demands would be greatly reduced given that the item and order information is readily available from LTM, so long as the first item in the sequence is held in memory. However, while the numerically ordered sequences are no doubt easier for participants to remember, there also may be a qualitative difference between the cognitive processes involved in immediate serial recall for these two types of lists. First, there may not be the same recall benefit for rehearsing sequences at an isochronous rate when the sequence order is very easy to remember. The participants' rehearsal may have been influenced by a cost-benefit analysis of the utility of isochronous rehearsal timing on STM for these familiar sequences, despite the explicit instructions to maintain precise sequence timing. This is supported by evidence that the number of spontaneous rehearsals decreases with memory load (Jantz et al., 2014), suggesting that rehearsal is employed as a strategy according to when it is needed, and that rehearsal pace/frequency changes according to the demands of the task.

Second, the participants may have been influenced by the non-isochronous timing that is inextricably linked to the numerically ordered sequences, which are highly practised, whereas sequences of digits in arbitrary order would not carry any rhythmic associations from LTM. Research has shown that spoken recall timing for auditory-verbal sequences tends to mimic input timing without explicit instructions to retain timing information (Farrell, 2008), so it is plausible that these timing patterns could be stored along with the item and order information for overlearned auditory-verbal sequences. Schulkind (1999) tested recognition for well-known songs with or without temporal alterations, and found that the unaltered versions of songs were better identified than those with any rhythmic alterations. While Schulkind's research dealt with specifically with LTM for songs, it seems likely that LTM for unsung auditory-verbal sequences are stored in a similar manner. Finally, evidence from Hebb learning paradigms has also shown that recall for repeated sequences within a block of trials benefits more when the temporal structure of the repeated sequences is the same across repetitions compared to when the same items in the same order are repeated with a different temporal structure (Bower & Winzenz, 1969; Hitch, Flude, & Burgess, 2009). Therefore, in the low load condition there may have been interference between the isochronous sequence presentation, which participants are instructed to maintain, and the non-isochronous rhythms previously associated with familiar digit sequences.

In addition to the significant effect of memory load, there was also a significant correlation between individual differences in digit span and mean CSD in the hypothesized direction. That is, individuals with a superior ability to recall sequences of digits in the correct serial order tended to show less temporal variability in the rehearsal-probe task. When the association between these measures was examined separately for each memory load condition, the correlation was only significant for the high load rehearsal-probe trials. One possible reason for the reduced correlation with low load trials relates to the earlier point discussed concerning the unintended effect of the low load manipulation; the use of overlearned ordered digit sequences to reduce the memory load during the task may have produced additional effects on the maintenance of isochronous sequence timing. Therefore, it is not possible to determine from this experiment whether the lack of a significant correlation between STM span and timing variability in the low load condition was observed because the association is specific to timing precision in high memory load conditions, or because of qualitative differences between the two specific load conditions used in this experiment.

In order to test whether the unexpected effect of load on timing variability was specific to the load manipulation used in Experiment 2, the next experiment was designed to examine the effect of memory load on timing precision without the use of over-learned item-order associations in the load manipulation. If the counterintuitive effect of memory load in Experiment 2 was due to an unforeseen effect of the specific memory load manipulation, then it was expected that the use of a different load manipulation in Experiment 3 would reveal that temporal variability is greater in the high memory load condition.

3.3 Experiment 3

In this experiment, the general aims were the same as those in the previous two studies; to establish and refine the rehearsal-probe task paradigm, and to determine where temporal variability of sequence representations during rehearsal differed as function of memory load. While the load manipulation did not appear to be valid in Experiment 2, the experimental design was successful and allowed for a straightforward comparison between the two load conditions. For that reason, no major changes were made to the task design in this experiment. Like in Experiment 2, in the present experiment the two memory load conditions were matched for number of items, presentation rate and probe timing. Matching these task parameters was critical to ensure that any differences in the timing variability measure between the two conditions was not an artifact produced by, e.g., effects of absolute rehearsal duration, number of item rehearsals, duration of silence between items, or probe frequency.

The memory load was varied in the present study using a STM preload. The preload paradigm allows the primary task to be matched in all aspects apart from the load on STM, which is varied via a separate preceding task. In a preload paradigm, participants are required to hold in memory the material from a first task (the preload) while performing a second task, and after completing the second task, participants are tested on the preload material. This paradigm has been used elsewhere (Baddeley & Hitch, 1974; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; FitzGerald & Broadbent, 1985; Halford, Maybery, O'Hare, & Grant, 1994; Morris, Gick, & Craik, 1988), often to determine the degree of resource-sharing between one type of STM or processing task and another. If performance on one task remains the same despite a change in the difficulty of the other, then the processes central to the two tasks are assumed to be independent. On the other hand, resource-sharing between the two tasks is assumed to take place when increasing the load in one task disrupts performance in the other. For example, FitzGerald and Broadbent (1985, Experiment 2) showed that, for visually-presented material, 6-digit preload lists produced an effect of load (where the preload digit lists were either fixed or variable across trials) on STM recall for 4-letter sequences.

The current experiment used a method similar to that used by FitzGerald and Broadbent's Experiment 2, except that letter stimuli were used for the preload task and digits were used for the rehearsal-probe task. The preload task was presented in the same modality (auditory) in order to ensure that the preload task loaded heavily on the resources required to perform the rehearsal-probe task. Memory load in this experiment was manipulated via the list length of the preload sequence. It was hypothesized that rehearsal timing variability would be higher in the high preload condition compared to the low preload condition. The second hypothesis was that, consistent with the results of Experiment 2, there would be a negative correlation between individual differences in rehearsal timing variability and STM capacity.

3.3.1 Methods

3.3.1.1 Participants

Participants were 25 undergraduate volunteers (18 women) recruited from the University of York with a mean age of 19.50 years (range 18 to 22 years). The compensation and informed consent details were the same as those from previous experiments. In addition to the eligibility criteria listed in Experiment 1, participants must have been able to correctly recall at least one 4-digit sequence in the digit span task in order to continue with the rehearsal-probe task.

3.3.1.2 Design

The experiment took a within-subjects design; all participants completed blocks of both high and low memory load trials. The dependent variables were the mean CSDs across serial positions (for each memory load condition and averaged across conditions), and the digit span score.

3.3.1.3 Materials

The digit stimuli used for the rehearsal sequences were those described in Chapter 2. In addition, the spoken letter names from the set {C, F, G, H, J, K, L, N, Q, R, S, T, X, Y} were used for the preload letter lists. The letter names in the set were selected based on phonological dissimilarity and equal number of syllables (one). The letter audio stimuli were digitally recorded in a sound-attenuated booth by a female native English speaker at 44100 Hz sampling rate as a single channel (mono). Post-recording processing on the stimuli was done using Audacity (Mazzoni & Dannenburg, 2000) and Praat (Boersma & Weenink, 1992) audio software. The letter recordings were adjusted to equal duration (400 ms) and maximum amplitude using the ‘tempo’ and ‘max amp’ features in Audacity. The letter stimuli were recorded and processed in the same manner as the digit stimuli, with the exception that the pitch contours were not reduced in the letter stimuli, and the files did not undergo any testing for p-centre adjustment.

3.3.1.4 Procedure

The rehearsal-probe task was administered according to the general procedures outlined in Chapter 2. The trials consisted of randomly-selected 4-digit sequences preceded by either a two-letter (low load) or six-letter (high load) preload list (see Figure 21). After responding to the rehearsal probe, participants were presented with a visual cue ("Letters?") to begin recalling the letters presented at the beginning of the trial. There was no time limit for responding to the rehearsal probe and the letter recall portions of the trials. For the letter recall, participants were able to correct their key press responses using the DELETE key, and they pressed the ENTER key to mark the end of their letter recall. The letters were scored correct or incorrect according to the presence or absence of the response in the letter sequence for that trial, i.e. the serial positions of the letter responses were ignored.

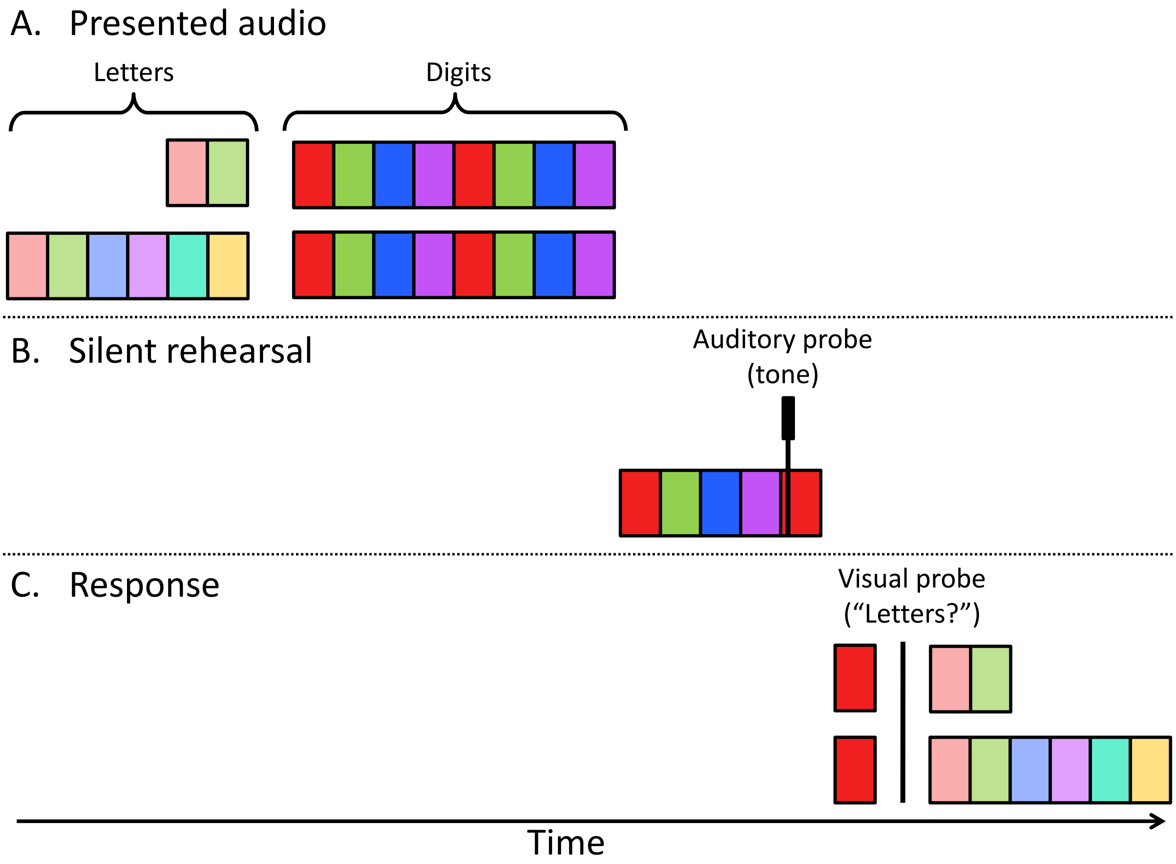

Figure 21. Representation of a single trial in the low preload (2-letter lists, top row in A and C panels) and high preload (6-letter lists, bottom row in A and C panels) conditions. Boxes represent the 400 ms letter/digit sequence items, and colours denote the serial positions within the letter and digit lists. The probe shown represents the earliest possible probe time, which was aligned with the first item in the second rehearsal cycle. Note that the digit rehearsal portion of the trial, between the letter sequence presentation and letter recall probe, was equivalent in the low (2-letter) and high (6-letter) preload conditions.

Participants completed 6 blocks of alternating low and high load condition trials, with the low load condition as the first block. Participants were informed about the length of the letter sequences at the start of each block. There were 16 trials in each block (2 trials for each of the 8 probe times) and 3 blocks per condition, resulting in 48 trials per load condition and 96 trials total per participant. In total there were 6 trials (2 trials per block x 3 blocks) for each of the 8 probe times within each condition. The probe times were selected randomly within blocks of trials. The blocks were separated by self-paced breaks, and participants were encouraged to rest for a few minutes between blocks.

The set of eight probe times ranged from 2083 ms to 3916 ms after the end of the last digit in the presented sequence. The set of probe times is provided on the x-axis of Figure 22 in the Results section. These probe times were selected such that the probes would occur during the beginning and end of each item in the second rehearsal cycle if the list was rehearsed exactly as presented during the silent rehearsal period.

These probe times were similar to those used in previous experiments, where probes were evenly spaced and corresponded to $^{1}/_{6}$, $^{1}/_{2}$ and $^{5}/_{6}$ time points within each 400 ms digit. However, there were two changes made in the probe timing relative to the previous experiments. First, the within-item probe times were determined based on the duration of the digit (400 ms) plus the inter-digit interval (100 ms). This was done because the two previous experiments showed substantial between-subject variation in rehearsal timing, including an apparent static offset in the rehearsal timing relative to the correct timing, so there was no reason to continue avoiding the presentation of probes during silent intervals in favour of biasing the probe timing toward individuals without such an offset. Second, the probe times in alignment with the halfway point of each digit were removed, leaving only the probe times in alignment with the beginning and end of each item during the second rehearsal cycle. This was done due to the practical constraints on the total number of trials that participants could be expected to complete within a single session, and to test whether a reduction in the frequency of probe times would have an appreciable effect on the results.

All participants were tested individually in quiet room by a single experimenter and in a single session. Digit span data was collected prior to the rehearsal-probe task. The digit span task was administered according to the procedures outlined in Chapter 2. In order to reduce the likelihood that participants would focus on their performance on one task at the expense of the other, the rehearsal-probe task instructions stated that their performance on the letter and digit portions of each trial were equally important.

3.3.2 Results

The responses to the rehearsal-probe portion of the task were first analysed according to response type; valid or invalid. Unlike the previous two experiments, there were no no-responses in this experiment because there was no time limit for the item response after the probe. There were 3 (0.3%) invalid responses in the low load condition and 10 (0.9%) invalid responses in the high load condition. The remaining responses for each participant were grouped by the two trial conditions (high or low memory load) and four serial positions of the responses (1st, 2nd, 3rd or 4th item in the sequence). One participant responded with the same serial position across all probe times. It was assumed that this participant either did not understand the task instructions or was heavily focused on the letter recall task, so the data from this participant was excluded.

For the remaining 24 participants, the valid item responses were subjected to a performance threshold for the letter recall portion of the task in order to ensure the validity of the memory load manipulation. In order for the rehearsal responses in the low load (2-letter sequence) trials to be included in the analysis, the participant had to correctly recall both of the letters presented before the rehearsal-probe portion of the trial, irrespective of serial order. For the high load (6-letter sequence) condition, the participant must have responded with at least three out of the six letters from the preload sequence in order for the rehearsal response to be included for that trial. The letter recall portion of the high load trials were also scored according to the number of letters present in the original sequence, irrespective of the serial order at recall.

The lower performance threshold (in terms of percentage correct) for the 6-letter preload trials compared to the 2-letter trials was used for two reasons. First, it was assumed that the 6-letter recall task would be considerably more difficult for participants, so using a higher performance threshold in this condition would have substantially reduced the number of trials included in the rehearsal analysis. Second, although the performance threshold for the high load trials was lower than the low load trials in terms of the percentage of letters correctly recalled, the absolute number of correctly recalled letters required for the high load trials was still higher than that for low load trials (three letters versus two letters). Thus, the performance threshold for the high load condition ensured that there was a minimum memory load difference of at least one letter between the two conditions.

The frequencies and percentages of accuracy scores for the 2-letter and 6-letter conditions are provided in Table 6. In the low load condition, 104 out of 1152 trials (9.0%) were excluded from analysis because fewer than two letters were correctly recalled at the end of the rehearsal portion of the trial. In the high load condition, 148 out of 1152 trials (12.9%) were excluded because fewer than three letters were correctly recalled.

Table 6. Frequencies and percentages of correct letters recalled across all trials in the low (2-letter preload) and high (6-letter preload) memory load conditions. Percentages are marginal by row. Bottom row shows the frequencies and percentages for the total responses (two groups combined). Rightmost column shows the total number of responses (trials) in each group and in total.

Load condition	Letters correct							Total
	0	1	2	3	4	5	6	
Low	16 (1.4%)	88 (7.6%)	1048 (91.0%)	-	-	-	-	1152
High	33 (2.9%)	35 (3.0%)	80 (6.9%)	165 (14.3%)	241 (20.9%)	323 (28.0%)	275 (23.9%)	1152
Total	49 (2.1%)	123 (5.3%)	1128 (49.0%)	165 (7.2%)	241 (10.5%)	323 (14.0%)	275 (11.9%)	2304

After eliminating the 252 (10.9%) of the 2304 total trials in which the letter recall performance fell below the threshold, the serial position response proportions at each probe time were calculated across the remaining trials in each load condition. Figure 22 shows the proportions of responses across probe times, averaged over all participants, for the low load (2-letter preload sequence, A) and high load (6-letter preload sequence, B) conditions.

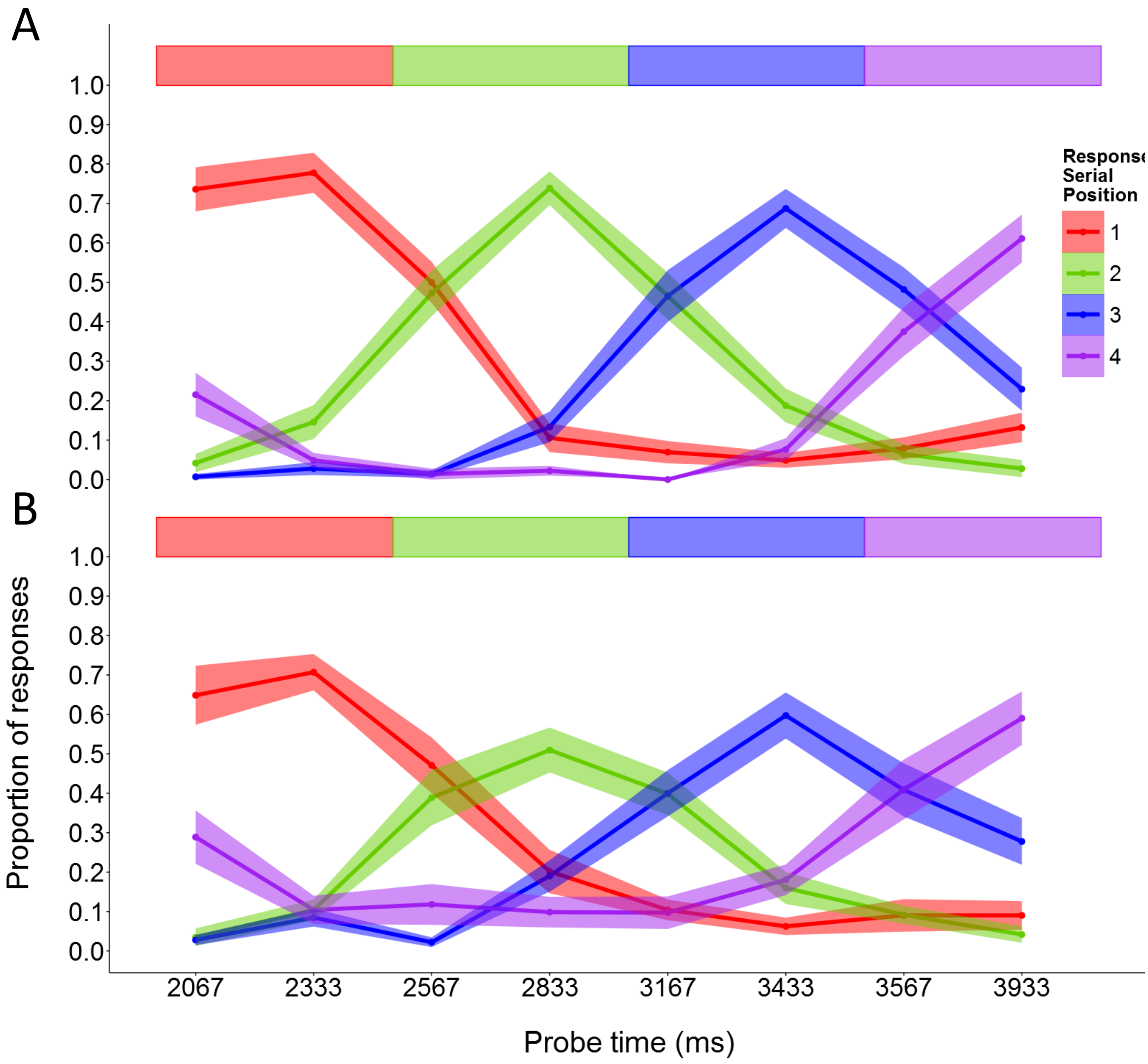

Figure 22. Mean proportions of responses for each serial position across probe times in the low (2-letter preload, A) and high (6-letter preload, B) load conditions. The widths of the ribbons at each probe time reflect the between-subject SEMs.

CSDs of the responses for the two load conditions and four response serial positions were calculated according to the procedure outlined in Chapter 2. The grand mean CSDs for the two memory load conditions, four response positions, and eight load by position combinations are shown in Table 7. Overall, the grand mean CSD was higher in the high load condition than in the low load condition, meaning that participants showed more variation in their responses to probes during rehearsal of 4-digit lists that were preceded by 6-letter sequences than to those preceded by 2-letter sequences. The grand mean CSDs across the serial positions of responses were similar with the exception of the last digit in the sequence (4th serial position), which was lower than the other positions.

Table 7. Mean CSDs (SEM in parentheses) for the 4 serial positions and the two memory load conditions; 2-letter (low) and 6-letter (high) memory preload. The maximum possible CSD value is 1.41, and the minimum possible CSD is 0.52 in this experiment.

Load condition	Serial Position 1	2	3	4	Mean
Low	0.73 (0.03)	0.70 (0.03)	0.70 (0.04)	0.62 (0.05)	0.69 (0.03)
High	0.73 (0.04)	0.76 (0.03)	0.83 (0.04)	0.76 (0.07)	0.77 (0.03)
Mean	0.73 (0.03)	0.73 (0.02)	0.76 (0.03)	0.69 (0.05)	

The grand mean CSD data were tested and found to have met the assumptions for a two-way load (2 levels) x serial position (4 levels) repeated-measures ANOVA. There was a significant effect of load condition on grand mean CSD, $F(1,21) = 10.28$, $p = .004$ (see Figure 23). The grand mean CSD of rehearsal-probe serial position responses in the high load condition (6-letter preload) was greater than that in the low load condition (2-letter preload). Although the mean CSDs varied across serial positions, and the effect of load on mean CSD seemed to vary with serial position (see Table 7), both the main effect of serial position and the interaction between load and position were not significant ($F(3,63) = 1.04$, $p = .383$, and $F(3,63) = 2.13$, $p = .105$, respectively).

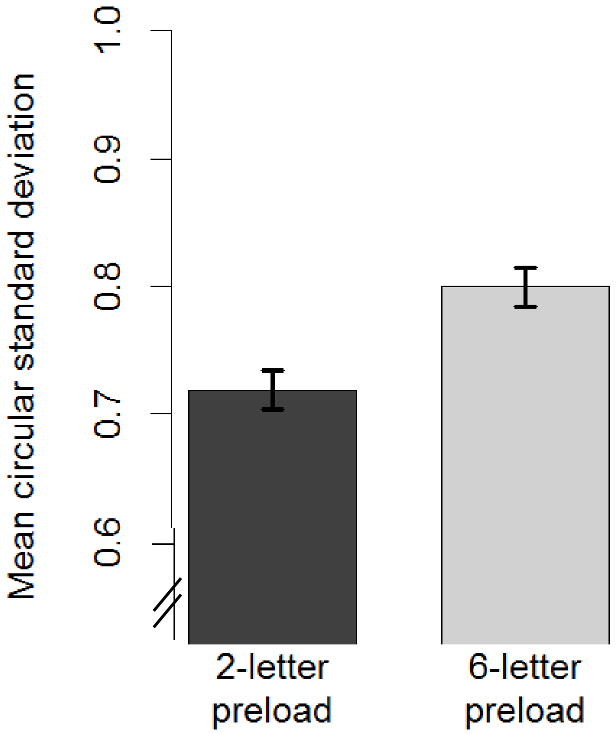

Figure 23. Mean CSDs for 4-digit rehearsal in the low preload condition (2-letter lists, dark gray) and high load condition (6-letter lists, light gray). Error bars show the 95% confidence intervals, adjusted for within-subject error using the method proposed by Cousineau (2005). The y-axis minimum reflects the lowest possible mean CSD score in this experiment.

The relationship between mean CSD and digit span was also examined. The digit span task was scored using the method described in Chapter 2. Digit span was significantly negatively correlated with individual mean CSD scores averaged over the two load conditions, *r* = -.59, *p* = .003. Correlations between digit span and the mean CSD scores for each load condition showed that this negative association was significant within the two conditions (digit span and low load mean CSD, *r* = -.57, *p* = .003; digit span and high load CSD, *r* =-.53 , *p* = .008). Figure 24 shows the negative associations between digit span scores and the mean CSDs in the high and low load conditions. Finally, there was no significant correlation between digit span scores and the mean CSD differences between the two conditions.

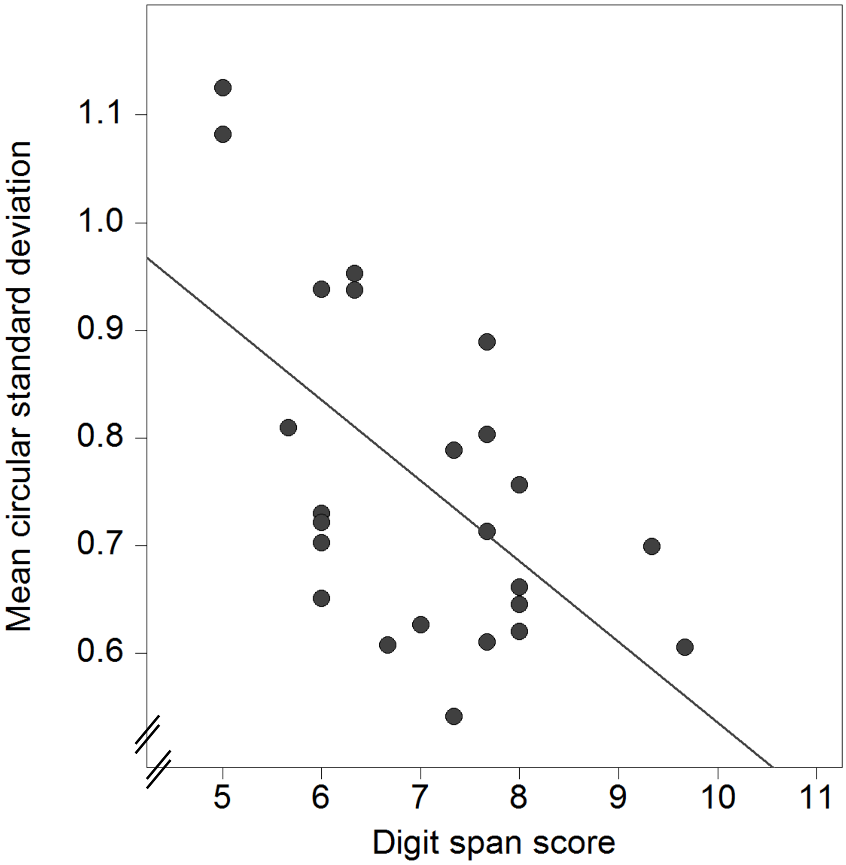

Figure 24. Scatterplot of the relationship between digit span scores and mean CSDs for 4-digit rehearsal lists for each participant, averaged over the two load conditions. Solid black line shows the linear regression model for mean CSD predicted by digit span score. The y-axis minimum reflects the lowest possible mean CSD score that can be obtained in this experiment.

3.3.3 Discussion

The results of this experiment have successfully replicated those of Experiments 1 and 2 in terms of the reliable patterns of response data. The response proportion functions were characterised by a rise and fall for each item over time. The modal probe times for each item response occurred in the correct serial order, and these mode times were roughly in line with the 'correct' presentation timing for each item. The mean CSD scores were normally distributed within the middle of the range of possible values, and they varied between participants and within participants under different conditions.

This experiment used a preload task paradigm to manipulate the STM load during paced subvocal rehearsal of 4-digit isochronous sequences, and the results showed a clear within-participant effect of STM load on the variability of sequence timing representations during rehearsal. Consistent with the hypothesis, timing variability during rehearsal was higher in the high STM preload condition compared to the low preload condition. Unlike Experiment 2, in this experiment the rehearsal digit sequences were identical in all aspects; the only difference between the two conditions was the number of letters presented before the rehearsal list, which then had to be remembered until the end of the trial. This provides strong evidence that the temporal precision of auditory-verbal sequence representations during a silent rehearsal period is constrained by shared resources. This result also supports the earlier conclusion that the use of LTM item-order associations as a STM load manipulation in Experiment 2 may have had a qualitatively different effect on the representation of sequence timing compared to load manipulations without a LTM component.

Also as hypothesized, the results of this experiment showed a significant negative correlation between individual differences in ISR capacity and temporal variability of sequence representations during rehearsal. The magnitude of this relationship was similar between the two load conditions, and the effect in the present experiment was similar to that observed in the high load condition (random 4-digit sequences) in Experiment 2. This provides support for the earlier suggestion that the introduction of LTM factors produced a qualitative difference in the maintenance of isochronous sequence timing compared to the random order condition in Experiment 2. Taken together, the correlational results from Experiments 2 and 3 suggest the ability to maintain the precise timing of auditory-verbal sequences is significantly related to ISR performance for the same type of material, but that this relationship applies primarily to situations where there are at least moderate demands on STM for serial order.

3.4 Experiment 4

The goal of Experiment 4 was to extend and generalise the findings of the previous three experiments by testing temporal variability during rehearsal with a different memory load manipulation, and to determine whether the task can be used with non-isochronous item timing. Memory load in the present experiment was manipulated via temporal grouping. Introspectively, temporal and rhythmic grouping is commonly thought to improve immediate memory for arbitrary sequences of spoken words; for example, telephone numbers tend to be rehearsed and recalled in predictable temporal groups or rhythmic patterns (Severin & Rigby, 1963). Empirical research has confirmed aspects of this intuition, showing that regular temporal grouping and stress patterns during sequence presentation and/or rehearsal produces better ISR performance overall, as well as different error patterns (Adams, 1915; Ryan, 1969a, 1969b; Severin & Rigby, 1963; Wickelgren, 1964, 1967). The effects of rhythm and timing on recall are especially apparent with auditory presentation, though grouping effects have also been observed in the visual modality (Collier & Logan, 2000; Frankish, 1985; Hitch et al., 1996). Recall for supraspan auditory-verbal lists has been found to reliably improve with regular temporal grouping compared to isochronous presentation (Frankish, 1985; Hartley, Hurlstone, & Hitch, submitted; Hitch et al., 1996).

Not only can recall benefit from regular grouping, but it can also be impaired by irregular grouping. Ryan (1969a) found that 9-digit auditory sequences were significantly better recalled when temporally grouped into a 3-3-3 pattern compared to all other possible permutations of three group sizes, and that recall performance was poorest with highly irregular grouping patterns (e.g. 1-7-1). These findings show that temporal grouping patterns can either improve or impair ISR performance, depending on the pattern regularity, relative to performance on isochronously-presented supraspan auditory-verbal sequences. Using ISR performance as an indicator of relative memory load, temporal grouping patterns that improve recall (i.e. regular groups of about 3 items) relative to ungrouped sequences can be considered as having a lower memory load than ungrouped sequences.

Because the effects of grouping on auditory-verbal ISR are generally investigated using supraspan list lengths to prevent ceiling effects, less is known about the effects of temporal grouping on STM for subspan auditory sequences. For instance, it is not clear whether the ceiling effect for serial recall of subspan sequences obscures an underlying beneficial effect of regular temporal grouping on auditory STM, or if temporal grouping only exerts an effect when the list exceeds span. There is some neuroimaging evidence to suggest that temporal grouping of subspan (6-item) lists affects patterns of neural activity during an ISR task, even when grouping effects are not identifiable from accuracy or reaction time measures due to ceiling effects (Henson, Burgess, & Frith, 2000; Kalm, Davis, & Norris, 2012). Kalm et al. (2012) observed decreased activity in auditory perceptual areas (bilateral posterior superior temporal gyrus) for grouped compared to ungrouped conditions during encoding of subspan sequences, and increased activity in the left inferior parietal lobe (supramarginal gyrus), which is known to be associated with storage and post-perceptual processing of phonological information (Henson et al., 2000; Kalm et al., 2012). Using visually-presented letter sequences in a full-list recognition task, Henson et al. (2000) found higher BOLD responses in the left dorsolateral premotor cortex for the isochronous compared to temporally grouped conditions. Because the left dorsolateral premotor cortex had previously been associated with rhythmic movement production, Henson and colleagues concluded that this area plays a role in the timing signals involved in aspects of serial order STM such as serial rehearsal and temporal grouping. It therefore seems plausible that temporal grouping produces quantitative and/or qualitative differences in STM processing, and while the effects of grouping on serial recall may only observable with supraspan lists, these effects may be detectable with span and subspan lists using the more sensitive temporal precision measure provided by the rehearsal-probe task.

Given the evidence that ISR for digit sequences presented in a regular temporal grouping pattern is superior to ISR for isochronous sequences, rehearsal-probe digit sequences presented in a 3-3 grouping pattern were considered to have a lower memory load than equal-length ungrouped lists. The hypotheses for this experiment were that temporal precision would be superior with 3-3 temporal grouping (low load) compared to ungrouped (high load) 6-digit sequences, and that individual differences in auditory digit span would be negatively correlated with temporal precision in both participant groups. The first hypothesis regarding the effect of temporal grouping on timing precision during rehearsal is consistent with a dynamic resource-sharing account of timing precision in auditory STM, as well as a continuous timing/context signal for coding serial order in STM. The hypothesis is also consistent with the results of Experiment 3, where temporal variability increased in response to a higher memory load. The second hypothesis concerning the relationship between individual differences in timing precision and STM span was based on the correlational results of Experiments 2 and 3, and is also consistent with a dynamic-resource allocation framework and a continuous signal used in the coding of serial order.

3.4.1 Methods

3.4.1.1 Participants

Participants were 28 undergraduate volunteers (23 women) recruited from the University of York with a mean age of 19.9 years (range 18 to 22 years). The details regarding ethics approval, participant eligibility criteria, compensation and informed consent were the same as those in previous experiments. In addition to the eligibility criteria listed in Experiment 1, participants must have been able to correctly recall at least one sequence of 6 digits in the digit span task in order to continue with the rehearsal-probe task.

3.4.1.2 Design

This experiment took a between-subjects design. Unlike Experiments 2 and 3, a between-participant design was used in order to avoid condition order effects that could occur in a within-participant grouping manipulation. Specifically, the concern was that the temporal grouping trials would impact the timing of rehearsal for the ungrouped sequences, as recall patterns are affected simply by instructing participants to rehearse or “think of” isochronously-presented sequences as being grouped (Farrell, Wise, & Lelièvre, 2011; Farrell, 2008; Ryan, 1969a; Wickelgren, 1964). Because Experiments 2 and 3 showed that timing variability is related to individual differences in STM span, it was therefore also necessary to ensure that the two participant groups did not significantly differ in STM span.

Participants were assigned to one of two sequence presentation groups; either isochronous or 3-3 temporal grouping of the 6-digit rehearsal lists. The order of group assignment was counterbalanced. The dependent variables were mean CSD and digit span scores.

3.4.1.3 Materials

The digit stimuli used were those described in Chapter 2.

3.4.1.4 Procedure

The rehearsal-probe task was administered according to the general procedures outlined in Chapter 2. Participants completed blocks of 6-digit rehearsal trials, where the sequences were either presented isochronously or they were temporally grouped. In the isochronous group, digits were presented at a rate of one digit per 500 ms. In the temporally grouped condition, the three-digit subgroups were presented at one digit per 410 ms, and subgroups were separated by 270 ms silent intervals (i.e. after the 3rd and 6th digits). The inter-subgroup silent interval was chosen by first restricting the overall sequence duration to 3 seconds in order to match the 500 ms presentation rate used in previous experiments and in the isochronous condition here. Then, the silent duration between subgroups was steadily increased until a clear temporal gap between subgroups was perceived without compromising the intelligibility of the items within subgroups as a result of the faster presentation rate. The temporally grouped sequences were checked to ensure that slight overlap of the 400 ms digits (as a result of variable p-centre offsets prior to acoustic onsets) did not cause masking or reduced intelligibility.

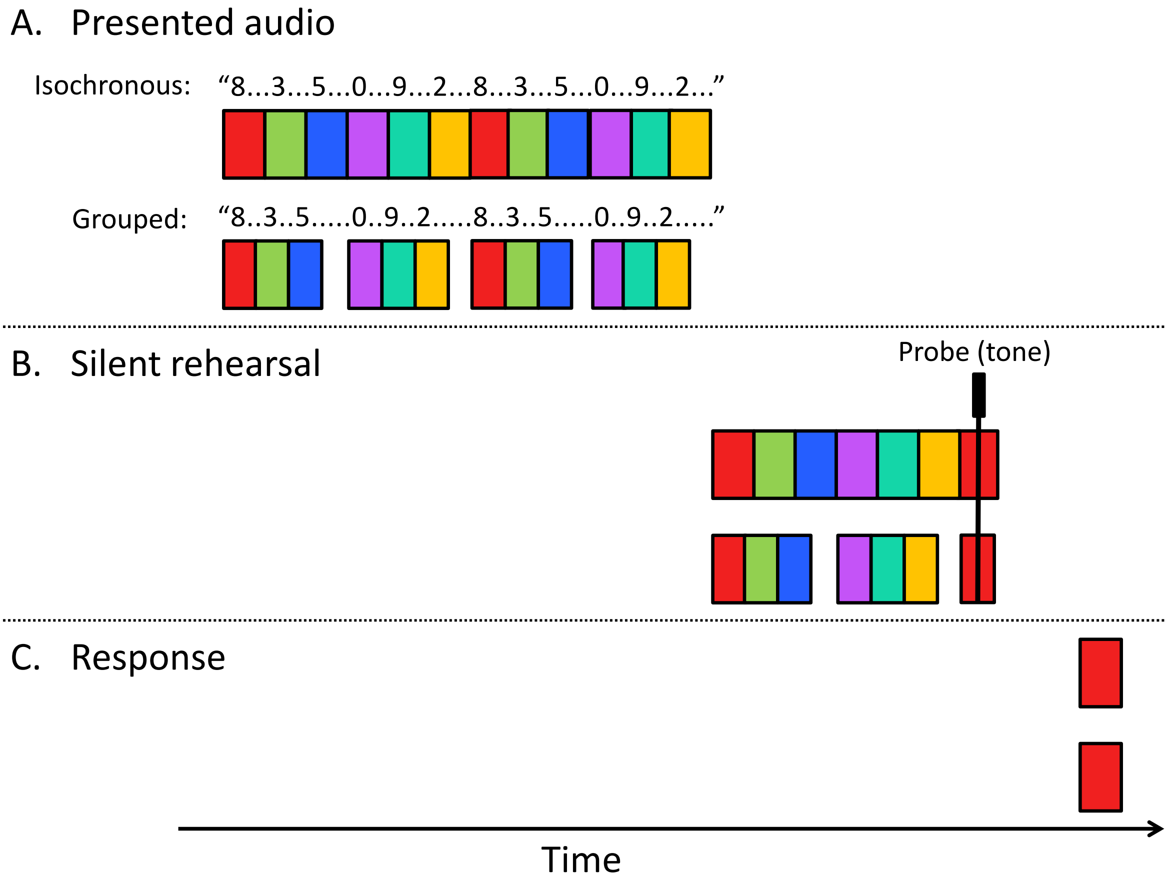

Figure 25. Schematic of the trial design. Six-digit sequences were either presented at an isochronous rate (A, top row) or in a 3-3 temporal grouping pattern (A, bottom row). After hearing two repetitions of the sequence, the participant began subvocally rehearsing the sequence exactly as it was presented (B) until hearing a tone (B, black vertical line). Upon hearing the tone, the participant was instructed to stop rehearsing and respond with the digit that was actively being rehearsed when the tone was presented (C). The probe shown represents the second earliest possible probe time, which is aligned with the beginning of the first item in the second rehearsal cycle.

There were 14 probe times in each experimental condition. The probe times corresponded to the beginning and end of each item during the second rehearsal cycle (if the list were rehearsed exactly as presented). In addition to these 12 probe times, there were two probe times added to extend the probe period beyond the 2nd rehearsal cycle; one was aligned with the end of 6th item during the 1st rehearsal cycle, and one was aligned with the beginning of the 1st item in the 3rd rehearsal cycle (see first and last probe times in Figure 26). Because the probe times were based on the presented item timing, the probe times were not always matched between the two conditions. There were three probes that were the same between the two conditions, which corresponded to the beginning of the 1st and 4th items in the 2nd rehearsal cycle, and to the beginning of the 1st item in the 3rd rehearsal cycle (shown in Figure 26). The two list types were matched for total duration, so the probe times did occur within the same range of durations (between 2.8 and 6.2 seconds after the end of the sequence presentation period). The reason for using different probe times was to avoid probing within the longer silent gaps between subgroups in the grouped condition, which would increase the variability of responses in that condition and thus bias the mean CSD measure.

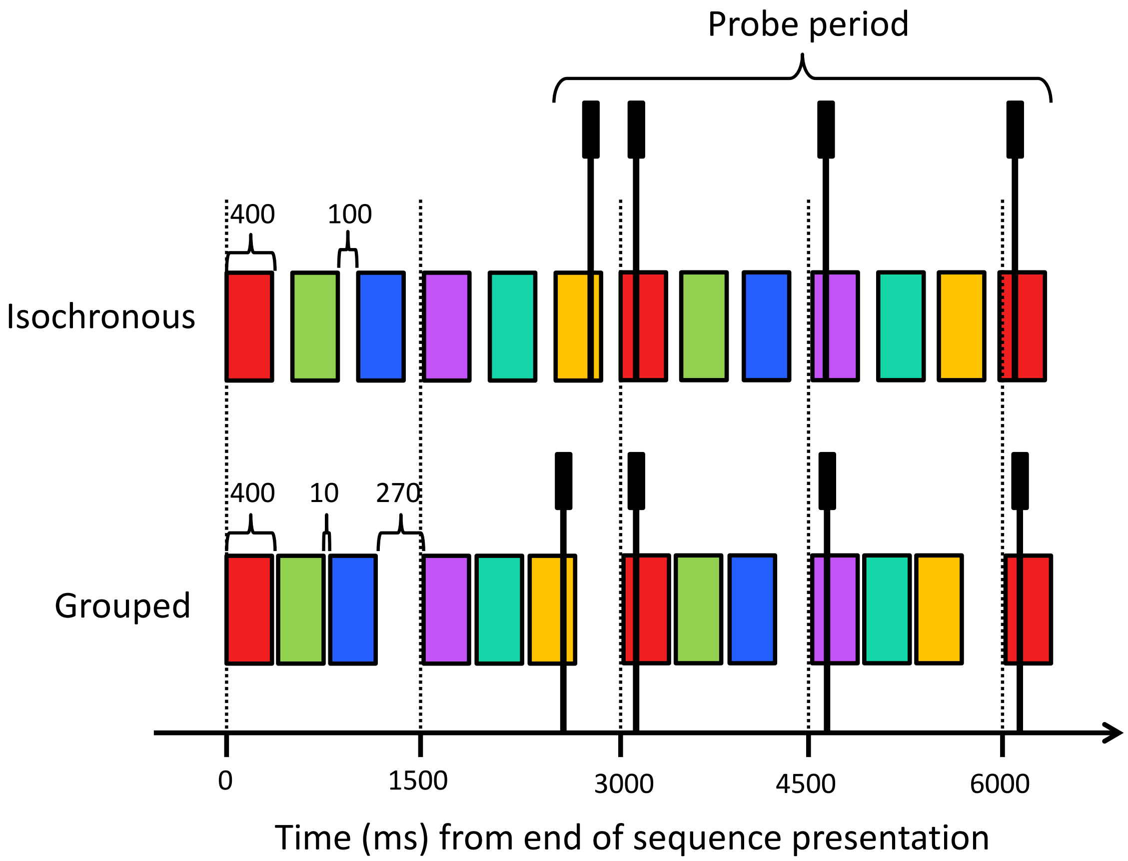

Figure 26. Depiction of the item durations (coloured boxes), inter-item durations and probe timing in the isochronous (top row) and temporally-grouped (bottom row) sequence presentation groups. Note that there were 14 probe times in total, but only four probe times are shown here (black vertical lines) for ease of visual display. The last three probe times shown, corresponding to the beginning of the 1st and 4th serial positions, were matched between the two conditions.

The two tasks were made up of three blocks of 42 trials. Probe times were selected randomly during each block, with each of the 14 probe times occurring three times per block. In total, there were 126 trials (9 trials per probe time x 14 probe times). Participants were given optional self-paced breaks between the blocks.

3.4.2 Results

The two experimental groups were analysed for differences in sex, age and digit span. A one-way chi-square test revealed no significant difference in the distribution of males and females between the two groups, $\chi^2(1) = 0.24$, $p = .622$. Table 8 shows the means and standard deviations for age and digit span distributions. Independent-samples *t*-tests confirmed that there were no significant differences between the two groups in these variables.

Table 8. Means, standard deviations and independent-samples t-tests for age and digit span data in the isochronous and temporal grouping experimental groups.

	Isochronous		Temporal Grouping				
Variable	Mean	SD	Mean	SD	*t*	df	*p*
Age (years)	20.1	1.1	19.8	1.2	0.43	26	.432
Digit span	6.93	0.88	7.24	1.20	0.78	26	.443

The item responses were categorised by response type; valid, invalid or no response. As in the previous experiments, invalid responses were those where the response item was not present in the digit sequence for that trial. A no response occurred when the participant did not respond within the two second response period after the probe was presented. Table 9 shows the frequencies and percentages of trials for each type of item response to the probe. A Fishers Exact Test on the response type by group frequency table showed that there was a significant relationship between the two factors ($p < .001$), and the standardised residuals ($z > +/- 1.96$) revealed that this difference was driven by the unequal distribution of no-responses. The number of no-responses was higher than expected for the participant group rehearsing temporally-grouped sequences ($z = 2.9$) and lower than expected in the group rehearsing isochronous sequences ($z = -2.9$).

Table 9. Response type frequencies and percentages within the two memory load groups; low (temporally-grouped) and high (isochronous). Percentages are marginal by row. Bottom row shows the frequencies and percentages for the total responses (two groups combined). Rightmost column shows the total number of responses (trials) in each group and in total.

	Response type			
Load condition	Valid	Invalid	No response	Total
Low	1542 (87.4%)	17 (1.0%)	205 (11.6%)	1764
High	1620 (91.8%)	13 (0.7%)	131 (7.4%)	1764
Total	3162 (89.6%)	30 (0.9%)	336 (9.5%)	3528

The no-response and invalid response trials (10.4% of total) were removed. Proportions of responses for each serial position were calculated across all probe times for each participant. These serial position response proportions were then averaged over participants in each group. Figure 27 shows the mean (solid lines) and SEM (width of ribbons) proportions of responses for each item at each probe time in the temporally-grouped (A) and isochronous (B) experimental groups. The pattern of responses in the isochronous group is very systematic and consistent with expectations; this pattern closely resembles the response probability functions in Experiments 1, 2 and 3. By contrast, the data from the temporally-grouped sequences appear qualitatively different from the isochronous sequences in this and previous experiments. The distributions of item responses within the temporally-grouped condition are more varied in shape; some distributions are multi-modal, and they differ in dispersion over probe times as well as maximum probability. Also, the between-participant variability in response proportions (indicated by width of ribbons) appears to be greater surrounding the response peaks.

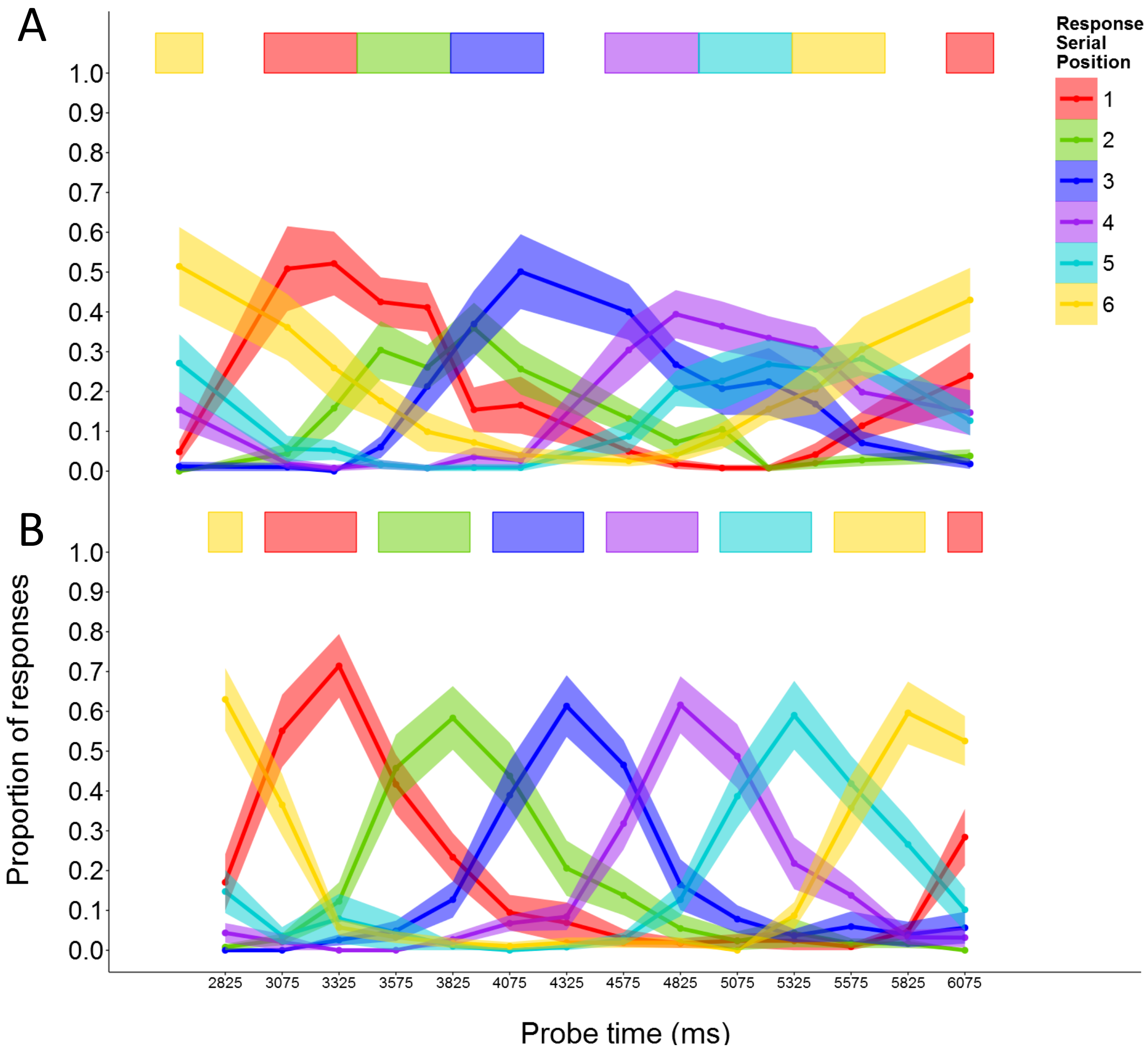

Figure 27. Mean proportions of responses for each serial position across probe times in the temporally-grouped (A) and isochronous (B) sequence presentation conditions. The widths of the ribbons at each probe time reflect the SEMs for mean response probabilities at each probe time.

Although the response probabilities in the temporal grouping condition are relatively more disorganised, there is evidence that these responses were somewhat systematic and that the participants generally performed the task as instructed. First, the modal probe time for each item occur in the expected serial order, which suggests that, on average, serial order errors were infrequent. Second, the mean response probabilities for each item approached or reached zero when the probe time was maximally distant from the modal probe time(s) associated with an item and rehearsal cycle. The between-participant variability was negligible for response probabilities that approached zero, meaning that participants were very consistent in not responding with an item that was highly out of phase with the modal probe time(s) for that item.

Grand mean CSDs were higher in the temporally grouped experimental group than in the isochronous group (see Table 10). CSDs tended to be higher for middle sequence items (positions 2, 3 and 4) than for items near the beginning or end of the sequence (1, 5 and 6), with the exception of serial position 6 in the temporally-grouped condition. Variances were similar across groups and serial positions, with the possible exception of item 6 in the isochronous condition, which showed the greatest amount of variance between participants.

Table 10. Mean CSDs (SEM in parentheses) for the two groups, six serial positions, and 12 group by position combinations. The maximum possible CSD value is 1.41, and the minimum possible mean CSD value is 0.26 in this experiment. The minimum possible CSD values for serial positions 1 and 6 are slightly lower than the others (0.25 vs. 0.26) due to the increase in probe times aligned with the 1st and 6th sequence items.

	Serial Position						
Group	1	2	3	4	5	6	Mean
Isochronous	0.58 (0.05)	0.60 (0.05)	0.63 (0.06)	0.65 (0.04)	0.58 (0.03)	0.49 (0.07)	0.59 (0.04)
Temporally Grouped	0.63 (0.05)	0.76 (0.06)	0.75 (0.05)	0.75 (0.04)	0.66 (0.04)	0.70 (0.05)	0.71 (0.04)
Mean	0.61 (0.04)	0.68 (0.04)	0.69 (0.04)	0.70 (0.03)	0.62 (0.03)	0.60 (0.05)	

The data were tested and found to have met the assumptions for a two-way mixed ANOVA, with group as a between-participant factor and serial position as a within-participant factor. There was no significant main effect of the experimental group on mean CSDs, $F(1,26) = 4.11$, $p = .053$ (see Figure 28). However, there was a significant main effect of serial position on mean CSD, $F(5,130) = 4.56$, $p < .001$. This effect was followed up with pairwise paired *t*-tests, which revealed that the CSDs for serial positions 1 and 4 were significantly different, $p = .009$ after Bonferroni correction for multiple comparisons (see Figure 29). No other differences between serial positions reached statistical significance. Finally, the interaction between the experimental group and serial position was non-significant, $F(5,130) = 1.95$, $p = .091$.

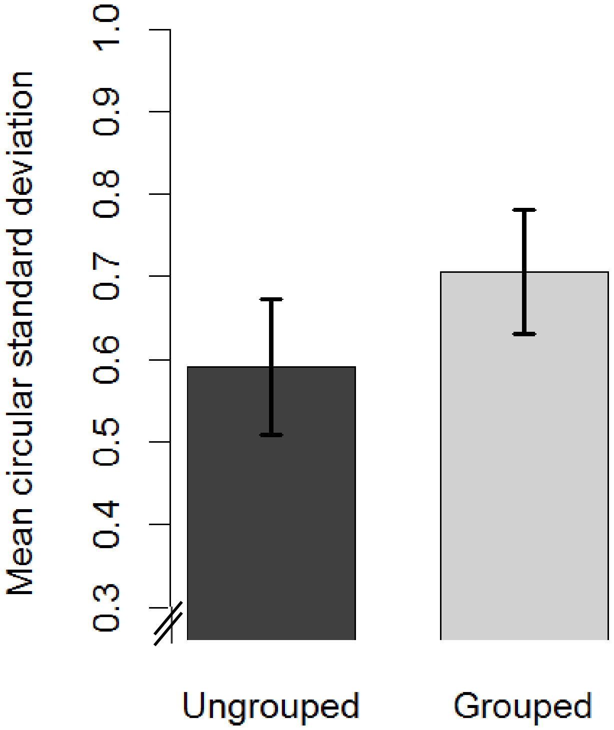

Figure 28. Mean CSDs for 6-digit rehearsal in the high load (ungrouped/isochronous, dark gray) and low load (3-3 temporal grouping, light gray) sequence presentation experimental groups. Error bars show the 95% confidence intervals. The y-axis minimum is adjusted to reflect the lowest possible mean CSD score in this experiment.

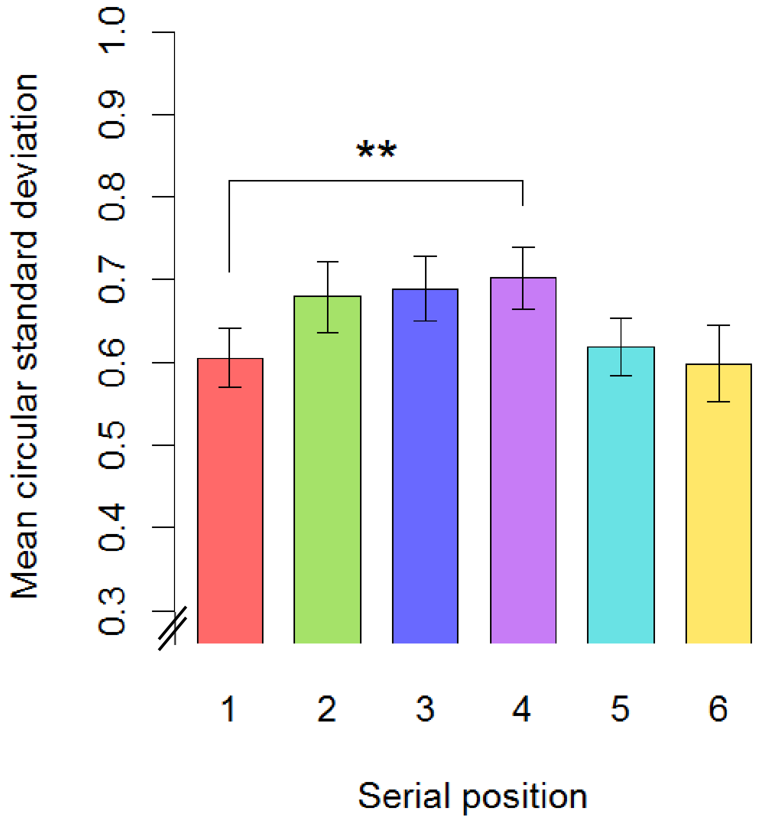

Figure 29. CSDs by serial position, averaged over the two experimental groups. Error bars show the 95% confidence intervals, adjusted for within-subject error using the method proposed by Cousineau (2005). **$p < .01$. The y-axis minimum has been adjusted to reflect the lowest possible mean CSD score in this experiment.

Finally, individual differences in mean CSD and digit span were examined. Across participants in both experimental groups, the association between mean CSD and digit span did not reach significance, $r = -.32$, $p > .05$. When the two experimental groups were examined separately, a distinct pattern of correlations emerged (see Figure 30). There was a significant negative correlation in the isochronous rehearsal group, $r = -.56$, $p = .036$ (Figure 30, A). There was no significant correlation across participants who rehearsed the temporally-grouped sequences, $r = .12$, $p > .05$ (Figure 30, B).

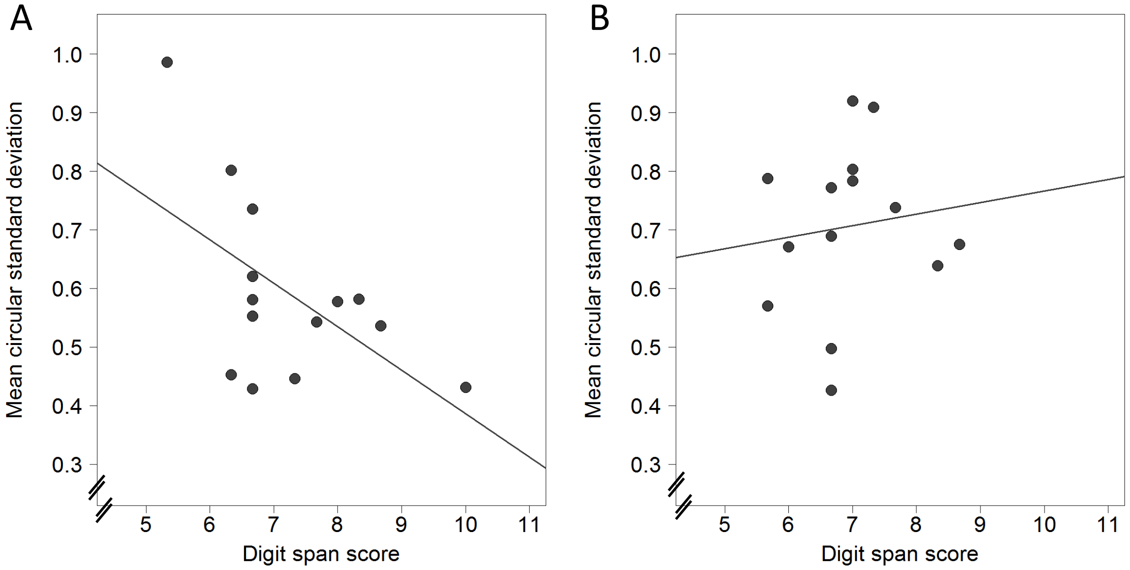

Figure 30. Scatterplots of the relationships between digit span scores and mean CSD in the isochronous (A) and temporally-grouped (B) sequence timing groups. Solid black lines represent the linear regression functions for mean CSD predicted by digit span in each group. The y-axis minima reflect the lowest possible mean CSD score in this experiment.

3.4.3 Discussion

One general aim of this experiment was to explore the possibility of using non-isochronous sequences in the rehearsal-probe task paradigm. Temporal precision during rehearsal was measured for isochronous and temporally-grouped digit sequences, where the two conditions were matched for number of items, number of probes and total sequence duration. The results in the isochronous groups showed lawful patterns of response proportions over time, and these replicated the patterns observed in previous experiments. However, the temporally grouped sequences produced a less reliable and more varied set of response data. While the pattern is still visible in this condition, the response probability functions for each item tended to be more variable than those in the isochronous condition here and in previous experiments. Also, in contrast to the response probability functions from isochronous conditions presented in this chapter, some of the response distributions in the temporally grouped condition had multiple probe time modes rather than a single clear peak. It is not clear whether the less structured response patterns observed for the temporally grouped sequences were a result of an inability for the general task paradigm to handle deviations from isochrony, or if they occurred as a result of specific issues with the task design. Later on in the discussion that follows, an issue with the task design will be outlined that may explain the less reliable pattern of results in the temporally grouped condition.

A comparison of the two memory load conditions revealed that, contrary to the hypothesis, temporal variability during rehearsal was not found to be lower for temporally-grouped (low load) sequences than for isochronous (high load) sequences. Instead, average temporal variability was higher in the temporally-grouped condition, meaning that participants who rehearsed the temporally-grouped sequences tended to show greater variability in responses to the same probes across trials. Despite the striking qualitative difference between the response probability functions (see Figure 27), the trend towards greater temporal variability in the temporally-grouped condition should be treated with caution as the difference between conditions was not statistically significant. Across both experimental groups, within-participant responses tended to be more variable for mid-list items (2, 3 and 4) relative to items near the ends of the list (1, 5 and 6), and this difference was significant for items in the 1st and 4th positions. It is interesting to note that the pattern of temporal imprecision across sequence items is similar to the pattern of recall errors for ungrouped lists, where recall is poorer for mid-list items. Finally, in the isochronous presentation group, individual differences in mean CSD were significantly negatively correlated with digit span. This result was consistent with the correlations observed in Experiments 2 and 3.

There are two main issues which may have contributed to the unexpectedly high temporal variability for the temporally grouped sequences. The first issue relates to the stimuli presentation timings, namely the lack of a metrical hierarchical structure in the presentation timing of the grouped sequences. The onsets of the two subgroups of three items were not aligned with the underlying beat produced by the rate of digits within subgroups. In the experiment, three digits within a subgroup were presented at a rate of one digit every 410 ms (average inter-digit silent interval[6] of 10 ms), with a silent interval of 270 ms between the two subgroups. In hindsight, the duration between the two subgroups should have been a factor of the underlying temporal structure formed by the within-subgroup item rate. The most straightforward hierarchical temporal structure for a 3-3 sequence is one where the three item subgroup timing is isochronous, and the inter-subgroup silent interval is either the duration of a single item or a multiple of the item duration. For example, for a within-subgroup rate of one digit per 410 ms, the between subgroup silent interval should have been 410 ms or a multiple of 410 ms (e.g. 820 ms). Indeed, this type of temporal structure was used by Ryan (1969a, 1969b) with 9-item lists, where one item was presented every 450 ms and the three subgroups were separated by 900 ms silent intervals. Hierarchical rhythmic structure results in a simpler ratio of rhythms, and results from finger-tapping studies have shown that performance is reliably better for simple versus complex rhythmic ratios, both initially and after extended practice (Collier & Wright, 1995).

---

[6] The inter-item silence was not exactly 10 ms due to the different p-centre silent offsets before each 400 ms digit. The pre-digit p-centre silent padding was subtracted from the post-digit silent interval, which resulted in p-centres occurring at 410 ms intervals.

The relatively irregular and complex presentation timing of the temporally grouped sequences in the present experiment may have increased the memory load relative to the isochronous lists, resulting in an additional unintended load difference between the experimental groups. Non-verbal auditory timing studies have shown that hierarchical metrical structure improves inter-beat interval precision on beat synchronisation (Madison, 2014) and STM tasks (Collier & Wright, 1995), and it reduces change detection thresholds for small perturbations of relative interval timing (Grube & Griffiths, 2009). A relative increase in task difficulty due to irregular timing in the grouped condition reported here would explain the higher amount of temporal variability, specifically because the silent duration between subgroups, corresponding to the second metrical level, might have been particularly vulnerable to imprecision both between participants and within participants across trials. This conclusion could be examined by repeating the irregular grouped timing and requiring full-list ISR at the end of each trial, however it is likely that recall would still be close to ceiling for these subspan lists, and therefore an increase in task difficulty might not be detected from recall.

A second issue to consider relates to a more general limitation of the task paradigm and the dependent variable, mean CSD. Like Experiment 1, which compared rehearsal timing for sequences of different presentation rates, it proved difficult in the current experiment to compare mean CSD scores between the two conditions directly because the sequences differed in item presentation and probe timings. Variability of isochronous serial interval production for ~400-800 ms intervals is known to increase as a function of the interval duration, and this is complicated by the fact that the relationship between variability and inter-beat interval may not be linear (Madison, 2001, 2014). This means that it is not possible to directly compare the response distributions for the same number of items that differ in relative timing, even if the probe times are the same. While it is possible to normalise the measures of temporal variability by dividing by the durations of intervals between the 'beats' (as in e.g. Madison, 2014), this solution should be treated with caution because it assumes a linear relationship between interval duration and variability of interval production.

Another way to deal with the limitation of the CSD measure for answering questions related to grouping in ISR would be to use non-temporal grouping cues to study the effects of grouping and memory load on temporal precision. Recall patterns showing signs of internal grouping have been produced from non-temporal grouping markers, such as pitch fluctuations and tones between subgroups (Frankish, 1989, 1995; Reeves, Schmauder, & Morris, 2000; Ryan, 1969a), as well as instructions to group isochronous (Wickelgren, 1964, 1967) and non-isochronous (Farrell et al., 2011) sequences during rehearsal. If the memory load of isochronously-presented sequences is reduced simply by providing non-temporal grouping cues or instructions, then it is possible that the temporal precision of the sequence representations in STM is improved by the addition of the cues. That is, it might be the case that subjective non-temporal grouping of isochronous sequences is itself enough to reduce the memory load without having any impact on the STM for isochronous temporal structure. On the other hand, if the non-temporal grouping cues improve STM only by inducing *temporal* grouping in the internal sequence representations, then these temporal grouping patterns would surface in the rehearsal-probe task. In the latter case, the response data would show either temporal grouping patterns or generally poorer temporal precision due to interference between temporally-grouped (induced) and the presented sequence timings.

A final point to consider is whether the poorer temporal precision in the temporal grouping condition occurred because the introduction of rhythm has an opposing effect on variability for auditory-verbal sequence timing compared to that of non-verbal auditory sequences. Studies by Madison (2014) and Grube and Griffiths (2009) have shown that temporal precision for auditory tone sequences improves when the tones are presented with hierarchical metrical rhythms compared to simple isochronous beats. In the case of auditory verbal lists, it is possible that rhythmic presentation actually introduces additional complexity when the lists must be remembered with exact timing, and thus these rhythms result in an increase in memory load rather than a decrease relative to isochrony. If this were true it would be in contrast to the effects of regular rhythmic presentation on ISR performance, as temporal grouping is known to improve memory for serial order. This question could be addressed by repeating the experiment using temporally grouped sequences with a regular hierarchical metrical structure, as suggested earlier.

In conclusion, this experiment has replicated the qualitative response patterns and correlational findings from previous experiments while exposing a limitation of the method, as well as highlighting the possible importance of metrical structure on precision of timing in STM. Further research is needed to determine whether the unpredicted result regarding increased response variability for temporally-grouped sequences was due to a lack of hierarchical metrical structure for non-isochronous auditory-verbal sequences, or if the result is an artifact of the task or dependent variable. The ability to draw conclusions regarding the possible importance of hierarchical metrical structure in this experiment was hindered by an inability to compare the dependent measure between sequence presentation conditions that differ in relative item timing. This experiment has extended the conclusions from Experiment 1 by showing that the CSD dependent variable becomes problematic when the item timing is not matched between conditions, even when the conditions are matched for number of items, total sequence duration and number of probe times.

3.5 General Discussion

The general aims in this chapter were to establish and refine the novel rehearsal-probe task, and to investigate the relationship between temporal precision in STM and memory load. The first aim was further broken down into specific predictions regarding the qualitative characteristics of the response proportion functions derived from the task, and the utility of the mean CSD measure as a quantification of temporal imprecision in sequence representations. This discussion section will evaluate these two specific predictions regarding the rehearsal-probe task itself, and then reflect on the effect of memory load on the temporal precision of auditory-verbal sequence representations in STM. Lastly, the emerging relationship between individual differences in temporal variability and STM capacity will be discussed.

3.5.1 Response proportion functions

Across four experiments, the rehearsal-probe task generally yielded systematic patterns of response data. Response proportions for each item were characterised by a rise and fall over the course of the rehearsal delay period. The periodicity of these distributions was revealed by multiple peaks for the distributions of responses corresponding to serial positions rehearsed near the edges of the probed period. The cyclical nature was most apparent for the first and last sequence items because, in the experiments presented here, the probed time period was centred on a single list cycle (beginning in alignment with rehearsal of the first serial position and ending with the last serial position). Furthermore, the modal probe times for each item occurred in the correct serial order, and these response distribution peaks were roughly in time with the expected rehearsal timing. While the response proportion functions retained this general character, they also proved to vary between individuals and across experimental manipulations of STM load. These observations are critical for confirming that participants are capable of performing the novel subvocal rehearsal task as expected, and that the properties of the response proportion functions are not invariant to task and individual factors.

In most cases the patterns of response proportions were clear, with the exceptions of 6-digit rehearsal in Experiment 1 and the temporally-grouped 6-digit rehearsal in Experiment 4. In these two cases, the mean distributions of serial position response proportions over probe times were not clearly unimodal, and the responses were spread more widely across probe times. In the 6-digit (faster rate) group in Experiment 1, the less systematic pattern was likely due to the increased item and probing frequency relative to the 3-digit (slower presentation rate) condition. The increased item rate may have meant that the distinction between adjacent items during subvocal was less clear compared to when the items were more temporally distinct. Also, the increased probe rate resulted in a finer-grained temporal resolution of rehearsal activity relative to the 3-digit group, which meant that the between-item ambiguity was more likely to be queried by a probe.

In Experiment 4, the more irregular pattern of responses for 6-digit temporally-grouped sequences may have been due to the combination of increased item and probe frequency within the 3-digit subgroups, as well as reduced precision during rehearsal of the temporal gap between subgroups. It is possible that the more variable patterns of response distributions that resulted from these two tasks were due to the additional items (6 digits vs. 3 or 4 digits), although this explanation does not account for the very systematic response probability functions in the Experiment 4 isochronous 6-digit group. This reason for this pattern of results could be examined in future work by examining the response proportion functions for rehearsal of sequences varying in number of items, where different conditions are matched for item presentation rate as well as either exact probe timing or number of item rehearsals before the probe.

3.5.2 Mean CSD as a dependent variable

Mean CSD, described in detail in Chapter 2, was used as the dependent variable and intended to be a coarse measure of response variability. The circular equivalent of the typical linear standard deviation was used in order to account for the cyclical nature of the serial position response distributions over probe times. The periodicity of the response proportions was expected due to the cyclical rehearsal enforced by the task, and this property was confirmed by the plots of the response proportions (discussed in the previous section). In all experiments, the mean CSD data varied both between-participants and within-participants across task situations, and this variance was normally distributed toward the lower end of the range of possible values. The mean CSD measure appears to capture the main performance attribute of interest for this investigation, which is the variability of serial position responses. Furthermore, statistical analysis of the mean CSD data produced findings that both did and did not support the prior hypotheses across the four experiments, which shows that the mean CSD variable is useful for testing hypotheses and is capable of providing unpredicted results.

While the mean CSD measure works well when the tasks are matched across all possible parameters, this dependent variable becomes problematic as soon as the rehearsal tasks are not matched for one or more key attributes. The critical task attributes include the number of sequence items, item presentation rate, number of probe times, and frequency of probe times. In Experiments 2 and 3, the rehearsal-probe task conditions were matched for all important task parameters, which allowed a direct comparison of the mean CSD data and straightforward interpretation of the results. However, Experiment 1 highlighted the fact that the mean CSD measure is affected by probe frequency, as evidenced by the difference in the lowest attainable mean CSD score between the two tasks. Experiment 4 further demonstrated that the mean CSD results may be biased when the exact item presentation and probe timings are not matched between conditions, even if the conditions are matched for sequence duration, number of items, number of probes, and the range of times over which the probes are presented.

The fact that the rehearsal-probe task conditions must be matched on a number of task parameters limits the flexibility of the rehearsal-probe task design. The use of different item presentation rates will affect measures of temporal variability due to the difference in the amount of inter-item silence, which in turn affects the amount of time during rehearsal where the responses to probes will be ambiguous and therefore likely to be more variable. Also, the duration of inter-item silence restricts the amount of local temporal variability that is possible while retaining the correct serial order. While these are not problems with the mean CSD variable per se, they do present a challenge for directly comparing measures of temporal variability between rehearsal-probe conditions where the item presentation rates differ.

3.5.3 The effects of memory load on temporal precision

The results of four experiments presented in this chapter provide limited support for the hypothesis that STM for the timing of subspan auditory-verbal sequence representations becomes more variable with increasing memory load. In Experiment 1 there was no effect of memory load on timing precision during rehearsal, where load was varied via the number of items in the rehearsal sequences. However, the high and low memory load tasks in this experiment were not matched for presentation rate or probe frequency, and further consideration of the mean CSD variable suggests this difference between the tasks produces a bias in the mean CSD measure (see previous section for more discussion on this topic). Therefore, while Experiment 1 was useful for establishing and refining the rehearsal-probe task, the result of this experiment with regard to the effect of memory load on temporal precision was inconclusive.

In Experiment 2, the within-participant memory load conditions were well-matched for the critical task properties so the mean CSD values could be directly compared. In this case the results showed that, contrary to the hypothesis, an increased memory load was associated with lower temporal variability during rehearsal. However, upon reflection, it seems that the specific method of manipulating load in this experiment was not valid. The memory load of 4-digit sequences in this experiment was reduced using over-learned item-order associations (i.e. counting) relative to random-order sequences. While this manipulation was expected to reduce the memory load for order information in the familiar order condition, it appears to have been detrimental to the isochronous timing aspect of the task. The use of sequences from LTM likely introduced non-isochronous rhythms that have been previously learned with the familiar item-order associations (Schulkind, 1999). Also, it is not clear whether the instructions to rehearse the counting sequences precisely and isochronously conflicted with the utility of doing so in this context, given that the item-order information was recoverable from LTM so long as the first item was maintained in STM. In light of these issues, it is difficult to draw conclusions about the influence of memory load on temporal precision based on the results of Experiment 2 without further research to clarify the influence of LTM on isochronous timing.

The unexpected finding from Experiment 2 could be explored in future studies using other auditory-verbal stimuli with characteristics similar to digits (i.e. closed set of items, familiar but arbitrary order, and no syntactic structure). For example, ordered and random sequences of letters or months could be used in a design similar to that used in Experiment 2 (e.g. “L M N O” vs. “O M N L”, “January February March April” vs. “March February April January”). Based on the results of Experiment 2, it would be predicted that the rehearsal timing for subspan sequences of this material is more variable when the words are presented in familiar order compared to random order. The effect of long-term item-order associations on timing could also be examined across languages, where languages with different prosodies would be expected to have different (non-isochronous) rhythmic structures linked to the same sequences. The different cross-linguistic rhythms would be predicted to arise during isochronous rehearsal of familiar sequences, but not for sequences of the same items in an unfamiliar order. Finally, it would be interesting to explore whether there is any effect of presentation modality, as it is not clear whether the encoding of isochronous, sequential visual presentation of overlearned sequences would automatically induce the rhythmic properties linked to their auditory representations.

The results of Experiment 4 on grouping showed no significant effect of memory load on temporal variability, but these results were also difficult to interpret for two reasons. First, like Experiment 1, the mean CSD measure may have been biased due to differences in the properties between the two tasks. In particular, it is possible that the mean CSD measure was affected by the irregular timing of the items and probes in the temporally-grouped condition. Second, like Experiment 2, a confounding factor introduced by the memory load manipulation may have produced counteracting effects that produced a critical deviation from the intended manipulation. Specifically, there was no hierarchical metrical structure to link the item and subgroup timing within the temporally grouped sequences, meaning that it may actually have been as or more difficult to retain the timing of the temporally-grouped sequences compared to the isochronous sequences. The results of Experiment 4 could be clarified by repeating the experiment with hierarchical grouping patterns and with multiple conditions, where the sequences would be matched for both within-subgroup item presentation rate as well as sequence duration. The issues with matching the temporally-grouped and isochronous sequences could be avoided using the rehearsal-probe task with only isochronous sequences, where in one condition there are instructions to mentally group the items in threes. Another possible method for exploring the effect of grouping as a memory load manipulation is to use non-temporal grouping cues (e.g. pitch fluctuations, auditory or visual signals) to mark the boundaries of subgroups within isochronously-presented sequences.

Of the four experiments in this chapter, Experiment 3 was the only one where the two tasks were well-matched across the critical task properties and the load manipulation was valid. In terms of the effect of memory load on timing precision, the results of this experiment are the most convincing. In Experiment 3, the memory load of 4-digit sequences was manipulated using a preload paradigm so that the rehearsal sections of the trials were equivalent between the two load conditions. The results of Experiment 3 showed a significant difference in the expected direction; within-participant temporal precision for rehearsal of random 4-digit sequences was superior in the low preload condition compared to the high preload condition.

The result of Experiment 3 is difficult to explain with a fixed/stochastic view of timing precision in STM, and it is more compatible with a dynamic resource sharing account of timing precision in STM. This finding suggests that the timing precision for STM of randomly-ordered auditory-verbal sequences is not fixed, but rather is task-dependent and constrained by the same limited resources that are required for the maintenance of item and/or order information. This appears to only be the case for sequences of items presented in a random, novel order because, as previously discussed, Experiment 2 showed that temporal precision suffers for the isochronous temporal maintenance of overlearned auditory-verbal sequences. The reliability and generalisability of the Experiment 3 result should be tested by replicating the experiment using different auditory-verbal stimuli and list lengths for both the preload and rehearsal sequences.

To summarise, the mean CSD measure was sensitive to the effects of sequence familiarity, memory preload and temporal grouping. While the effects of these manipulations were not always in the expected directions, they do show that timing precision is not fixed but rather is sensitive to the task and to individual differences. Also, it is clear that the manipulations originally labelled as 'memory load' across experiments were not all representing the same underlying factor. Instead, these manipulations resulted in inadvertent but nonetheless interesting effects on timing variability.

3.5.4 Relationships between ISR capacity and temporal precision

While the effects of memory load on timing variability were the main focus of the experiments presented in this chapter, correlations between individual differences in timing variability and STM span were also examined in Experiments 2, 3 and 4. Temporal variability in the representation of isochronous, sub-span digit sequences was found to be significantly negatively correlated with auditory digit span in these three experiments. This relationship is consistent the broader notion that timing precision for isochronous sequence representations in STM is variable rather than fixed, both between individuals as well as within individuals across task situations. In other words, the data presented here provide initial support for the view that temporal variability is a function of both the task context, such as memory load, as well as individual differences in ISR capacity. That the negative correlation between temporal variability and digit span was replicated in all three experiments where digit span data was collected, as well as for two different list lengths, suggests it is not likely to be a false positive result. However, the experiments presented in this chapter were not designed for sensitivity to individual differences. These experiments were underpowered to reliably detect such correlations and did not control for possible mediating variables, so these results should be treated with caution. A dedicated experiment with a larger sample is needed to confirm the correlations found in Experiments 2, 3 and 4, and to examine the role of possible mediators. An examination of individual differences in auditory-verbal STM capacity and temporal variability is presented in Chapter 4.

3.6 Chapter Summary

The first four experiments using the novel rehearsal-probe task demonstrated its promise as a method for measuring rehearsal timing. Participants were able to perform the task as instructed, and the qualitative patterns of responses were consistent with expectations. The mean CSD variable provided a quantification of temporal precision in STM, and the measure was particularly useful in experiments where the memory load conditions were well matched for probe timing and number of items. When task conditions were not matched for these factors, the results were difficult to interpret due to either a bias in the mean CSD measure or in the task design. Results from two experiments showed that temporal precision for auditory-verbal sequences was significantly responsive to memory load manipulations, and in three experiments, temporal variability for isochronous, random digit sequences was correlated with auditory-verbal ISR performance. These results are consistent with a dynamic shared resource account of auditory timing information in STM.

CHAPTER 4

INDIVIDUAL DIFFERENCES IN SHORT-TERM MEMORY FOR SERIAL ORDER AND TEMPORAL PRECISION

The results of three experiments presented in Chapter 3 revealed a correlation between individual differences in ISR performance and temporal variability during the rehearsal-probe task. The aims of this chapter are to 1) investigate this correlation with a dedicated individual differences experiment, and 2) estimate the correlation coefficient with a meta-analysis of the correlational results presented in this thesis. With regard to the first aim, the goal of the individual differences experiment is to further elucidate the relationship between timing variability and STM for serial order by using a larger sample size and measuring possible mediating factors. After discussion of the individual differences experiment, this chapter will present a meta-analysis of the correlation results across experiments from this thesis. The meta-analysis was conducted in order to estimate the true effect size of the relationship between individual differences in timing variability in STM for subspan sequences and auditory-verbal ISR capacity.

4.1 Experiment 5

Generally speaking, research on auditory timing precision and STM for serial order has been sparse. One study has shown that individual differences in the ability to accurately perceive and maintain auditory timing may relate to serial order STM for auditory-verbal sequences (Saito, 2001). Saito presented 57 participants with irregularly-timed sequences of between 4 and 7 tones, where sequences were either 1960 or 2520 ms in total duration. After a 5-second unfilled delay, participants tapped out the pattern with the index finger of their dominant hand. The study reports a highly significant positive correlation between digit span scores and two separate measures of rhythmic reproduction accuracy; the absolute precision of the reproduced sequence and the relative accuracy, defined as the ratio of each interval produced to the whole pattern duration. In addition, Saito reported that the correlation between digit span and the two measures of rhythm STM remained significant after partialling out reading speed, and there was a significant correlation between auditory digit span and reading speed that was not mediated by the rhythmic precision scores. The double-dissociation of these correlational results suggest that there may be a timing mechanism involved in STM for serial order, and that this mechanism is unrelated to the previously-established contribution from rehearsal speed in the articulatory loop to STM span.

In addition to correlations with digit span performance, precision in reproduction of auditory rhythms has also been shown to relate to intelligence (Holm, Ullén, & Madison, 2011; Lorås, Stensdotter, Ohberg, & Sigmundsson, 2013; Madison, Forsman, Blom, Karabanov, & Ullén, 2009; Ullén, Forsman, Blom, Karabanov, & Madison, 2008; Ullén et al., 2012). Because digit span also correlates with intelligence measures (Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002), it remains unclear whether the relationship between timing precision and digit span, such as that reported in Experiments 2-4 (see Chapter 3) and by Saito (2001), is mediated by general intelligence. In the current study, a measure of non-verbal intelligence is included in addition to auditory digit span with the hypothesis that rehearsal timing variability will account for more variance in digit span than will the intelligence measure. Furthermore, a measure of visual STM with a minimal serial ordering component was included in order to account for modality-general STM factors that may be common to both the rehearsal-probe and digit span tasks.

There are two key reasons why testing the relationship between digit span and variability during the rehearsal-probe task will be informative beyond results obtained by studies using overt speech or motor-timing tasks to measure temporal precision. First, the rehearsal timing task does not involve any overt motor execution to indicate timing. Individual differences in motor-timing precision contribute an independent source of temporal variability to overall measures of variability in finger-tapping tasks (Wing & Kristofferson, 1973a; Wing, 1977), and temporal precision in these tasks changes as a function of movement trajectory (Balasubramaniam, Wing, & Daffertshofer, 2004). As a method of eliminating or at least reducing these sources of noise, the rehearsal-probe task was developed with the aim of indexing internal representations of the timing of events separately from one's ability to execute the remembered timing pattern. A second, related point is that the rehearsal-probe task measures one's temporal precision for speech sounds in the absence of auditory feedback. This is in contrast to motor-tapping and overt rehearsal methods for measuring temporal accuracy for speech and rhythms, as these methods provide multi-sensory feedback that is used to monitor and correct behaviour.

The present experiment is expected to replicate the correlation between digit span and temporal variability found in earlier experiments (see Chapter 3), and to clarify the results of Saito (2001) by using a test of timing precision during subvocal rehearsal of auditory-verbal sequences. It is predicted that temporal precision in auditory-verbal STM, as indexed by response variability during the rehearsal-probe task, will be negatively correlated with auditory STM capacity for serial order, as measured using the digit span task. Further, it is hypothesized that the relationship between rehearsal timing precision and STM span will remain statistically significant after controlling for individual differences in non-verbal intelligence and visual STM. Both the non-verbal intelligence and visual STM control tasks should also tap into non-specific cognitive factors such as modality-general attention and task motivation, therefore any variance shared with both of the control tasks will indicate a mediating effect of these factors.

4.1.1 Methods

4.1.1.1 Participants

Participants were 40 undergraduate and postgraduate volunteers (29 female) recruited from the University of York with a mean age of 19.8 years (range 18 to 24 years). The eligibility criteria, ethics approval, compensation and informed consent details were the same as those reported in previous experiments (see Chapter 3, Experiment 1). In addition, the participants must have recalled at least one digit span sequence at a list length of 6 digits in order to continue with the rehearsal-probe task.

4.1.1.2 Design

The experiment took a within-subjects design; all participants completed the same four tasks. The calculation of the dependent variables for the rehearsal-probe task (mean CSD) and the digit span task are described in detail in Chapter 2.

In addition to the rehearsal-probe and digit span tasks, there were two other tasks administered that have not been used in previous experiments. First, non-verbal IQ scores were measured using the Matrix Reasoning subset of the Weschler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) IQ battery. The Matrix Reasoning task consists of a set of visuospatial pattern completion matrices ordered in increasing difficulty, and the dependent variable for this task was the average of the three highest matrix numbers that the participant responded to correctly. Second, the Visual Patterns Test (VPT; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999) was used to estimate individual differences in non-verbal STM without a serial ordering component. The dependent variable for the VPT was the average of the last three largest grid sizes that were correctly recalled.

4.1.1.3 Materials

The audio stimuli used in the rehearsal and digit span tasks were the digits zero to nine described in Chapter 2.

The stimuli in the VPT[7] were images of test grids and response grids presented on a 1280 x 768 LCD monitor display. Each grid was a set of black-outlined squares on a gray background. In each test grid, half of the squares were filled in blue and the other half were unfilled (same as background colour). The blue/gray colours and low contrast levels were chosen in order to minimize the likelihood of an afterimage of the test grid on the blank screen display. For each test grid, there was a corresponding response grid consisting of unfilled squares. There were 14 grid sizes, ranging from 4 to 30 squares and increasing in steps of two squares. The predetermined patterns of filled squares were unpredictable and differed across the three trials for each grid size.

The Matrix Reasoning task is a subtest of the WASI used to estimate non-verbal intelligence. The task is made up of 35 sets of tiled abstract images (matrices), and the sets are ordered from 1-35 in increasing difficulty. Each matrix presents a pattern or relationship among the tiled abstract images, with a missing piece indicated by a "?" in a single image tile location. Below each image matrix is a set of five images, one of which would correctly complete the matrix if inserted in the location of the missing piece. The test includes two sample matrices used to demonstrate the task, and these are easier than the first test stimulus.

4.1.1.4 Procedure

The general procedures for administering and scoring the digit span and rehearsal-probe tasks are described in Chapter 2. The trial structure and probe times used in the rehearsal-probe task here were the identical to those used in the Experiment 4 isochronous condition (see Chapter 3). In each trial, the participant heard an isochronous, random 6-digit sequence presented twice. Figure 31 shows a representation of a single trial in the rehearsal-probe task. There was a two-second time limit for responses after the probe, and the trials were separated by a two-second inter-trial interval that began after the a response was entered or after the time limit had passed.

---

[7] I would like to thank Dr Stephen Darling for providing a computerized version of this task.

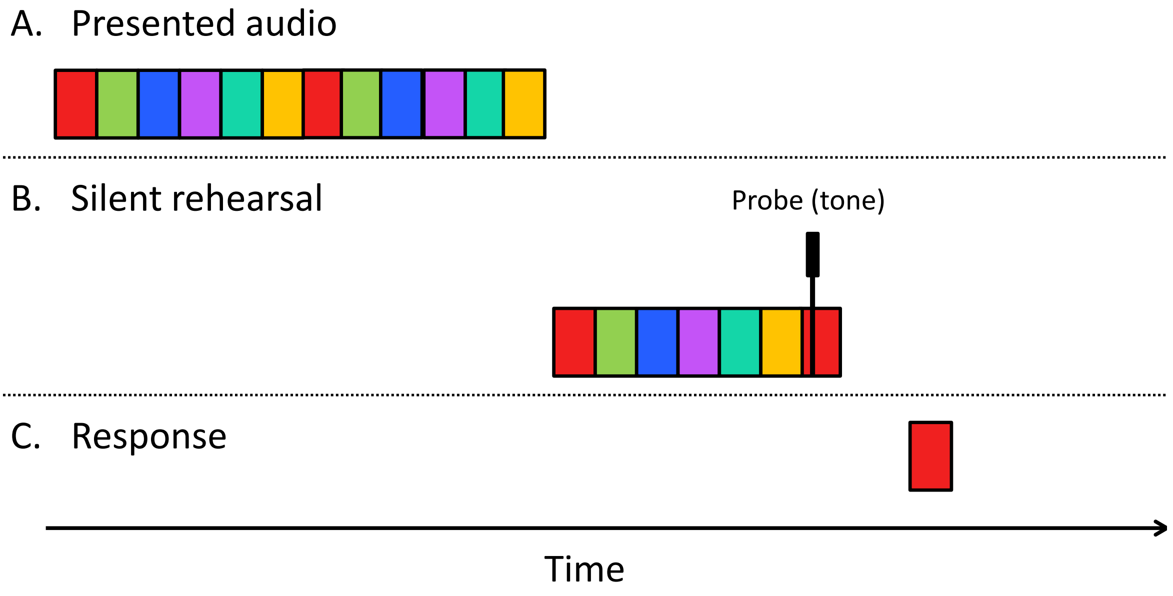

Figure 31. Representation of the Experiment 5 rehearsal-probe task design. The six colours represent the serial positions of the digits. Row A represents the presented audio sequence, where participants heard two repetitions of a random 6-digit list. Participants silently rehearsed the list exactly as it was presented until hearing a probe (B). Upon hearing the probe, participants responded with the item that s/he was rehearsing when the probe was presented (C).

The 14 probe times were selected such that, if the sequence were perfectly replicated during the rehearsal delay, the probes would occur during the beginning or end of each of the six items during the second list cycle. Also, two probes were added to extend the probe period beyond the second rehearsal cycle to reduce the predictability of probe timing. These additional probes were placed at the end of the first rehearsal cycle and the beginning of the third rehearsal cycle. For each of the 400 ms digits, the first probe was set to occur 75 ms after the digit onset, and the second probe was set to occur 325 ms after the digit onset. The probes were evenly spaced in time (occurring every 250 ms). The probe times, measured from the final presented digit stimulus offset time to the probe onset time, are shown along the x-axis in Figure 33.

Digit span was administered prior to the rehearsal-probe task. After receiving written and verbal instructions, observing the rehearsal-probe task demonstration and completing three practice trials, participants completed 126 rehearsal-probe trials (14 probe times x 9 trials per probe time). There were three blocks of 42 trials, with three trials per probe time selected at random within each block. Participants were given optional self-paced breaks between the three blocks of trials.

Following the rehearsal-probe task, the WASI Matrix Reasoning task was administered as an estimate of non-verbal intelligence. Participants were asked to indicate which of five numbered images below the image matrix would complete the pattern set up in the matrix if it were inserted in the location of the missing piece. The verbal responses were recorded by the experimenter. As per the WASI administration instructions for adult participants, each participant began the test at the 7th matrix in the set and continued to respond to each matrix in succession until s/he completed the most difficult matrix number (35), or until s/he responded with 4 consecutive incorrect answers or 4 incorrect out of 5 consecutive answers. Before the task, participants were asked to respond to two sample matrices to illustrate the task and ensure that participants understood the instructions. The dependent measure was the average of the matrix numbers of the last three correct responses.

The final task of the session was the VPT, which was used to estimate non-verbal STM capacity with a minimal serial order component. The stimuli were presented via a custom Python (version 2.7) program on a Windows 7 computer. Participants were seated at a comfortable distance from the screen. Each trial in the VPT began with an “Are you ready?” prompt, which required a keyboard response, followed by a 0.5 second (s) fixation cross. For the first grid size (4 squares), the test grid was shown for 1 second, then a blank screen was shown for 1 second. An empty ‘response grid’ appeared after the 1 second blank screen delay, and participants recalled the pattern from the test grid by using the computer mouse to click on the corresponding squares in the response grid which were originally coloured blue. After the first grid size, the presentation duration of the test grid increased by 0.5 seconds with each increase in grid size up to a maximum of 7.5 seconds for the 30-square grid (see Figure 32 for a sample stimuli sequence). The presentation durations of the fixation cross and blank screen displays stayed constant throughout. Participants completed three trials per grid size, starting with the smallest size (4 squares) and continuing to increase in difficulty until all trials for the largest size (30 squares) were completed or until the participant responded incorrectly to all three trials of a given grid size. The dependent measure was the average grid size of the last three correct responses.

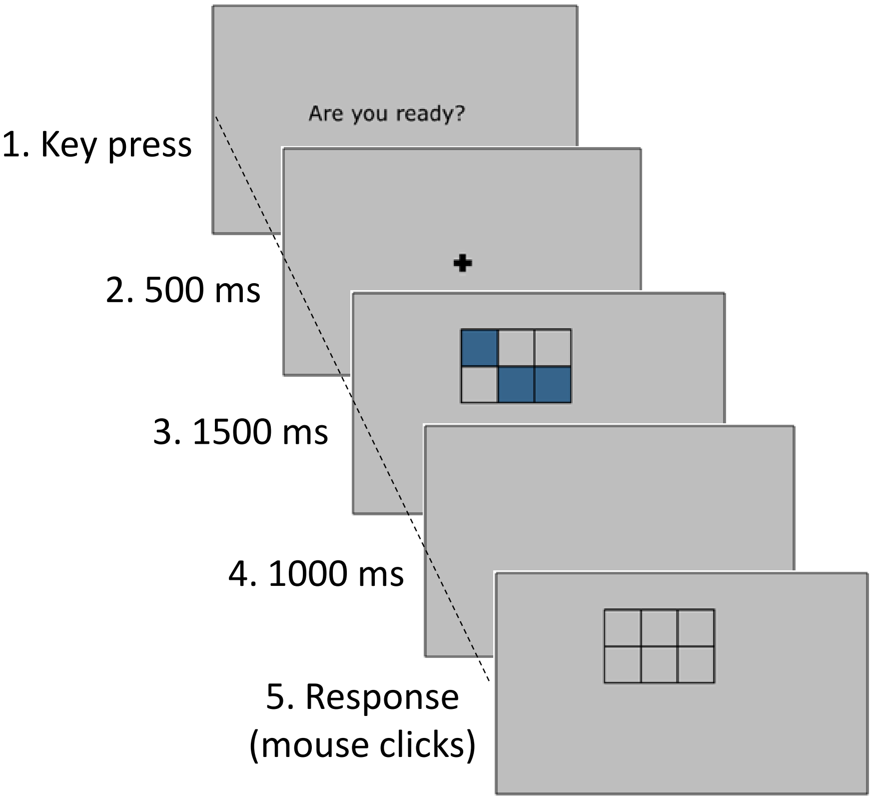

Figure 32. Sample trial sequence at the second grid size level (6 squares) from the Visual Patterns Test. Note that the duration of the test grid presentation (3) increases with increasing grid size.

4.1.2 Results

Responses to probes were categorised according to the serial position of the item (digit) within the presented sequence for that trial. Digit responses that were not part of the presented 6-digit sequence on that trial were considered invalid responses. There were 5040 total trials (126 trials per participant x 40 participants), with very low percentages of no-response (37/5040, 0.73%) and invalid response (49/5040, 0.97%) trials. The invalid responses were removed from the data, leaving the remaining 4954 valid responses (98.29% of total trials) for analysis.

The means, standard deviations minima and maxima of raw scores for all measures are presented in Table 11. CSDs were calculated according to the procedure outlined in Chapter 2. The dependent variable calculation for the WASI Matrix Reasoning and VPT tasks are provided in the Procedure section. As shown in Table 11, the CSD data were centred within the lower-middle of the range of possible CSD scores. There was no evidence of floor or ceiling effects based on the minima and maxima of these scores, although histograms of the data showed positive skewness for some of the CSD distributions. This skew was driven by 1-3 cases (2.5-7.5% of total) with *z*-scores greater than 2.58 (i.e. above the 99th percentile) on the right tails of the distributions, but with all other *z*-scores within +/- 1.96. Kolmogrov-Smirnov tests of normality confirmed significant deviations from normality for the mean CSD measure ($p < .01$), and for the CSDs of serial positions 1 ($p < .05$), 4 ($p < .01$) and 6 ($p < .05$). Digit span, VPT and WASI MR data were normally distributed (all *p*s > .05). With the exception of one case in the serial position 1 and 5 data, there were no cases beyond +/- 3 standard deviations of the means. Therefore data from all subjects was included in subsequent analyses.

Table 11. Means, standard deviations, minimums and maximums of raw scores on all measures. The lowest and highest possible scores are listed in the 'Range possible' column.

Measure	Mean	SD	Min	Max	Range possible
Mean CSD	0.67	0.19	0.41	1.22	0.26-1.41
SP 1 CSD	0.61	0.20	0.31	1.37	0.26-1.41
SP 2 CSD	0.72	0.22	0.36	1.30	0.26-1.41
SP 3 CSD	0.73	0.23	0.37	1.26	0.26-1.41
SP 4 CSD	0.73	0.21	0.36	1.21	0.26-1.41
SP 5 CSD	0.66	0.20	0.35	1.27	0.26-1.41
SP 6 CSD	0.58	0.21	0.32	1.18	0.26-1.41
Digit span average	6.93	1.08	5.00	9.33	5.00-12.00†
WASI MR average	29.73	2.71	24.33	33.33	0.00-34.00
VPT average	10.58	1.72	7.67	13.67	0.00-14.00

Note: SP = serial position; WASI MR = Weschler Abbreviated Scale of Intelligence Matrix Reasoning; VPT = Visual Patterns Test; average = average of last three correct trials.
† Digit span lowest possible score is based on the eligibility criteria of at least one correct response to a 6-digit trial. One correct response to list lengths of 4, 5 and 6 (and no correct responses for 7-digit trials) in the digit span task results in a digit span average score of 5.

Figure 33 shows the mean response proportions according to serial position across probe times. The averaged response proportions over probe times exhibited the characteristic periodicity seen in previous experiments, and the modal probe times for each position were roughly in time or lagged slightly behind the expected rehearsal time. The peaks of each serial position response probability function occurred in the correct serial order, and the distributions were unimodal.

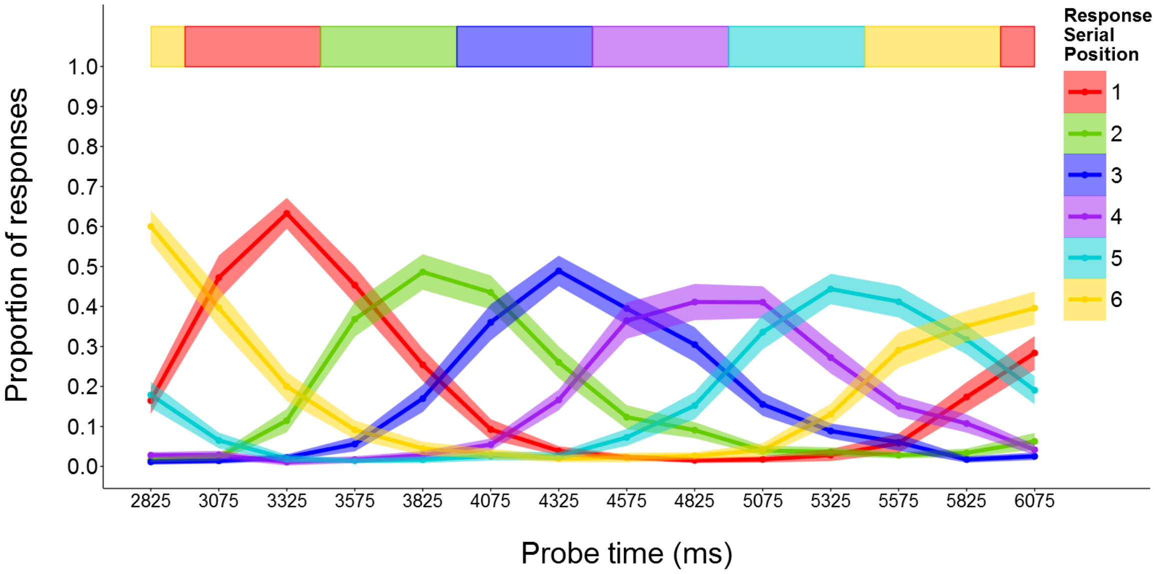

Figure 33. Mean proportions of responses across all participants for each serial position 1-6 at each probe time. Colours indicate the serial position of the response. The widths of the ribbons indicate the standard error of the mean at each probe time.

On average, participants' responses with items in the middle of the list (serial positions 2, 3 and 4) were more variable than items at the start or end of the list (positions 1, 5 and 6; see SP CSDs in Table 11). A one-way repeated-measures ANOVA showed that the effect of serial position on CSD was significant, $F(5,195) = 14.68$, $p < .001$. Pairwise comparisons revealed that there were significant within-participant differences between CSDs of multiple serial positions after using a Bonferroni correction for multiple comparisons (see Figure 34). The significant differences were between serial position 1 and positions 2-4, and between serial position 6 and positions 2-5.

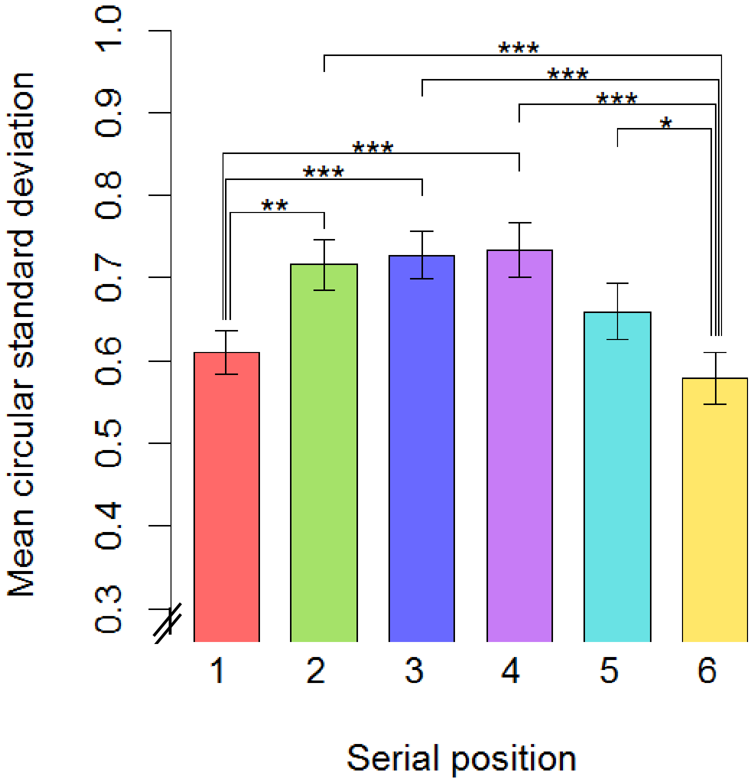

Figure 34. Mean circular standard deviations by serial position. Error bars show 95% confidence intervals adjusted for within-participant error using the method proposed by Cousineau (2005). The y-axis minimum reflects the lowest possible CSD value attainable in this experiment. Asterisks mark the significant differences between positions after Bonferroni correction for multiple comparisons; * *p* < .05, ** *p* < .01, *** *p* < .001.

Relationships between measures were examined using two-tailed Pearson's *r* correlations (see Table 12). There was a significant negative correlation between mean CSD and the digit span scores (see Figure 35). There was a significant positive correlation between the VPT and digit span scores, though this relationship was not as strong as that between digit span and mean CSD. Finally, mean CSD was negatively correlated with WASI MR, and the VPT was positively correlated with WASI MR, however neither of the *p*-values corresponding to these correlations reached the .05 significance threshold.

Table 12. Pearson's *r* correlations between measures.

Measure	Mean CSD	Digit span	VPT
Mean CSD	1.00		
Digit span	-.57***	1.00	
VPT	-.13	.32*	1.00
WASI MR	-.25	.05	.31†

Note: **p*<.05, ***p*<.01, ****p*<.001, †*p*=.056; CSD = circular standard deviation; WASI MR = Weschler Abbreviated Scale of Intelligence, Matrix Reasoning task; VPT = Visual Patterns Test.

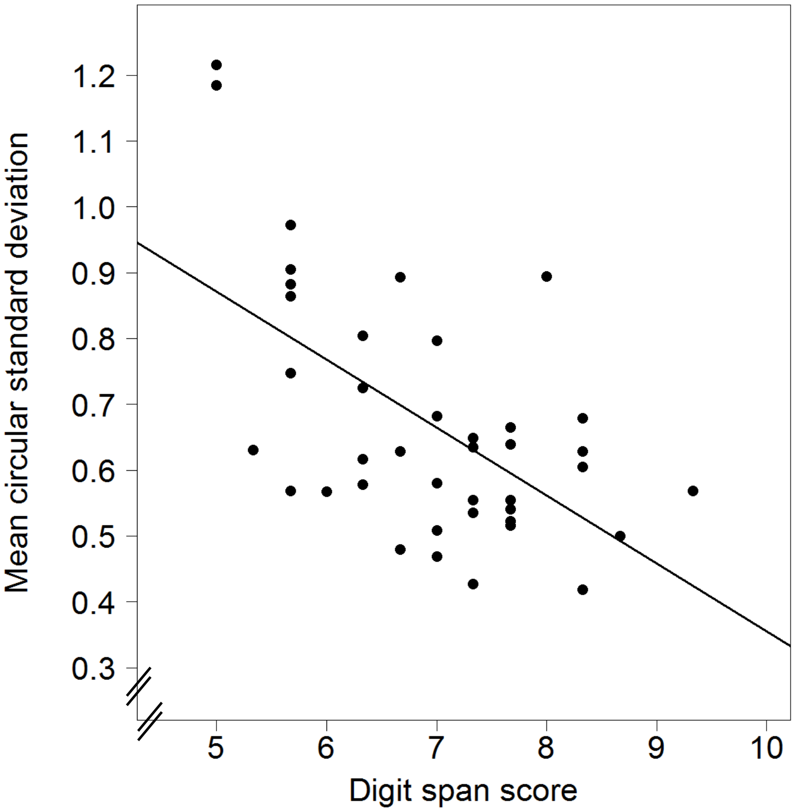

Figure 35. Scatterplot with simple linear regression line for mean circular standard deviations and digit span scores. The y-axis minimum reflects the lowest possible mean CSD score in this experiment.

A multiple linear regression analysis was conducted to determine the amount of variance in digit span scores uniquely accounted for by the mean CSD. The data were tested and found to have met the assumptions of independence of errors (Durbin-Watson = 1.82), homoscedasticity and no multicollinearity (maximum VIF = 1.16, minimum tolerance = 0.86). The results of the linear regression analysis are shown in Table 13. The regression method used was forced entry, with WASI MR entered first, VPT entered second and mean CSD entered third. The rationale for this method was that general intelligence and common STM factors should be controlled for in order to make a more conservative estimate of the amount of variance in STM capacity accounted for by individual differences in mean CSD. There were no cases with undue influence over the model parameters (all Cook's Distance values < 1).

Table 13. Results of multiple linear regression analysis of factors predicting the variance in digit span scores

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
WASI MR	.02	.06	.06	-.02	.06	-.04	-.07	.05	-.18
VPT				.20	.10	$.33^{\dagger}$	.18	.08	.30*
Mean CSD							-3.37	.74	-.59***
R		.06			.32			.66	
R^2		<.01			.10			.43	
F for ΔR^2		.14			$4.04^{\dagger}$			20.84***	

Note: **p*<.05, ****p*<.001, $^{\dagger}p$=.052; WASI MR = Weschler Abbreviated Scale of Intelligence, Matrix Reasoning task; VPT = Visual Patterns Test; CSD = circular standard deviation

The results of Model 3 in Table 13 show that individual differences in intelligence and visuospatial STM performance did not moderate the relationship between mean CSD and digit span. Visuospatial STM was a marginally significant predictor of digit span scores after accounting for intelligence (Model 2), but the shared variance between visuospatial STM and digit span appears to be separate from that shared between mean CSD and digit span. That these sources of variance are separable is supported by the fact that the two variables were both statistically significant predictors of digit span in Model 3, and that the corresponding effect sizes for the relationships with each predictor were not markedly reduced when the two predictors were considered in conjunction with one another.

4.1.3 Discussion

The aim of the present study was to investigate whether the ability to recall a sequence of spoken items in correct serial order relates to the temporal precision with which spoken material can be encoded and maintained in STM. The results of this experiment clearly supported the hypothesis that auditory-verbal STM for serial order would be negatively correlated with timing variability, and that this relationship would not be moderated by individual differences in visuospatial STM or intelligence. The inverse relationship between timing variability and ISR performance found in Experiments 2, 3 and 4 (see Chapter 3) was replicated here using a larger sample size and with the inclusion of possible mediating variables. Multiple linear regression analysis showed that timing variability during the rehearsal-probe task accounted for considerable variance in participants' auditory STM spans after controlling for non-verbal IQ and visuospatial STM. Non-verbal IQ did not account for a significant amount of variance in digit span, while visuospatial STM accounted for a low-to-moderate amount of digit span variance.

The correlation between timing precision and auditory-verbal serial order STM is consistent with the results reported by Saito (2001). The correlations between timing variability and digit span presented here extend Saito's finding by showing that a similar correlation exists with verbal material and without the overt production of the whole sequence timing. Also, this experiment demonstrates that the correlation between timing precision and ISR performance is not due the mediating effects of IQ, non-serial STM, or general task-relevant cognitive factors such as modality-general attention and motivation.

One interesting finding from Saito (2001) was that visual digit span was more strongly associated with rhythmic precision than was the auditory version of the span task. Based on this and the other patterns of correlations, Saito argues that the relationship between STM for rhythms and verbal serial order STM might be mediated by a timing mechanism that is separate from the phonological store and articulatory loop. According to this hypothesis, the increased correlation between visual versus auditory digit span occurs because a timing mechanism must be more strongly and actively involved to retain sequential visual information compared to auditory stimuli, which automatically induces the activation of timing mechanisms.

Saito makes two other arguments in support of this hypothesis. First, the timing mechanism is assumed to be necessary for the production of regularly-timed speech, for example during articulatory suppression, but independent from the articulatory loop itself. This interpretation explains why the correlation between digit span and temporal precision was not mediated by articulatory programming efficiency or subvocal rehearsal speed, as indexed by reading speed. While the correlation between speech rate and STM span has been established by a number of previous studies (e.g. Baddeley et al., 1975; Hitch, Halliday, & Littler, 1989), the notion that this relationship is limited is also consistent with Tehan and Lalor's (2000) finding that that variables intended to tap rehearsal speed (e.g. reading and counting rates) account for less digit span variance compared to other factors, such as access to lexical memory. A second, related point is that rhythm reproduction is disrupted by articulatory suppression (Saito & Ishio, 1998), which is consistent with the hypothesis that a timing mechanism is involved in regularly-timed speech. In future experiments it would be interesting to see if the pattern of results by Saito can be replicated using the rehearsal-probe paradigm with both auditory and visual digit span tasks, and with speech rate measures.

The lack of a significant correlation between timing variability and non-verbal IQ in this experiment is inconsistent with the results of Madison and colleagues (Holm et al., 2011; Madison et al., 2009; Ullén et al., 2012). However, the correlation coefficient observed here (r = -.25) is not far from the weighted mean effect size of the relationships between intelligence and temporal variability across three studies (r = -.30, Ullén et al., 2012), so there may have been insufficient power in the present experiment to detect this relationship. Another possible explanation for the difference between these findings is that Madison, Ullén and colleagues used the Raven SPM Plus to measure intelligence, which is a longer (60-item) version of the WASI Matrix Reasoning task used in this experiment and therefore may be a more sensitive measure. Nonetheless, the central question for this experiment regarding the relationship between timing precision and digit span was clearly not mediated by the shared variance between IQ and timing, as the variance in digit span predicted by timing variability did not change at all after accounting for individual differences in non-verbal IQ.

Finally, it is interesting to note that there was an overall effect of serial position on temporal variability reminiscent of the serial position gradients typically seen in ISR tasks for supraspan lists. Serial position gradients are the patterns of correct ISR, where each item response is scored as correct only if the item is recalled in the original serial position. The characteristic bowed or U-shaped serial position gradients for supraspan isochronous lists reflect the greater tendency for serial order errors to occur in mid-list positions in comparison to early-list positions (primacy effects) and positions near the end of the list (recency effects). In Experiment 5, the responses for items in the first and last serial positions (1 and 6) were significantly less variable over probe times than responses for items in mid-list positions. It is notable that this pattern of temporal variability over serial positions has arisen for subspan lists, where ISR performance for isochronous 6-digit lists should be at or near ceiling for the participants. This finding has implications for the interpretation of the mechanism(s) underlying response variability in the rehearsal-probe task and STM for serial order, which will be explored in more detail in the General Discussion section.

4.2 Meta-Analysis of Correlations between ISR Capacity and Timing Variability

Experiment 5 firmly established the existence of a linear association between digit span and temporal precision in a subspan rehearsal-probe task. However, to arrive at a more robust estimate of the size of this effect, data from the same or similar experiments and populations can be incorporated. Therefore a meta-analysis of the correlations between mean CSD and digit span was conducted across the behavioural experiments presented thus far. Included in the meta-analysis were the correlations from Experiments 2-4 (Chapter 3) and Experiment 5 (present chapter). Experiment 1 could not be included because digit span data was not collected in that experiment. Among the remaining behavioural experiments (2-5), there were two criteria for inclusion in the meta-analysis that were chosen to ensure that the experiments were sufficiently similar in terms of the underlying theoretical effect size being estimated. First, the participants must have been sampled from the same population, namely, adults without any developmental language or learning disorders. All experiments met this criterion. Second, the meta-analysis was restricted to the results pertaining to isochronous sequence presentation. This criterion was based on the conclusions from Experiment 4, namely that the temporal-grouping of rehearsal sequences may have introduced an influential confounding factor. Thus the participant group that rehearsed temporally-grouped sequences was excluded from the analysis, but the isochronous participant group was included. The correlations included in the analysis were independent; there was one correlation from each of the five experiments included in the meta-analysis, where the mean over groups/conditions within each experiment were used as a summary of the effect found in that experiment. The one exception to the use of group/condition average was with Experiment 4, because only one of the two groups was included.

A random effects model was used to estimate the true correlation, as well as a margin of error for that value, based on the correlations obtained in the five relevant experiments. The random effects model is similar to the fixed effect model in that a mean population effect size (the population correlation coefficient ρ) is estimated based on the sampled means and their corresponding sampling variances. However, unlike the fixed effect model, the random effects model does not assume that all of the studies are estimating the same underlying population parameter. Rather, the random effects model includes an additional assumption that the studies are samples of a normally-distributed population of effect sizes. If the variability in effect sizes across the experiments can be adequately explained by sampling variation, then the experiments are assumed to be samples of the same underlying population (i.e. they are homogenous). On the other hand, if sampling variability cannot fully account for the variation in the effect sizes across the experiments, then the remaining variability likely reflects the fact that the studies are estimates of effect sizes from underlying populations that differ to some extent (i.e. they are heterogeneous). Cumming (2012) and others argue that random effects models are always preferable over fixed effects models because the assumption of heterogeneity is valid in most cases, and when the experiments are homogenous, the result of the random effects model is the same as that of a fixed effect model.

The model was run using the statistical programming language R (R Core Team, 2014) and the R package 'metafor' (Viechtbauer, 2010). The correlations were converted to *z*-values using Fisher's *r*-to-*z* transformation, which allowed the computation of the sampling variance for each correlation using the standard formula. The random effects model was run on the *r*-to-*z* transformed data using the restricted maximum-likelihood (REML) estimator of population heterogeneity (Viechtbauer, 2005). The transformed means and variances (95% confidence intervals) were converted back to correlation coefficients (*r*) using the inverse of Fisher's *r*-to-*z* transformation.

The result of the random effects model is shown in Table 14. The model was run on the correlations between digit span and mean CSD scores. Where the experiment had a within-participant design (Experiments 2 and 3), individual mean CSD scores were the average over conditions. As shown in the table, the meta-analysis revealed a significant negative estimate of the population correlation coefficient between digit span scores and mean CSD scores across the five experiments. Further, the results of this model showed that there was no significant heterogeneity among the correlations, suggesting that the differences among the correlations obtained in the experiments can be accounted for by sampling variance.

Table 14. Results of meta-analyses of correlation coefficients between digit span and mean CSD scores from Experiments 2-5, with isochronous presentation only.

Modelled data	k	$\hat{\mu}_\rho$	C.I. Upper	C.I. Lower	p	$\hat{\tau}^2$	Q	p_Q
Isochronous sequences, Experiments 2-5	4	-.56	-.69	-.40	<.001	0.00	0.18	.981

Note: k = number of studies included; $\hat{\mu}_\rho$ = estimated average correlation in the population distribution; C.I. = 95% confidence interval for $\hat{\mu}_\rho$; p = p-value for the null hypothesis (that $\hat{\mu}_\rho$ = 0); $\hat{\tau}^2$ = estimated heterogeneity in the sample correlations; Q = Q-test for heterogeneity with k – 1 degrees of freedom; p_Q = p-value for Q.

Figure 36 shows a forest plot of the correlation coefficients and confidence intervals from each experiment, including those for conditions in within-participant experiments. Fisher's *r*-

to-*z* transformation for calculation of 95% confidence intervals results in asymmetrical upper and lower bounds. This asymmetry is particularly evident when the correlation coefficient is far from zero, and when N is low (i.e. sampling variance is high). Below the experimental correlation coefficients is the random effects model estimate of the population correlation coefficients and margin of error for the model. That the model estimates of heterogeneity (i.e. Q in Table 14) among the experimental correlations are very low is evident from the overlapping confidence intervals surrounding the correlation values across experiments.

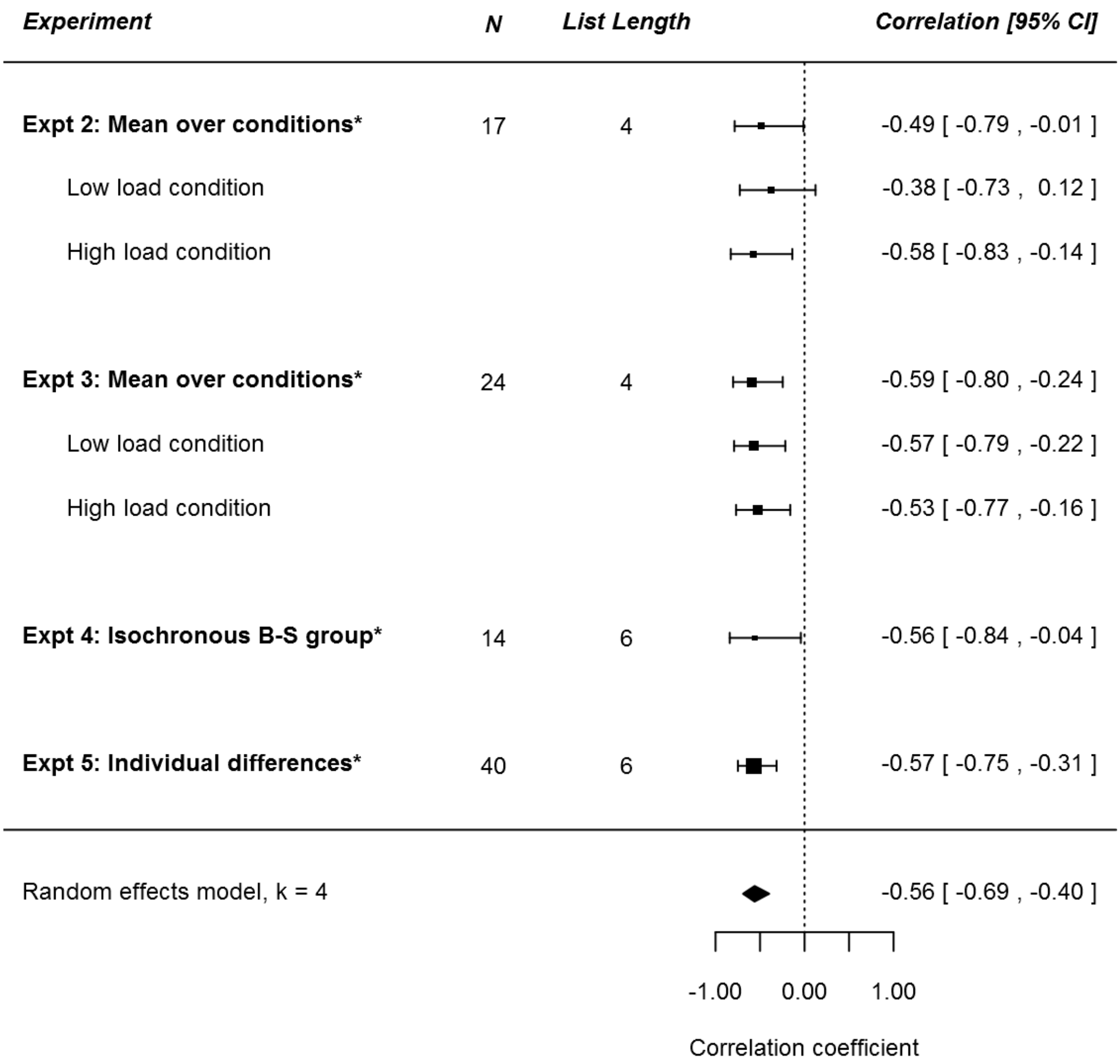

Figure 36. Forest plot of correlations between digit span and mean CSD in Experiments 2-5. 95% confidence intervals were calculated using Fisher's *r*-to-*z* transformation. Bottom row shows the random effects model estimate of the true correlation based on the subgroup of experimental results (adult participants and isochronous timing). The model includes four independent correlations; these are marked with a * in the leftmost column.

As can be seen in the forest plot (Figure 36), the correlation coefficients across and within experiments (i.e. conditions/groups) were generally similar in direction and magnitude, but the confidence intervals for these coefficients varied more widely depending on the sample size. The lower bounds of correlation coefficient 95% confidence intervals from experiments with smaller samples (i.e. Experiments 2 and 4) tended to approach or cross zero, whereas those with larger samples (Experiments 3 and 5) were more clearly negative.

The model produced an estimate of the true effect size of the correlation coefficient that is negative and of a moderate magnitude, and with a lower bounded confidence interval that is comfortably below zero (estimated r = -.56, 95% C.I. [-.40, -.69]). It is important to note that this population correlation estimate only applies to the type of data reflected in the samples (experiments) included in the model. That is, the model estimate of the correlation between auditory digit span and temporal precision in the rehearsal-probe task applies to young adult participants rehearsing isochronous subspan (4-6 digit) sequences. The results of this meta-analysis cannot be extended to rehearsal of sequences with more complex timing patterns (e.g. the temporally-grouped participant group in Experiment 4, Chapter 3), or to other developmental stages where ISR capacity and rehearsal use are known to differ from the young adult participants in the experiments included in the meta-analysis.

4.3 General Discussion

The results of Experiment 5 and the meta-analysis replicate the previous results observed in Experiments 2-4 (Chapter 3), which showed a negative correlation between timing variability in the rehearsal-probe task and digit span. In Experiment 5, temporal precision during rehearsal of 6-digit sequences accounted for a significant portion of the variance in ISR for auditory-verbal digit sequences, and the relationship between these measures was independent from the correlations with both non-verbal IQ and visuospatial STM. The meta-analysis revealed that the correlation results across Experiments 2-5 were homogenous and therefore well accounted for with a random effects model. The model produced a significant negative estimate of the true population correlation coefficient between auditory digit span and temporal variability for isochronous subspan digit auditory sequences in young adults. Taken together, these results provide convincing evidence of an inverse relationship between auditory STM capacity for serial order and the variability in STM for temporal representations of auditory-verbal sequences. In addition, the replicability of the effect was demonstrated across a range of different task variations, suggesting that these do not greatly affect the association.

While the correlation between temporal precision for isochronous subspan auditory-verbal sequences and auditory-verbal ISR capacity does not provide sufficient information to draw conclusions about causality, it is possible at this point to reflect on the plausibility of hypothetical causal relationships. It appears that temporal precision is a function of both task context and individual differences in ISR capacity, so any explanation must be able to account for relationships among all three factors. There are at least two conceptual frameworks that could provide a plausible account for the patterns of rehearsal-probe results presented in this thesis. One possibility is that timing imprecision is a precursor to order errors. Within this framework, an increase in the individuals' variability of temporal representations causes a decrease in ISR performance due to the increased probability of order errors. If timing precision limits ISR performancethen there would also have to be a load-dependent increase in timing variability to explain the within-participant effects of memory load on temporal precision, as well as the persistence of the correlation effect within load conditions. Thus from this perspective, individual differences in temporal precision are a causal factor in determining auditory-verbal STM for serial order, where temporal precision is directly influenced by task memory load (as in Experiment 3) as well as other contextual factors such as LTM for item-order associations (Experiment 2) and rhythmic structure (Experiment 4).

The notion that temporal variability is a precursor to serial order errors is supported by the effect of the serial position of the responses on mean CSDs in Experiment 5. In this experiment there was a significant main effect of serial position on CSD, which was driven by lower CSD values for responses corresponding to items at the beginning and end of the list compared to the items in mid-list positions. The patterns of temporal precision resembling serial position gradients observed in subspan lists is evidence for the view that temporal variability is a precursor to serial order errors because ISR performance would not be likely to show serial position effects for subspan sequences (due to performance at or near ceiling). That serial position effects were present in experiments where the number of digits was closer to a typical adult digit span (6 digits; Experiments 4 and 5) but the effect was weaker or absent in experiments where the rehearsal lists were shorter (4 digits; Experiments 2 and 3) is further evidence that temporal variability is predictive of serial order errors. This pattern of results shows that patterns of temporal variability across serial positions are not static across all list lengths but instead emerge as the list length approaches the individual's STM span, which is consistent with the correlations observed between digit span and timing variability.

Another possible framework for interpreting causal relationships between timing variability and STM for serial order is that these are both limited by shared resources. Within this framework, temporal precision and serial order STM do not directly affect one another, but instead load on a shared pool of resources, which in turn constrains the capacity of both the timing and serial order maintenance systems. Changes in the memory load of the sequences would affect the shared resources, rather than either serial order coding or STM for temporal structure directly. One potential issue with this explanation is that it may not be parsimonious to assume that timing and serial order mechanisms are completely independent, given that serial order depends on timing to an extent (albeit coarse, relative timing). This framework would predict a trade-off between temporal precision and number of sequence items, but it is not clear how are the shared resources would be redistributed as the memory load changes. For example, if there is an increase in memory load, it is not clear whether either temporal precision or STM for serial order performance could suffer alone as a result of the increased demands on shared resources, or if both deteriorate to a similar degree. A further difference between this framework and that previously discussed is that this one would predict a trade off in both directions; that is, an improvement in temporal precision as a result of decreased load on STM for serial order. The earlier attempt to reduce the load on STM for serial order in Experiment 2 prompted the realisation that the use of long-term knowledge about serial order likely introduces an additional influence on sequence timing. Thus, other methods of minimising the load on serial order STM are needed to test this prediction.

Implicit in both hypothetical frameworks is the notion that the STM process allows the encoding, maintenance and recall of fine-grained temporal information. This assumption is of theoretical importance, given ongoing debates regarding the precise nature of the processes involved in serial order STM. There are many computational models of STM for serial order that aim to explain ISR in terms of underlying processes (reviewed in depth in Chapter 8), some of which posit a temporal coding system while others implement positional coding of the items. The correlational results established in Experiment 5 and the meta-analysis are relevant to the debate over whether or not timing information is encoded and retained in STM for serial order. In particular, the fact that timing precision during rehearsal can vary for subspan sequences is more naturally compatible with the notion that timing is encoded in memory, and it is not clear how these results would arise within a memory system where items were only encoded in terms of their relative positions. The issue of the nature of coding in STM is deferred until Chapter 8, where models of STM for serial order are given a more thorough treatment.

In addition to the two frameworks described, there is at least one alternative explanation for the correlation, which is that individuals with lower digit spans were more likely to make serial order errors during rehearsal as these people were closer to their STM capacities. By this account, the mean CSD measure would capture the frequency of serial order errors, rather than timing imprecision. However, there are both theoretical and empirical reasons to suggest that this is not the case. First, as previously mentioned, there a number of existing computational models of serial order STM, and many of these models converge on the general principle that items are represented in parallel until the items are recalled, at which point they are selected discretely[8]. Therefore it is not obvious that each item is selected discretely during rehearsal, which would be a requirement for serial order errors to occur. Because it is also possible that items are activated in parallel during rehearsal, with no discrete selection process occurring until a response is required to the probe, no assumptions should be made at this point about the possibility of order errors during rehearsal.

Even under the assumption that discrete order errors can occur during rehearsal, there are a few reasons why this is not a likely explanation for the correlations observed across experiments. First, all participants had a maximum digit span greater than or equal to the rehearsal list length, so all participants demonstrated the ability to recall at least one digit list at the rehearsal list length in the correct serial order. The possible influence of serial order errors during rehearsal seems especially unlikely in Experiments 2 and 3, where the lists were only 4 digits long and therefore well within the capacities of all participants. Second, the rehearsal sequences were repeated twice whereas the digit span lists are only presented once. Given the greater opportunity for encoding in the rehearsal-probe task, it is expected that listeners can rehearse sequences at span length in this context without making order errors. In future work, one way to further investigate this issue would be to use computational methods to model the patterns of responses that would be predicted from rehearsal with and without serial order errors, in order to determine whether specific attributes of the response data can be used to determine the likelihood of discrete serial order errors as distinct from continuous temporal imprecision.

In conclusion, experiments using the rehearsal-probe task have provided consistent and convincing evidence for a negative correlation between individual differences in auditory-verbal serial recall capacity and temporal variability during subvocal rehearsal of subspan sequences. The correlations between digit span and temporal variability are notable given that ISR tasks do not demand the precise encoding and maintenance of sequence timing beyond the broad, relative timing of digits necessary to distinguish serial order. This finding is relevant to current debates surrounding models of serial order STM in that it may have implications for the nature of item encoding, and it highlights the need for more concrete descriptions about the mechanisms involved in rehearsal. The individual differences results, when considered in conjunction with the results of manipulations of contextual factors (Chapter 3), offer insight into the possible causal relationships among auditory-verbal STM load, ISR performance, and temporal precision. Lastly, the link between individual variation in the temporal precision of rehearsal and auditory-verbal ISR opens the door to

---

[8] Again, this issue is outlined in more depth in Chapter 8.

investigations of developmental disorders of language development, which are characterised by deficits in both auditory timing and STM for serial order.

4.4 Chapter Summary

Experiment 5 was designed as a dedicated individual differences experiment to investigate the previously observed negative correlations between auditory-verbal ISR performance and timing variability in Experiments 2-4 (Chapter 3). Using a larger sample size and the addition of intelligence and visuospatial control measures, this experiment firmly established the negative association between digit span and mean CSD. The relationship between the two variables was not mediated by individual variance in intelligence or visuospatial STM measures, although the latter variable was independently related to span scores. A meta-analysis was conducted in order to incorporate data from earlier experiments into a more robust estimate of the effect size. A random effects model estimated r at -.56 with 95% C.I. bounds of -.40 and -.69, and the correlations across experiments were well-explained by a single model. The associations between memory load, STM capacity and timing precision are considered in terms of possible causal relationships as well as their implications for models of STM for serial order.

CHAPTER 5

TEMPORAL VARIABILITY IN DEVELOPMENTAL LANGUAGE DISORDERS

In this Chapter I will review the evidence for increased variability in the representation of auditory sequence timing in developmental language disordered groups before reporting two experiments with the rehearsal-probe task. The literature related to the more general role of auditory temporal processing in speech perception and language development was reviewed in Chapter 1. Here, the specific focus will be on the evidence that individuals with developmental language disorders have deficits in perception of auditory timing and rhythm, and on the argument that these deficits play a central role in the disorders. The evidence for links between auditory-verbal sequence timing and STM for serial order will also be explored. The literature and rationale reviewed here and presented in Chapter 1 will form the basis of the hypotheses for two experiments reported in this Chapter, where the rehearsal-probe task was used to measure the precision of auditory-verbal sequence timing in dyslexic adults (Experiment 6) and language impaired children (Experiment 7). It was expected that the dyslexia and SLI groups would show greater variability in their representations of sequence timing in the rehearsal task, and that, as in previous experiments, timing precision would be correlated with STM capacity for serial order.

5.1 Experiment 6

Developmental dyslexia is a diagnosis given to children who show severe difficulties in learning to read despite adequate instruction and an absence of any perceptual or neurological problems that would account for the reading impairment. A wealth of research on dyslexia over the past four decades has reliably shown that dyslexia is characterised by poor phonological awareness and processing, which is thought to be a proximal cause of the reading problems (Hatcher, Snowling, & Griffiths, 2002; Wagner & Torgesen, 1987). Phonemes are the smallest units of speech that affect meaning (e.g. the spoken words ‘at’ and ‘mat’ differ by one phoneme, the /m/ sound). Reading requires that these phonemes be mapped onto letters, which explains the source of reading problems in people who have difficulty identifying phonemes in spoken language and representing them abstractly. This has been the basis of the ‘phonological deficit’ hypothesis of dyslexia, which is the view that dyslexia is caused by poor phonological representations. While widely accepted, the phonological deficit hypothesis has not gone unchallenged. For instance, there is some evidence that individuals with dyslexia have intact phonological representations but impaired access to those representations (Ramus & Szenkovits, 2008; Ramus, 2014). Many other theories of dyslexia do not dispute the presence of phonological representation or access deficits, but attempt to explain these deficits in terms of underlying neurological or perceptual causes.

In addition to phonological processing deficits, another reliable cognitive marker for dyslexia is impaired auditory-verbal ISR capacity. Impaired digit span performance has been reported very consistently across a large number of studies with dyslexic children and adults (Ahissar, Lubin, Putter-Katz, & Banai, 2006; Brady, 1986; Corkin, 1974; Hatcher et al., 2002; Helenius, Uutela, & Hari, 1999; M Laasonen, Service, & Virsu, 2001; Marja Laasonen et al., 2012; Law, Vandermosten, Ghesquiere, & Wouters, 2014; Lehongre, Morillon, Giraud, & Ramus, 2013; Leong & Goswami, 2014b; Oganian & Ahissar, 2012; Pasquini et al., 2007; Rapala & Brady, 1990; Soltész, Szűcs, Leong, White, & Goswami, 2013; Wang & Gathercole, 2013). According to the phonological hypothesis, problems with verbal and phonological STM in people with dyslexia are a direct consequence of impaired phonological representations or access to these representations. However, it is not clear how phonological deficits alone can explain the extent and severity of serial order STM problems observed in people with dyslexia. While it seems reasonable that poorly specified phonological representations should interfere with STM for low frequency words and nonwords, it is less obvious how phonological deficits could explain STM impairments for sequences of highly familiar words, such as digits. Differences in vocabulary and verbal fluency are also unable to account for digit span deficits, because the names of single digits (i.e. “zero”, “one”, “two” etc.) represent a small, closed set of simple, semantically-related words. For these reasons, digit span is often considered a test of STM for serial order and should therefore be one of the least likely verbal STM tasks to be affected by either phonological deficits, or by non-phonological language factors such as vocabulary, awareness of phonotactics and morphology, and speed of lexical access.

In an effort to explain auditory-verbal ISR deficits, some researchers argue that people with dyslexia have specific deficits in STM for serial order, and that serial ordering impairments could play a causal role in reading impairments (Corkin, 1974; Wolff, Cohen, & Drake, 1984). Perhaps one of the first proposals that people with dyslexia have serial-ordering deficits was by Corkin (1974), who referenced Lashley’s (1951) insights into the inherent serial organisation of language. Corkin found that the performance of average readers surpassed that of inferior readers on auditory digit span and visuospatial (Knox Cube, i.e. Corsi block) tasks, and that the differences were amplified when the sequences had to be retained for longer periods of time. More recent evidence for serial order impairments in dyslexia comes from investigations using tasks that aim to emphasize memory for serial order while minimizing the STM load for item information. Using a serial reconstruction task with monosyllabic words, for instance, Martinez Perez and colleagues (2013) found that the dyslexic group performed more poorly overall and the difference was more pronounced for order errors than item errors. Serial reconstruction tasks are meant to maximise STM for order information and minimise the load on item information because the digits in the sequence are provided during the response period. These authors also reported that the dyslexic group made more errors in an order probe task (yes/no response to indicate correct or incorrect order of a two-item sequence probe), but the dyslexic and control groups did not differ in an item probe task (yes/no response to a single nonword item probe, which was either present in the sequence or differed by a single phoneme). Experiments using similar methods to isolate STM for serial order have shown that STM for order is uniquely related to vocabulary development as well as reading abilities and dyslexia in children (Leclercq & Majerus, 2010; Martinez Perez, Majerus, Mahot, et al., 2012; Martinez Perez, Majerus, & Poncelet, 2012). Thus, serial ordering deficits appear to be closely related to language and literacy impairments, and these findings are difficult to explain through phonological representation or access deficits alone.

As discussed in Chapter 1, there is also a growing body of literature showing that entrainment to auditory rhythms on specific timescales may be critical for speech perception, and that people with dyslexia have specific difficulties with the perception of and/or entrainment to these rhythms. For instance, adults with dyslexia show increased temporal variability in motor synchronization-continuation to isochronous beats presented around 1.5-2 Hz (Flaugnacco et al., 2014; Overy, Nicolson, Fawcett, & Clarke, 2003; Thomson et al., 2006; Thomson & Goswami, 2008; Wolff et al., 1984; Wolff, 2002), and inter-tap interval variability at these rates is also related to reading skills in typically-developing adolescents (Tierney & Kraus, 2013b). Also, some research has shown that children and adults with dyslexia perform poorly on tasks requiring the discrimination of temporal modulations in the amplitude envelopes (i.e. 'rise times') of simple tones using forced choice paradigms (Fraser, Goswami, & Conti-Ramsden, 2010; Hämäläinen et al., 2012; Huss et al., 2011; Muneaux et al., 2004; Pasquini et al., 2007). These apparent low-level auditory deficits have also been observed using more ecologically valid tests of timing and rhythm perception, such as the identification of familiar nursery rhymes based on the tone vocoded segments that were amplitude modulated at stress, syllable and sub-beat timings (Leong & Goswami, 2014b). There is additional evidence that poor synchronisation to auditory timing and rhythm in dyslexia can be detected on a neurophysiological level. For instance, M/EEG data collected from adults with dyslexia show reduced phase-locking to slow (~2 Hz) amplitude modulations (Hämäläinen et al., 2012; Power et al., 2013; Soltész et al., 2013). Taken together, these findings provide strong evidence for an auditory timing impairment in developmental dyslexia.

The evidence for auditory timing impairments in dyslexia forms the basis of the temporal sampling framework (Goswami, 2011; see Chapter 1 for a review). Studies of speech perception show that oscillations track stimuli at different frequencies relevant to speech processing, and Goswami argues that temporal sampling of speech via these low frequency, amplitude-modulated oscillations could explain phonological deficits in dyslexia. Amplitude-modulated beats in continuous speech are closely linked to syllable and vowel timing, as well as to the perceptual occurrence ('p-centre') of words. Goswami hypothesizes a specific impairment with oscillatory entrainment at low Theta and Delta (1.5-10 Hz) frequencies. In the rehearsal-probe task used in the present study, the digits are presented at 2 Hz, so the temporal sampling framework predicts that entrainment to the word-level timing of the rehearsal task sequences would be impaired in the dyslexia group.

While the temporal sampling framework addresses the potential link between low-level auditory perceptual deficits and phonological impairments, one remaining question relates to the co-occurrence of serial order STM deficits in dyslexia. Both children and adults with dyslexia have demonstrated auditory timing imprecision and deficits in STM for serial order, and the perception and entrainment to auditory timing information has also been linked to literacy measures in non-dyslexic populations. In many of the aforementioned studies showing that people with dyslexia present auditory timing and rhythm perceptual deficits, performance on the auditory timing tasks has been correlated with both STM for serial order (e.g. digit span) as well as language and literacy measures (Banai & Ahissar, 2013; Flaugnacco et al., 2014; Law et al., 2014). For instance, Flaugnacco et al. (2014) reported a factor analysis where digit span performance loaded on a factor they describe as "broad auditory temporal processing", which also included motor-timing variability, rhythm reproduction, meter perception and pseudoword repetition tasks. Banai and Ahissar (2013) found that, not only did children with poor reading skills show poorer performance in temporal interval discrimination tasks, but children's digit span and temporal-interval discrimination abilities were correlated after controlling for non-verbal IQ.

Perhaps the simplest explanation for these patterns of timing and serial order STM deficits in dyslexia is that auditory timing perception and auditory-verbal STM for serial order are directly related. This seems plausible given that the encoding of serial order depends on at least a coarse perception of relative event timing. Within this view, poor temporal precision and extraction of auditory rhythm underpins the weak phonological representations that are characteristic of dyslexia. This provides a parsimonious explanation for three specific non-phonological deficits in individuals with dyslexia; 1) auditory STM for serial order (e.g. digit span, serial reconstruction), 2) auditory perception tasks requiring the tracking and integration of stimuli changes over time (e.g. amplitude rise time perception), and 3) timing precision at relatively slow rates (e.g. auditory-motor synchronisation at ~2 Hz).

In this experiment, timing precision is tested using the rehearsal-probe task in a sample of young adults with dyslexia enrolled in higher education. In adults with developmental dyslexia, early reading problems are often remediated through instruction and exposure to orthographic-lexical associations so that later on, reading skills are buttressed by long-term orthographic representations of familiar words. Once long-term mappings are formed between familiar words and orthographic word forms, phonological decoding is no longer critical for reading. Therefore reading performance can improve to normal or near-normal accuracy levels in adults with a childhood diagnosis of dyslexia, despite the continued presence of other deficits such as poor phonological processing, impaired verbal/phonological STM and slower reading rates (Hatcher et al., 2002). If auditory timing plays a causal role in the formation of phonological representations, with knock-on effects for reading development, then the timing and serial ordering deficits should not be expected to improve due to the remediated reading abilities. According to this view, the auditory timing deficit should remain present in adults who have compensated for early reading problems, and therefore these individuals should show poorer temporal precision while maintaining digit sequences in the rehearsal-probe task.

The aims of the present study were as follows. The first aim was to determine whether young adults with developmental dyslexia show increased temporal variability in the rehearsal-probe task relative to age- and non-verbal IQ-matched individuals without a history of language or literacy problems. Previous research on temporal variability in dyslexia has been tested mainly using perceptual discrimination and timing reproduction tasks. While it may be possible to detect auditory perceptual difficulties in dyslexia via discrimination thresholds using forced choice tasks, it is not obvious that these impairments would lead to more variable STM for event timing, which is the question of interest here. Also, as argued in Chapter 1, the use of overt responses to mark timing (as in motor synchronisation and continuation tasks) introduces confounds. For instance, overt responses to mark timing involve a motor planning and execution component, which adds another potential source of temporal variability. In addition, overt responses necessarily have sensory and motor consequences. The availability of sensory and/or motor feedback to monitor and adjust ongoing timing performance is an uncontrolled factor that could explain individual and group differences in these tasks. In the present study, the rehearsal-probe task was used in order to measure auditory-verbal sequence timing in STM without the confounds associated with overt responses to indicate ongoing event timing. It was hypothesised that the dyslexia group would show greater dispersion of serial position responses across probe times relative to the control group.

A second aim of this experiment was to explore the relationship between temporal precision and auditory-verbal STM for serial order. The hypothesis here was that the negative correlation between digit span and mean CSD scores that has been observed in previous experiments will also be present in the current sample of young adults with and without developmental dyslexia. In addition to the rehearsal-probe and digit span tasks, the participants were given a battery of dyslexia screening tasks in order to confirm group differences in phonological, language and reading measures. Mean CSD was expected to relate more closely to performance on the dyslexia screening tasks that place higher demands on phonological awareness and decoding (e.g. nonword reading) compared to tasks that index either processing speed (e.g. rapid automatic naming) or long-term lexical-orthographic knowledge (e.g. word reading). This prediction was based on the view that STM for auditory timing is particularly relevant for the formation of robust phonological representations.

5.1.1 Methods

5.1.1.1 Participants

Thirty-six native English speaking adults were recruited from the University of York and York St John University. There were 18 participants (3 men) in both the dyslexic (mean age 20.83 years, SD = 3.05) and control (M = 20.30, SD = 2.17) groups. Each of the participants in the dyslexia group was diagnosed by a professional psychologist during childhood (before age 16). All of the participants in the dyslexia group also had a re-assessment and confirmation of their dyslexia diagnosis within the past 5 years during their enrolment in further education (A-levels) or higher education (University). None of the participants in either the control or dyslexia group had self-reported history of hearing problems, nor any other language or learning disorders (e.g. ADHD). In addition, the participants must have recalled at least one digit span sequence at a list length of 5 digits in order to minimise the likelihood of errors during the 4-digit rehearsal task. Participants were given a small cash reward or course credit for their participation. All participants provided written consent, and the experiment was approved by the Departmental Ethics Committee at the University of York.

Independent samples *t*-tests were used to identify differences in age, years of education, and non-verbal skills between the dyslexic and control groups (see Table 15). All variables met the assumption of homogeneity of variance using Levene's test of equality of variances. There were no significant differences between the dyslexic and control groups among these variables.

Table 15. Means and standard deviations of age, academic background and non-verbal IQ in the dyslexic and control groups. Results of independent samples *t*-tests with 34 degrees of freedom and corresponding *p*-values are also given.

	Dyslexic Group		Control Group			
Variable	Mean	SD	Mean	SD	*t*	*p*
Age (years)	20.83	3.05	20.39	2.17	-0.50	.618
Years of education	13.67	1.03	13.94	0.94	0.85	.278
Non-verbal intelligence [a]	28.15	4.40	29.74	3.57	1.19	.242

[a] Average of the last three correct trials on the WASI Matrix Reasoning task, maximum possible score = 34.

5.1.1.2 Design

The experiment had a between-subject design, where participants were assigned during recruitment to either the control group or the dyslexic group on the basis of a diagnosis of dyslexia received during childhood. The two groups completed all of the same tasks.

5.1.1.3 Materials

Digit span task. This task was used to assess auditory-verbal immediate serial recall abilities. The digit stimuli, trial procedures and scoring criteria used for the digit span task are the same as those described in Chapter 2 and used in previous experiments.

Rehearsal-probe task. The general task procedure and digit stimuli are described in detail in Chapter 2, and these are the same as those used in previous experiments. Figure 37 shows a schematic diagram of the structure of a single trial in this experiment. The rehearsal lists consisted of four digits selected randomly without replacement from the set of digits 0-9. The rehearsal list length was set to 4 digits in this experiment in order to ensure that the rehearsal lists were subspan in the dyslexia group, as average STM span in the dyslexia group is expected to be lower than that of the control group. The dependent variable for this task is the mean of the CSDs for each serial position response distribution (see Chapter 2 for details about mean CSD calculation).

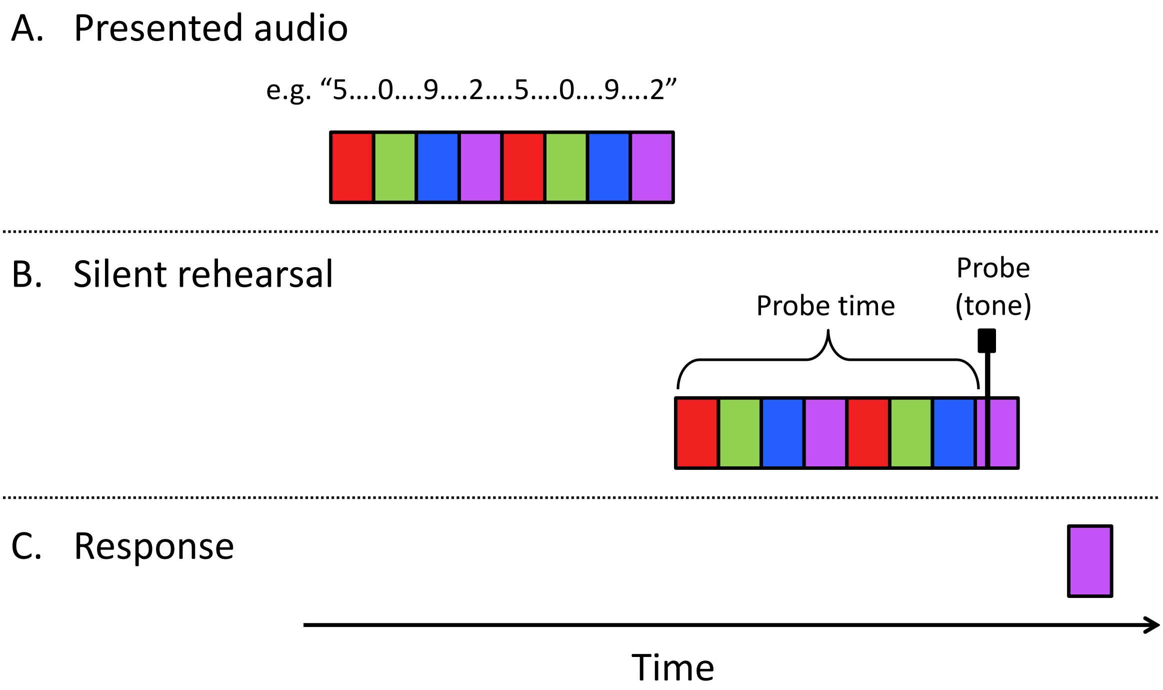

Figure 37. Schematic of a single trial in the Experiment 6 rehearsal-probe task. In each trial, a random 4-digit sequence was presented twice in succession (A), then the participant subvocally rehearsed the sequence exactly as it was presented until hearing the probe (B). Upon hearing the probe, the participant responded with the digit that s/he was currently rehearsing at the time of the probe by pressing the corresponding digit key on the computer keyboard (C). In this example, the probe time shown is the 8th of 10 possible times, occurring at 3575 ms and corresponding to the beginning of the 4th item during the 2nd rehearsal cycle (if the list was rehearsed exactly as it was presented). More details about the probe times are provided in the Procedure section.

WASI Matrix Reasoning. Non-verbal IQ was measured using the WASI Matrix Reasoning task (Wechsler, 1999). The task stimuli and scoring procedures are the same as those described in Experiment 4. In brief, participants are presented with visual pattern matrices containing a single missing piece, and they are asked to select the piece that completes the pattern from a set of five options. The pattern matrices are presented in order of increasing difficulty. Participants continue to respond to each matrix up to matrix number 35 unless they meet the stop criteria. The stop criteria are responding incorrectly to four consecutive matrices or to four out of the five previous matrices. The task is scored by taking the average matrix number of the last three matrices that the participant responded to correctly, giving a maximum score of 34 (where the last three correct matrices were numbers 33, 34 and 35).

There were six additional measures used to confirm group differences in characteristics of dyslexia. These tasks were chosen because they have been shown to be the most sensitive to differences between young adults in higher education with and without a history of reading difficulties (Snowling, Nation, Moxham, Gallagher, & Frith, 1997). Three tasks were used from the revised York Adult Assessment battery (YAA-R; Warmington, Stothard, & Snowling, 2013) for assessing dyslexia in higher education: Spoonerisms, and rapid automatic naming (RAN) for digits and objects. Two tasks were used from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999): Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE). Finally, the phoneme elision task was used from the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999).

YAA-R Spoonerisms. In each trial in this task, the experimenter pronounces the name of a familiar person (e.g. "Michael Jackson") and the participant must swap the beginning sounds of the person's first and last names (e.g. "Jichael Mackson"). Participants are given two practice trials, and then they are timed on their responses for 12 trials. The experimenter started a stopwatch at the end of the pronunciation of the name for that trial, and the time stopped as soon as the participant finished pronouncing the second name. Trials are scored for accuracy, rate, and total time. For the accuracy measure, up to two points are awarded for each of the 12 trials, one point for each correctly pronounced first/last name with the swapped phoneme. Rate is calculated as the average response time per correct response; that is, the sum of the response times for correct responses divided by the total number of correct responses. Total time is the sum of the response times over all of the 12 trials (regardless of the participants' accuracy).

YAA-R RAN for Digits and Objects. In the RAN tasks, participants must speak the names of an array of digits or objects as quickly as possible. For both the digit and object arrays, there were 50 items arranged into 10 rows. The digits were the symbols 0-9, and the objects were pictures of the following items: duck, shoe, car, frog. Participants are given a shorter sequence of the digits and objects to read aloud before the task, in order to give them a chance to practice retrieving the digit/object names and to ensure that they understood the task. Participants are explicitly instructed to read the digits and objects in the same way that they read text (i.e. left to right, top to bottom). The experimenter started the stopwatch when the participant began naming and stopped the stopwatch when the participant finished naming the last item. This task is scored as the rate of words per second, i.e. 50 divided by time to completion (in seconds).

TOWRE SWE and PDE. In both the TOWRE Sight Word Efficiency (SWE) and Phonological Decoding Efficiency (PDE) tasks, participants are presented with a sheet of words printed in columns on a single page and asked to read them as quickly as possible. In SWE, the list is comprised of 104 familiar words, starting with one-syllable words and continuing with multi-syllable words of increasing difficulty. In PDE, the task is structured in the same way except that the list is comprised of 63 non-words. Participants were given a short column of one-syllable words to allow them to practice and to familiarize them with the task. The experimenter instructed participants to read the words in order down each column from left to right, and that if they encounter a word that they cannot read, they should skip that word and go to the next one. The experimenter kept the time by starting a stopwatch when the participant read the first word. Incorrectly read words were marked by the experimenter during the task. The total score was the number of words read in 45 seconds minus the number of incorrect words. The maximum score was 104 for SWE and 63 for PDE.

CTOPP Phoneme Elision. In the phoneme elision task, participants are asked to say a word, and then to say that word without a specific phoneme. For example, the experimenter would instruct the participant to “Say the word ‘mat’”, then, “Now say ‘mat’ without saying /m/”. There were three practice trials followed by 19 assessed trials. In all trials, the deleted phoneme was a consonant, and the remaining phonemes formed a word. The position of the deleted phoneme within the word varied randomly across trials. The task was scored as the number of correct responses, with a maximum score of 19.

5.1.1.4 Procedure

Participants were tested in individual sessions by a single experimenter in a quiet room. The session lasted approximately 45-55 minutes, depending on the length of self-paced breaks taken during the rehearsal task and the duration of debriefing after testing. After the initial consent process, the tasks were administered in the following order: digit span, rehearsal-probe, Matrix Reasoning, Spoonerisms, RAN, SWE, PDE, phoneme elision.

The digit span and rehearsal-probe tasks were administered and scored according to the same general procedures described in Chapter 2 and used in previous experiments. In this version of the rehearsal-probe task, there was no time limit for a response after the probe. Participants were provided with written and oral task instructions, followed by three demonstration trials by the experimenter with overt rehearsal, and three vocal and subvocal practice trials for the participant. Participants then completed 90 rehearsal trials over 3 blocks. Each block was made up of 3 trials for each of the 10 probe times (30 trials per block), where probe times were selected randomly within blocks. The blocks were separated by self-paced breaks, and participants were encouraged to rest between blocks.

The 10 probe times were the following (in msec): 1825, 2075, 2325, 2575, 2825, 3075, 3325, 3575, 3825, 4075. If the presented sequence were continued as presented into the rehearsal delay, the 8 rehearsal durations from 2075 to 3825 would correspond to the beginnings and ends of the digits in serial positions 1-4 during the second rehearsal cycle. To discourage participants from adjusting their rehearsal timing and/or responses as a result of statistical learning of the probe timings, the first and last rehearsal delays were extended such that they fell outside of the second rehearsal cycle (if the timing of the sequence were continued into the rehearsal delay as presented). The first delay time (1825) would align with the end of the digit in serial position 4 in the first rehearsal cycle, and the last delay (4075) would align with the beginning of the digit in serial position 1 in the third rehearsal cycle. All 10 of the probe times were evenly spaced at 250 ms intervals in order to avoid biasing the response variability measure against participants whose rehearsal phase fell out of alignment with the presented sequence timing[9].

The WASI Matrix Reasoning, YAA-R Spoonerisms, YAA-R RAN, TOWRE SWE and PDE, and CTOPP Phoneme Elision tasks were administered according to the procedures outlined in the corresponding test manuals. Each of these tests included practice trials. The administration and scoring procedures for these tasks are listed in the Materials section.

5.1.2 Results

Responses to probes were categorised according to the serial position of the item (digit) within the presented sequence for that trial. Digit responses that were not part of the presented 4-digit sequence on that trial were considered invalid responses. The frequencies and percentages of valid and invalid responses[10], within each participant group and across all participants, are shown in Table 16. The percentage of invalid responses was very low (0.2% overall), and the frequencies of valid and invalid response types were exactly the same between the two participant groups. The 6 invalid responses were removed from the data before further analysis.

---

[9] This general approach to setting the probe times (i.e. extending the probed period beyond a single rehearsal cycle, and selecting evenly-spaced probe times) is the same as that used in the previous two experiments; Experiment 4 (Chapter 3) and Experiment 5 (Chapter 4).

[10] Unlike the earlier versions of the rehearsal-probe task used in Experiments 1, 2 and 4, in the present experiment there was no time limit for a response after the probe. For this reason, omissions (i.e. no-responses) were not possible in this experiment.

Table 16. Response type frequencies and percentages within the dyslexia and matched control groups. Percentages are marginal by row. Bottom row shows the frequencies and percentages for the total responses (two groups combined). Rightmost column shows the total number of responses in each group and in total.

	Response type		
Group	Valid	Invalid	Total
Dyslexia	1617 (99.8%)	3 (0.2%)	1620
Control	1617 (99.8%)	3 (0.2%)	1620
Total	3234 (99.8%)	6 (0.2%)	3240

For each participant, the CSDs were calculated across valid responses using the method described in Chapter 2. Descriptive statistics (means, SDs, minima and maxima) for all measures are listed in Table 17. For the digit span task, all participants demonstrated the ability to correctly recall at least one 5-digit sequence in the correct forward serial order. This was important to ensure that the dyslexia group in particular was not more likely to make serial order errors during the 4-digit rehearsal task as a result of having to rehearse lists at or above their STM spans. As in previous experiments, the digit span task was scored as the average list length of the last three correct trials. Thus, although all participants were required to be able to correctly recall at least one digit span sequence at a length of 5 digits, the minimum possible digit span average score was 4.00 (one correct trial at 3, 4 and 5 digits).

Table 17. Descriptive statistics and independent samples *t*-tests (34 degrees of freedom, unless noted) for group differences across all measures.

			Dyslexic Group		Control Group			
Variable	Min. Possible	Max. Possible	Mean (SD)	Min. – Max.	Mean (SD)	Min. – Max.	*t*	*p*
Digit span average	4.00	-	5.65 (0.77)	4.33 – 7.33	7.33 (0.95)	6.00 – 9.00	5.84	< **.001**
WASI MR average	0.00	34	28.09 (4.33)	20.00 – 34.00	29.74 (3.57)	18.00 – 33.00	1.25	.221
YAA-R Sp accuracy †	0.00	24	19.61 (3.78)	10.00 – 23.00	22.56 (1.29)	19.00 – 24.00	3.13	**.005**
YAA-R Sp time † (sec)	0.00	-	77.49 (92.00)	23.78 – 402.22	32.37 (17.08)	14.73 – 93.28	-2.05	.056
YAA-R Sp rate † (sec/word)	0.00	-	3.64 (5.53)	0.99 – 24.70	1.34 (0.71)	0.61 – 3.89	-1.75	.097
YAA-R Digit RAN (word/sec)	0.00	-	2.52 (0.54)	1.51 – 3.99	3.35 (0.58)	2.26 – 4.51	4.47	< **.001**
YAA-R Object RAN (word/sec)	0.00	-	1.63 (0.22)	1.18 – 1.97	2.00 (0.21)	1.67 – 2.47	5.16	< **.001**
TOWRE SWE †	0.00	104	83.11 (12.91)	61.00 – 104.00	95.44 (7.01)	82.00 – 104.00	3.56	**.001**
TOWRE PDE †	0.00	63	44.00 (8.67)	24.00 – 57.00	57.61 (4.06)	50.00 – 63.00	6.03	< **.001**
CTOPP Phoneme Elision accuracy	0.00	19	16.33 (2.43)	10.00 – 19.00	16.44 (2.06)	13.00 – 19.00	.148	.883
Mean CSD	0.38	1.41	0.80 (0.13)	0.60 – 1.07	0.71 (0.08)	0.55 – 0.89	-2.51	**.018**

Note: WASI MR, Weschler Abbreviated Scale of Intelligence Matrix Reasoning; YAA-R, York Adult Assessment Revised; Sp, Spoonerisms; RAN, Rapid Automatic Naming; TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency (timed word reading); PDE, Phonemic Decoding Efficiency (timed non-word reading); CTOPP, Comprehensive Test of Phonological Processing. Significant group differences indicated by bolded *p*-values.

†These variables violated the assumption of homogeneity of variance. Degrees of freedom in corresponding *t*-tests were adjusted to account for unequal variance.

As listed in Table 17, independent-samples *t*-tests revealed significant group differences in auditory digit span, YAA-R Spoonerisms accuracy, YAA-R digit and object RAN, TOWRE SWE (timed word reading) and TOWRE PDE (timed non-word reading). These results confirm that the sample of adults with dyslexia differed significantly from non-dyslexic controls on standardized measures of auditory-verbal ISR, phonological skills, timed item naming, reading ability, and phonological decoding skills.

Results from the rehearsal-probe task showed that mean CSDs for the dyslexia group were larger than the mean CSDs in the control group, indicating more response variability in the dyslexia group on average. The dyslexia group also showed more within-group variability in the mean CSD measure. The increased response variability in the dyslexia group relative to the control group can be seen by the mean response probability curves (solid lines) in Figure 38. The wider and flatter distributions of the responses for each serial position in the dyslexia group are indicative of more variability in the timing of the responses for items in each position. The between-subject variability (standard error of the mean) within each group is indicated by the width of the semi-transparent ribbons around the solid mean lines in Figure 38. For reference, the coloured rectangles above the response probability curves in Figure 38 represent the timing of the items (colour-coded according to serial position) if the presented sequence were continued into the rehearsal delay. In both groups, the peaks of the serial position response distributions were aligned with, or lagging slightly behind, the expected rehearsal timings. This shows that, on average, the participants were rehearsing the digits in the correct order and at roughly the presented rate.

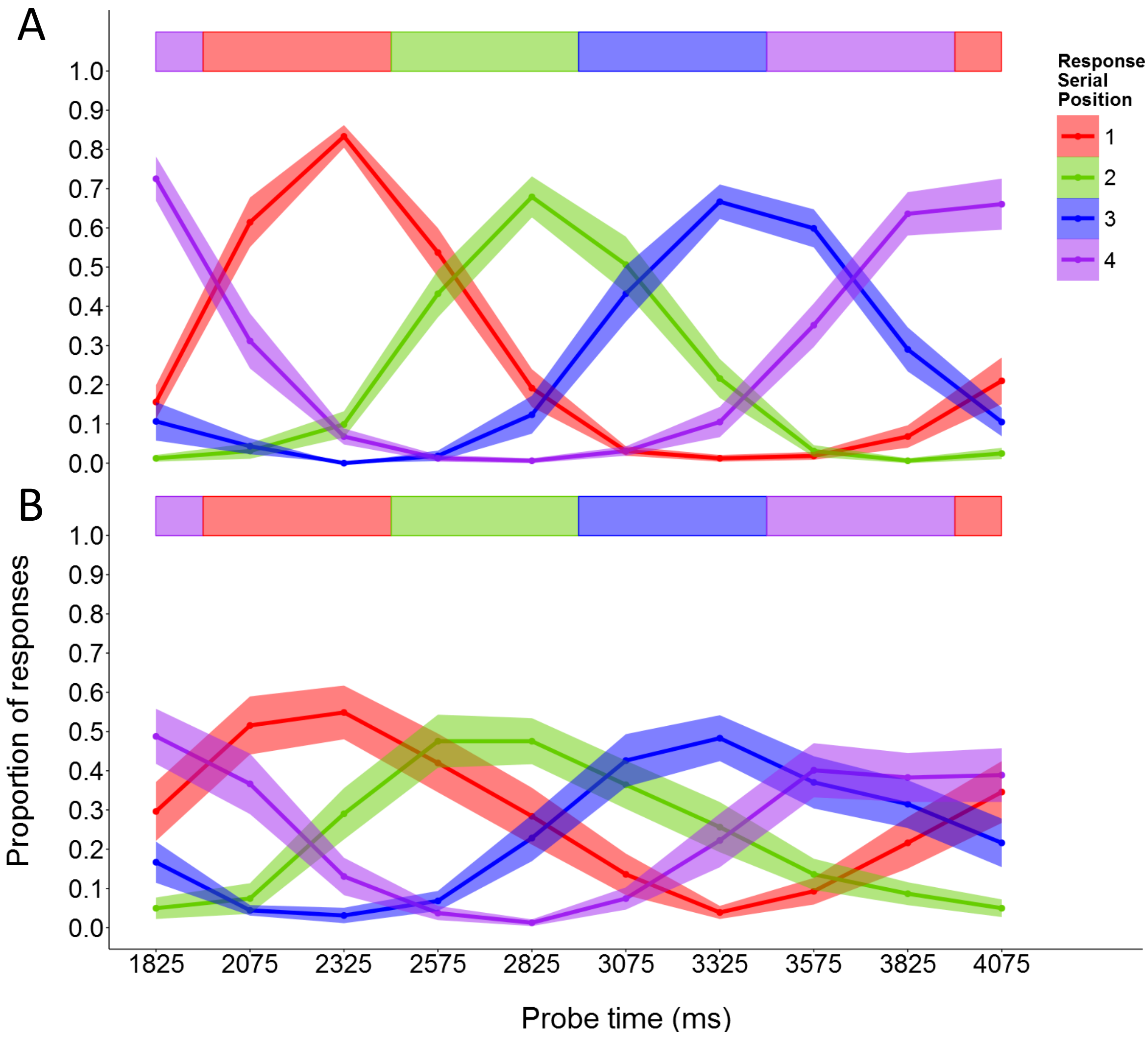

Figure 38. Mean proportions of responses in the control group (A) and dyslexia group (B) for each serial position across probe times. The four colours denote response serial positions. The widths of the ribbons represent the standard error of the means. Boxes above the plots show the 'correct' digit timing, i.e. the timing of subvocal rehearsal if the sequences were rehearsed exactly as presented.

A mixed-design ANOVA was used to test for differences in CSD between the dyslexic and control groups, where group (2 levels; control or dyslexic) was the between-subjects factor and serial position (4 levels; positions 1-4) was the within-subjects factor. The between-subjects factor was tested across all levels of serial position and found to have met the assumption of homogeneity of variance using Levene's test for equality of variances (all *ps* > .05). The within-subjects factor, serial position, violated the assumption of sphericity (*p* = .007), so a Greenhouse-Geisser correction was applied to the degrees of freedom where appropriate.

The main effect of group was significant, *F*(1,34) = 6.28, *p* = .017 (see Figure 39), indicating that there were higher mean CSD scores (i.e. more response variability) in the dyslexia group (M_{dys} = 0.80, SD_{dys} = 0.13) relative to the control group (M_{con} = 0.71, SD_{con} = 0.08, $M_{difference}$ = 0.09, 95% $CI_{difference}$ [0.02, 0.17]). There was also a main effect of serial list position on mean CSD, $F(2.215,75.297) = 17.03$, $p < .001$ (see Figure 40). The main effect of serial position was followed up with planned contrasts between each pair of adjacent serial positions. These contrasts revealed that mean CSDs for the 4th serial position responses were significantly lower than those for the 3rd ($p < .001$) and 1st ($p = .23$) serial position responses. The interaction between group and serial list position was not significant, $F(2.215,75.297) = 1.22$, $p = .305$.

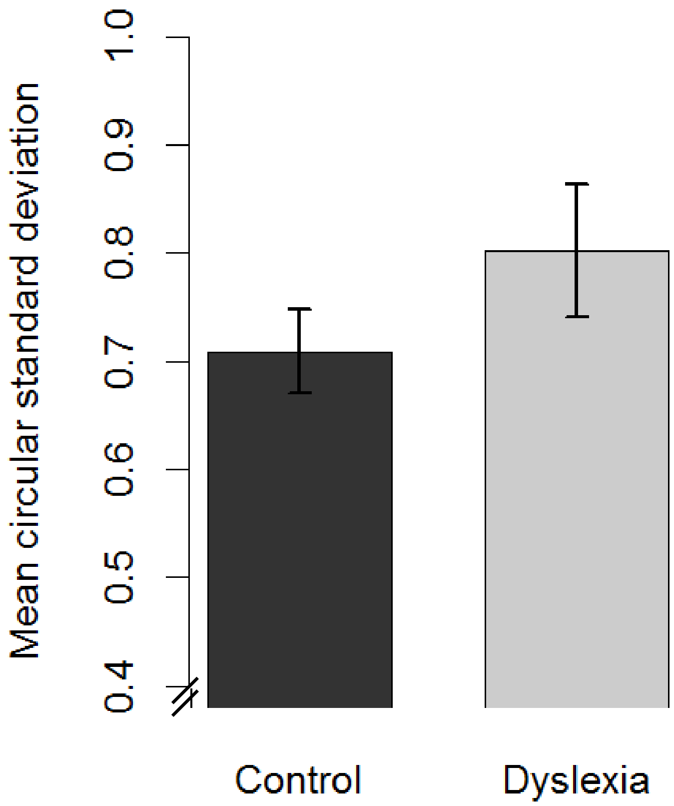

Figure 39. Mean CSDs in the dyslexia and control groups, averaged over the four response serial positions. Error bars indicate 95% confidence intervals. The y-axis minimum is adjusted to the lowest possible mean CSD score in this experiment (0.38).

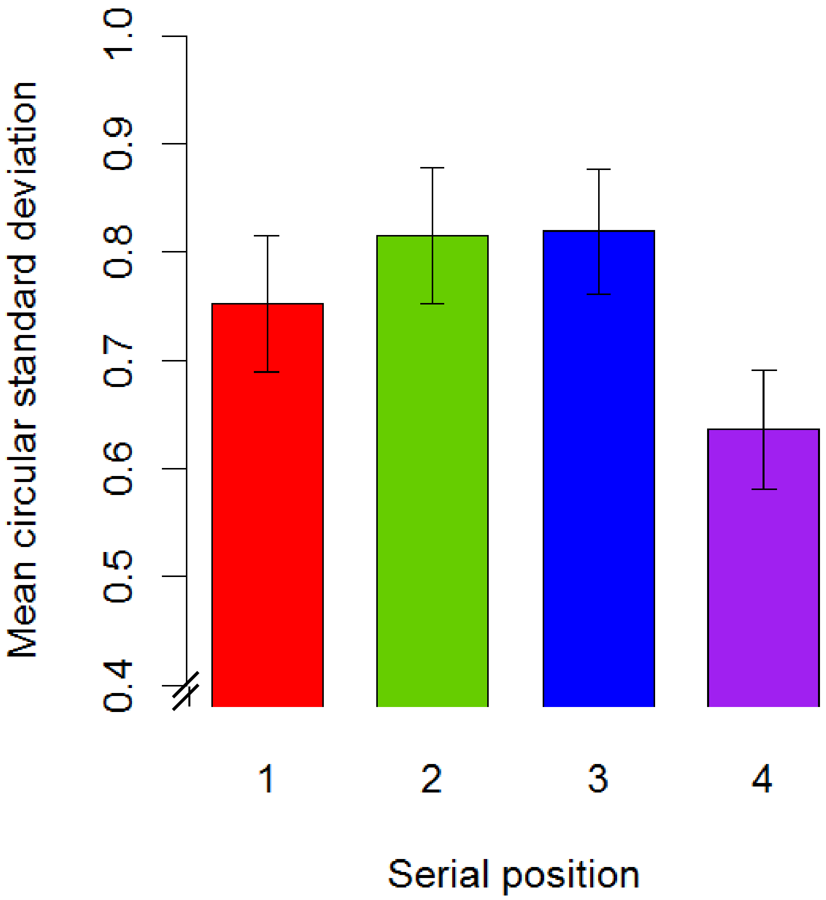

Figure 40. Mean CSDs for each serial position response, averaged over the two participant groups. Error bars indicate 95% confidence intervals. The y-axis minimum is adjusted to the lowest possible mean CSD score in this experiment (0.38).

Correlations between measures are listed in Table 18. Across all participants, mean CSD was significantly negatively correlated with digit span, *r* = -.35, *p* = .036. The correlations between mean CSD and non-verbal IQ (*r* = -.27, *p* = .117) and between mean CSD and non-word reading (*r* = -.26, *p* = .119) showed trends toward negative correlations, but these did not reach the .05 alpha threshold. There were no other significant correlations between mean CSD and other variables.

Table 18. Pearson's *r* correlations between measures and corresponding statistical significance (Bonferroni-corrected for multiple comparisons), n = 36.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Mean circular SD											
2. Digit span	-.35										
3. WASI MR	-.27	.31									
4. YAA-R Sp Accuracy	-.15	.49	.40								
5. YAA-R Sp Time	.11	-.42	-.48	-.77*							
6. YAA-R Sp Rate	.10	-.36	-.46	-.69*	.98*						
7. YAA-R Digit RAN	-.23	.64*	.24	.38	-.27	-.22					
8. YAA-R Object RAN	-.22	.64*	.31	.63*	-.57*	-.50	.51				
9. TOWRE SWE	-.16	.56*	.46	.64*	-.59*	-.54	.75*	.66*			
10. TOWRE PDE	-.26	.76*	.36	.77*	-.61*	-.52	.57*	.71*	.74*		
11. CTOPP Phoneme Elision	.09	.22	.13	.56*	-.39	-.32	.02	.24	.19	.38	

Note: WASI MR, Weschler Abbreviated Scale of Intelligence Matrix Reasoning; YAA-R, York Adult Assessment Revised; Sp, Spoonerisms; RAN, Rapid Automatic Naming; TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency (timed word reading); PDE, Phonemic Decoding Efficiency (timed non-word reading); CTOPP, Comprehensive Test of Phonological Processing. * $p < .001$, statistically significant after Bonferroni correction for multiple comparisons.

Digit span scores showed the strongest relationship with non-word reading, followed by rapid naming, word reading, and Spoonerisms. Most of the dyslexia screening measures were intercorrelated. This is to be expected given that these tasks tap common underlying factors such as phonological awareness, letter-sound mapping and decoding abilities, and speed of lexical access. Phoneme elision showed little relation to other variables apart from two other phonological tasks; Spoonerisms and non-word reading.

To examine the possibility that the correlation between digit span and mean CSD could be accounted for by variance shared with non-verbal IQ, a partial correlation was calculated between digit span and mean CSD after controlling for non-verbal IQ. The partial correlation revealed that the relationship between digit span and mean CSD was no longer significant, $r = -.29$, $p = .089$. This result means that, while ISR capacity is a significant sole predictor of mean CSD, the ISR performance measure shares some variance with non-verbal IQ, and ISR capacity does not account for significant unique variance in mean CSD scores after the variance shared with non-verbal IQ is taken into account.

Because of the group differences in both digit span and mean CSD scores, it was necessary to examine this correlation separately within each group in order to determine whether the correlation observed across all participants was due entirely to the group differences on each continuous variable. The correlation coefficients indicated a negative and slightly stronger relationship between digit span and mean CSD in the control group, $r = -.26$, $p = .29$ (Figure 41, A), compared to a weaker positive relationship in the dyslexia group, $r = .12$, $p = .624$ (Figure 41, B), though both within-group correlations were not significant. These results must be interpreted with caution given 1) the reduced statistical power within the smaller subgroups to detect significant individual differences effects, and 2) the presence of group differences in both measures, which limits dispersion across the two scales. However, it is nonetheless interesting to note that the previously observed relationship between digit span and mean CSD may be stronger in individuals without any history of developmental dyslexia. Based on the scatterplot of the dyslexia group, it is apparent that there are substantial individual differences in mean CSD beyond what would be predicted by digit span within this sample.

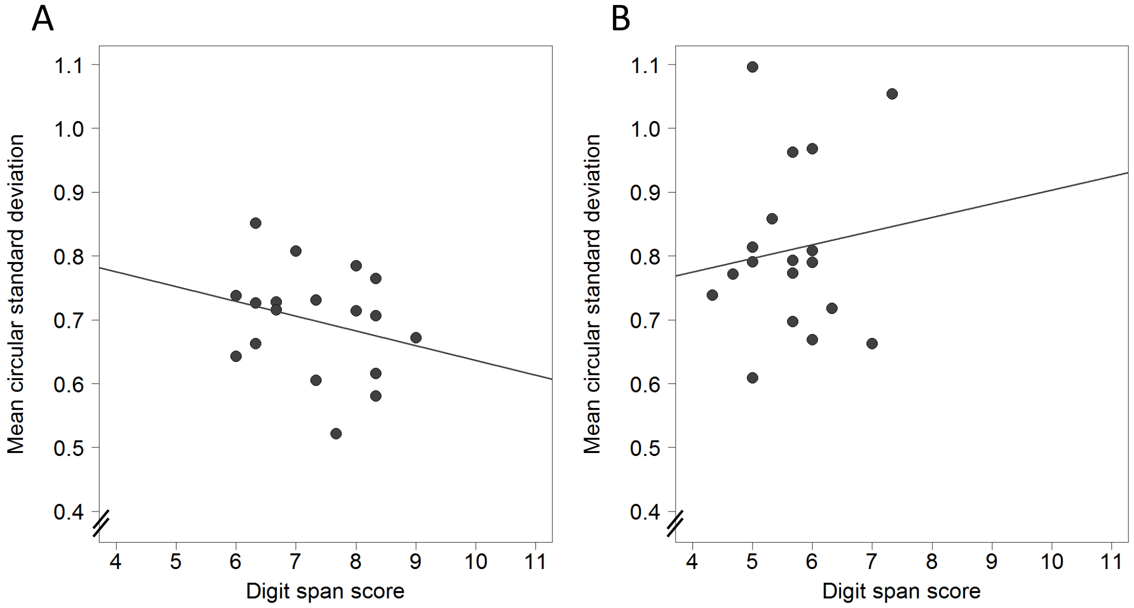

Figure 41. Scatterplots of the relationships between digit span scores and mean CSD in the control (A) and dyslexia (B) participant groups. Solid black lines represent the least squares line of best fit for mean CSD predicted by digit span in each group. The y-axis minima reflect the lowest possible mean CSD score in this experiment.

5.1.3 Discussion

The response proportion functions for the two participant groups replicated the general patterns observed in previous experiments and, as hypothesised, the distributions of responses for each serial position were more variable in the dyslexia group than in the control group. On average, both groups performed the task as expected. The response patterns indicate that the items were rehearsed in the correct serial order. Also, the correspondence between the modal probe times for each response distribution and the expected timings show that the digits were rehearsed at roughly the presented rate. The most striking difference in the response proportion functions between the two groups is the widths of the response position distributions. In the dyslexia group, the response distributions for each serial position were flatter and more overlapping. Extra-list responses were extremely rare, which suggests that item errors during rehearsal are very unlikely to account for the group difference in CSDs. Taken together, these attributes of the data strongly suggest that the higher mean CSD scores in the dyslexia group were not simply a result of the dyslexics' inability to accurately maintain the digit sequences during the rehearsal delays, nor were they due to abnormal task performance in the dyslexia group (e.g. extremely slow or fast rehearsal rates, frequent order errors).

In line with the second hypothesis, there was a significant negative correlation between mean CSD and digit span across all participants, though this was likely a result of gropu differences in the two measures as neither correlation within groups was significant. The negative correlation between mean CSD and digit span is consistent with the results of Experiments 2, 3, 4 (isochronous group) and 5, although the strength of the relationship observed here was weaker compared to previous experiments[11]. The scatterplots of the relationships between digit span and mean CSD within the two participant groups hint at a possible reason for the weaker correlation coefficient here compared to previous experiments, which is that timing precision may not be as closely linked to digit span performance in the dyslexia group. In particular, there were a few participants in the dyslexia group whose temporal variability appears to be higher than expected given their digit span scores. The present study is not sufficiently powered to determine whether the correlations within dyslexic and non-dyslexic samples are significantly different, but future research with larger samples could investigate this possibility.

Although the association between mean CSD and non-verbal IQ scores was not significant in this experiment, individual differences in non-verbal IQ performance moderated the relationship between digit span and mean CSD scores. The results of a partial correlation showed that the relationship between digit span and mean CSD was reduced slightly once shared variance with non-verbal intelligence was taken into account, and this reduction resulted in a non-significant association between the two variables. In Experiment 5, there was also a non-significant negative correlation between non-verbal IQ and mean CSD of a similar magnitude (r = -.25 in Experiment 5 versus r = -.27 in the present study). One key difference is that, in Experiment 5, the individual variance in non-verbal IQ did not mediate the relationship between mean CSD and digit span. One possible reason for this inconsistency is related to the difference in the populations sampled. For instance, it may be that non-verbal IQ is more closely associated with timing precision and serial order STM in people with dyslexia, so IQ had a greater moderating effect in this experiment. Also, as previously mentioned, there was a weaker relationship between digit span and timing variability in the present experiment compared to that in Experiment 5, so the moderating effect of non-verbal IQ might have been more substantial in the present experiment simply because there was less of an association between digit span and mean CSD to begin with.

The finding that non-verbal IQ shares some variance with mean CSD in this study, albeit not a statistically significant amount, is consistent with previous research showing correlations between non-verbal IQ and inter-tap interval variability in motor synchronisation tasks. Correlations reported by Madison and colleagues (Holm et al., 2011;

---

[11] A summary of the correlations between digit span and mean CSD is given in Chapter 4. A meta-analysis of the correlations from Experiments 2-5 (excluding the results from temporally-grouped sequences in Experiment 4) revealed a correlation estimate of r = -.56, 95% CI [-.40, -.69]. In this experiment, the observed correlation (r = -.35) was just outside of the lower bound of the meta-analysis estimate.

Madison et al., 2009; Ullén et al., 2012) are of about the same magnitude as those in the present experiment and in Experiment 5. In a summary of the correlations between intelligence and inter-tap interval variability in isochronous interval serial production tasks across five samples, Ullén et al. (2012) reported a weighted mean correlation of $r = -.30$. Therefore the associations between timing variability and intelligence observed in Experiments 5 and 6 are similar to those found by Madison et al. The differences in statistical significance among these results can be attributed to greater statistical power in the auditory-motor synchronisation studies, where samples ranged from 30 up to 112 participants.

Timing variability was predicted to correlate with phonological skills tasks, but apart from an association with digit span, there were no significant correlations between mean CSD and other measures. Therefore the results of this experiment do not support the hypothesis that timing precision is closely related to phonological skills. However, this null finding may be partly due to problems with the validity and sensitivity of certain phonological skills tasks when used with skilled adult readers. In particular, performance on both the phoneme elision task and the Spoonerisms task (accuracy measure) is likely to have been heavily influenced by the participants' lexical-orthographic knowledge. For example, in the phoneme elision task participants are asked to say a familiar word (e.g. "split") without one phoneme (e.g. the /p/ sound). This task is meant to tap the participant's ability to identify phonemes in continuous speech and remove them while keeping the other phonological segments of the word intact. However, adults with well-formed lexical-orthographic representations are able to complete this task without having to segment the continuous auditory signal. They can do this by first spelling out the word, then using letter-sound knowledge to map the auditory phoneme to the letter, then removing the letter from the word, and finally pronouncing the word made up of the remaining letters. The same method can be used in the Spoonerisms task, where participants must swap the initial phonemes of two familiar names. This method might be somewhat taxing on working memory processes, but it results in a correct response without having to extract and manipulate the acoustic features of auditory input. For this reason, future studies with experienced readers in higher education might consider using phonological skills tasks that requiring the segmentation or pronunciation of unfamiliar word forms.

Compared to mean CSD scores, digit span was more closely related to performance on nearly all phonological skills and reading tasks, and digit span was significantly related to word reading, non-word reading and rapid automated naming for digits and objects. Because digit span and mean CSD scores were correlated, it appears that the rehearsal timing measure may tap a different source of individual variability in digit span than that shared with the other measures. Digit span performance is known to be influenced by automatic, bottom-up processing as well as more attention-driven top-down control. It is possible that the timing variability measure indexes the unique contribution of bottom-up temporal entrainment to tests of ISR for auditory-verbal sequences. The relationships between digit span and other phonological skills and reading tasks (e.g. word reading, rapid automatic naming) are perhaps attributable to common factors that are less relevant to auditory-verbal temporal precision, such as executive control, lexical knowledge and processing speed.

As previously mentioned, there were significant group differences in both digit span and timing precision in this study, as well as a significant correlation between the two variables across all participants. Therefore the most straightforward explanation for the higher temporal variability in the dyslexia group may be that these individuals had lower digit spans. That is, the dyslexia diagnosis itself may not be necessary to explain the group difference in mean CSD scores, and instead this difference could be explained by the group difference in auditory-verbal ISR. This interpretation of the results is supported by the lack of any significant relationships between CSD and phonological skills tasks. In order to determine this is the case, future studies could compare groups of dyslexic and non-dyslexic individuals matched for auditory digit span. It would be interesting to see whether any difference in timing precision exists in between span-matched groups with and without dyslexia, and whether phonological skills tasks predict any variance in timing variability beyond individual span scores in this case.

At this point it is worth considering some of the possible mechanisms underlying relationships among serial order, auditory timing variability, and phonological and reading skills. As mentioned in the Introduction to this experiment, individuals with dyslexia generally present deficits in STM for serial order and auditory temporal entrainment, and one possible explanation for the co-occurrence of these two deficits is that they are causally related. More precisely, it is possible that the encoding of serial order information is dependent, at least in part, upon the precision of a more fine-grained timing signal. While the correlations between digit span and mean CSD observed here and in previous experiments cannot speak to causality, they are consistent with this possibility. Auditory-verbal ISR tasks such as digit span involve a number of cognitive factors that are unrelated to the perception and maintenance of timing information. As mentioned previously, digit span performance is no doubt influenced by top-down factors. For instance, digit span can be dramatically improved with training (Ericcson, Chase, & Faloon, 1980; Kliegl, Smith, Heckhausen, & Baltes, 2009), meaning performance can have a powerful strategic component. Likewise, performance on the rehearsal-probe task is probably influenced by cognitive factors other than bottom-up entrainment to auditory timing, such as sustained attention. Therefore, under the hypothesis that auditory-verbal ISR is causally related to temporal precision for auditory-verbal sequences, it is not surprising that performance on the digit span and rehearsal-probe tasks is not perfectly correlated. Moreover, this explanation, while speculative, has the advantage of accounting for the co-occurrence of dyslexics' deficits in both auditory timing perception and in auditory-verbal STM for serial order.

Martinez Perez, Majerus and Poncelet (2013) present a different explanation for serial order STM impairments in people with dyslexia. These authors propose a specific role of STM for serial order in acquiring reading skills, which is that reading requires the sequencing of grapheme-phoneme associations and these associations must be held in STM until the word has been identified. This explanation is also consistent with the results of the present experiment because, according to this view, adults with developmental dyslexia who have remediated previous reading problems would still be expected to show persistent deficits in STM for serial order. This is because STM for serial order would not be expected to change when reading abilities improve due to the shift from phonological decoding to the recognition of orthographic word forms from LTM. It is important to note that the explanation for the association between STM for serial order and dyslexia given by Martinez Perez and colleagues does not rule out a possible causal association between representations of sequence timing and serial order. Rather, these two possibilities are mutually compatible.

In conclusion, this experiment replicated the expected response patterns seen in previous experiments with the novel rehearsal-probe task while demonstrating that adults with dyslexia show more variability in STM for auditory-verbal sequence timing compared to age- and IQ-matched controls. The group difference found here provides evidence that previously reported differences between dyslexic and control groups in motor-timing variability are not simply due to variability in motor planning and execution, or to differences in the use of overt sensorimotor feedback to adjust continuous temporal interval reproduction. More research is needed to examine whether poor temporal precision is associated with dyslexia beyond what would be predicted by equivalent auditory digit span performance in non-dyslexic individuals. Also, future investigations with greater statistical power have the potential to determine whether dyslexic and non-dyslexic adults differ in terms of the variance shared among non-verbal IQ, auditory-verbal ISR, and timing variability.

5.2 Experiment 7

This study was designed to determine whether the novel rehearsal-probe task can be used with children, and to test whether differences in timing precision exist between groups of children with and without specific language impairment. Specific language impairment

(SLI) is a developmental language disorder typically defined as a severe impairment in receptive and expressive vocabulary and oral language skills despite intelligence and hearing abilities in the normal range, no neurological damage that would explain the impairment, and no other obvious explanation for the language problems (Bishop & Snowling, 2004). Children with SLI often present similar cognitive profiles to children with developmental dyslexia, such as poor phonological awareness and phonological/verbal STM (Casalini et al., 2007; Gathercole & Baddeley, 1990; Majerus, Vrancken, & Van der Linden, 2003). Also, many children with SLI go on to develop dyslexia, particularly the children who have lower non-verbal IQ and/or more severe phonological deficits (Catts, Adlof, Hogan, & Weismer, 2005; Snowling, Bishop, & Stothard, 2000), and the co-occurrence of the two disorders has been estimated at around 50% for 6 to 13 year old children (McArthur, Hogben, Edwards, Heath, & Mengler, 2000). Given the similarities between patterns of deficits in developmental dyslexia and SLI, as well as the high rates of comorbidity, it seemed plausible that the same evidence and rationale for auditory timing and rhythm deficits in dyslexia might also apply to SLI.

There is ongoing debate about how best to characterise SLI and its relationship to dyslexia, owing in large part to inconsistent results about the predominant cognitive deficits in children with language impairments. For instance, studies have shown that verbal and phonological STM is severely impaired in children with SLI (Casalini et al., 2007; Majerus et al., 2003), leading some researchers to suggest that deficits in phonological STM capacity may be a cause of language impairments (Gathercole & Baddeley, 1990). However, others have contested this view, arguing that apparent deficits in verbal/phonological STM performance are due (primarily or in part) to deficits in linguistic factors that typically support verbal STM, such as long-term vocabulary knowledge, knowledge of morphology and phonotactics, and/or access to lexical representations (Archibald & Gathercole, 2007; Bowey, 1997; Coady, Mainela-Arnold, & Evans, 2013; Gathercole & Adams, 1994; Gathercole, 1995; Leclercq, Maillart, & Majerus, 2013; Mainela-Arnold & Evans, 2005; Snowling, Chiat, & Hulme, 1991). Similarly, some studies have shown that children with SLI show certain low-level auditory processing deficits (Corriveau, Pasquini, & Goswami, 2007; Ferguson, Hall, Riley, & Moore, 2011), while other researchers point out that problems with auditory temporal processing are absent in most individuals with developmental language disorders (Rosen, 2003), co-occur but are not causally linked (Rosen, 1999), or only occur in linguistic contexts (C. R. Marshall, Harcourt-Brown, Ramus, & van der Lely, 2009). Given these and other examples of conflicting evidence surrounding the primary deficits in SLI, as well as the precise relationships among verbal STM, oral language and reading impairments, further research is clearly necessary to resolve ongoing debates.

Some studies have shown auditory timing and rhythmic processing deficits in children with SLI, similar to those observed in children and adults with dyslexia. Corriveau et al. (2007) found that children with SLI performed worse on rise time perception and tone duration discrimination tasks compared to both age- and language-matched controls. Crucially, both groups performed equally well on non-temporal auditory perception tasks, such as intensity discrimination, which rules out a general auditory perception impairment as the cause of the group differences. In addition, Corriveau and colleagues showed that performance on tasks requiring the integration of auditory temporal information predicted unique variance in a number of language and literacy measures after accounting for variance shared with age, non-verbal IQ and an attentional control task. In another investigation, Corriveau and Goswami (2009) found that inter-tap-interval variability in 1.5-2 Hz paced tapping task accounted for significant variance in spelling, single word reading and nonword reading after accounting for individual differences in age, non-verbal IQ and manual dexterity. Corriveau and Goswami propose the possibility of a common aetiology in SLI and dyslexia, namely that the disorders could arise from insensitivity to low-level auditory rhythm and timing cues. If impairments in the perception and maintenance of auditory timing are present in children with SLI, then this group would be expected to show increased temporal variability in the rehearsal-probe task.

Emerging evidence for a relationship between auditory timing imprecision and developmental language disorders is made more complicated by the fact that non-verbal IQ has also been linked to temporal precision in motor timing tasks, as well as to language and literacy measures. McAuley et al. (2006) used a synchronised finger-tapping task to test the performance of a wide range of age groups across multiple tapping rates. These authors found that children with higher non-verbal IQ could successfully synchronise their tapping to a wider range of rates (with optimal rates around 2 Hz). Waber et al. (2000) tested groups of 7-11 year old children with and without learning disorders on paced finger-tapping tasks, and found that inter-tap-interval variability higher in the group with learning impairments. Regression analyses across all children showed that timing variability was significantly and independently related to reading, spelling and non-verbal IQ scores. These findings highlight the importance of controlling for individual and group differences in non-verbal intelligence when considering links between language, STM and auditory temporal precision.

As previously mentioned, auditory-verbal STM problems in SLI are often attributed to deficits specifically within the language system, such as poor vocabulary, impaired phonological skills and ineffective access to lexical stores. In accordance with this view, children with SLI are reliably impaired on nonword repetition tasks (Gathercole & Baddeley, 1990; Majerus et al., 2009), which are considered to place a relatively low load on STM for serial order and high load on the sublexical (i.e. phoneme, syllable) 'item' information. However, there are two potential problems with the use of nonword repetition performance to support the view that verbal STM deficits originate in the language system. First, it is not clear that the nonword repetition task carries a markedly reduced memory load for serial order information compared to typical serial order tasks, such as digit, word and sentence span. This is because nonword repetition requires STM for an unfamiliar sequence of sublexical units, such as phonemes and syllables. Based on studies of adults and children, Gupta (2003) demonstrated robust correlations between digit span and nonword repetition, and argued that the two tasks are related due to an underlying serial order maintenance component.

A second issue with the argument that all STM deficits in children with SLI are a result of the primary language problems is that these children also show impairments with auditory-verbal STM tasks that load heavily on serial order, such as digit span, where the words are highly familiar and come from a semantically-related, closed set. In the case of digit span, it is not clear how long-term linguistic knowledge (e.g. vocabulary, morphology, phonotactics) or deficits in lexical access could explain the poorer performance observed in children with SLI. In addition to performance deficits in ISR tasks such as digit span, children with SLI have also been shown to perform more poorly on other tasks designed to load heavily on STM for serial order, such as serial order recognition and reconstruction (Majerus et al., 2009; Nithart et al., 2009). One suggestion that has been raised for why children with language impairments struggle with verbal STM for serial order is that they do not use subvocal rehearsal, or that their rehearsal is slower and less efficient (Gill, Klecan-Aker, Roberts, & Fredenburg, 2003; Kirchner & Klatzky, 1985). However, this explanation does not seem tenable given that children with SLI show evidence of rehearsal via the word length effect (Gathercole & Baddeley, 1990), and that these children use vocal rehearsal with the same frequency as children without language impairments (Alt & Spaulding, 2011). More generally, there are a number of reasons to doubt that rehearsal differences are a major determinant of performance on auditory-verbal ISR tasks such as digit span (Dempster, 1981; Jarrold & Hall, 2013; Tehan & Lalor, 2000). Another potential explanation for reduced auditory-verbal ISR capacity in developmental language disorders is that individuals with these disorders have a specific difficulty in perceiving and/or maintaining auditory rhythms, which underlies deficits in both the accurate extraction of sublexical units in continuous speech as well as auditory-verbal STM for serial order.

The present experiment used the rehearsal-probe task to address the remaining questions about the possible auditory-verbal STM and auditory timing deficits in children with SLI. However, it was not clear from the outset whether children would be able to perform the rehearsal-probe task, which up to this point has only been used with adult participants. Two additional challenges were considered that had the potential to prevent the rehearsal-probe task from being successfully used with child participants. First, the children may have difficulty understanding the instructions, and second, even if the task is properly understood, the task itself may be too difficult for children to perform. Regarding the first issue, the task requires a certain amount of metacognition in that the participant is expected to know (or 'guess' with a certain degree of accuracy) what s/he is subvocalising throughout the rehearsal period. The notion that the children in the present experiment are capable of metacognition is supported by a study by Flavell, Greeen and Flavell (2000) showing that, while 5 year olds have a clear difficulty with introspection, 8 year old children performed either at an intermediate level or as well as adults on metacognition tasks. Another potential issue with the task instructions is that participants are asked to remember multiple instructions simultaneously (i.e. silently rehearse the list at the same pace, continue to rehearse until hearing the tone, do not whisper or tap to the beat, etc.), which could be especially difficult for children with limited verbal STM/WM abilities and language comprehension problems. This potential problem was addressed through changes to the task administration, which are described later in this section as well as in the Methods section.

Regarding the second issue, if the task itself is too difficult for children to perform then floor effects may occur and the data could be too noisy to be meaningful. However, there are a few reasons why it was expected that children would be able to perform the rehearsal-probe task successfully. Research has shown that children begin to show signs of using subvocal rehearsal spontaneously after the age of about 5 to 7 years (Alt & Spaulding, 2011; Gathercole & Baddeley, 1990; Hitch, Halliday, & Littler, 1989; Hitch et al., 1991; Lehmann & Hasselhorn, 2010; Tam et al., 2010)[12], and at age 10, children's reports their own inner speech are consistent with other indicators (Flavell et al., 1966). It was therefore assumed that the children participating in the present experiment, who are aged 8 to 11, should be familiar with subvocalisation. Also, research has shown that the majority of 7-8 year old children with and without SLI use vocal rehearsal spontaneously when learning new words, and that the children with SLI use rehearsal with the same frequency as their typically-developing peers (Alt & Spaulding, 2011). This is important given that the children with SLI

---

[12] See Jarrold and Hall (2013) for a critique of the evidence presented in these and other studies to support the claim that there is a qualitative change in the use of subvocal rehearsal during development. These authors argue that some of the effects used to indicate a presence or absence of rehearsal can be explained by other phenomena (e.g. output timing, limited power to detect effects in young children, phonological complexity) and that presentation rates and response initiation times are too fast to allow cumulative rehearsal in ISR tasks. Importantly, Jarrold and Hall do not suggest that young children cannot or do not rehearse, rather they call into question the conclusions that 1) the use of rehearsal changes qualitatively during early or middle childhood, and 2) age-related increases in ISR capacity can be explained by the use or efficiency of subvocal rehearsal.

could be at a particular disadvantage in the rehearsal-probe task due to a reduced propensity to use inner speech and/or phonological coding for memory maintenance. In terms of the timing aspect of the task, children as young as 8 years can synchronise to a range of paced beats with about the same level of precision as adults (McAuley et al., 2006). Thus, it is expected that the 8-11 year old participants in the present study would be able to perform the rehearsal-probe task based on the evidence that these children use inner speech, and subvocal rehearsal in particular, and that they are able to synchronise to paced auditory input with sufficient precision to produce systematic responses.

In response to these two methodological considerations, the rehearsal-probe task was adapted in this experiment to make it more appropriate for children. The instructions were altered to incorporate a short storyline and character in order to provide a pragmatic motivation for completing the task as directed. Also, the multiple task instructions were addressed intermittently during the demonstration and practice trials in order to allow the children to familiarise themselves with each aspect of the task, and to allow the experimenter to verify the child's understanding of each instruction before starting the task. The rehearsal delays were shortened relative to the versions of this task used with adult participants, and there were fewer trials and shorter trial blocks. These modifications were made because it was suspected that children would have more difficulty with sustained attention compared to adults, especially since the task provides little variation or visual interest, and therefore the children's performance would likely be more vulnerable to boredom and fatigue effects. Finally, visual feedback was given at the end of each trial block to reward the children for their effort and to encourage them to continue.

There were two main aims of this experiment. The first objective was to determine whether the rehearsal-probe task can be successfully conducted with child participants. Some of these participants have language comprehension impairments, so these children may have particular difficulty understanding the task instructions. The success of this aim will be evaluated with regard to the evidence of either understanding or misunderstanding the task instructions based on the response proportion functions, as well as the presence or absence of floor effects. The response proportion functions should appear systematic and lawful if the task is performed correctly, whereas these functions are otherwise expected to be either unsystematic (random responses with respect to probe times) or systematic but non-cyclical (e.g. responses across all probe times correspond to a single serial position). Of interest is whether a substantial portion of the children show difficulty understanding the task as indicated by the extent to which individual response proportion functions are systematic and cyclical.

The second main aim was to explore whether the rehearsal-probe task is sensitive to individual and developmental language disorder differences in children. There were two hypotheses regarding the results of group and individual differences. First, it was expected that the group of children with SLI would, on average, show poorer timing precision compared to a group of age- and IQ-matched children with typical language development. The second prediction was that, consistent with previous investigations with adult participants, there would be significant correlations between temporal variability and auditory-verbal STM for serial order. Based on the results of Experiment 6, timing variability was also expected to be separately associated with non-verbal intelligence. Finally, it was predicted that timing variability would be relatively less correlated with performance on tasks that do not contain an auditory temporal processing or sequencing component, such as visuospatial STM, or with tasks that are more heavily influenced by executive function (i.e. WM tasks).

5.2.1 Methods

5.2.1.1 Participants

An opportunity sample of 12 children with SLI and 15 IQ-matched control children were tested at two primary schools in Cambridge, England. One child from the control group was excluded from the analysis due to environmental distractions during testing[13]. One child from the SLI group was excluded from the analysis because the data revealed no systematic pattern in the responses, suggesting that either the child's STM for timing was severely impaired or the child did not understand the task[14]. After these exclusions there were 11 children (5 girls, 6 boys) in the SLI group and 14 children (6 girls, 8 boys) in the control group. The children were 8-11 years old (M = 9 years, 9.6 months; SD = 8.9 months), and in school years 4-6 (7 children in year 4, 11 in year 5, 7 in year 6). Group descriptions and differences on a number of measures are reported in Table 19.

These children were previously selected for a working memory training study conducted by the MRC Cognition and Brain Sciences Unit (CBSU) in Cambridge. Before the start of the

---

[13] Due to a room scheduling conflict, this child was tested in an open plan area with open corridors to nearby rooms. After the first block of trials, there were audible conversations in the nearby rooms. During the fourth trial block, a teacher and group of students entered the room and began a reading lesson. These events appeared to be very distracting for the child, who reacted by looking in the direction of the noise and watching the reading lesson. All other children were tested in a quiet, enclosed room.

[14] This child responded with the item in the first serial position in 67 out of the 70 trials (95.7%). The child responded with the item in the second position in 2 trials (2.9%), and one response was invalid (1.4%).

working memory training, the MRC CBSU researchers conducted a number of STM, WM, IQ, reading and language measures with the children (see Table 19 for a list of tasks). Only the rehearsal-probe task was administered as part of the present experiment; all other data was collected by the MRC CBSU researchers prior to the rehearsal-probe task administration, and the results of all other measures were provided for the purpose of analysis in the present study. The children were about 18 weeks through the 20-week working memory training program when the current research was conducted.

Table 19. Group means (SDs in parentheses) and differences in age, sex and cognitive pre-training measures (data provided by MRC CBSU researchers). Significance value *p* refers to the outcome of independent samples *t*-tests with 23 degrees of freedom unless noted.

Variable	Construct	Control (N=14)	SLI (N=11)	*p*
Age (years; months)	-	9; 9.4 (9.5)	9; 9.9 (8.5)	.881
Sex † (female; male)	-	6; 8	5; 6	.897
CELF Recalling Sentences	Lang	10.79 (1.58)	3.82 (1.67)	**<.001**
CELF Comprehension	Lang	9.64 (1.65)	5.91 (2.81)	**<.001**
PPVT	Lang	103.29 (11.26)	80.73 (4.08)	**<.001**
WIAT Word Reading	Lang	93.59 (14.04)	85.46 (11.63)	.136
WIAT Comprehension	Lang	118.43 (15.62)	98.91 (16.15)	**.006**
CNRep [a]	pSTM	31.14 (3.78)	20.91 (6.99)	**.001**
Digit Recall	vSTM	106.72 (11.31)	85.68 (9.44)	**<.001**
Word Recall	vSTM	104.18 (11.24)	87.38 (11.33)	**.001**
Listening Recall	vWM	102.43 (14.21)	94.95 (14.03)	.202
Backwards Digit Recall	vWM	100.64 (13.09)	92.09 (9.67)	.083
Dot Matrix	vsSTM	93.31 (11.14)	94.99 (6.50)	.662
Block Recall	vsSTM	91.69 (9.08)	89.90 (12.34)	.680
Mr. X	vsWM	99.86 (14.48)	98.82 (16.12)	.867
Spatial Recall	vsWM	103.12 (9.99)	97.86 (17.05)	.344
WASI Vocabulary	vIQ sub	51.71 (9.55)	42.09 (7.69)	**.012**
WASI Similarities	vIQ sub	51.57 (8.05)	43.82 (7.57)	**.022**
WASI Verbal IQ	vIQ	102.71 (12.45)	89.18 (8.41)	**.005**
WASI Block Design	nvIQ sub	44.07 (6.16)	44.27 (7.67)	.943
WASI Matrix Reasoning	nvIQ sub	43.00 (7.02)	40.46 (6.65)	.367
WASI Performance IQ [b]	nvIQ	89.93 (5.66)	88.18 (9.25)	.590
WASI Full IQ	Full IQ	95.71 (8.69)	87.46 (8.34)	**.025**
WIAT Numerical Operations	Maths	96.79 (12.77)	88.18 (11.21)	.091

Note: † Chi-square test for differences. Letter subscript indicates violation of homogeneity of variance; [a] df = 14.53, [b] df = 15.71. Bolded *p* values are significant at α = .05. Task abbreviations: CELF, Clinical Evaluation of Language Fundamentals screening test; PPVT, Peabody Picture Vocabulary Test; WIAT, Wechsler Individual Achievement Test; CNRep, Children's Test of Nonword Repetition; WASI, Wechsler Abbreviated Scale of Intelligence; STM, short-term memory; WM, working memory. Construct abbreviations: Lang, language; p, phonological; v, verbal; vs, visuospatial; nv, non-verbal; sub, subtest.

5.2.1.2 Design

This experiment took a between-subjects design; children were assigned to either the SLI group or control group, depending on the criteria for inclusion/exclusion used by the MRC CSBU research group conducting the memory training study. All children in the experiment completed the same tasks.

5.2.1.3 Materials

Only the rehearsal-probe task was administered as part of the present experiment. The general rehearsal-probe task paradigm was the same as those used in previous experiments, with some modifications made to the task in order to make it more appealing to children (see the Procedure section). The auditory digit stimuli used were the same as those described in Chapter 2 and used in previous experiments. The list of cognitive and language tasks administered by the MRC CBSU research group is given in Table 19, and details of the methods for these tasks are provided in the Appendix.

5.2.1.4 Procedure

The children were tested individually by a single experimenter, who was blind to the children's group status. The testing was conducted in quiet rooms within the children's schools. The instructions to the task were read to the child by the experimenter, and were integrated with the practice trials in a custom E-Prime program. The instructions introduced a cartoon character called "Eleanor the Elephant" (see Figure 42) who would be saying the numbers during the task. The instructions explained that, after Eleanor was done saying the numbers aloud (Figure 42, slide 2), she would begin thinking the numbers to herself (Figure 42, slide 3), and the child's task was to guess the number that Eleanor is thinking when they hear the "beep" (auditory probe; Figure 42, slide 4).

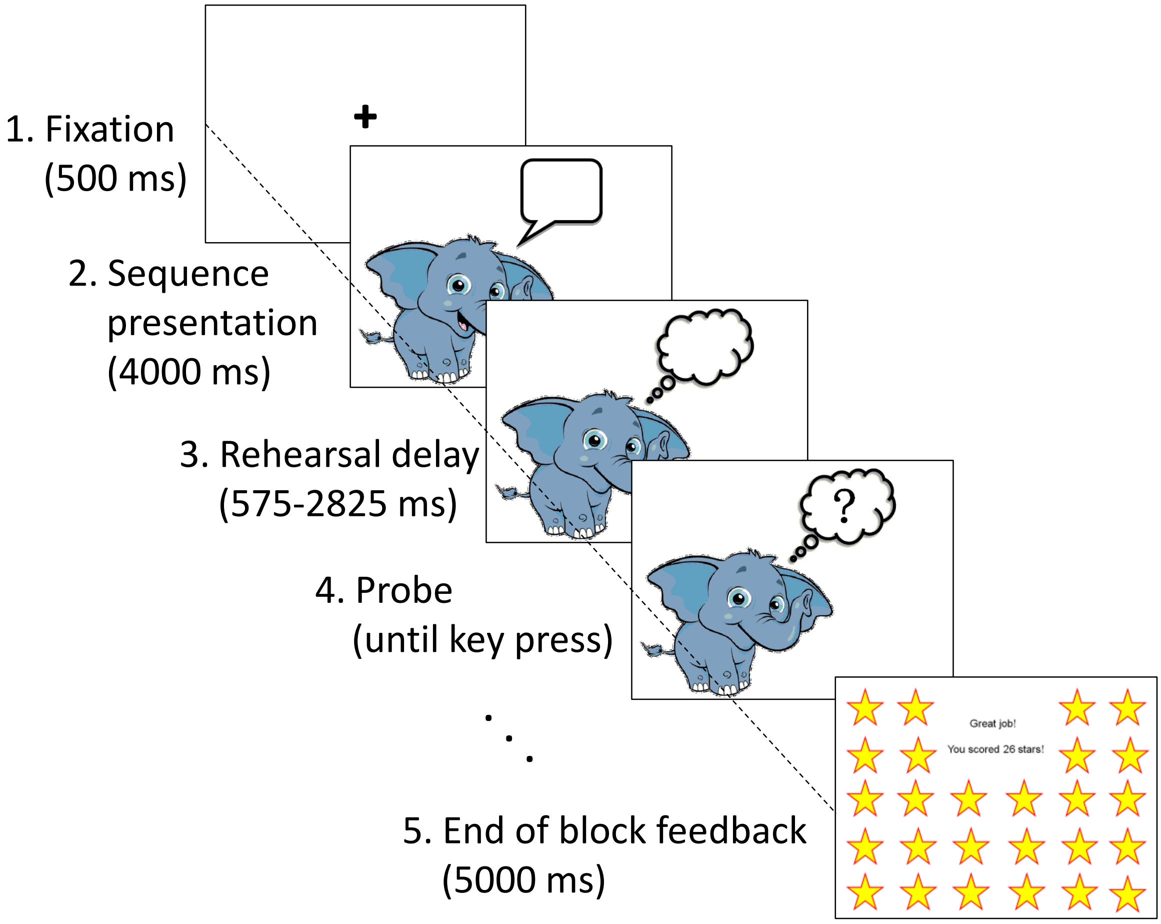

Figure 42. Sequence of visual stimuli from the rehearsal-probe task, adapted for children. Each trial begins with a 500 ms fixation cross (1), followed by the 'Eleanor the Elephant' character, who is 'speaking' the 4000 ms digit sequence (2). The speech bubble converts to a thinking bubble during the rehearsal delay (3), which lasts 575-2825 ms, depending on the probe time for that trial. When the auditory probe (tone) is presented, a question mark appears in the thought bubble (4), and this image remains until the child responds with the digit s/he was rehearsing at the time of the probe. At the end of the block, a feedback slide displaying the number of stars the child received for that block is shown for 5000 ms (5).

The children were then told that they should try this first by saying the numbers aloud during the rehearsal delay. At this point the experimenter demonstrated three trials with overt rehearsal, and the child completed three practice trials with overt rehearsal. Following the overt practice trials, the instructions then asked the child to do the same thing by "saying the numbers in your head instead of aloud" and clarified this point further: "You should think the numbers silently to yourself. That means only in your head, without moving your lips, hands or feet to keep the beat". The child then completed three silent rehearsal trials, and the experimenter watched carefully in order to correct any articulatory or rhythmic movements such as whispering, tapping or nodding.

The children's practice and experimental trials were followed by performance feedback which was presented in the form of text and a number of stars displayed on the screen after the trials ended, e.g. “Great job! You scored 26 stars” (Figure 42, slide 5). The number of stars displayed after each block was calculated as follows: the child automatically received 1 star for each trial, 2 stars if the response was adjacent to the correct item (e.g. a response position of 1 or 3 if the correct item was in position 2), or 3 stars if the child responded with the correct item. The purpose of the feedback was primarily to provide encouragement and motivation during the task, rather than to modify the children’s rehearsal timing or responses. For this reason, the children always received as many stars as there were trials in the block (in order to prevent any child from receiving 0 stars). Because the children were not aware of how many stars it was possible to receive, they could not have been affected by poor feedback about their performance. The feedback served a secondary purpose of signalling the experimenter when the child was making a high number of invalid (extra-list) responses. This was particularly important during the two sets of practice trials, which were repeated if the child made more than one invalid response.

Figure 43 shows a schematic of the auditory stimuli from a single trial. The rehearsal lists were random 4-digit isochronous sequences. The task was made up of 70 trials in total, divided up over 5 blocks of 14 trials. There were 10 probe times, and 7 trials per probe time in total. The 10 probe times were: 575, 825, 1075, 1325, 1575, 1825, 2075, 2325, 2575 and 2825 ms. These times corresponded to the beginning and end of items in positions 2, 3 and 4 in the first rehearsal cycle, and items in positions 1 and 2 in the second rehearsal cycle. The probe times were selected at random without replacement from the list of 10 throughout the experiment, and the full-list randomisation began anew after multiples of 10 trials.

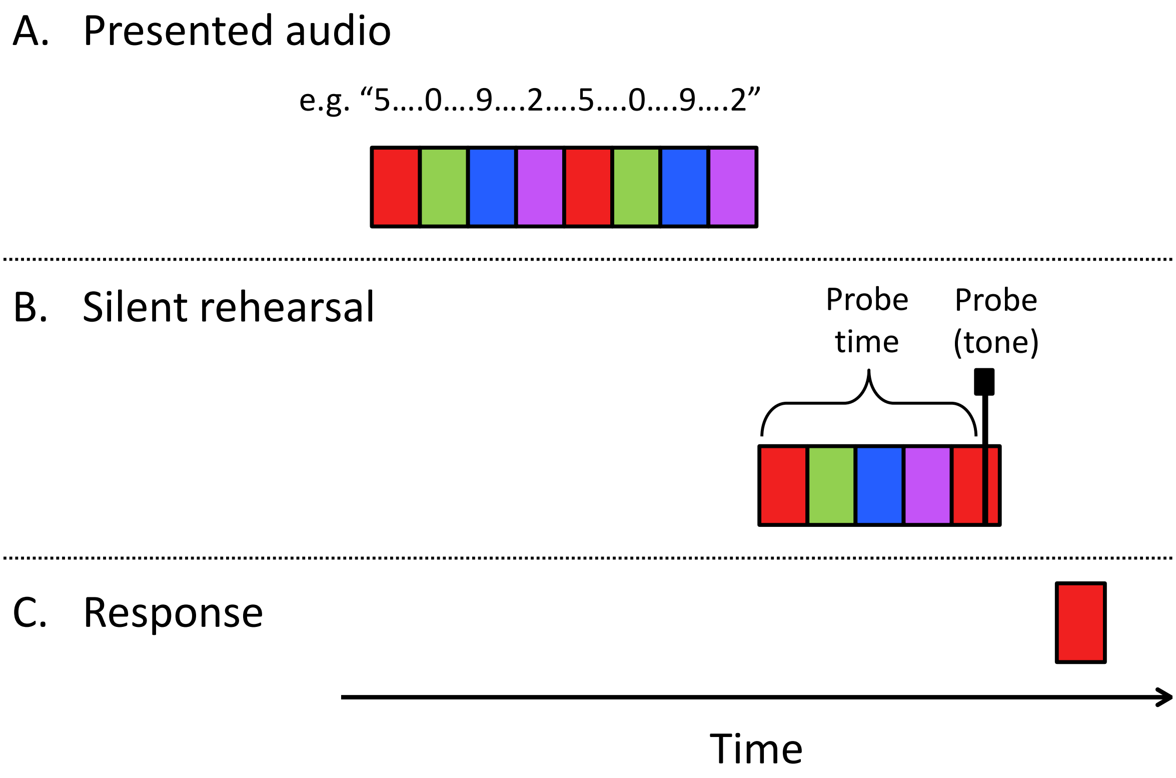

Figure 43. Schematic of a single trial in the rehearsal-probe task from Experiment 7. The children heard a random, isochronous 4-digit sequence repeated twice (A) and then silently rehearsed the sequence exactly as it was presented until hearing a tone (B). Upon hearing the probe, the child responded with the digit s/he was rehearsing at the time of the probe (C). The 10 probe times were aligned with the beginnings and ends of items in positions 2, 3 and 4 during the 1st rehearsal list cycle, and positions 1 and 2 in the 2nd rehearsal cycle. This example shows a probe presented during the end of the item in the 1st serial position during the 2nd rehearsal list cycle, which corresponds to the 8th of 10 possible probe times. More details about probe times are given in the Procedure section.

5.2.2 Results

Responses to probes were categorised according to the serial position of the digit within the presented sequence for that trial. Digit responses that were not part of the presented 4-digit sequence on that trial were considered invalid responses. Because there was no time limit for a response following the probe in this experiment, there were no trials classified as a ‘no-response’. The frequencies and percentages of valid and invalid responses, within each participant group and across all participants, are shown in Table 20. The percentage of invalid responses was very low (1% overall). A Fishers Exact Test showed that the distributions of valid and invalid responses did not differ significantly between participant groups, ($p > .05$, all standardised residuals were $|z| < 1.1$). The 18 invalid responses were removed from the data before further analysis.

Table 20. Response type frequencies and percentages within the two participant groups; children with specific language impairment (SLI) and the age- and IQ-matched controls. Percentages are marginal by row. Bottom row shows the frequencies and percentages for the total responses (two groups combined). Rightmost column shows the total number of responses in each group and in total.

	Response type		
Group	Valid	Invalid	Total
SLI	759 (98.6%)	11 (1.4%)	770
Control	961 (99.3%)	7 (0.7%)	968†
Total	1720 (99.0%)	18 (1.0%)	1738

† The total number of trials in this group is lower than the expected total of 980 trials (14 participants x 70 trials per participant = 980) because a technical problem during data collection for one participant in this group resulted in an inability to run the last 12 trials of the experiment. The data that was successfully collected from this participant (58 trials total) was included in the analysis.

Figure 44 shows the proportions of responses for each serial position across probe times in the control (A) and SLI (B) participant groups. It is worth noting the qualitative aspects of the response proportion functions in the two groups before analysing the quantitative measures derived from the data. As in previous experiments, the modal probe times associated with each serial position occurred in the expected order in both groups. These peaks in the serial position response distributions are roughly in line with the presented sequence timing, as indicated by the correspondence between the serial position response distribution peaks and the boxes above the plots that show the ‘correct’ timing (if the sequence was rehearsed exactly as it was presented). These attributes suggest that, on average, the children in both groups were able to maintain the items in the correct serial order, and the rate of subvocal rehearsal was generally accurate. The response proportion functions of the two groups differ more clearly in terms of their dispersion and overlap. The wider and flatter distributions of the responses for each serial position in the SLI group are indicative of more average variability in responses for each position over time. The between-subject variability (standard error of the mean) within each group, indicated by the widths of the semi-transparent ribbons around the solid mean lines, is also noticeably larger in the SLI group.

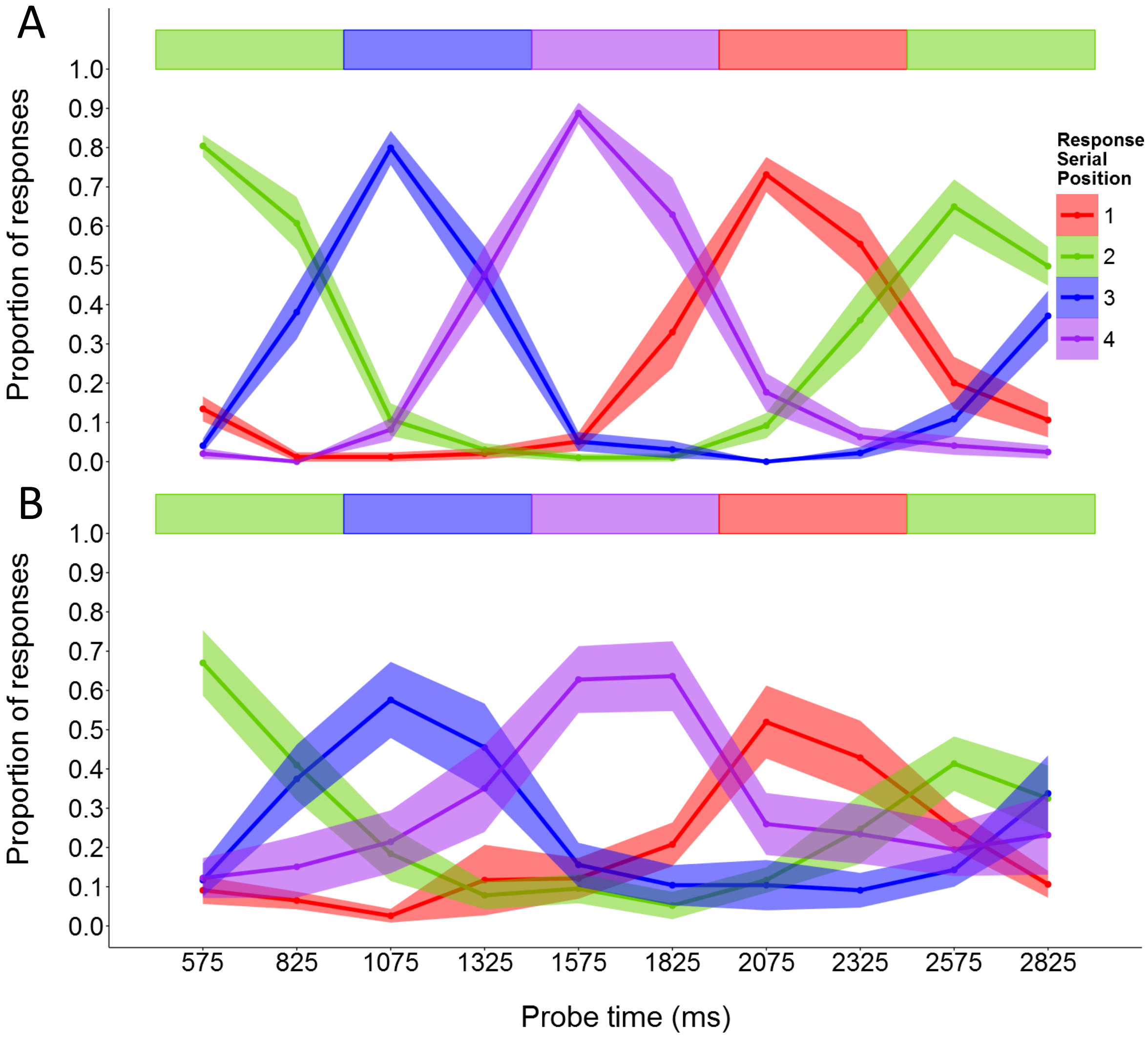

Figure 44. Mean proportions of responses in the control (A) and SLI (B) participant groups for each serial position across probe times. The four colours denote serial positions of the responses. The widths of the ribbons represent the standard error of the means. Boxes above the plots show the ‘correct’ digit timing, i.e. the timing of subvocal rehearsal if the sequences were rehearsed exactly as presented.

CSDs were calculated across valid responses from each participant using the method described in Chapter 2. Means and standard errors for the CSD data are shown in Table 21 according to the two participant groups, four serial positions, and 8 group-by-position combinations. On average, the mean CSD scores, as well as the between-participant variance in these scores, were higher in the SLI group compared to the control group. When averaged over serial positions across all participants, CSDs were highest for the 4th serial position and lowest for the 2nd serial position. This pattern is in contrast to most of the results from previous experiments where the highest CSDs tended to be in mid-list positions and lowest CSDs were usually associated with the positions at the start and/or end of the sequence.

Table 21. Mean CSDs (standard error in parentheses) for the 2 participant groups, 4 serial positions, and 8 group-by-position combinations. The maximum possible CSD value is 1.41, and the minimum possible mean CSD value is 0.39 in this experiment.

	Serial Position				
Group	1	2	3	4	Mean
SLI	0.81 (0.07)	0.78 (0.08)	0.90 (0.08)	0.93 (0.10)	0.84 (0.07)
Control	0.73 (0.04)	0.61 (0.03)	0.67 (0.03)	0.77 (0.04)	0.69 (0.02)
Mean	0.77 (0.04)	0.68 (0.04)	0.77 (0.04)	0.84 (0.05)	

Unlike the previous experiments with adult participants, in this case it is likely that the children's age differences account for some of the variance in task performance, as ISR performance is known to increase progressively with age up to adulthood. Therefore the data were examined using a mixed ANCOVA, with age (in months) as a covariate, group as a between-participant factor (2 levels; SLI or control), and serial position as a within-participant factor (4 levels; positions 1-4). The CSD distributions for each serial position were normally distributed both within and across the two participant groups. The data met the assumption of independence of the age and participant group (i.e. covariate and independent variable), as revealed by a non-significant independent-samples *t*-test (see Table 19). The data also met the assumptions of linearity, no collinearity, homoscedasticity, normality of residuals, and absence of any influential cases. For the within-participant factor, serial position, the data met the assumption of sphericity (Mauchly's test, $p > .05$). The assumption of homogeneity of variances was violated (Levene test $p < .05$ for differences in between-group variances across CSDs for serial positions 2-4), due to larger between-subject variance in the SLI group compared to the control group. However, the ANCOVA is robust to violations of homogeneity of variance, and the Levene test is not reliable with small samples and unequal group sizes (Field, 2013; Zimmerman, 2004). Therefore the ANCOVA was carried out on the data.

The covariate, age, was not significantly related to mean CSD scores, $F(1,22) = 0.44$, $p = .512$. After adjusting for individual differences in age, mean CSD scores in the SLI group were significantly higher than those in the control group[15], $F(1,22) = 6.60$, $p = .017$ (adjusted

---

[15] Due to the fact that Levene test of equal variances was significant for the distributions of CSD scores between the two groups, this result was followed up with a more conservative Welch two-sample *t*-test of mean difference, where between-group variances are not assumed to be equal. The Welch test also showed that the group difference in mean CSD scores is significant, $t(12.03) = -2.34$, $p = .037$, although this is without adjusting for age differences. Unfortunately it is not possible to control the

$\text{mean}_{\text{difference}}$ = 0.16, 95% $\text{C.I.}_{\text{difference}}$ [.03, .29]). There was no significant main effect of serial position after adjusting for age differences, $F(3,66) = 0.81$, $p = .491$, and there was no significant interaction between serial position and group, $F(3,66) = 1.09$, $p = .359$. With regard to the assumption of homogeneity of regression slopes, there was no significant interaction between the age and serial position ($F(3,66) = 0.70$, $p = .558$), however there was a significant interaction between age and group, ($F(1,21) = 5.32$, $p = .031$)[16].

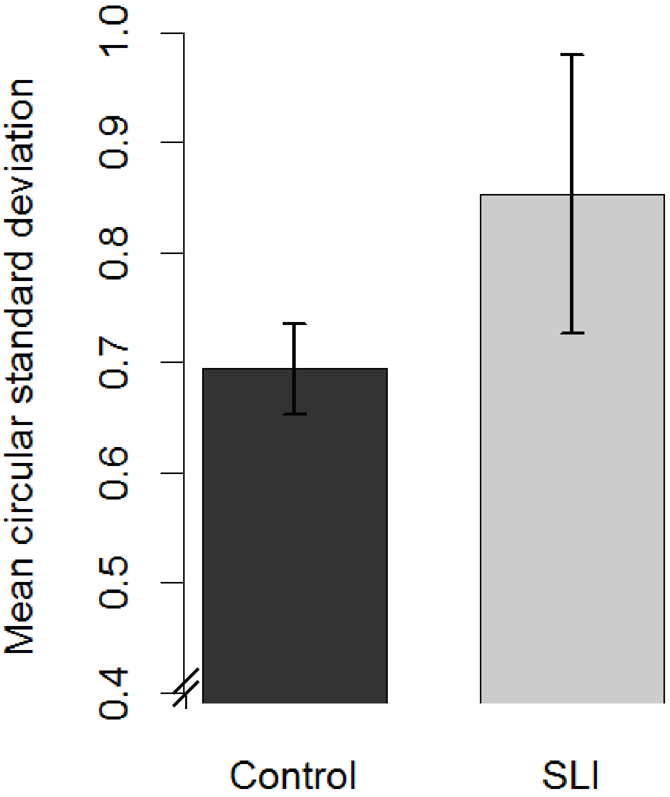

Figure 45. Mean CSDs for the control and SLI groups. Error bars show 95% confidence intervals, without accounting for the additional variance attributable to age differences within groups. The y-axis minimum is adjusted to reflect the lowest possible mean CSD score attainable in this experiment.

The correlation between auditory digit recall and mean CSD was of particular interest given the relationships between these variables observed in previous experiments with adult participants.. Across all participants, the correlation between mean CSD and digit recall was significant, $r = -.47$, $p = .019$. It was also necessary to test for correlations within each group

---

effects of a covariate using the Welch test, but this result is nonetheless informative in that it shows that the main effect of group in the ANCOVA is not due to the between-group difference in mean CSD variances.

[16] This interaction indicates a violation of homogeneity of regression slopes. The interaction between age and group occurred due to a positive regression line slope within the control group and a negative slope in the SLI group. However, because there was no main effect of age on CSD scores and the correlations within groups were not significant, it is unlikely that any differential effects of age on mean CSD scores would have a meaningful effect on the overall ANCOVA results.

in order to determine whether the correlation across all participants was a result of the group differences on both measures. The correlation between mean CSD and digit recall in the control group was moderately negative, *r* = -.50, *p* = .068 (black circles in Figure 46). In the SLI group, the correlation was much weaker, *r* = -.10, *p* = .782 (gray triangles in Figure 46), reflecting the considerable variance in mean CSD scores that is not accounted for by digit recall scores across participants in this group .

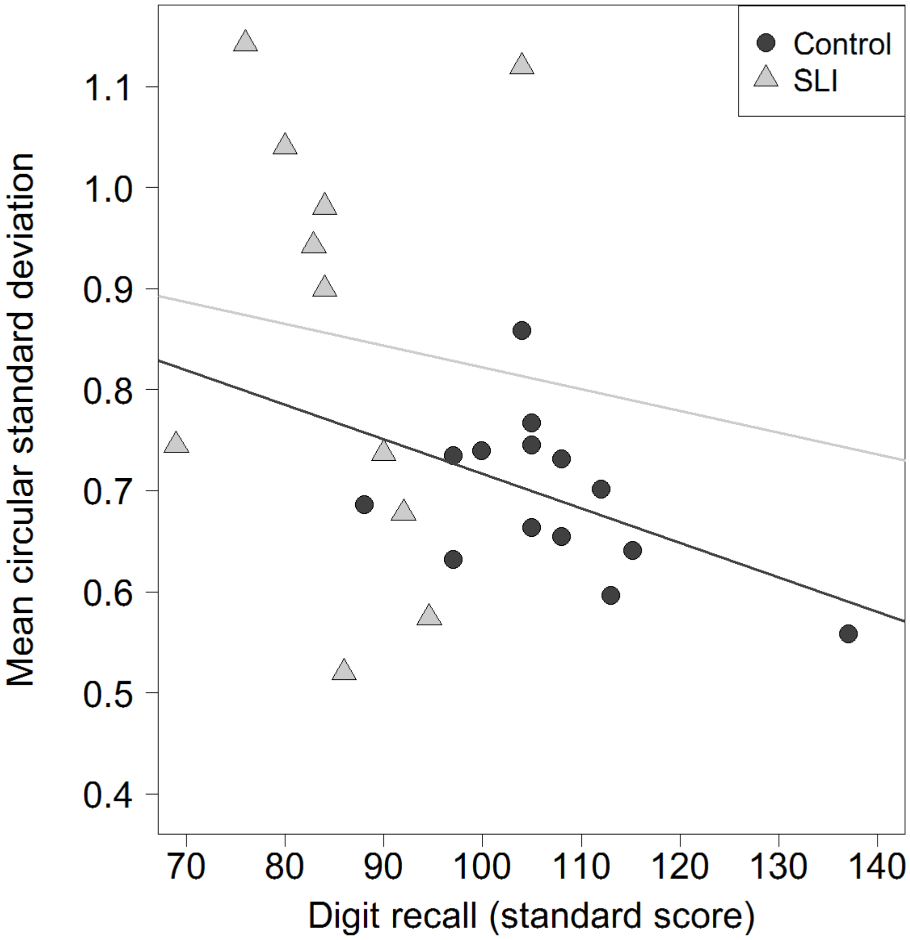

Figure 46. Scatterplot of digit recall and mean CSD scores across participants in the control (black circles) and SLI (gray triangles) groups. Solid black line represents the least squares line for mean CSD predicted by digit recall in the control group, and the solid gray line shows the same line in the SLI group. The y-axis minimum reflects the lowest mean CSD score attainable in this experiment.

Because of the potential influence of development on mean CSD and digit recall scores, partial correlations controlling for age were conducted. Across all participants, the partial correlation between mean CSD and digit recall was significant after controlling for age, *r* = -.47, *p* = .021. Within the control group, the partial correlation between mean CSD and digit recall was again similar to that across all participants, though the relationship was not statistically significant, *r* = -.47, *p* = .075. Within the SLI group, the partial correlation coefficient was non-significant, *r* = .13, *p* = .701. Relationships between mean CSD and the other available cognitive and language measures were also examined, again using partial correlations to control for age differences (see Table 22).

Table 22. Partial correlations (*r*, df = 22) between mean CSD and other measures, controlling for individual differences in age (months), n = 25. Statistical significance evaluated after Bonferroni correction for multiple comparisons.

Measures	Correlation with mean CSD	Measures	Correlation with mean CSD
CELF Sentence Recall (L)	-.56	Mr. X (vsWM)	-.26
CELF Comprehension (L)	-.02	Spatial Recall (vsWM)	-.38
PPVT (L)	-.41	WASI Vocabulary (vIQ-s)	-.31
WIAT Word Reading (L)	-.23	WASI Similarities (vIQ-s)	-.45
WIAT Comprehension (L)	-.27	WASI Verbal IQ (vIQ)	-.42
CTNP (pSTM)	-.48	WASI Block Design (nvIQ-s)	-.18
Digit Recall (vSTM)	-.47	WASI Matrix Reasoning (nvIQ-s)	-.55
Word Recall (vSTM)	-.56	WASI Performance IQ (nvIQ)	-.49*
Listening Recall (vWM)	.03	WASI Full IQ	-.53
Backwards digit recall (vWM)	-.41	WIAT Numerical (maths)	-.39
Dot Matrix (vsSTM)	.11	AV STM Composite	-.56
Block Recall (vsSTM)	-.18	VS STM Composite	-.06

Note: no correlations reached statistical significance using $p < .002$ for Bonferroni correction for multiple comparisons.
Abbreviations: CELF, Clinical Evaluation of Language Fundamentals screening test; PPVT, Peabody Picture Vocabulary Test; WIAT, Wechsler Individual Achievement Test; CTNP, Children's Test of Nonword Repetition. WASI, Wechsler Abbreviated Scale of Intelligence; WIAT, Wechsler Individual Achievement Test; AV, auditory-verbal; VS, visuospatial; STM, short-term memory.

In addition to the performance measures from tasks administered by the MRC CBSU researchers (listed in Table 19 and Table 22), there are two additional composite measures in Table 22 which were computed to index auditory-verbal (AV) STM and visuospatial (VS) STM performance. These composite measures were calculated by averaging individual standard scores on the two AV STM tasks (Digit Recall, Word Recall) and the two VS STM tasks (Dot Matrix, Block Recall) in order to produce a more robust summary measure of these cognitive factors for use in a regression analysis. Consistent with the results of tests of group differences for performance on individual tasks (see Table 19) and the expected cognitive profiles of children with SLI, independent samples *t*-tests showed that the control group AV STM composite scores (M = 105.45, SD = 9.49) were significantly higher than those of the SLI group (M = 86.53, SD = 9.20), $t(23) = 5.02$, $p < .001$. There was no significant difference in VS STM composite scores between the two groups ($M_{Con} = 92.50$, $SD_{Con} = 7.98$, $M_{SLI} = 92.45$, $SD_{SLI} = 7.15$; $t(23) = .02$, $p = .986$).

A regression analysis was used to determine the unique relationships between timing variability, auditory-verbal STM, and non-verbal IQ, while controlling for age and visual STM abilities. Specifically, the aim was to determine whether non-verbal IQ and auditory-verbal STM accounted for unique variance in mean CSD scores. Regression was computed using a hierarchical method, with mean CSD as the dependent variable, age as the first predictor and visuospatial STM (VS STM Composite) as the second predictor. The third predictor variable was either auditory-verbal STM (AV STM Composite) in model 1, or non-verbal IQ (WASI Performance IQ) in model 2. In models 3 and 4, the remaining variable was added as the fourth predictor to models 1 and 2. This method allowed the examination of the amount of variance in mean CSD scores accounted for by auditory-verbal STM and non-verbal IQ (in models 1 and 2, respectively), and the additional variance accounted for by these two variables after the other is added to the model. The standardised regression coefficients (β) for each predictor variable and the overall fit of the resulting models (R, R^2, F) are shown in Table 23. In addition, the change in the amount of variance in mean CSD accounted for by the model after adding the final predictor variable is given as ΔR^2, and the statistical significance of this change is listed as F for ΔR^2.

Table 23. Results of multiple linear regression analysis of factors predicting individual differences in mean CSD scores, n = 25. Age and visuospatial STM were entered as the 1st and 2nd predictors in all models. Auditory-verbal STM and non-verbal IQ were entered separately as the 3rd predictor variables (models 1 and 2), and in alternate order as the 3rd and 4th predictors (models 3 and 4). The R-value and *F*-value refer to the multiple correlation coefficient between predictors and mean CSD, and the significance of the overall model, respectively. The *F*-value for ΔR^2 refers to the significance of the change in R^2 (variability in mean CSD accounted for by predictors) after adding the final (3rd or 4th) predictor variable, relative to the model with the same order of predictor variables in the non-final positions.

	Model 1	Model 2	Model 3	Model 4
Predictor variable	β	β	β	β
1. Age	-.11	-.13	-.13	-.13
2. VS STM	.05	.08	.14	.14
3. AV STM	-.56**		-.48*	
4. NV IQ			-.40*	
3. NV IQ		-.51*		-.40*
4. AV STM				-.48*
R	.57	.50	.68	.68
R^2	.32	.25	.46	.46
F †	3.33*	2.34	4.32*	4.32*
ΔR^2	.31	.24	.14	.21
F ‡ for ΔR^2	9.49**	6.57*	5.26*	7.94*

Note: *$p < .05$, **$p < .01$. Abbreviations: VS STM, visuospatial short-term memory (composite); NV IQ, non-verbal IQ (WASI Performance IQ); AV STM, auditory-verbal short-term memory (composite).
† Degrees of freedom = (3,21) for models 1 and 2, and (4,20) for models 3 and 4.
‡ Degrees of freedom = (1,21) for models 1 and 2, and (1,20) for models 3 and 4.

Across all models, the assumptions of multiple collinearity, influential cases, independence of errors, homoscedasticity, and linearity were tested. None of the models violated any assumptions of regression (Field, 2013). For model 1, the statistics used for the evaluation of the assumptions were as follows – multiple collinearity: max VIF = 1.04, average VIF = 1.03, minimum tolerance = 0.96; influential cases: two cases (8%) with standardised residuals $2 < |z| < 2.5$, all Cook's Distance values < 1; independence of errors: Durbin Watson = 1.47. In model 2, max VIF = 1.09, average VIF = 1.06, minimum tolerance = 0.92, no cases with a standardised residual $|z| < 1.94$, all Cook's Distance values < 1, Durbin Watson = 0.99. In models 3 and 4, max VIF = 1.14, average VIF = 1.09, minimum tolerance = 0.88, all standardised residuals $|z| < 1.67$, all Cook's Distance values < 1, Durbin Watson = 1.57. Homoscedasticity and linearity were evaluated based on the scatterplot of standardised predicted values versus standardised residual values.

As can be seen in Table 23, the significant β values in models 1 and 2 show that auditory-verbal STM and non-verbal IQ both predict variance in mean CSD scores after accounting for individual differences in age and visuospatial STM. When auditory-verbal STM was entered as the 4th predictor variable after non-verbal IQ, there was a larger increase in the variance explained than when non-verbal IQ was entered as the 4th predictor. This is shown by the larger change in R^2 and the significant F-value for this change after adding auditory-verbal STM as the 4th predictor (model 4) compared to adding non-verbal IQ (model 3). When all four predictor variables are entered into the model, the variance in mean CSD accounted for by non-verbal IQ and auditory-verbal STM are both significant, as shown by the significant β values for these predictor variables in models 3 and 4. Therefore, auditory-verbal STM and non-verbal IQ are each unique predictors of individual differences in mean CSD scores after controlling for the shared variance accounted for by age and visuospatial STM, and all together these variables predict about 46% of the variance in timing variability.

To explore whether the relationship between mean CSD and non-verbal IQ was similar or different between the two experimental groups, partial correlations (controlling for age) were calculated separately for the two groups. There was evidence of an association between non-verbal IQ and mean CSD in the SLI group, $r = -.78$, $p < .001$, but not in the control group, $r = .14$, $p = .643$ (Figure 47). This pattern of correlations is opposite to that between mean CSD and digit recall, where there was stronger evidence for an association in the control group and no association in SLI group. These partial correlations should be interpreted with caution given the substantial loss of power that comes with computing correlations in small samples. However, the potentially differing pattern of correlations between groups is nonetheless interesting to note. These preliminary results may be signs of qualitative differences between the two groups in terms of relationships between temporal precision, STM for serial order and non-verbal IQ.

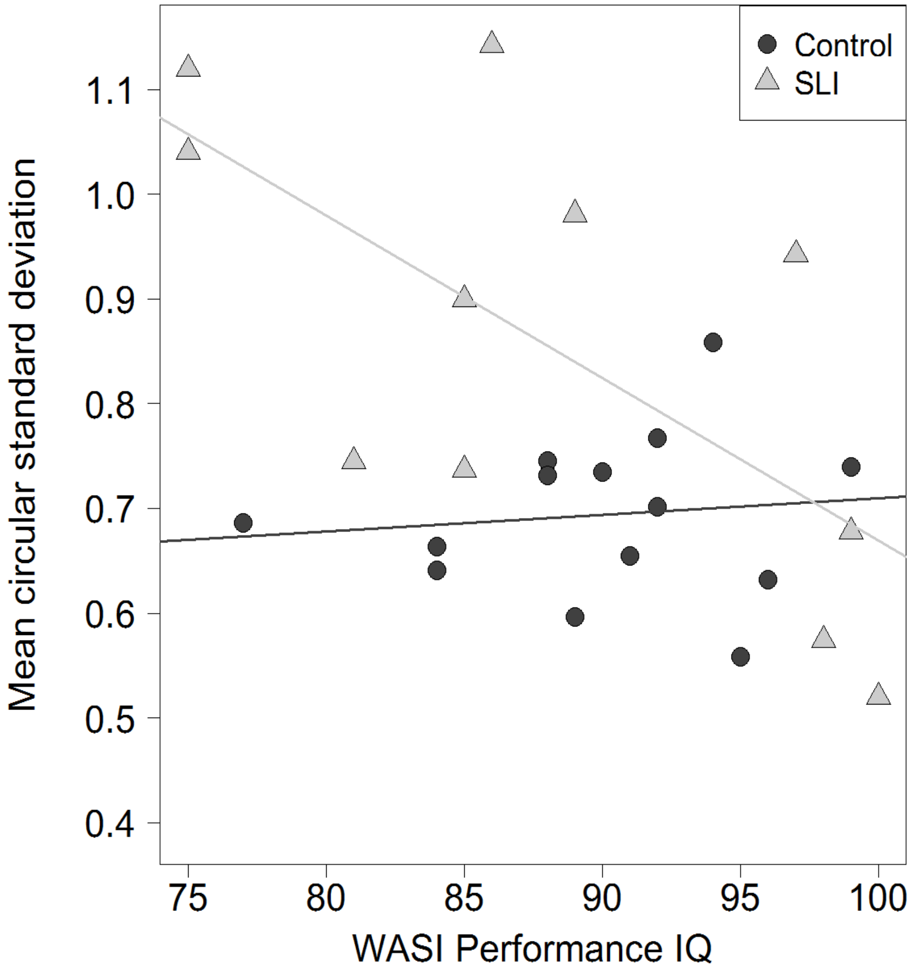

Figure 47. Scatterplot of non-verbal IQ and mean CSD in the control group (black circles) and SLI group (gray triangles). Solid black line represents the least squares regression line in the control group, and the solid gray line shows the regression line in the SLI group. The y-axis minima reflect the lowest mean CSD score attainable in this experiment.

5.2.3 Discussion

There were two primary aims of the present study; first, to determine whether the rehearsal-probe task can be successfully conducted with child participants, including children with language difficulties, and second, to explore whether the rehearsal-probe task is sensitive to differences in language development and auditory-verbal STM abilities among these children. With regard to the first aim, the experiment has demonstrated that 8-11 year old children can perform the rehearsal-probe task as expected. A large majority of the children in the SLI group (11/12 children, 91.7%) and all participants in the control group (14/14 children, 100.0%) produced data that was consistent with an understanding of the task instructions. One child in the SLI group was excluded due to responding to nearly all trials with the same serial position, which would indicate either very slow rehearsal or, more likely, a misunderstanding of the task instructions. It may be advisable to take this into consideration for future experiments with children, for example by increasing the period of training and practice before the task, reducing the rehearsal list length, or increasing the age of participants. On average, however, the response proportion functions for the two groups were cyclical and systematic, showing evidence of generally accurate maintenance of the sequence presentation rate. In addition, the correct serial order of the modal probe times associated with each serial position response distribution, as well as the low frequency of invalid responses, provides strong evidence that the children maintained the correct serial order and item information when performing the rehearsal-probe task. Finally, despite the higher and more variable mean CSD scores in the SLI group, there was no evidence of floor or ceiling effects in either group.

Within the second aim, there were two hypotheses. The first prediction was that children with SLI would, on average, show poorer timing precision compared to a group of age- and IQ-matched children with typical language development. Consistent with this hypothesis, the SLI group showed significantly greater temporal imprecision compared to the control group after accounting for the age differences across participants. The mean CSDs in the SLI group were larger on average, indicating poorer temporal precision, and there was more between-participant variability in these scores in the SLI group compared to the controls. This result is visible in the serial position response functions (Figure 44), where the averaged response distributions in the SLI group are wider with flatter peaks, and the standard error ribbons are wider. Given the interaction between age and group, more research is needed to examine the possible difference in the developmental trajectories of STM for fine-grained timing in children with and without language impairments.

The second hypothesis was that, consistent with previous investigations with adult participants, there would be a significant correlation between temporal variability and auditory-verbal ISR capacity. In addition, it was expected that timing variability may also be closely associated with non-verbal IQ, but that correlations between timing variability and auditory-verbal ISR performance would not be entirely mediated by variance shared with non-verbal IQ. The results from this experiment revealed that, as hypothesized, there was a significant negative association between digit recall and mean CSD across all participants. The regression analysis also showed that a substantial portion of the variance in mean CSD scores (46%) across all participants is explained by a linear model with age, visuospatial STM, auditory-verbal STM and non-verbal IQ as predictors. Non-verbal IQ and auditory-verbal STM each predicted a significant amount of unique variance in individual mean CSD scores after accounting for age and visuospatial STM. It therefore seems that both auditory-verbal ISR capacity and non-verbal IQ are closely and separately linked to temporal precision in STM for auditory-verbal sequences, and that these relationships cannot be explained by age-dependent development of general cognitive skills, or by the cognitive processes shared with visuospatial STM or WM tasks, such as general attention and executive function.

In addition to the specific hypothesis regarding timing precision and ISR performance, the pattern of correlations between mean CSD and performance on other tasks is informative in that it provides construct validity for the mean CSD measure. As discussed in Chapters 1 and 2, mean CSD was intended to be a proxy for the timing precision of internal sequence representations, and temporal precision was hypothesized to be specifically relevant to the maintenance of serial order information. Because auditory information is necessarily presented sequentially, it was expected that temporal precision in STM would be especially relevant to tasks where stimuli is presented aurally, and that this relationship would be strongest for simple auditory STM tasks with relatively little additional demands on executive function. Across all participants in this experiment, mean CSD was negatively associated with performance on tasks requiring STM for auditory-verbal or phonological sequences (digit, word and sentence recall, nonword repetition), as well as non-verbal intelligence tasks, though these relationships were not statistically significant after correcting for multiple comparisons. Mean CSD was also correlated with vocabulary measures (Peabody Picture Vocabulary Test, WASI Similarities), which suggests that timing precision in auditory STM might play a role in word learning[17]. Crucially, correlation coefficients were lower between mean CSD and measures of working memory, visuospatial STM, reading, or language comprehension. Overall, the correlational results among tasks in this experiment are consistent with the notion that mean CSD indexes cognitive processes closely associated with auditory STM for serial order, and that these processes are relatively distinct from those involved in visuospatial STM and WM.

Correlations within the two participant groups suggest that there may be qualitative differences in the patterns of cognitive associations between the language disordered and typically-developing groups of children. It appears that, while both ISR capacity and non-verbal IQ show close coupling with temporal variability across participants when the two groups are combined, each participant group shows a particular association mainly with one of the two variables. In the SLI group, there may be a relationship between non-verbal IQ and mean CSD, however no such relationship is apparent in the control group. By contrast, the scatterplot of mean CSD scores and digit recall shows a stronger link in the control group, and very little association in the SLI group. It is important to note that the sample sizes used here

---

[17] There is an interesting parallel between the relationship between timing precision and vocabulary observed here, and the argument put forward by Baddeley, Gathercole and Papagno (1998) that primary purpose of phonological STM is for vocabulary acquisition. The data presented here may indicate a specific role for a timing signal in STM for novel auditory sequences that is associated with word learning in children. This issue will be revisited in Chapter 8, the General Discussion.

are small and therefore not ideal for analysis of individual differences within groups. Nonetheless, the disparate pattern of relationships among mean CSD, digit span and non-verbal IQ in SLI and control groups is striking and merits further investigation with larger samples.

Across all children, there was a correlation between full scale IQ and timing variability, with a particularly strong link between temporal precision and the non-verbal IQ subtests. This pattern of correlations differs from those among the same measures observed in young adults with and without dyslexia (Experiment 6), and in a separate sample of adults without a history of language or reading impairments (Experiment 5, Chapter 4). In Experiment 5 there was no significant correlation between non-verbal IQ and mean CSD, and the relationship between digit span and mean CSD was not mediated by shared variance with non-verbal IQ. In Experiment 6, there was also no significant correlation between non-verbal IQ and digit span, although in this case non-verbal IQ did have a mediating effect on the span-CSD correlation. One potential explanation for this discrepancy is that non-verbal IQ is more closely linked to the mechanisms underlying temporal precision during childhood than during adulthood. Also, as previously mentioned, in the present study there is some suggestion that the link between non-verbal IQ and timing variability across all participants is due in part to a stronger association within the SLI group, and that within the control group, timing variability may be more closely associated with auditory-verbal ISR capacity.

One preliminary conclusion is that auditory-verbal ISR and temporal variability are only closely associated in individuals without language difficulties. In language impaired groups, where participants are matched for non-verbal IQ but show poorer performance on auditory-verbal ISR tasks, perhaps non-verbal IQ becomes a more important predictor of timing precision. It is interesting to note that the control and SLI groups did not differ across measures of auditory-verbal WM or visuospatial STM/WM, and that these measures were not significantly correlated with timing variability. These findings preclude an initially straightforward explanation that the reason for an apparent lack of correlation between auditory-verbal ISR and timing variability in the SLI group is due to compensatory executive function and general (non-verbal) STM/WM abilities in these children. Instead, there appears to be common variance between timing precision and non-verbal IQ that cannot be accounted for by cognitive skills critical for performance in WM and visuospatial STM tasks.

The group difference in temporal precision observed here is consistent with previous reports that children with language impairments show greater inter-tap interval variability in motor timing synchronisation tasks (Corriveau & Goswami, 2009). The present results extend the findings by Corriveau and Goswami (2009) by revealing that the increased temporal variability cannot be entirely attributed to differences in motor execution timing or sensorimotor feedback, both of which can affect performance on motor synchronisation tasks. Corriveau and Goswami also found that paced tapping variability was correlated with word and nonword reading, spelling, and non-verbal IQ performance after controlling for age, and associations between motor timing precision and reading and spelling scores remained significant after accounting for shared variance with non-verbal IQ. The results of this experiment are generally consistent with the findings reported by Corriveau and Goswami. In the present study, timing variability was associated with IQ as well as phonological/verbal STM tasks, although it was not significantly correlated with word reading. Also, in this experiment, individual differences in non-verbal IQ did not mediate the relationship between auditory-verbal STM composite and mean CSD scores. Given the evidence of similar timing deficits in SLI and developmental dyslexia, the result of the present experiment provides further support for a common underlying deficit in timing and rhythm processing in the two disorders.

The results of the present study are also consistent with the previous literature showing that non-verbal IQ is related to motor-timing variability (Holm et al., 2011; Madison et al., 2009; Ullén et al., 2012). While most research on intelligence and motor timing has focused on adults without language difficulties, correlations between non-verbal IQ and motor timing variability have been demonstrated in children, and in these instances timing variability was also related to reading abilities (Waber et al., 2000; Wolff et al., 1984). However, at least one study with adolescents reported no association between non-verbal IQ and motor timing precision, and in this study timing variability was again found to be specifically related to reading skills, as well as to attention and auditory temporal processing (Tierney & Kraus, 2013b). Again, it appears there may be a complex evolution of associations among auditory rhythm and timing processing, intelligence, and language and reading skills over development. However the lack of research on these topics from a developmental perspective precludes a clear understanding of how the relationships between these factors may change with age.

As previously mentioned in this section, the major limitation of this experiment was the small sample size. The small number of participants limited the power to detect group differences as well as the ability to make inferences about the complex relationships among multiple variables within participant groups. Longitudinal studies would also be useful for examining the developmental trajectories of associations among these measures. In addition to using larger samples or longitudinal designs, future studies might benefit from including a younger language-matched control group in order to determine whether the poorer timing precision in SLI is a result of this group's more limited language skills, rather than the mechanisms underlying language deficits. However, the inclusion of a younger language-matched control group might be problematic because timing precision improves with age, especially between 4 and 7 years old (McAuley et al., 2006). Future investigations aiming to address these issues would benefit from multi-level modelling techniques to account for the complex developmental, individual and group level associations between timing variability and language skills, non-verbal IQ, and auditory-verbal memory for serial order.

In conclusion, the present experiment has shown that 8-11 year old children with and without language impairments are capable of performing an adapted version of the rehearsal-probe task. The children's data were consistent with expectations and similar to data from previous experiments with adult participants; the serial position response functions were systematic and cyclical. Both participant groups showed evidence of accurate STM for item, order and presentation rate during the rehearsal task, with the main difference between groups being the amount of dispersion in the sets of serial position responses over probe times. This difference in variability was reflected in the mean CSD dependent measure, which was sensitive to group differences in language impairment. Furthermore, auditory-verbal ISR abilities and non-verbal IQ were found to be unique predictors of individual differences in mean CSD across all children. Patterns of correlations differed within the two groups, where mean CSD was moderately negatively associated with digit recall in the control group, albeit to a non-significant degree, and negatively correlated with non-verbal IQ in the SLI group. Research in these areas would benefit from future studies with larger samples to increase power to detect individual differences, and to disentangle the time course of relationships among timing variability, serial order STM, intelligence, and language abilities.

5.3 General Discussion

Two experiments using the rehearsal-probe task presented in this Chapter addressed the extent to which STM for auditory sequence timing may be impaired in adults and children with developmental language disorders. The main findings fall into two categories; first, the utility of the novel rehearsal-probe paradigm in these populations, and second, the presence of group and individual differences in temporal variability and other cognitive and language measures. These two findings will be addressed in turn, followed by a discussion of the wider implications for theories of developmental language disorders.

With regard to the first finding, both experiments yielded systematic patterns of serial position responses. The qualitative characteristics of the response proportion functions observed in previous experiments were replicated here with two different populations. The results of the rehearsal-probe task do not appear to be influenced by difficulties maintaining the item and order information during rehearsal, given that serial position response proportions peaks occurred in the correct order, and that extra-list responses were very infrequent. The serial position response patterns were systematic and cyclical, showing that the 8-11 year old children in particular had the meta-cognitive skills necessary for identifying the contents of their own inner speech. In general there was no evidence that the participants had trouble understanding and performing the task, with the exception of one child in the SLI group. The rehearsal-probe methods might benefit from further adaptations for use with younger children and individuals with severe language comprehension problems. But, in general, the results of the two experiments presented in this Chapter show that the rehearsal-probe task can be used successfully with children as young as 8 years old, and with individuals who have poorer auditory-verbal STM and developmental language disorders.

The second major finding was that there are significant group and individual differences in temporal precision. On average, both the adults with dyslexia and the children with SLI responded more variably over probe times compared to their age- and IQ-matched peers. The differences between groups are immediately visible from the response proportion functions, and these wider response proportion curves are reflected in the significantly higher mean CSD scores for both groups with developmental language disorders. The timing of response proportion peaks for each item suggest that the rehearsal rates were in time with, or slightly slower than, the sequence presentation rate, and the rehearsal rates did not appear to differ between groups. The mean response proportion curves in the language disorder groups were noticeably wider than those of the control groups, and this difference is reflected by the higher mean CSD scores in the dyslexia and SLI groups compared to controls.

In addition to group differences in temporal variability, the results of Experiments 6 and 7 reveal an interesting pattern of relationships among timing precision, auditory-verbal STM and non-verbal intelligence. Consistent with the results of Experiments 2-5, there were significant negative correlations between auditory-verbal ISR performance and timing variability, although the correlation coefficients here were somewhat weaker than those found with adults without a history of language or reading problems (cf. Chapter 4, meta-analysis). Non-verbal IQ was a strong predictor of individual differences in mean CSD performance in children, whereas there were only weak, non-significant correlations between these variables in adults (Experiments 5 and 6). Relationships between ISR performance and timing variability appeared to be mediated by non-verbal IQ in the sample of adults with and without dyslexia, however this was not true for children with and without SLI.

The patterns of correlations among variables were consistent with predictions and provide evidence of construct validity for mean CSD as a dependent variable. Mean CSD tended to be most closely associated with auditory-verbal STM (Experiments 5, 6 and 7) and phonological STM (Experiment 7) compared to other memory and language measures. Importantly, timing variability was not related to performance on all cognitive tasks. Individual differences in mean CSD were only weakly and non-significantly associated with performance on tasks that do not involve an auditory timing or serial ordering component, such as visuospatial STM (Experiments 5 and 7) and WM tasks (Experiment 7). The specificity of these correlations is important in order to establish that mean CSD indexes timing variability in the maintenance of auditory-verbal sequences, rather than primarily reflecting individual differences in cognitive processes common to other tasks such as sustained attention and motivation. That the measure of timing precision in the rehearsal-probe task is not sensitive to top-down factors common to many tasks is consistent with evidence from Ullén and colleagues (2012) showing that the relationship between motor-timing precision and non-verbal IQ was not influenced by a cash incentive to manipulate motivation. Based on their findings, Ullén et al. argue that isochronous serial interval production is an index of relatively low-level, bottom-up temporal precision for auditory sequences. The mean CSD measure may tap into the same bottom-up auditory timing perception and rhythmic entrainment system that drives performance in auditory-motor synchronisation tasks. More research is needed, for instance with manipulations of motivation (as in Ullén 2012), to confirm that the temporal precision measure derived from the rehearsal-probe task is not heavily influenced by top-down factors.

The findings presented here contribute new evidence to the larger picture surrounding the aetiological and cognitive correspondence between SLI and dyslexia, and in particular regarding similarly impaired perception of auditory timing and maintenance of serial order. The response patterns in the SLI and dyslexia groups are qualitatively comparable, and the quantitative analysis also shows a general correspondence in rehearsal-probe performance in the two disorders. These results point to a potentially causal underlying deficit related to the processing of auditory rhythm and timing information that is common to the development of both oral language and reading problems.

There have been different causal pathways proposed that separately link aspects of auditory and speech perception, serial order STM and language and reading development. Several of these accounts are compatible with the results of Experiments 6 and 7, and with one another. For instance, the temporal sampling framework (Goswami, 2011) explains how impaired entrainment to slow (theta and delta) amplitude modulations in auditory stimuli may be causally related to poorer discrimination of syllable and syllable-stress boundaries, with knock-on effects for the integration of acoustic features within phonemes. Results from Experiments 6 and 7 provide new evidence that is consistent with the predictions of the temporal sampling framework, in particular because the p-centres of the digits in the rehearsal sequences were presented at a relatively slow rate (2 Hz) and timing precision was assessed at this same word-level timing.

In addition, evidence from the rehearsal-probe experiments fits in well with arguments about the role of rhythm and timing in speech perception and in auditory-verbal STM for serial order. Boucher (2006) presents a linguistic-based argument for the link between stress rhythms in speech and grouping in serial order STM, backed by evidence that the size (number of syllables) of prosodic units in natural speech correspond to the number of unstressed arbitrary syllables that can be reproduced from memory, and that serial recall benefits from rhythmic grouping in arbitrary syllable and digit sequences (see also Ryan, 1969b). Finally, these ideas are consistent with the dynamic attending theory and models of auditory timing perception, where it is argued that the bottom-up perception of auditory timing allows the formation of prediction and directs the focus of attention to the most salient aspects of incoming continuous auditory stimuli (Large & Jones, 1999; Schwartze, Tavano, Schröger, & Kotz, 2012). Taken together, these hypotheses appear to converge on the view that low-level perception of auditory timing is critical for the segmentation of continuous speech, and thus development of robust phonological representations, as well as for the efficiency of STM processes required for serial recall.

If differences in the perception of auditory timing and rhythm reflect an underlying cause of the disorders, it is likely to be only one of multiple causal factors. Dyslexia and SLI are known to be heterogeneous disorders, and the diagnoses are applied to individuals with a wide range of cognitive profiles who differ markedly in terms of severity and developmental trajectories. For this reason, an emerging view is that there a number of genetic and cognitive risk factors for developmental language disorders, and that individuals can experience similar language and reading problems with different aetiologies. Therefore, an impairment in auditory temporal processing is likely to be only one piece of this puzzle, and there are remaining questions about the relative importance of auditory temporal precision in the development of language disorders. Results from Experiments 6 and 7 show that timing variability does not predict all, or even most variance in auditory-verbal ISR or phonological skills, so any causal influence of entrainment to auditory timing on these abilities must exert an influence in conjunction with other factors. What is clear, however, is that characterisations of developmental language disorders that point solely to deficits within the language system (e.g. quality of, or access to, phonological representations) are not sufficient to explain the typical co-occurrence of deficits in auditory timing and serial order.

5.4 Chapter Summary

In two experiments, temporal precision in STM for auditory-verbal sequences was compared between groups of children and young adults with and without developmental language disorders. Consistent with expectations and with results from earlier studies, the patterns of response data from both experiments using the novel rehearsal-probe task were systematic and cyclical. The results from both experiments supported the hypothesis that young adults with dyslexia and children with SLI would show increased temporal variability relative to age- and IQ-matched controls. In both experiments, temporal variability was negatively correlated with auditory-verbal STM capacity for serial order, which has now been shown to be a very robust finding across experiments. Interestingly, both experiments also showed that non-verbal IQ scores share some variance with timing precision, and this association was more pronounced in the sample of children with SLI. More research is needed to determine whether individual differences in auditory-verbal ISR performance alone are capable of explaining the group differences in timing, and whether patterns of associations among timing precision, STM for serial order and intelligence differ over the course of development and between typically-developing and language-disordered groups.

CHAPTER 6

A COMPUTATIONAL MODEL OF THE REHEARSAL-PROBE DATA

The rehearsal-probe task provides a rich set of data, consisting of distributions of item response proportions that vary over time. Thus far, the mean CSD dependent measure has been used to extract the amount of variability in these serial position response distributions over time. However, CSDs of the response distributions are a rather coarse measure and may be influenced by a combination of temporal properties. Subvocal rehearsal may differ along a number of temporal dimensions, such as the overall rate of rehearsal or change in variability over time. In order to gain a clearer understanding of how STM for auditory sequence timing relates to other aspects of cognition, it would be useful to estimate the distinct temporal properties that are reflected by the rehearsal-probe data, and examine them separately. Furthermore, estimates of these properties could be used to make inferences about rehearsal timing at unprobed rehearsal times.

In this chapter, I will present a preliminary model of the rehearsal probe data that was developed with the aim of extracting the main temporal properties of the rehearsal-probe response data. I will begin by reviewing some of the relevant timing literature, where certain temporal properties of serial interval reproduction data have been successfully modelled and shown to be important contributors to overall performance. The rehearsal model is then described, followed by a demonstration of the model's ability to fit observed data across experiments. In the discussion section, I will highlight the insights that arose from modelling the data, and the limitations of the current model will be addressed along with recommendations for further development.

6.1 Properties of Sequence Timing Reproduction

Much of our current understanding of temporal reproduction is based on motor-timing tasks, where individuals are asked to move in time to a beat (referred to as 'synchronisation') and/or reproduce the temporal pattern in the absence of the auditory stimuli ('continuation'). One very evident temporal property is the overall rate, which is often operationalised as the mean of inter-response intervals (Wing & Kristofferson, 1973b). Within the motor-timing literature, however, there is typically a greater focus on variability in data analysis, which is often quantified using the coefficient of variability measure. The coefficient of variability is computed by taking the standard deviation of inter-tap intervals and dividing by the mean interval duration (Hiscock, Cheesman, Inch, Chipuer, & Graff, 1989; Kee, Morris, Bathurst, & Hellige, 1986). The variance of inter-response intervals is scaled by the mean interval because response variability is known to increase with longer target intervals (Wing & Kristofferson, 1973b). This aspect of timing reproduction performance can be separated from other temporal characteristics, such as the rate. The coefficient of variability is widely used as a measure of timing precision (e.g. Hiscock et al., 1989; Madison, Karampela, Ullén, & Holm, 2013; Schwartze, Keller, Patel, & Kotz, 2011) as it allows comparison of timing precision across different tapping rates[18].

In addition to the coefficient of variability, there is sometimes a further distinction made between 'local variability' and 'drift' in motor synchronisation and continuation tasks. Local variability is computed as the squared sums of differences between each interval and that interval plus two, then scaled to the number of intervals and the mean interval (Holm, Ullén, & Madison, 2013). The differences are computed between temporally proximal but non-consecutive intervals in order to avoid the negative correlation between consecutive intervals, which is likely due to self-correction in the subsequent interval when an error is made. Because the local variability measure is the sum of these temporally proximal deviations, gradual changes in interval durations are not reflected in this measure. Instead, gradual changes are measured using the drift component, which is computed by subtracting the local variability from the total variance, where the total variance term should increase with the amount of drift that has occurred. Large differences in the inter-response intervals at the start and ends of the task would not be reflected in the local variability measure, but these would be reflected in the drift measure. As with the coefficient of variability, local variability and drift are scaled to the target interval duration (i.e. rate). Research by Madison and colleagues (Holm et al., 2013; Madison et al., 2009; Ullén et al., 2012) has shown that both local variability and drift are important and unique measures of timing performance, and that these measures relate to individual differences in intelligence.

Performance in rhythm reproduction tasks has also been described in terms of absolute and relative accuracy. For instance, Saito and Ishio (1998) and Saito (2001) computed absolute accuracy by scoring the inter-response interval durations as correct if they were within 15% of the target duration. By contrast, for the relative accuracy measure, inter-response interval durations were scored as correct if the ratio of this duration to the total reproduced sequence duration was within 15% of the actual ratio between the interval and total sequence duration. In other words, the relative accuracy measure captured instances where the relative sequence structure was preserved despite a change in the sequence reproduction rate. The two accuracy measures were shown to be closely but not perfectly

---

[18] Although, it is worth noting that comparisons of the coefficient of variability across rates assumes a linear relationship between the interval duration and inter-response interval variability, and this relationship may not apply to certain interval durations, e.g. less than 250 ms (Wing & Kristofferson, 1973b).

correlated (Saito, 2001), and they were affected differently by manipulations of dual-task and rhythmic structure conditions (Saito & Ishio, 1998). This work has shown the importance of considering rate accuracy separately from other timing performance measures.

Unfortunately, the formulae used to compute the temporal properties of motor-timing performance cannot be directly applied to the data from the rehearsal-probe task. This is because the data collected from the rehearsal-probe task are categorical item responses rather than continuous inter-response interval durations. However, similar theoretical principles about the critical aspects of timing can still be applied to the rehearsal-probe task. For instance, it is possible to conceive of a mean rehearsal rate that could deviate from the presented sequence rate, and that might vary between individuals and across situations. In addition, the motor-timing literature shows that timing can be described in terms of some static amount of error in interval production tasks (as measured using e.g. the coefficient of variability), as well as a separable source of error that builds up over time (e.g. drift). These two components of timing variability are referred to here as ‘constant’ and ‘cumulative’ error. One final temporal property relevant to the rehearsal data is a temporal offset (i.e. early start or lag), which can account for a static phase shift between rehearsal timing and the expected item timing. An offset might occur between the transition from sequence presentation to subvocal rehearsal, or as a static temporal forward jump or lag at the response stage.

Some examples of the different properties of timing as applied to the rehearsal-probe task are shown in Figure 48. The different rows of coloured boxes in the rehearsal section represent the different temporal trajectories leading to a response as a result of increases in constant error, rate and offset relative to the expected (“as presented”) rehearsal timing. Because only one data point (an item response) is collected on each trial in the rehearsal-probe task, there are a number of different temporal paths that can give rise to each response. For instance, in the examples in Figure 48, a response corresponding to the 4th sequence item (purple) at the probe time shown could result from an initial offset, where rehearsal timing lags behind the expected timing, or from increased constant error, where the temporal boundaries between items are more difficult to discern. Distinctions among these possibilities are impossible to discriminate from a single response, but taken together, the set of responses within and across probe times provides more information about the most likely underlying temporal properties.

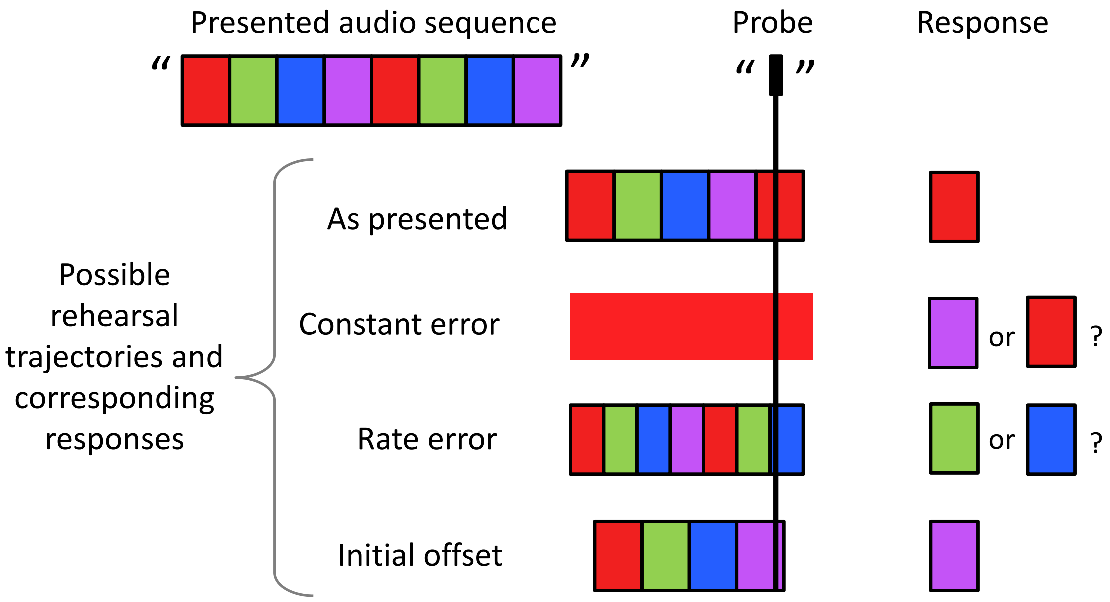

Figure 48. Schematic of different possible temporal trajectories during rehearsal and corresponding behavioural responses to the same probe. Coloured boxes represented the four sequence items. The four rows representing rehearsal timing depict the expected rehearsal timing ("as presented"), followed by the effects of increased constant error, rate and an initial offset. Coloured boxes in the 'Response' column show the response(s) that would result from each temporal trajectory at the time of the probe.

6.2 A Model of Rehearsal Timing

The model of rehearsal timing presents a method for describing and quantifying the relationship between real time and mental rehearsal time. This is accomplished by constructing a family of distributions and allowing them to be fit to the data. The set of parameters that describe these distributions are theoretically-motivated, with minimal assumptions about the mechanisms driving the separable and variable attributes of the response proportion data. There are four parameters in the model that are used to fit the data – rate, constant error, cumulative error and offset – and these will be explained in this section.

One of the simplifying assumptions of the model is that the relationship between real and mental time can be described as a line. A depiction of this relationship is shown in Figure 49. It is possible to imagine that the state of mental rehearsal at any time corresponds to a point in the figure, where the real elapsed time is shown on the x-axis, and the item being rehearsed at a given point on the x-axis is indicated by the distance along the y-axis. Because the y-axis represents rehearsal ('mental') time, the y-coordinate of any point in the figure determines the serial position of the response if the individual is probed. In an ideal situation, the rehearsal process follows linearly along the diagonal. This would mean that, for any probe time along the x-axis, the rehearsal time would always correspond perfectly to the presented sequence timing, and the response would correspond to the ‘expected’ response (i.e. given perfect rehearsal), which is indicated by the coloured item onset/offset boundaries along the x-axis.

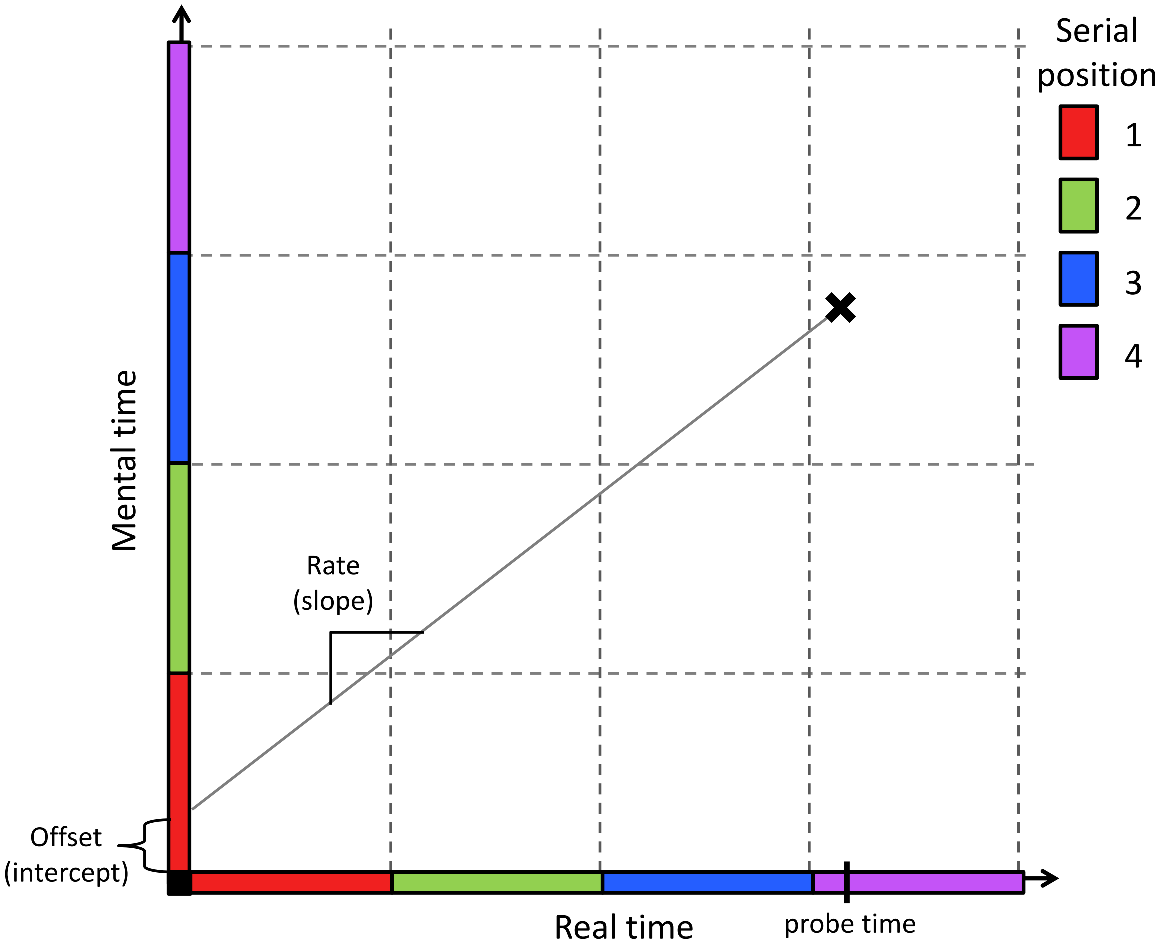

Figure 49. Schematic representation of the relationship between real time (x-axis) and ‘mental’ time (y-axis). A response is indicated by the ‘X’, where the participant responds with the item s/he is currently rehearsing (represented by coloured boxes on the y-axis), at the time of the probe shown on the y-axis. The response corresponds to the 3rd item (in blue, y-axis) while the 4th item (purple, x-axis) is the expected response. The slope of the line shows the rate of mental rehearsal relative to real time, which, in this example, is slightly too slow (slope < 1). The offset (intercept) is positive, mental time is shifted slightly ahead of real time.

Assuming a constant rate of subvocal rehearsal within participants and task situations, the set of responses across all trials is expected to be scattered near a positively-sloped line, where the slope indicates the rate of rehearsal. For instance, if rehearsal proceeds more slowly or rapidly than expected, this would correspond to a slope less than or greater than 1, respectively. However, in a given trial, and across trials in a given condition, or across participants, the relationship between real time and mental time is expected to deviate from the ‘ideal’ rehearsal situation. Thus the responses are not expected to lie perfectly along the line, but instead are expected to be dispersed around it. The dispersion is dependent on the degree of precision in the representation of timing, which was estimated using the mean CSD measure in previous experiments. By quantifying the rate of rehearsal separately from precision, the aim is to separately quantify the key properties of rehearsal timing in order to gain a more nuanced understanding of task performance.

In the model, the rehearsal *rate* is the slope of the line previously described, where a rate value of 1 indicates a perfect correspondence between presented and rehearsal rates, and values less than or greater than 1 correspond to rehearsal that is slower or faster than the presented rate, respectively. The *offset* corresponds to the positive or negative intercept in this equation, which is a static shift in mental time relative to real time. In the example linear relationship shown in Figure 49, the rehearsal rate is slightly slower than the presented rate (i.e. slope is less than 1), and there is a positive temporal offset (y-intercept is a positive value). The slower rehearsal rate results in a response for the item in the 3rd serial position at a probe time associated with the 4th serial position. This is indicated by the coordinates of the response (X), where the x-value (real time) falls within the 4th serial position on the x-axis, while the y-value falls within the 3rd serial position on the y-axis.

Given a linear relationship between mental time and real time, as described by the rate and offset parameters, there should be some timing imprecision around the mean line. The amount of noise around the mean can be thought of as a general measure of timing variability, like the coefficient of variability metric used in motor-timing tasks. In the model, this amount of variability is referred to as *constant error*, and it is represented as Gaussian noise around the mean line. There may also be a time-dependent increase in the amount of deviation from the mean rehearsal time. This is referred to as *cumulative error* in the model, and it is analogous to the drift measure in the motor-timing literature. These two error parameters can be combined to predict the overall amount of variability surrounding the mean line, as shown in Figure 50. Constant error defines a static amount of response variability around the mean line (light yellow ribbon in Figure 50), whereas cumulative error defines the amount of response variability that increases linearly over time (light blue ribbon in Figure 50). In this way, the relationship between real time and mental rehearsal time can be described by the four parameters introduced in the previous section; rate (denoted as m_r in the model), constant error (s_{con}), cumulative error (s_{cum}), and offset (m_{off}).

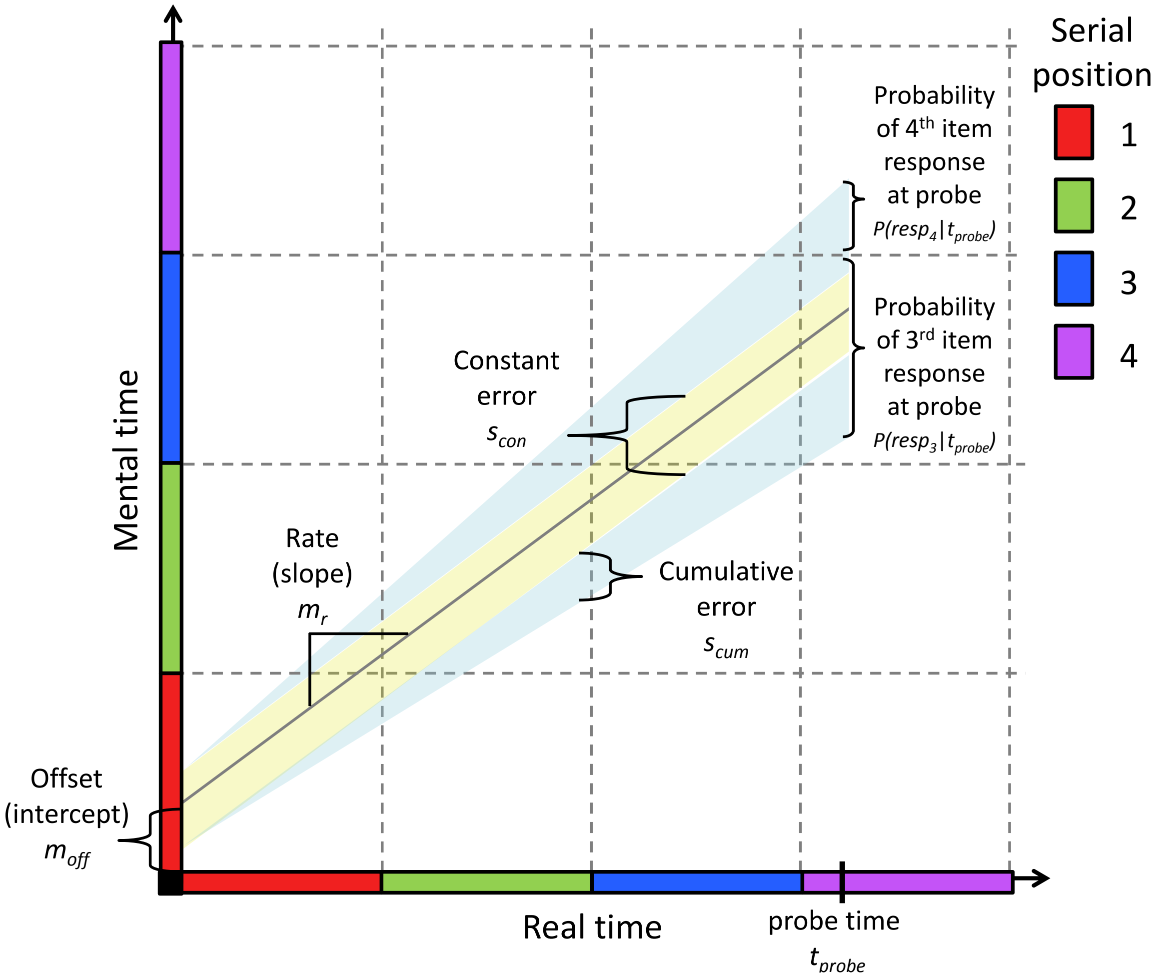

Figure 50. Depiction of how the four model parameters are used to predicted rehearsal timing in the model. Real time is shown on the x-axis, and mental time on the y-axis. At any probe time, the model parameters describe the corresponding distribution of predicted mental time (widths of ribbons around the line at the probe time), and this predicted time distribution maps on to serial position response probabilities (corresponding coloured boxes on the y-axis).

When the four parameter values are combined, the resulting line and error values form a probability distribution of mental time at any given real time, and this distribution can be translated into corresponding predictions about the number of responses for each item that would be observed at a particular probe time given the model. This is accomplished by integrating the area under the probability curve of mental time, which is defined by the mean (determined by rate and offset) and standard deviation (determined by constant and cumulative error) within the boundaries of the mental start and end times of each item. For example, in Figure 50, variability around the predicted mental time at the probe time x-value is depicted by the widths of the light yellow (constant error) and blue (cumulative error) semi-transparent ribbons above and below the y-coordinate at the centre line. Most of the predicted distribution at this probe time corresponds to the 3rd serial position in mental rehearsal (blue portion of the y-axis), meaning most of the responses are predicted to correspond to this item. This is labelled in Figure 50 as $P(resp_3|t_{probe})$ (i.e. response probability for item in the 3rd serial position at the probe time shown), and it is computed as the integral of the of the distribution bounded by the start and end times of the 3rd item on the y-axis. One tail of the predicted mental time distribution also falls into the 4th serial position in mental rehearsal (within the purple portion of the y-axis), meaning that the model predicts some corresponding proportion of responses for the 4th item at this probe time. In Figure 50, the predicted portion of responses for the 4th sequence item at the probe time is labelled as $P(resp_4|t_{probe})$. Finally, there is no portion of the distribution of mental time at this probe time that falls in to the 1st (red) or 2nd (green) serial positions in mental time, which means that this model predicts 0 response probabilities for items in the 1st and 2nd serial positions.

More formally, the equation for calculating the probability P of responding with a given item i at a given probe time $probe_t$, can be expressed as follows:

Equation 1. Probability of responding with a given item at a given probe time.

$$P(resp_i \mid t_{probe}) = \int_{start_i}^{end_i} N(m_r\, t_{probe} + m_{off}, s_{con}^2 + \; s_{cum}^2\, t_{probe})$$

where N is a normal probability density function defined by mean: $m_r\, t_{probe} + m_{off}$ and variance: $s^2_{con} + s^2_{cum}\, t_{probe}$. The variable t_{probe} corresponds to the probe time, and the integral of the probability density function is computed between $start_i$ and end_i. The latter two terms refer to the start and end times of the item i (corresponding to item start/end times shown on the y-axis in Figure 50). Considering first the normal probability density function N, this defines the estimated rehearsal ('mental') time at probe time t_{probe}. The means of the distributions result from the combination of rate and offset values, and the variances of the distributions are defined by the combination of constant and cumulative error values. For each item i and probe time t_{probe}, the integral is computed for the mental time probability density function between the bounds of the item start and end times.

Once the expected response probabilities have been computed for all items and probe times, the expected probabilities can be compared to the observed data. The model uses maximum likelihood estimation (MLE), which is a commonly used method of quantifying the likelihood of the data given the model (Lewandowsky & Farrell, 2011). To calculate the log likelihood, first the observed responses corresponding to each item at each probe time are summed, and then this value is multiplied by the natural log of the predicted probability for that item and probe time. These values are summed across all combinations of items and probe times resulting in an overall value for the discrepancy function, which is the log likelihood of the data given the model and parameter values. Thus the log likelihood of the data given the model and parameters can be expressed as:

Equation 2. Log likelihood of the data given the model and parameters.

$$L(data \mid model(m_r, s_{con}, s_{cum}, m_{off})) = \sum_{i=1}^{j}\sum_{t=1}^{k} f(o_{it})\ ln(P(resp_i \mid t_{probe}))$$

where items are indexed by subscript *i*, ranging from 1 to the number of sequence items *j*, and where probe times are indexed by subscript *t*, ranging from 1 to the number of probe times *k* . The term $f(o_{it})$ is the number of observed responses for item *i* at probe time *t*, and it is multiplied by the natural log (*ln*) of the predicted probability of responses for item *i* and probe time *t* (i.e. $P(resp_i \mid t_{probe})$) as defined in Equation 1.

When the model is fit to a set of observed response proportions, the fit is first computed with neutral starting parameter values (e.g. $m_r = 1.0$, $s_{con} = 0.1$, $s_{cum} = 0.1$, $m_{off} = 0.0$). Then, the log likelihood of the data given the model and starting parameters is computed, which is used as the measure of goodness-of-fit. The parameter values are then adjusted iteratively using the Nelder-Mead Simplex algorithm until the goodness-of-fit (log likelihood) value is maximised[19]. Within the parameter adjustment algorithm, the constant and cumulative error parameters were bounded at 0 to reflect the fact that it is not possible to have negative variance. Also, the rate value was bounded to reduce highly improbable rate values that may occur when the observed responses are highly unsystematic or close to chance[20]. Finally, once the parameter values have been determined for a given data set, Equation 1 can be used to predict the probability of item responses at any time.

Figure 51 shows an example of modelled response data predictions based on a given set of rate, constant error, cumulative error and offset values, as well as the effects of changing each parameter in isolation. As in the previous experiments with observed data, the model predictions are shown here as response proportions for each of the four sequence items over time. First of all, it is worth noting the qualitative similarities between the response proportions simulated in the left column and those observed across experiments; the

---

[19] This is implemented using the 'fminsearchbnd' Matlab function. In this implementation, the log likelihood of the data given the model is converted to a negative value, and the multidimensional minimisation function finds the set of parameters that minimise the negative log likelihood value. More details on this function and on the model implementation in general are provided in the Appendix.

[20] The model typically arrives at a best-fit solution to the individual data, but occasionally when the response proportions were close to chance or very non-cyclical/unsystematic, the model does not find a solution or it produces haywire best-fit values. The use of rate bounds from 0.05 to 1.95 was a measure of convenience to allow the model to complete the fitting process in these rare cases. In these instances, the resulting parameter values were not included in the analysis described in this Chapter.

responses for each serial position gradually rise, peak and fall in a systematic and cyclical fashion. The fact that the responses for each item are not uniform (rectangular) distributions over time reflects the response variability in the data and is represented in the model by a non-zero constant error value.

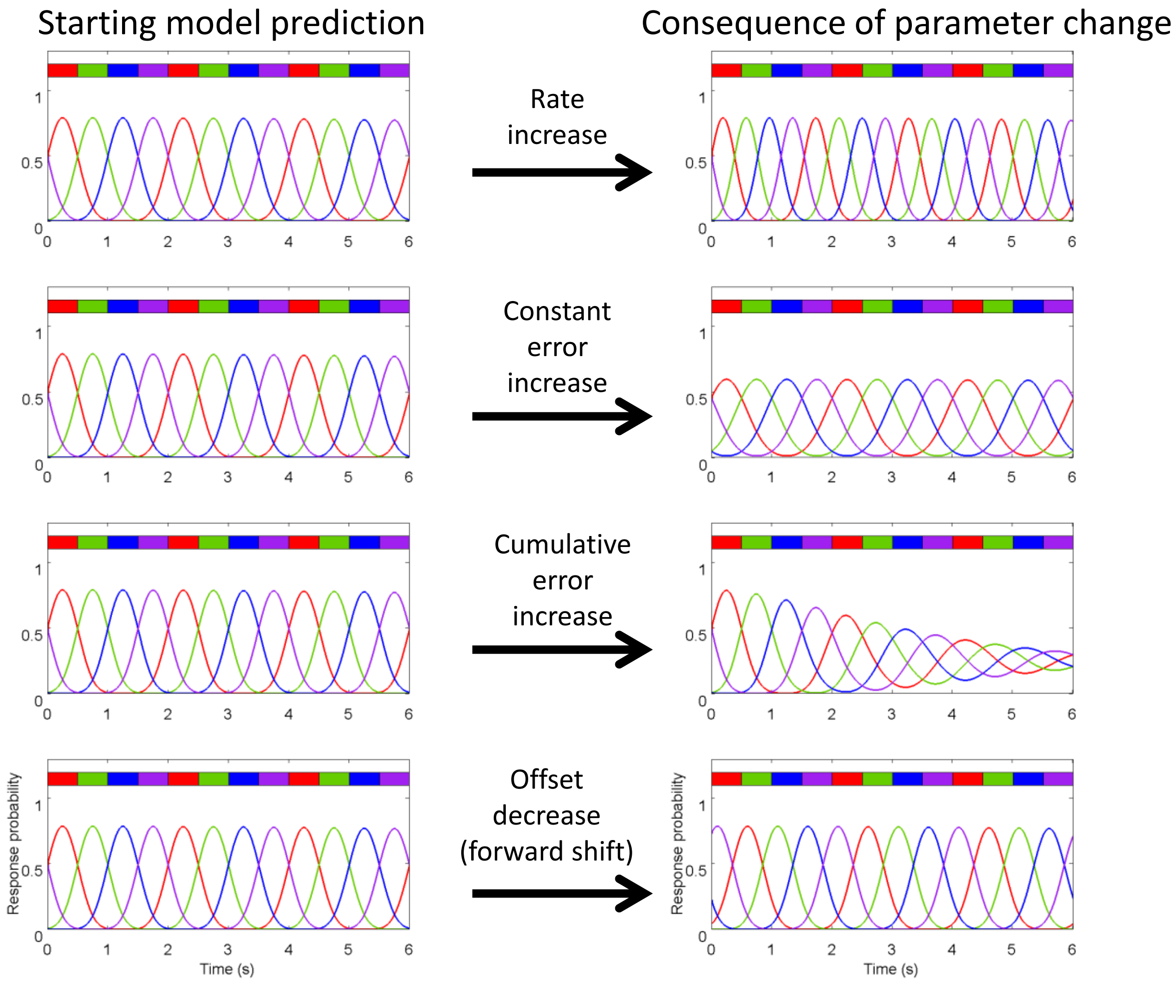

Figure 51. Examples of the effects of model parameter changes on the predicted response proportion data. As in previous plots, the x-axis represents time since the end of the sequence presentation period, the y-axis shows the proportion of responses for each item (colour) over time, and the boxes above the plots show the expected sequence timing (i.e. exactly as it was presented). A starting model prediction is given in the left column, which is the same for all rows. The right column shows the isolated consequence of a rate increase, constant error increase, cumulative error increase, or offset decrease (i.e. forward shift with respect to real time) on the model predictions.

In the right column, the effect of a change in each parameter value is shown, with the same set of parameters shown for all plots in the left column as a reference for the parameter changes. The serial position response distributions become more compressed in time when the rate is increased. When the constant error value increases, the response proportions become wider and more overlapping in time, but note that the peaks of the distributions are still aligned with the expected timings, reflecting the accurate rate. When cumulative error is increased, the response proportion distributions become wider over time. It is clear from the plot demonstrating the effect of cumulative error that the response proportions will approach the chance response rate (1/number of items) over time. In effect this means that, with non-zero cumulative error and a sufficiently long rehearsal interval, rehearsal responses cannot be predicted at a better than chance level. Finally, decreasing the offset parameter (i.e. negative y-intercept value) results in a forward shift in rehearsal time relative to real time. This would reflect, for instance, a delay between the end of the sequence period and the start of rehearsal. A positive offset, on the other hand, would result in a backward shift in the mental time relative to real time, and this might reflect an earlier than expected rehearsal start.

6.2.1 Fitting the model to the data

In order to test the ability of the model to fit real data from the rehearsal-probe task, the model was fit to the response proportions averaged over groups and conditions from Experiments 1-7. The following starting parameters were used: rate = 1.0, constant error = 0.1, cumulative error = 0.1, offset = 0.0. Once the set of best-fit parameter values was identified for each data set, the model was then used to generate predicted response probabilities for rehearsal cycles before and after that probed in the experiment. Plotting the observed response probabilities alongside the model's predictions allowed an assessment of the qualitative fit of the model to the data, as well as the model's predicted response proportions for the unprobed periods. Examples of the model fits and predictions are shown in Figure 52 (Experiment 5), Figure 53 (Experiment 6, control group) and Figure 54 (Experiment 6, dyslexia group). In each figure, the observed data is shown in the top panel, and the model's predicted response proportions and best-fit parameter values are given in the bottom panel.

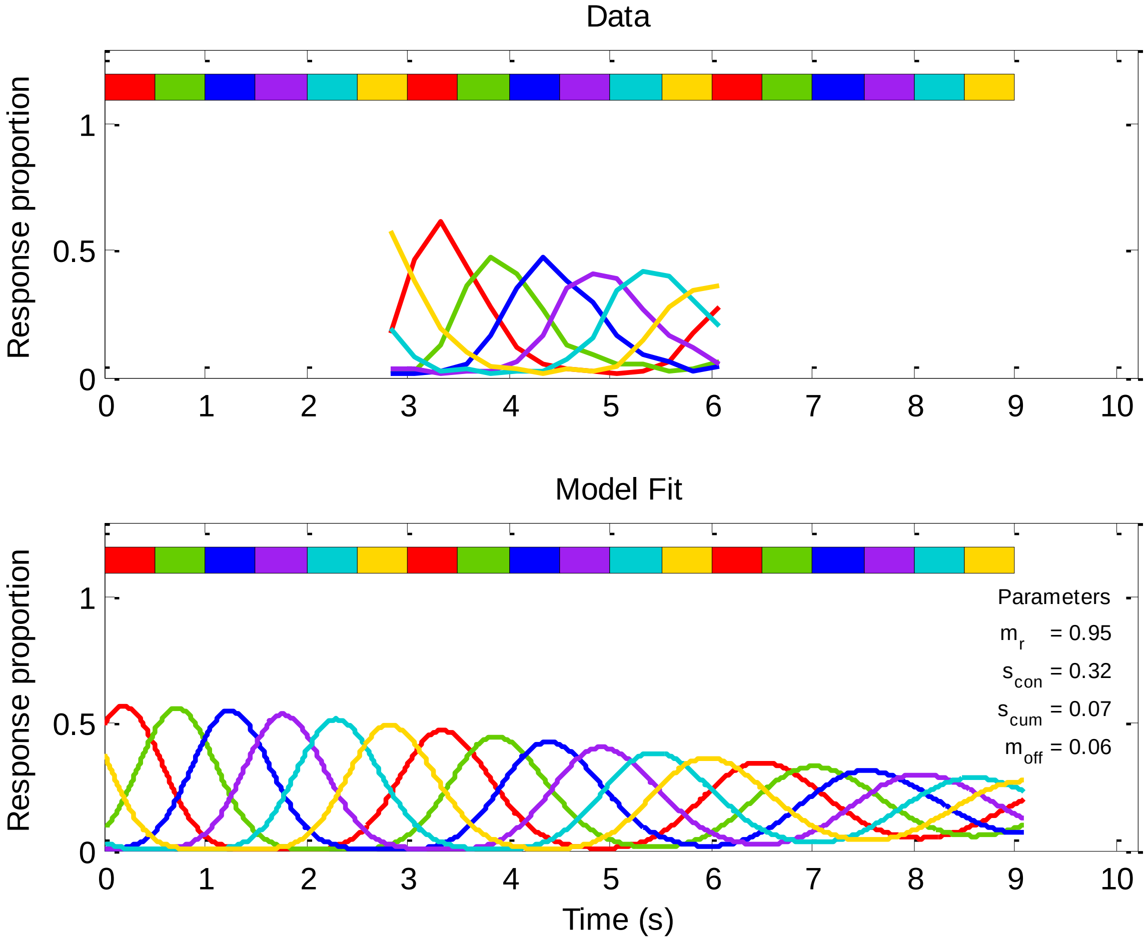

Figure 52. Experiment 5 average response proportion data from the individual differences study (6-digit rehearsal; top panel) and the model fit to this data (bottom panel). The model parameter values are given on the right side of the bottom panel; rate (m_r), cumulative error (s_{cum}), constant error (s_{con}) and offset (m_{off}). The x-axis shows the time starting at the end of the sequence presentation period (i.e. the start of the rehearsal period). Coloured boxes above the plots show the timing of the items if they are rehearsed exactly as presented.

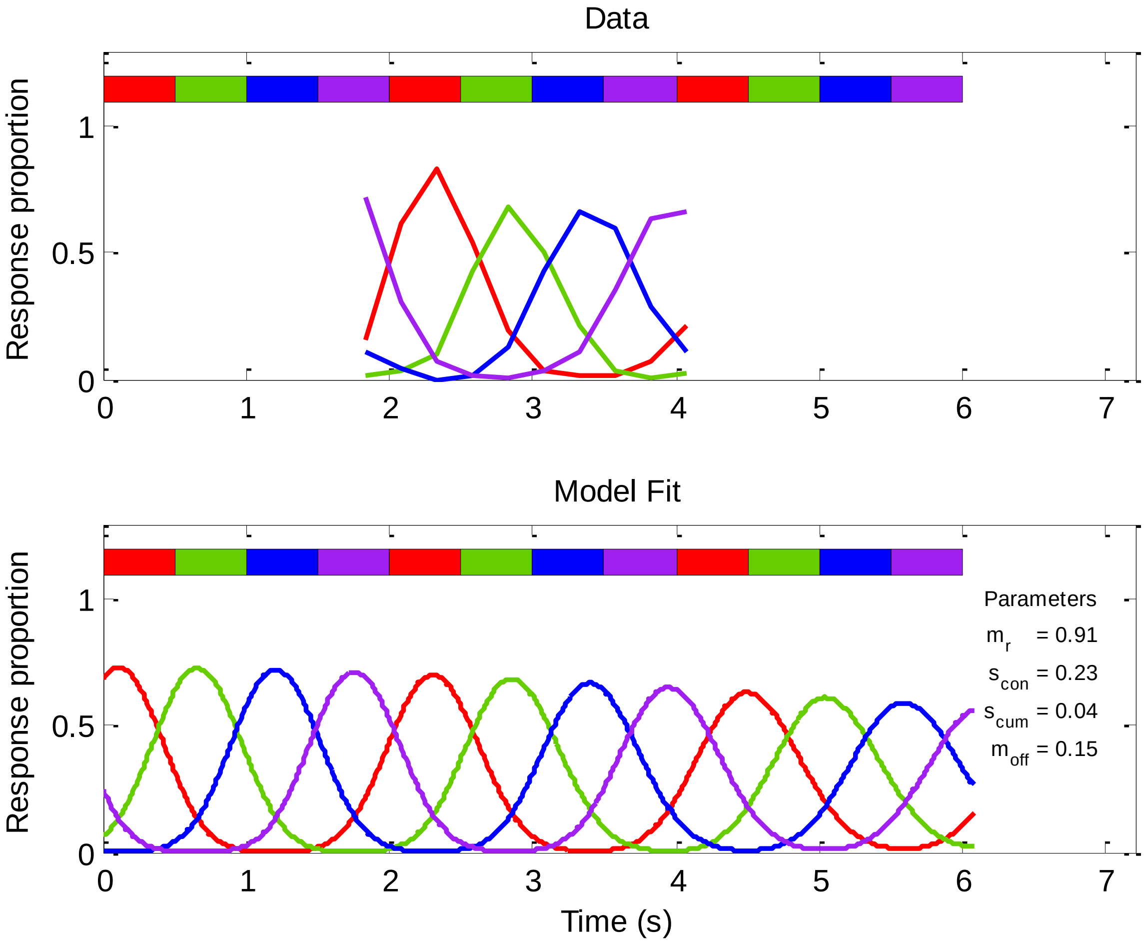

Figure 53. Experiment 6 average response proportion data from the control group (top panel) and the model fit to this data (bottom panel). The model parameter values are given on the right side of the bottom panel; rate (m_r), cumulative error (s_{cum}), constant error (s_{con}) and offset (m_{off}). The x-axis shows the time starting at the end of the sequence presentation period (i.e. the start of the rehearsal period). Coloured boxes above the plots show the timing of the items if they are rehearsed exactly as presented.

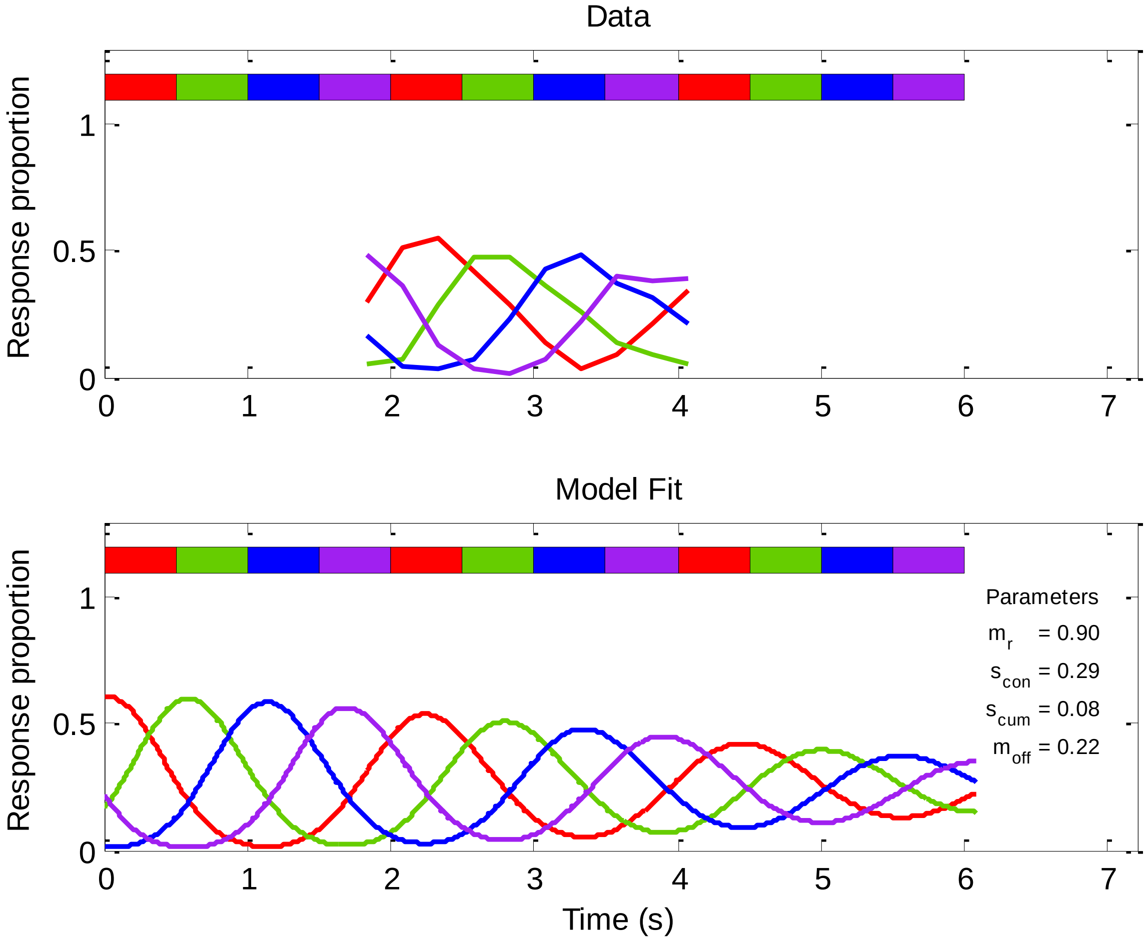

Figure 54. Experiment 6 average response proportion data from the dyslexia group (top panel) and the model fit to this data (bottom panel). The model parameter values are given on the right side of the bottom panel; rate (m_r), cumulative error (s_{cum}), constant error (s_{con}) and offset (m_{off}). The x-axis shows the time starting at the end of the sequence presentation period (i.e. the start of the rehearsal period). Coloured boxes above the plots show the timing of the items if they are rehearsed exactly as presented.

Figure 52, Figure 53, and Figure 54 show that, overall, that the model is capable of producing a very good qualitative fit to the observed data. The model's fitted response proportion distributions share most characteristics of the observed data, such as alignment of the centres of the distributions for each serial position and similar amounts of dispersion and overlap in response distributions over time. In terms of the parameters, the response variability in the data was captured by non-zero constant error and cumulative error values. Constant error tended to be higher than the starting parameter value (0.1), while cumulative error was lower than this value. In addition, all of the model estimates of the rate parameter were less than 1, reflecting average rehearsal rates that were slower than the expected (presented) rate. These measures of slower rates are reasonable given that the peaks of the response probability distributions in the observed data tended to fall behind the expected rehearsal timing. The model estimated offset values that also differed from the starting parameter value (0), showing that the model fit included a constant shift in mental time relative to real time. This might occur, for instance, because participants did not begin rehearsing exactly when they were expected to. In general, the results of fitting the model to these data sets show that the model is capable of providing parameter estimates that are reasonable and supported by the qualitative aspects of the data.

Having obtained the best-fit parameter estimates for the response proportion data from all experimental conditions and groups, it was possible to compare these values in order to examine how the timing parameter estimates did or did not change between task conditions and participant groups. The parameter values are provided in Table 24. Of particular interest in the present research was the fine-grained precision of item timing during subvocal rehearsal, that is, the consistency of word-level (i.e. p-centre) timing. The constant and/or cumulative error parameters were expected to capture this timing precision, so one important question was whether these parameter values would differ between groups and conditions in the expected directions (i.e. in the same direction as the mean CSD values). Table 24 shows an overview of the model parameter estimates for all groups/conditions within each experiment, as well as differences between these parameters. As indicated in the table, the constant error parameters tended to differed in the expected directions for the averaged response proportion data between groups and conditions, and this was sometimes also the case for cumulative error. These differences in error parameter values are consistent with the results from Chapters 3-5 using mean CSD as the dependent variable.

Table 24. Overview of the parameter values that resulted from fitting the model to averaged response proportions within groups/conditions across experiments.

Experiment	Group/Condition	Rate	Con. Error	Cum. Error	Offset
1	Low	0.97	0.16	0.11	-0.07
	High	0.89	0.29	0.07	0.22
	Difference	0.08	-0.13 †	0.04	-0.29*
2	Low	0.88	0.28	0.06	0.23
	High	0.90	0.27	0.06	0.17
	Difference	-0.02	0.01 †	<0.00	0.06
3	Low	0.84	0.16	0.07	0.38
	High	0.88	0.35	0.02	0.23
	Difference	-0.04	-0.19 †	0.05	0.15
4	Grouped	0.94	0.33	0.09	-0.08
	Isochronous	0.94	0.30	0.06	0.04
	Difference	<0.00	0.03 †	0.03 †*	-0.12*
5	-	0.95	0.32	0.07	0.06
6	Control	0.91	0.23	0.04	0.15
	Dyslexia	0.90	0.29	0.08	0.22
	Difference	0.01	-0.06 †	-0.04 †	-0.07
7	Control	0.98	0.20	0.08	0.15
	SLI	0.88	0.34	0.14	0.31
	Difference	0.10	-0.14 †	-0.06 †	-0.16

† indicates a numerical difference in constant or cumulative error values between groups/conditions that is consistent with the difference in mean CSD.

* indicates a statistically significant difference as using paired or independent-samples *t*-tests as appropriate. Where data did not meet the assumptions for these tests, Wilcoxon signed-rank or Mann-Whitney *U* non-parametric alternatives were used.

Across all experimental comparisons, the constant error value differed between groups and conditions in the expected direction (i.e. consistent with differences in mean CSD) as indicated in the 'Difference' rows in Table 24. This shows that, as was intended in the model development, the constant error parameter captures the differences in timing variability that were previously measured with the mean CSD variable. The cumulative error values, on the other hand, did not differ as expected between load groups/conditions in Experiments 1-3. Interestingly, cumulative error was higher in the dyslexic and SLI groups compared to controls, and it was higher for temporally-grouped compared to isochronous sequences. While there were no prior expectations about differences in rate and offset values as a function of the experimental manipulations, the data provided here might be nonetheless useful for generating hypotheses in future work. For instance, in two experiments (1 and 4), there were statistically significant group differences in offset, and in Experiment 7, the rehearsal rate for the SLI group was considerably lower than that in the control group. Together, these results show the potential benefit for using the model to quantify separable properties of the data.

6.2.2 Meta-analysis of model parameters with non-dyslexic adult participants

In addition to comparing parameter values between groups/conditions within experiments, another goal of the rehearsal model development was to estimate rehearsal timing parameter values at the population level. For this reason, the parameter values across experiments were incorporated into a single analysis in order to obtain more robust estimates of the typical timing properties observed more generally across the rehearsal-probe experiments. Random effects models for meta-analyses were used here for the reasons described in Chapter 4, where the same procedure was used to estimate the correlation between digit span and mean CSD across experiments. The models were implemented in R software (R Core Team, 2014) using the 'metafor' package (Viechtbauer, 2010). The goal of this meta-analysis was to generate robust estimates of the typical rehearsal rate, offset and error parameters, along with 95% confidence intervals for these values. These estimates will be useful in describing the general rehearsal timing properties within the sampled population and for the specific types of rehearsal-probe tasks administered.

In order to restrict the parameter value estimation to a single population, only the experiments with adult non-dyslexic participants were included in the meta-analysis of model parameter fit values. These criteria excluded the data from Experiment 7, where children were tested, as well as the data from the adult dyslexic group from Experiment 6. Because there were no statistically significant differences in parameter estimates between the two load conditions in the two experiments with within-subject designs (Experiments 2 and 3), the parameter best fit values were averaged across conditions within participants in these experiments in order to ensure independence of the data included in the meta-analysis.

The results of the random effects models for each parameter are shown in Table 25, and the forest plots for these models are shown in Figure 55. There were 6 independent samples (k) included in each model, where one of these was the control group only from Experiment 6 (N = 18). The estimated parameter values are given by $\hat{\mu}_\rho$ in the table, along with upper and lower bound 95% confidence intervals for these estimates. These estimates and confidence intervals are also shown at the bottom of the forest plot in Figure 55, in the form of a diamond to mark the random effects model estimate and precision. The $\hat{\tau}^2$ value in Table 25 shows the estimated heterogeneity, while the Q and associated p_Q values give the results of the significance test for heterogeneity.

Table 25. Results of meta-analyses of model parameter fit values from Experiments 1-6, with non-dyslexic young adult participants.

Modelled data	k	$\hat{\mu}_\rho$	C.I. Upper	C.I. Lower	$\hat{\tau}^2$	Q	p_Q
Rate	6	0.91	0.93	0.89	0.001	18.81	.002
Constant error	6	0.15	0.19	0.12	0.002	25.65	<.001
Cumulative error	6	0.04	0.04	0.05	<0.001	4.49	.481
Offset	6	0.15	0.21	0.09	0.003	13.17	.022

Note: k = number of studies included; $\hat{\mu}_\rho$ = estimated average parameter value in the population distribution; C.I. = 95% confidence interval for $\hat{\mu}_\rho$; $\hat{\tau}^2$ = estimated heterogeneity in the sample correlations; Q = Q-test for heterogeneity with k – 1 degrees of freedom; p_Q = p-value for Q.

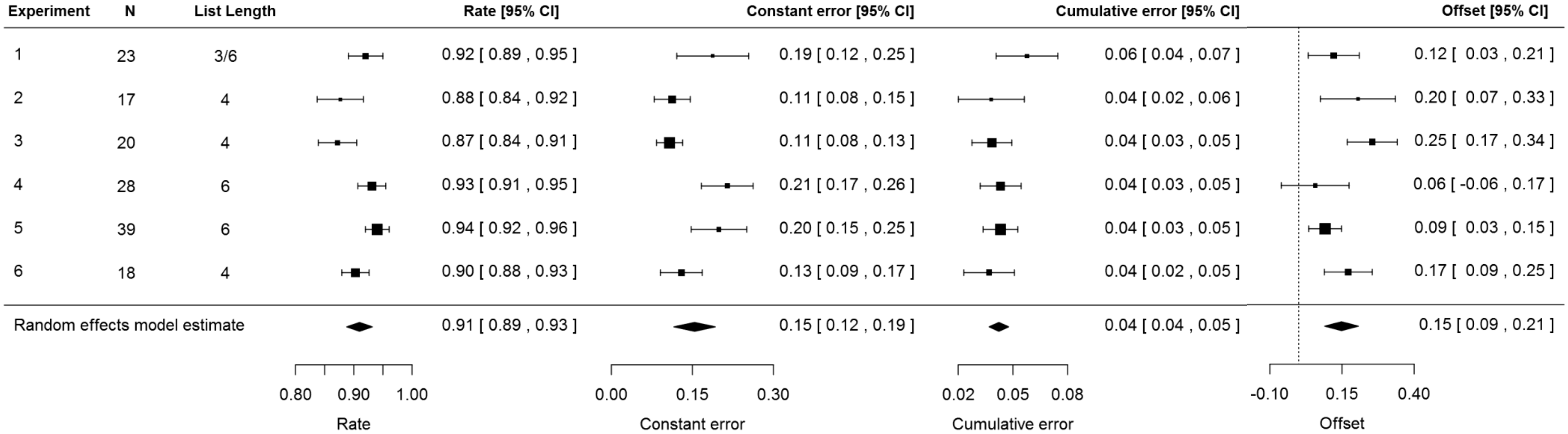

Figure 55. Forest plot of the means and 95% confidence intervals of the four model parameters obtained from fitting the data from young adults without dyslexia in Experiments 1-6. The diamonds below the experiment-level results show the mean parameter estimates using a random effects model. Point sizes for individual experiments are proportional to the precision of the estimates, which is calculated as the sampling variance divided by the number of observations.

Results of the meta-analysis showed that, across the variations on the rehearsal task paradigm used in Experiments 1-6, young adult non-dyslexic participants tended to rehearse at a slightly slower rate (~91% of the presented rate). Overall, these participants showed some amount of constant error (0.15 second standard deviation), reflecting rehearsal timing precision that was not perfect but not terribly poor either. The cumulative error value was relatively small in these participants (0.04 second standard deviation). Finally, these individuals tended to have a positive offset (0.15 seconds), meaning that their rehearsal tended to start slightly ahead of the expected start. These results provide a more generaliseable estimate of similar patterns of parameter values observed within experimental groups/conditions in the previous section.

However, the significant p-values for Q-tests for heterogeneity in Table 25 indicate that, in three out of the four random effects models, there was significant heterogeneity across the six included samples. That is, the parameter values observed across the six samples are not well-accounted for by a single model due to remaining experiment-specific differences in the parameter distributions. Upon examination of the estimates and variances within experiments (see Figure 55), it is clear that these estimates are not well-aligned in the rate, constant error and offset parameter values. Instead there is a snake-like pattern in the estimates across experiments within the rate and constant error parameters, with a similar but reversed pattern in the offset data. The heterogeneity appears to be mainly a result of higher rate and constant error values in Experiments 1, 4 and 5, and lower offset values for these same experiments. However, in the case of cumulative error, the estimates and variances are fairly consistent across experiments, which accounts for the non-significant Q-test for heterogeneity and more precise estimate (shorter width of the diamond) for these data.

Due to the significant heterogeneity in the random effects models for rate, constant error and offset parameters, a further exploratory analysis was conducted in order to determine whether the data could be better described by reducing the differences between experiments within meta-analysis models. One obvious difference between the rehearsal-probe task methods across the six experiments was in the number of items in the rehearsal lists; Experiments 2, 3 and 6 used 4-digit lists while Experiments 4 and 5 used 6-digit lists. Experiment 1 involved rehearsal of either 3-digit or 6-digit lists. It seems plausible that the 4- and 6-digit rehearsal lists might differ on important dimensions such as memory load and duration of the rehearsal delays, and that these differences would perhaps account for some of the between-experiment variability in model parameter estimates. The separate random effects models for 4-digit and 6-digit rehearsal sequences provide more accurate estimates of the parameter values than when the data from the different list lengths are included in a single model. Therefore, it appears that there are significant differences between typical rate, constant error and offset parameter values across different variations of the rehearsal-probe task. However, the splitting of experiments based on list length may have been confounded by other differences between the tasks, such as sequence presentation rate (Experiment 1), sequence familiarity (Experiment 2), the use of preloads (Experiment 3), and temporal grouping (Experiment 4). Therefore, in future work, an analysis of the parameter estimates resulting from more systematic manipulations of these factors would help to specify their possible effects on rehearsal timing estimates more precisely.

6.2.3 Individual differences in model parameters and digit span

A final analysis was conducted with the modelling data in order to determine whether auditory-verbal ISR performance is associated with any particular modelling parameter or combination of parameters. For this analysis, only Experiments 2-7 were included as digit span data was not collected in Experiment 1. It was predicted that the significant correlations between mean CSD and digit span across these experiments would be reflected by correlations between digit span and the constant error and/or cumulative error parameters.

The individual constant and cumulative error parameters were not normally distributed, so a non-parametric alternative was used to test for correlations. The results of Kendall's τ correlations between digit span scores and the four model parameters are shown in Table 26. Correlations were computed across all experiments (means over conditions/groups) and the within group/condition correlations were also computed to assess the possibility that these relationships differed markedly in terms of direction and/or strength. None of the correlations reached statistical significance after correcting for multiple comparisons. However, strikingly, digit span scores were negatively related to constant error values in all but one instance, and this pattern was apparent but slightly less consistent for correlations between digit span and cumulative error.

Table 26. Kendall's τ correlations between the four model parameters and digit span scores across Experiments 2-7. Statistical significance is Bonferroni-corrected for multiple comparisons ($p < .001$).

Experiment	Group/Condition	N	Rate	Constant Error	Cumulative Error	Offset
2	Mean over conditions	17	.11	-.11	-.22	.15
	Low load	17	-.05	.14	-.23	.20
	High load	17	.30	-.24	-.14	.09
3	Mean over conditions	20	-.22	-.19	.05	.11
	Low load	22	.03	-.03	-.07	-.16
	High load	21	-.20	-.19	-.13	.00
4	All participants	28	.08	-.10	-.15	-.16
	Isochronous	14	.05	-.23	-.44	-.14
	Temporally-grouped	14	.27	-.07	.15	.04
5	-	39	.20	-.30	.04	-.01
6	All participants	36	-.06	-.09	-.04	-.15
	Control	18	-.01	-.25	-.30	-.07
	Dyslexia	18	-.06	-.08	.15	-.26
7	All participants	23	.20	-.34	-.16	-.15
	Control	14	.17	-.26	-.18	-.43
	SLI	9	.31	-.48	-.17	-.14

Note: no τ values were statistically significant after correcting for multiple comparisons ($p < .001$).

Figure 56, Figure 57 and Figure 58 show scatterplots of the relationships between digit span and the constant error parameter for data from Experiments 2-5, 6 and 7, respectively. Figure 59, Figure 60 and Figure 61 show relationships between digit span and the cumulative error parameter in Experiments 2-5, 6 and 7, respectively. Experiments 6 and 7 were plotted separately in order to reveal possible differences between populations.

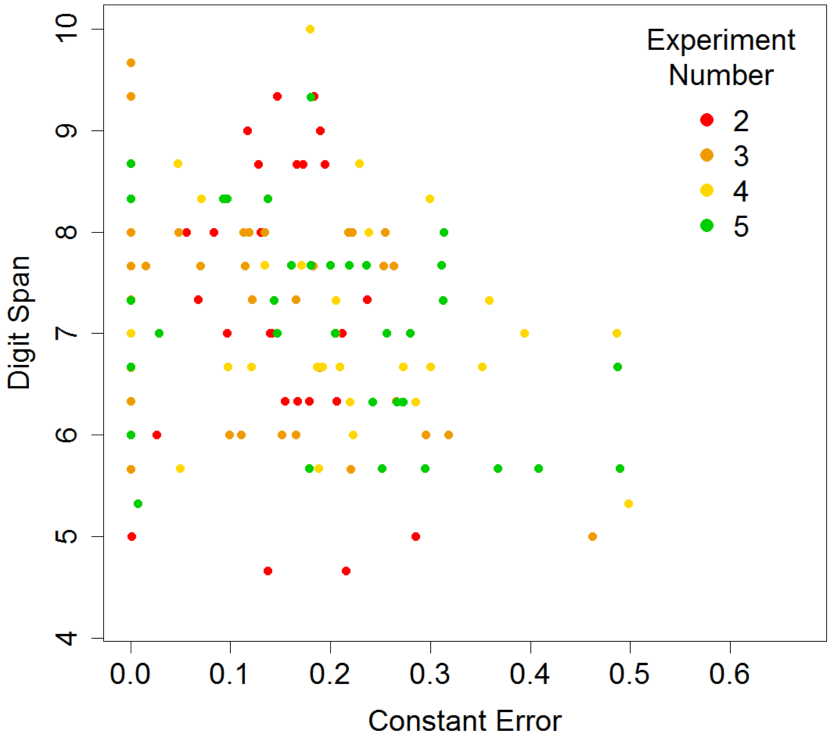

Figure 56. Scatterplot of the relationship between constant error model parameter values and digit span scores in young adult participants from Experiments 2-5.

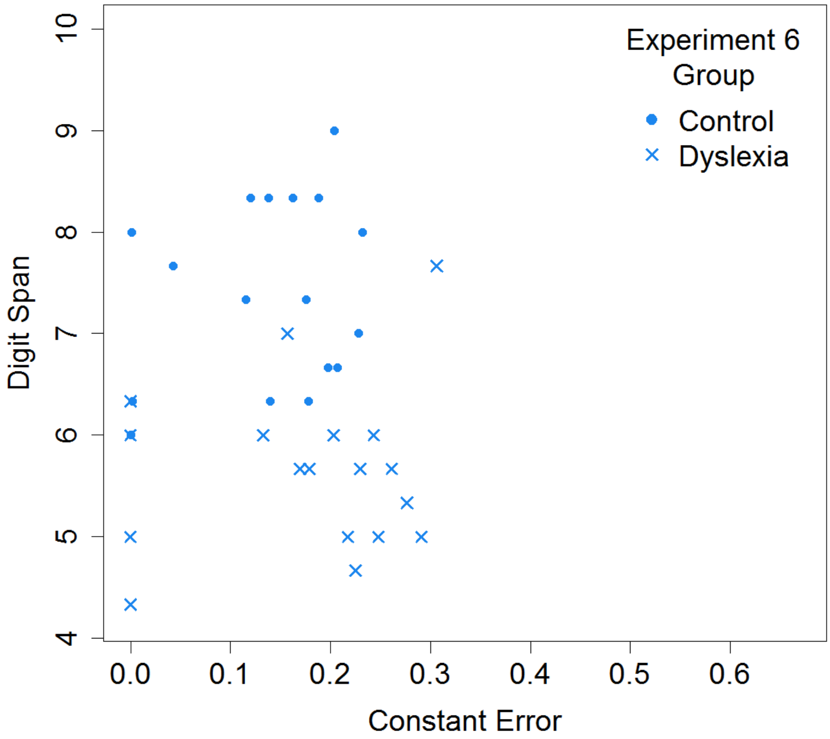

Figure 57. Scatterplot of the relationship between constant error model parameter values and digit span scores in young adults with and without dyslexia from Experiment 6.

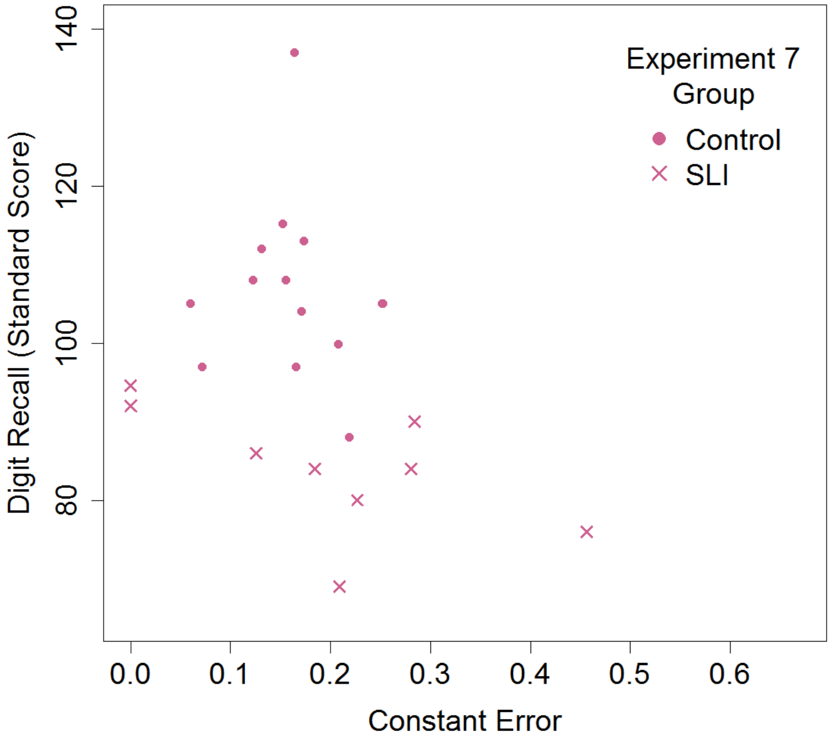

Figure 58. Scatterplot of the relationship between constant error model parameter values and digit recall standard scores in 8-11 year old children with and without SLI from Experiment 7.

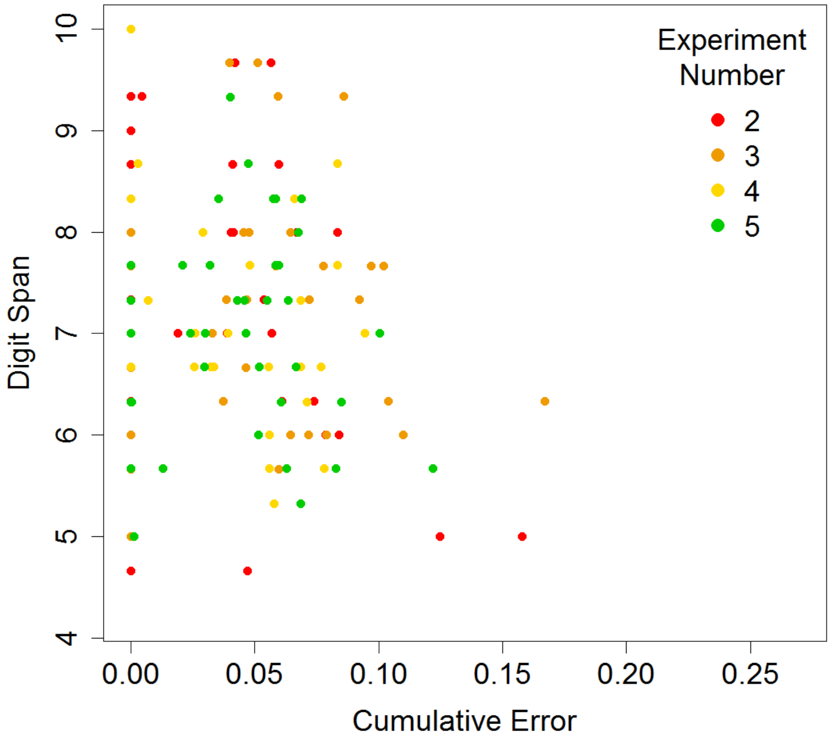

Figure 59. Scatterplot of the relationship between the cumulative error model parameter values and digit span scores in young adult participants from Experiments 2-5.

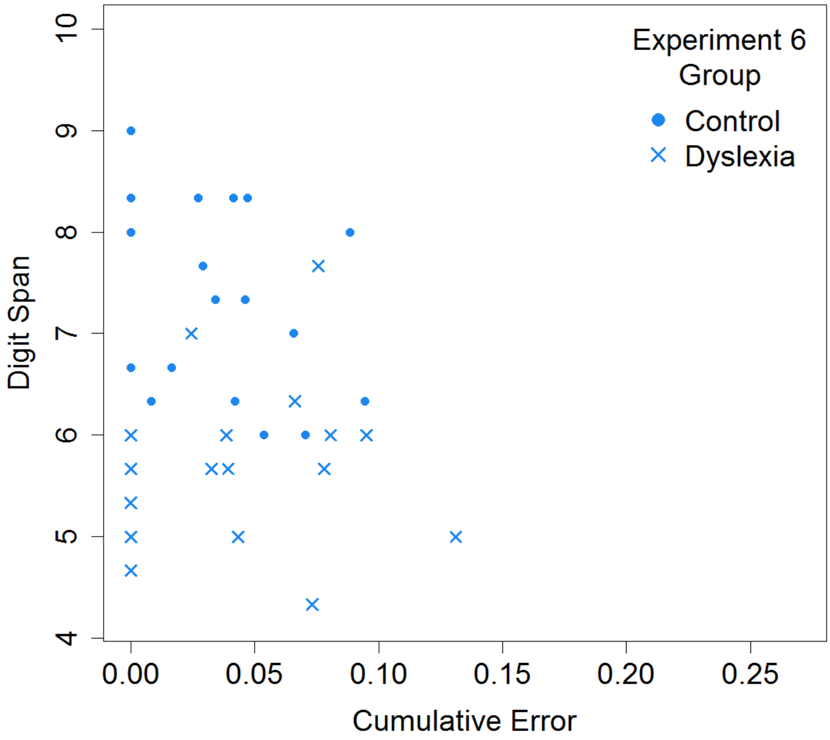

Figure 60. Scatterplot of the relationship between cumulative error model parameter values and digit span scores in the group of young adult participants with and without dyslexia from Experiment 6.

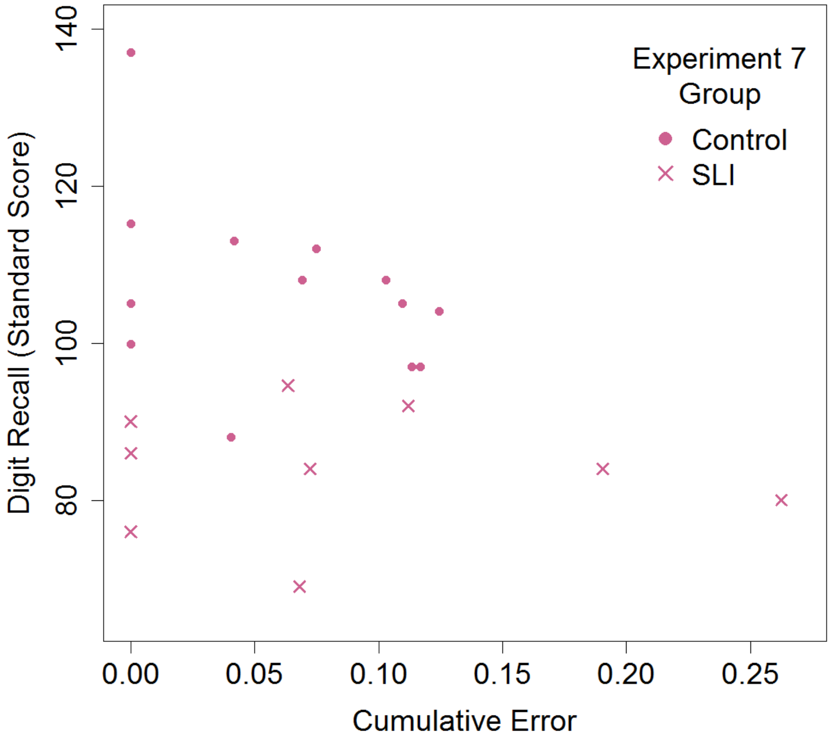

Figure 61. Scatterplot of the relationship between cumulative error model parameter values and digit recall standard scores in 8-11 year old children with and without SLI from Experiment 7.

The fact that the constant and cumulative error parameters were not normally distributed across participants is immediately visible in the scatterplots. The distributions of constant and cumulative error parameters were often bi-modal, with a tendency for the values to fall into separate groups of zero and more positive values. The scatterplots show high numbers of participants with zero error values in each distribution, where the zero error data points are evenly dispersed in terms of digit span performance. However, within the data points corresponding to non-zero error parameter values, there appears to be a negative linear relationship between the error parameters and digit span performance. One possible reason for the bi-modal distributions is that the model accounts for the variability in responses with *either* the constant or cumulative error term, at least in some data sets. This may mean that the model's fit with either a non-zero constant or cumulative error parameter may be only trivially different from the fit with the a non-zero value for the other error parameter. If this is indeed the case then it would reduce any systematic effects of task condition on one or both error parameters, and it would weaken the correlations between either or both error parameters and digit span performance.

Given that the distributions of the two error parameter values are non-normal, with both zero and non-zero peaks, it is possible that digit span scores will not be significantly correlated with one or the other error parameter but instead with the combination of the two predictors. This possibility was investigated with regression, where digit span scores were predicted by both constant and cumulative error values. A regression analysis was conducted on data from all experiments where digit span was collected (Experiments 2-7). Digit span was the dependent variable and the two error parameters were the predictors, which were entered in the same step.

Table 27. Results of linear regression applied to data from Experiments 2-7, with constant and cumulative error values as the predictors and digit span as the dependent variable.

Experiment	N	Constant Beta	Cumulative Beta	R	*F*	df	df	*p*
2	17	-.36	-.53*	.54	2.80	2	14	.095
3	20	-.55*	-.10	.51	3.04	2	17	.074
4	28	-.30	-.36	.38	2.16	2	25	.137
5	39	-.51**	-.19	.46	4.77	2	36	.015
6	36	-.06	-.14	.13	0.27	2	33	.766
7	23	-.44*	-.41†	.51	3.54	2	20	.048

Note: beta values statistically significant at *$p < .05$, **$p < .01$, ***$p < .001$. † $p = .052$

The regression results show that, across all experiments, the regression coefficients for both constant and cumulative error predictor variables were negative. Also, in four out of the six experiments examined, one of the two error parameter values accounted for a significant amount of variance in digit span scores. Constant error was significantly related to digit span scores in Experiments 3, 5 and 7, while the cumulative error value was a significant predictor in Experiment 2. Across the four experiments where one of the two error parameters was a significant predictor, the other predictor was non-significant (although the cumulative error value was marginally significant in the data from Experiment 7). In Experiments 5 and 7, the overall amount of variance in digit span scores predicted by the regression model was significant, and in Experiments 2 and 3 the regression model approached significance.

The results of the individual differences analysis generally supported the hypothesis that constant and/or cumulative error parameter values would be negatively associated with digit span scores across participants. Correlation coefficients were in the expected direction between constant error and digit span across experiments though these correlations did not reach statistical significance. The scatterplots of relationships between these measures suggest that there may be some inverse linear relationship between digit span and both constant and cumulative error, but these associations are accompanied by 1) a tendency for zero values to occur across the range of individual differences in digit span performance, and 2) a strong tendency for individuals with high spans to have low error values, but a wider range of high and low error parameter values among individuals with low spans. In the regression analysis, one of the two error parameters (usually constant error) was sufficient to account for variability in digit span scores in most experiments, though the regression coefficient for the other error parameter was also negative and sometimes of a moderate effect size. The results show that when response variability was divided into separate constant and cumulative error components by the model, the correlations between response variability and digit span performance are reduced within the two individual error terms. This appears to be due to the model's ability to account for response variability in the data with either a non-zero constant or cumulative error value (and sometimes with both). Possible reasons for this and potential solutions are raised in the discussion section that follows.

6.3 General Discussion

The aim of the work presented in this chapter was to develop a computational model to describe and quantify the dissociable temporal properties of the rehearsal-probe data, and to determine whether the resulting best-fit parameter estimates can provide any new insight into subvocal rehearsal timing in auditory-verbal STM. The results of fitting the model to aggregate response proportion curves show that the model is able to produce a very good fit to the data. This demonstrates that the serial position response proportions are well-described in terms of four parameters; two that determine the centres of the serial position response distributions, and two that determine the variability of the responses around the centres. In terms of differences between parameter values between experimental groups and conditions, the constant error parameter values differed in the expected directions, and this was sometimes the case with cumulative error as well. The constant error parameter value in particular was higher for the averaged response proportions from higher memory load conditions compared to the data from low load conditions, and it was higher in groups with developmental language disorders relative to the control groups.

The ability of the model to provide new insights about the data was further demonstrated by a meta-analysis of parameter values obtained across experiments with participants sampled from the same population (young adults with typical language development). Results of this analysis showed a general tendency for participants to rehearse more slowly than the presented rate, and with less cumulative than constant error. Also, in general, the model estimated a positive offset value, meaning that there was a static forward shift in mental time compared to real/expected time. This shift might occur as a result of, for example, a tendency to begin rehearsing earlier than expected, or a tendency to estimate the time of the probe as occurring slightly later than it is presented, resulting in a forward shift occurring at the time of the responses to probes. However, it should be noted that significant heterogeneity remained in the random effects models of rate, constant error and offset, which means that the distributions of these values differed across experiments. In future work, this type of analysis could prove useful for adjusting the task design, for instance by slightly decreasing the item presentation rate, and for examining similarities and differences in timing parameters across experimental designs and participant populations.

Regarding individual differences, the scatterplots and results of Kendall's τ correlations indicated a negative correlation between constant error and digit span, and regression analyses showed that both constant and cumulative error values were negatively-signed predictors of digit span scores. This is consistent with the previous findings that auditory-verbal ISR was negatively correlated with timing imprecision, as measured with mean CSD, and that this relationship is also present when imprecision is estimated with the model. Correlations between digit span and the error parameters were not statistically significant, which was probably due to the high frequency of zero error parameter values for participants across the range of digit span scores, as well as the use of less powerful non-parametric tests. Thus the individual differences analysis was consistent with expectations about the direction of the relationships, but it also revealed the model's apparent inability to reliably distinguish between the two error types in the data.

It could be argued that the zero and non-zero pairings of constant and cumulative error values suggest that only one of the two error terms is necessary to account for the data. This possibility could be tested by fitting the data with only one error parameter and then assessing the cost to the goodness-of-fit relative to the benefit of having fewer parameters in the model. However, there is a theoretical reason to support the inclusion of both types of error in the model. It is safe to assume that, no matter how accurate one's rehearsal timing, the response proportions would eventually decay toward asymptote at chance response probability (1 divided by the number of items) after a sufficiently long rehearsal interval. The constant error parameter alone would not predict this decay, and the cumulative error alone does not account for the substantial widths of response proportion distributions that remain static over time, particularly at shorter rehearsal delays.

Assuming that the two error parameters should be included in the model for the reason mentioned, another possibility arises for explaining the distributions of error parameters, which is that the rehearsal-probe task designs used were not optimal for differentiating between the two error types. Across experiments, the probes were presented in a somewhat narrow window, and this may have limited the model's ability to detect time-dependent changes in particular (i.e. cumulative error). The model's current difficulty in reliably distinguishing between the two parameters could be resolved by probing earlier and later in time. The use of a wider range of probe times should lead to more independent and reliable estimates of constant and cumulative error. This would perhaps reveal clearer and more consistent differences in error values between experimental conditions and groups, as well as more consistent correlations with auditory-verbal ISR performance.

While the model has shown to be useful in describing and quantifying the data, and in generating predictions about rehearsal timing, it of course limited in its ability to fit the data and to address certain questions about response timing. The model assumes that temporal imprecision is either constant or time-dependent. Therefore the model does not predict, for example, differences in timing precision as a function of serial position. Another important limitation is that the model is not capable of distinguishing among the multiple possible mechanisms that can generate the properties of the response proportion distributions. Indeed, the model was not designed to produce process-level explanations, as it was intended to be a theory-neutral method for describing the data. Explanations for serial order STM mechanisms have been addressed by a number of computational models of STM for serial order, so a next step is to determine how the data and parameters from the rehearsal-probe task and rehearsal model could be accounted for by these more detailed existing models of serial ordering processes. This issue will be explored in Chapter 8, the General Discussion.

The mean CSD dependent measure, described in detail in Chapter 2, was intended to be simple metric of response variability over time. The advantages of this variable are that 1) it is fairly straightforward, and 2) it does not depend on any theoretical assumptions. The mean CSD variable can be considered a 'blunt instrument'; it provides a broad, summary statistic for the variability of responses. The experiments presented in this thesis have shown that mean CSD values are robust and sensitive to task manipulations and individual differences, which shows that it is a meaningful and useful measure. On the other hand, the computational model is less transparent and requires more assumptions about the defining temporal properties of rehearsal. However, the model also provides a potentially sharper tool for understanding more specific temporal properties, with the ability to quantify the subtle aspects of the response proportion data and make concrete predictions about response proportion distributions at any given time.

There remains plenty of scope for further development of the model of rehearsal timing, as well as promising directions for future empirical work. With regard to model development, it would be useful to test the necessity of all parameters. Also, because of the heterogeneity in meta-analysis estimates of model parameter values across experiments, it would be interesting to study the effects of task designs more systematically in order to parse out the specific effects of different task factors on rehearsal timing. As previously mentioned, future rehearsal-probe experiments might benefit from using a wider range of probe times. Specifically, the use of later probe times would help to resolve the possible multiple combinations of the two error parameter values that result in similar fits to the data because the cumulative error in particular would have more of an opportunity to build with longer rehearsal delays. This latter direction for future empirical work shows that the model can be valuable for optimising the rehearsal-probe task design.

6.4 Chapter Summary

This chapter presents a computational model for the describing temporal properties of the rehearsal-probe data. Data from all behavioural experiments were modelled using four parameters; rate, constant error, cumulative error and offset. The model captured the signature features of the serial position response proportions and it produced an impressive fit to the observed data across different experiments. The resulting model parameters differed between groups and conditions in the expected directions (i.e. higher constant and/or cumulative error values corresponded to higher mean CSD in group/condition comparisons), though most of these differences were not statistically significant. A meta-analysis of parameter distributions across all experiments allowed the estimation of population-level values, but it also revealed that there was significant heterogeneity across experiments in estimates of rate, constant error and offset parameters. Overall, the model proved to be useful in generating testable predictions about rehearsal timing at any point in time, and in revealing more optimal rehearsal-probe task designs.

CHAPTER 7

INVESTIGATING NEURAL CORRELATES OF SUBVOCAL REHEARSAL

Thus far, the rehearsal-probe task has shown to be useful for measuring the timing of subvocal rehearsal, and it has demonstrated that timing precision varies as a function of memory load, individual auditory-verbal STM capacity, and language abilities. The aim of the work presented in this chapter was to combine the rehearsal-probe task with electroencephalography (EEG) in order to investigate the neural correlates of subvocal rehearsal. The rehearsal-probe task is unique in that it demands rehearsal, rather than assumes its occurrence, and the responses provide information about rehearsal timing. These attributes make it ideal for examining neural activity linked to the presence and rate of rehearsal. EEG is a method of recording the changes in summed electrical potentials at the surface of the scalp, and it is particularly suited for examining the time-dependent changes in neural activity due to its high temporal resolution. There are two general methods for examining EEG data; event-related potential (ERP) analysis, which measures changes in electrical potentials time-locked to specific events, and frequency analysis, which involves the decomposition of a time series into its component frequencies and associated strengths. In the context of the rehearsal-probe task, ERP analysis can be used to investigate the neural activity elicited by the probe, which thus relates to the content of memory (similar to the overt responses in earlier experiments), and frequency analysis can be used to investigate rhythmic activity associated with rehearsing a regularly-timed sequence during the maintenance interval.

Regarding the ERPs to probes, how might these responses be related to the content of memory during the rehearsal-probe task? The answer to this depends critically on a modification to the task for the purpose of ERP analysis. In the present experiment, unlike in previous experiments, one of the sequence items (a digit) was presented as a probe during rehearsal, rather than a tone. It was hypothesised that ERP analysis would reveal changes in brain activity related to whether item probe matched or did not match the contents of inner speech during rehearsal. This hypothesis was based on previous work showing similar effects in other contexts, but before reviewing the previous findings it is worth establishing the background information necessary for interpreting ERP results.

ERPs to auditory stimuli are characterised by a reliable series of amplitude fluctuations, with nomenclature indicating the direction of change in electrical potential at the vertex (N, negative; P, positive) and the peak latency (time in ms or order). The latency of the response reflects the level of processing, where earliest responses (< 10 ms after stimulus onset) are generated at the level of the brainstem, mid-latency responses (10-50 ms post stimulus) tending to reflect subcortical activity, and long-latency responses (50+ ms post stimulus) typically at or near the cortex (Picton, Hillyard, Krausz, & Galambos, 1974; Simpson & Prendergast, 2013). The obligatory (i.e. bottom-up driven) effects of lower-level auditory perception (e.g. pitch, intensity) are visible in the long-latency N1, P2, and N2[21] components, which correspond to amplitude deflections in fronto-central electrodes around 100, 150 and 250-300 ms post-stimulus, respectively. The longer-latency components reflect relatively higher-level and more abstract cognition, and those occurring ~250-400 ms post-stimulus (e.g. P300, N400) generally reflect top-down, attention-dependent processing.

The ERP method has been used in previous work to detect changes in electrical brain potentials as a function of the contents of inner speech (Numminen & Curio, 1999; Tian & Poeppel, 2013), though not in the context of a serial order memory or rehearsal paradigm. In these previous studies, a sound is presented during or after subvocal speech, where the resulting ERPs to auditory stimuli differ according to whether or not the sounds match the contents of inner speech. This method is based on the hypothesized neural process by which inner speech is produced in the brain. As discussed in Chapter 1, there is evidence to suggest that the neural mechanisms involved in subvocal speech are highly similar to those implicated in the production of overt speech in terms of the planning stages, and that these two processes diverge just prior to the execution of motor commands (Tian & Poeppel, 2012).

One of the processes that occurs during overt speech production is *corollary discharge*, which is the auditory feedback predicted from self-produced speech. This is thought to occur as a result of an internal forward model process whereby an *efference copy* is generated in the motor system and a sensory prediction sent to auditory perception areas (Niziolek et al., 2013; Scott, Yeung, Gick, & Werker, 2013; Scott, 2013; Sperry, 1950; von Holst & Mittelstaedt, 1950). The purpose of efference copies is to distinguish between the sensory consequences of self-produced versus externally-produced actions, and to assess any discrepancy between the predicted and actual sensory consequences of a self-produced action (Grüsser, 1995). A consequence of efference copies is that the brain response to stimulation from a self-produced action results in an attenuated neural response relative to the same sensory input produced by an external agent. In terms of speech production, this means that the ERPs to speech stimuli in left hemisphere areas are attenuated and delayed when the sound is self-produced compared to when one's own speech is replayed via a tape recorder (Curio et al., 2000; Ford, Roach, & Mathalon, 2010; Heinks-Maldonado et al., 2005; Houde et al., 2002).

---

[21] In magnetoencephalography (MEG), these components sometimes referred to with an "m" suffix, as in "N1m" or prefix, as in "M100".

The use of ERPs for categorising the contents of inner speech depends critically on the assumption that efference copies are produced even when the speech motor actions are not executed. The processes involved in both overt and covert speech production are not fully understood, and so a key remaining question concerns the motor planning or execution stage at which an efference copy is sent. Recently, Timm et al. (2014) recorded auditory ERPs to sound-initiating button presses, where the button presses were either voluntary or TMS-evoked involuntary finger movements. These researchers found that N1-P2 complex was attenuated in response to sounds generated from voluntary button presses, but not those evoked by self-executed but involuntary button presses resulting from stimulation to the motor cortex. This suggests that efference copies are produced as a result of processes involved in the planning of motor execution, rather than the motor execution itself.

If internal motor simulation relies, at least in part, on the same neural mechanisms that control overt motor execution, then inner speech should modulate the auditory cortex in the same or similar manner as overt speech. Support for this comes from magnetoencephalography (MEG) studies showing an attenuation of neural responses to specific sounds during silent speech. One of the first demonstrations of this effect in humans was presented in a study by Numminen and Curio (1999), who showed that MEG-recorded evoked responses to auditory stimuli during covert speech differed according to whether the external stimulus was the same or different as the contents of inner speech. These authors compared the neural responses to the tape-recorded playback of participants' own articulations of different vowel sounds (/a/ or /i/) during the concurrent subvocalisation of one of the two sounds. The results of this study showed that there was a delayed M100 (i.e. N1m) response in left hemisphere temporal sites for presented vowel sounds that matched the content of inner speech compared to vowels that did not match.

The contents of subvocal speech have also been shown to affect ERPs in the absence of concurrent probe presentation, and with stimuli that was not recorded with the participants' own voice. Tian and Poeppel (2013, see also Tian & Poeppel, 2012) measured responses to probe (syllable) sounds after a short period of one of the following adaptor (priming) conditions: overt articulation, imagined articulation, auditory presentation, or imagined auditory presentation. These authors found that the adaptation intervals of overt articulation produced a priming effect in response to the target sounds, with a greater percent change from the reference condition (unprimed auditory ERPs) in response to 'repeated' (match) compared to 'novel' (mismatch) target stimuli at 200 ms after the target presentation. Importantly, the same priming effect was also present in the imagined articulation condition, albeit to a lesser degree. The authors also showed that the only condition that produced a difference in the earlier M100 response was the overt perceptual (hearing) adaptation interval, which the authors highlight as being the only condition influenced by bottom-up (stimulus driven) factors. The effect of subvocal speech priming on evoked potentials shows that ERPs are enhanced to match probes around 200 ms after probe presentation, in addition to the earlier (~100 ms) suppression effect reported by Numminen and Curio (1999). Unlike the match suppression effect, which is attributed to corollary discharge, in the case of priming, the match probe enhances the ERP peak amplitude and occurs at a later (~200 ms) latency. This difference reflects the fact that, in the priming experiment, the match probe differed acoustically from the participants' own voices and was presented after a subvocal speech period rather than simultaneously. Therefore the effect observed by Tian and Poppel was likely driven by more abstract, higher-order processes as opposed to modulation of auditory perceptual areas via corollary discharge.

In addition to the ERP analysis, the use of EEG with the rehearsal-probe task was motivated by the goal of investigating the neural oscillatory features of rehearsal. Cortical oscillations are known to track the phase of the low-frequency (i.e. 4-10 Hz, syllable or stressed syllable rate) amplitude envelope in speech and speech-like stimuli (Doelling, Arnal, Ghitza, & Poeppel, 2014; Ghitza, 2011, 2012, 2013; Saoud et al., 2012). This phase-tracking is apparent from the increased power at relevant frequencies in the averaged time series, and the magnitude of this effect is thought to be related to speech perception and comprehension (Peelle & Davis, 2012; Peelle et al., 2013). However, at present there is no previous research to suggest whether increased phase-locking might occur during subvocal rehearsal at frequencies relevant to the rehearsal rate. While there has been some research to show possible changes in frequency dynamics resulting from inner speech, such as increased power in the high delta band (~3.9 Hz; Harmony et al., 1999), this work has only been done where subvocal speech was unpaced and was assumed to take place on the basis of the verbaliseable nature of the memory material.

To summarise, the previous ERP studies on the effects of the contents of subvocal speech to auditory match/mismatch probes show two main effects. First, the presentation of an auditory probe that matches the contents of inner speech produces a suppression in the negative peak over left temporal lobes at about 100 ms after probe presentation, but this may only occur when probe is presented simultaneously with the expected sensory feedback and/or when the probe matches the expected consequence on the level of specific acoustic properties (Numminen & Curio, 1999; Timm et al., 2014). Second, when subvocal speech occurs before the probe, then ERPs to content-matched probes are enhanced in the M200 (i.e. P2m), around 200 ms after presentation (Tian & Poeppel, 2012). Thus the effect of probe type (match or mismatch) depends on the degree of temporal and/or feature overlap. The earlier N1 suppression effect for match probes occurs when the timing of the external stimulus is consistent with that of the expected timing of the sensory feedback from motor commands, and this perhaps occurs at the level of acoustic features. By contrast, the later P2 enhancement for match probes is seen when the contents of inner speech match the content of a subsequent auditory stimulus, and this likely occurs at the level of more abstract phonological representations. Finally, with regard to frequency dynamics, previous research shows that neural oscillations entrain to the prominent slow rhythms in the amplitude envelope of continuous auditory stimuli. However, there is a lack of research on changes in oscillatory power as a result of inner speech, particularly in the context of paced subvocal rehearsal.

In the present experiment, the rehearsal-probe task was used to determine whether an effect of probe type could be observed in the context of a memory task with cyclic paced rehearsal. Unlike previous rehearsal-probe experiments where the probe was a tone, in the present experiment, the probe presented during rehearsal was a sequence item (digit). The presentation of the item probe was timed to occur at the onset of the same item (match) or a different item (mismatch) during rehearsal. Based on the findings reported in the literature and on the results of a pilot study with a separate participant group (N = 13), it was expected that the ERP response to match probes would be attenuated in left temporal channels at the N1 (~100 ms) peak due to the close correspondence between the stimulus onset and the timing of subvocal item rehearsal. Also, it was hypothesised that the ERPs to match probes would show a greater P2 (~200-250 ms) peak in comparison to ERPs to mismatch probes.

Regarding the frequency analysis in the present experiment, based on previous research it was predicted that the evoked (phase-locked) EEG signals during the sequence presentation period would show increased power at the digit presentation rate of 2 Hz. However, there is a lack of research to inform predictions about whether or not an enhancement in 2 Hz phase-locked power might also be apparent during the paced subvocal rehearsal period. This question was addressed in the current experiment through an analysis of the frequency-power distributions of sequence presentation, rehearsal and baseline (inter-trial interval) periods. Crucially for this analysis, ongoing activity was decoupled from any overt action, and the trial periods were matched for duration in order to equate the frequency resolution. It was predicted that there would be relative increases in 2 Hz power in the sequence presentation and rehearsal periods, but not in the baseline period.

7.1 Methods

7.1.1 Participants

Participants were 20 right-handed native English speakers (10 women, mean age 22.6 years, SD 4.2 years). The consent process was the same as that described in previous behavioural experiments, with further information provided about the EEG recording process. Participants were compensated with either a small cash reward or credit toward course requirements.

7.1.2 Design

The EEG rehearsal task took a within-participant 2 x 2 factorial design, where one factor was probe type (match or mismatch) and the other was probe time (A or B). In the 'match' probe condition, one of the 4-digit sequence items (the probe) was presented again during the rehearsal period, and this item probe corresponded to the item that the participant should be rehearsing at the time of the probe presentation. In the 'mismatch' probe condition, the item probe presented during rehearsal corresponded to the sequence item +/- two positions away from the current item. The item probe corresponded to either the 4th digit in the first rehearsal cycle (probe time 'A', 1.5 seconds after the start of rehearsal), or to the 2nd digit in the second rehearsal cycle (probe time 'B', 2.5 seconds after the start of rehearsal). Thus, for mismatch probe types, the presented item was the item in the 2nd position for A probe times and the item in the 4th position for B probe times. The probe times in this experiment are earlier than those used in previous experiments because it was expected that rehearsal timing would be more accurate with shorter delays.

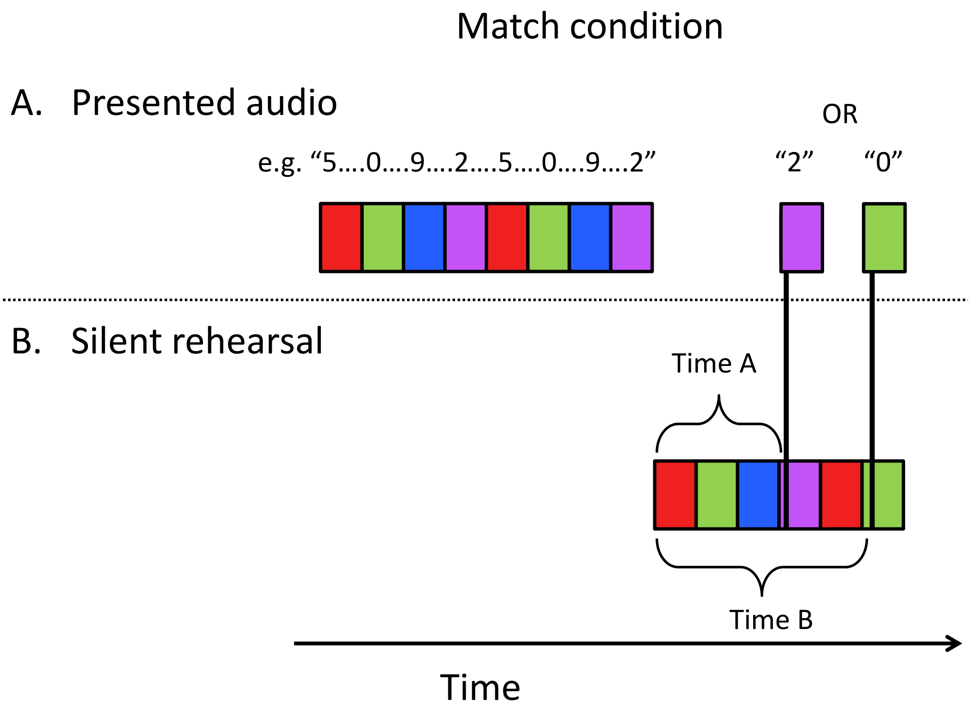

Figure 62. Schematic of Time A and Time B probes in the Match condition. An item probe is either presented during the 4th item in the 1st rehearsal (Time A), or the 2nd item in the 2nd rehearsal cycle (Time B). The item probe matches the item that the participant should be rehearsing (given accurate rehearsal timing).

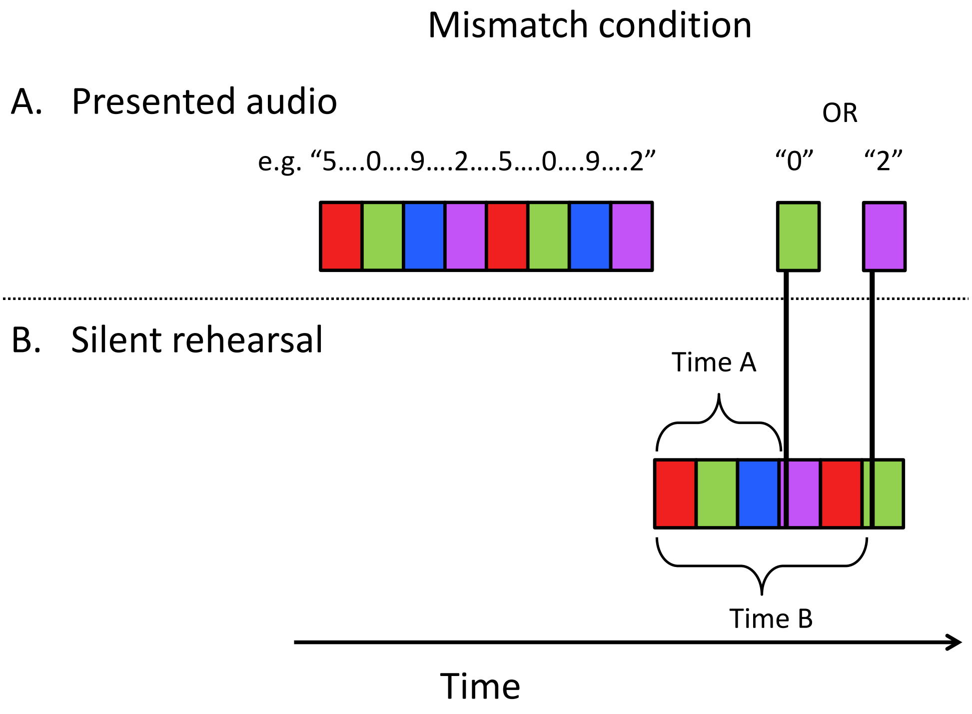

Figure 63. Schematic of Time A and Time B probes in the Mismatch condition. An item probe is either presented at either Time A or Time B. The item probe does not match the item that the participant should be rehearsing (given accurate rehearsal timing).

Because the goal was to induce changes in the ERP only as a function of the correspondence between the presented stimuli and the contents of inner speech, the probe type was presented with equal probability. This is in contrast to other common ERP match/mismatch paradigms, such as MMN and oddball (P3) tasks, where responses are dependent upon the trial frequency statistics, e.g. 80% for standard stimuli and 20% deviant/target stimuli. Here, to avoid conflating the intended match/mismatch response with a response generated by the different probabilities in trial types, the match and mismatch probe types were each presented with 50% probability.

7.1.3 Materials

The digit stimuli used in the digit span and rehearsal tasks were those described in Chapter 2 and used in previous behavioural experiments.

7.1.4 Procedure

Digit span was administered according to the procedure outlined in Chapter 2 and used in previous experiments. Following the digit span task, participants were set up in EEG caps and impedance values were checked. This process lasted 20-30 minutes. Participants were then given the task instructions with demonstration and practice trials. After this, the participants completed the experimental task. The task lasted 45-55 minutes, depending on the duration of breaks between trials. The entire testing session lasted 1.5 to 2 hours.

The rehearsal sequences were random 4-digit lists presented at two digits per second. Unlike the previous behavioural experiments, where the probe was a simple tone, here a sequence item was presented as the probe. The other major difference in this task compared to those used in previous experiments is that in this experiment, 2/3 of the trials were passive. That is, no overt response was required from the participant on these ‘standard’ trials. One third of trials were ‘catch’ trials, where a response to the probe was required (as in the previous experiments). In pilot work it was established that participants had difficulty sustaining attention throughout the task, so catch trials were added to the protocol in order to reduce the passive nature of this version of the task. Also, a short (0.5 second) delay was introduced between the item probe presentation and visual response cue on catch trials, which allowed the catch trials to be included in the ERP analysis without contamination from visual or motor ERPs. This reduced the total number of trials, and thus the task duration, because the ERPs to catch trials would otherwise have to be discarded. Finally, the responses to catch trials allowed an analysis of the validity of the match and mismatch probes, as well as an analysis of relationships between behavioural responses and EEG measures.

The instructions, demonstration and practice trials were similar to those used in the previous behavioural experiments. After the general introduction to the task, participants were told that they would hear a lone number some point while they were rehearsing, and this was their cue that the trial is over and they can stop rehearsing. The instructions also stated that, on some trials, the lone number would be followed by a visual cue ("????"), which meant that they should respond with the digit they were currently rehearsing at the time that the lone digit was presented. Further instructions stated that the lone digit they heard could be the same or different as the one they were rehearsing at the same time, but that they should respond with the number they were rehearsing at the time it was presented, regardless of whether it was the same or different.

Following these instructions, the experimenter demonstrated four trials with overt rehearsal. The demonstration trials were Match A, Mismatch B, Mismatch A (catch) and Match B (catch). The participant then completed four trials with overt rehearsal: Match A, Match A (catch), Match B, and Mismatch A (catch). If the participant struggled with these trials or showed evidence of misunderstanding the task instructions, then the set of four trials was repeated. Then, the participant was given additional instructions specific to the EEG recording (e.g. eyes open and focused in the middle of the screen, avoid blinking).

There were 180 trials in total, which were made up of 120 standard (no-response) trials and 60 catch (response) trials. The four conditions (Match A, Match B, Mismatch A, Mismatch B) were equally represented within the standard trials (30 trials per condition) and catch trials (15 trials per condition). The task was separated into five blocks of 36 trials. Within each block, there were 6 standard trials for each of the four conditions (24 standard trials per block) and 3 catch trials per condition (12 catch trials per block). The order of conditions and trial types was randomised within sets of 12 trials (i.e. 2 standard trials per condition and 1 catch trial per condition).

Within blocks, the trials were separated by jittered inter-trial intervals (ITIs). Each ITI was selected at random from a set of 26 durations on each trial. These durations were linearly-spaced from 1.5 to 2 seconds (in steps of .2 seconds). The minimum ITI duration (1.5 seconds) ensured that all ITIs were at least as long as the shortest rehearsal intervals, which was critical for using the ITIs as a baseline period for comparison with rehearsal periods in low frequency dynamics. Also, the jittering of trial onsets prevented confounding neural activity as a result of predictable trial onsets or synchronisation to auditory rhythms across all trials.

7.1.5 EEG Recording

The EEG data was collected using 64-channel WaveGuard caps with active shielding Ag/AgCl electrodes and an ANT-Neuro amplifier with an input impedance of 10^{12} Ω.

Electrode locations are shown in Figure 64. The ground was placed between the Fpz and Fz sites. Impedances at electrode sites were kept below 10 kΩ. Sampling rate was 1000 Hz, and the online reference was the common average reference. Bipolar electrodes were used to record vertical eye movements (VEOG). The data was recorded with ASA software on a separate data acquisition computer. Event markers were sent to the amplifier pulses aligned with physical sound onsets in one channel of the audio stimuli. The channel containing the digit sequence stimulus was converted from mono to stereo and sent to the headphones, while the channel containing square-wave pulses was sent to a microcontroller (Arduino), where it converted from analogue to a digital 5V TTL and sent to the amplifier. The event marker latency was tested and found to be accurate within +/- 1 ms, thus it was within the acceptable range for data recorded at a 1000 Hz sampling rate.

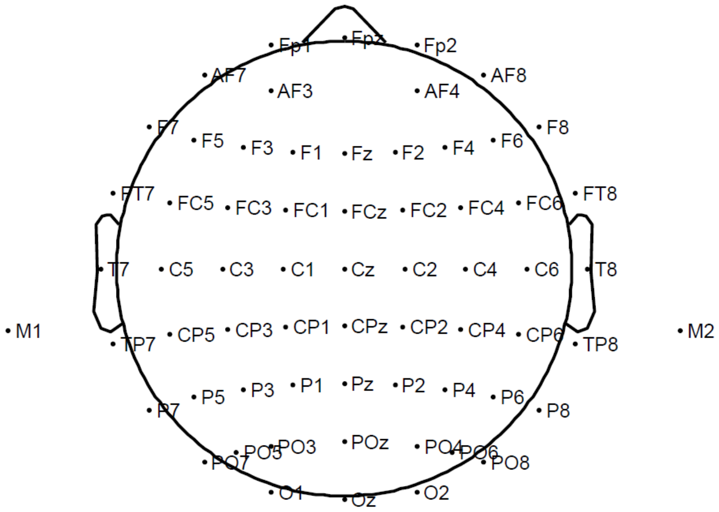

Figure 64. Two dimensional representation of the 64-channel electrode locations and labels within the International 10-20 system.

7.1.6 *ERP data processing and analysis*

All processing and analysis of the EEG data was done using the EEGLAB toolbox for Matlab (Delorme & Makeig, 2004) and a combination of standard functions and custom analysis scripts. The continuous data was high-pass filtered using a non-causal FIR filter with a 0.01 Hz cut-off frequency, then epoched at -100 to 500 ms relative to the probe onset. Next, noisy or flat channels were identified visually, and rejected and interpolated if necessary. The data was then re-referenced to the common average. Artifact detection was applied with a +/- 75 μV threshold and visual inspection, and epochs containing artifacts were rejected. The remaining epochs were low-pass filtered at 30 Hz, high-pass filtered at 1.7 Hz and averaged within the four conditions. One participant was excluded during this process because too few trials remained after artifact rejection, so the analysis was based on ERPs from 19 participants. Of these participants, the number of artifact-free trials analysed ranged from

178 to 93 (M = 152.63, SD = 25.78). The number of trials included in the analysis did not differ across the four probe conditions ($F(3,54) = 0.74$, $p = .531$).

Visual inspection of the grand averaged waveforms revealed four deflections at the vertex (Cz) and surrounding sites: a small positive peak with a latency of 50 ms, a prominent negative peak around 100 ms, a prominent positive peak around 200-250 ms, and a smaller negative peak around 400 ms. These peaks were consistent with the standard auditory P1, N1, P2 and N400 components (Simpson & Prendergast, 2013). The peaks have central and centro-parietal distributions, with polarity reversals over pre-frontal and occipital sites. The lateral and mastoid electrodes were characterised by a series of deflections with slightly different latencies: a small negative peak around 50 ms, a positive peak around 100 ms, and a negative peak around 175 ms. This series of deflections over the temporal sites reflect a subcomponent of the N1-P2, known as the T-complex (Hämäläinen, Fosker, Szücs, & Goswami, 2011; Shafer, Schwartz, & Martin, 2011). The T-complex is comprised of the Na (first negative peak), Ta (first positive peak) and Tb (second negative peak). Because the ERPs at temporal sites were of primary interest in the present research, analysis was conducted on two of the T-complex peaks with latencies that corresponded to those hypothesised: Ta (75-125 ms, i.e. N1) and Tb (150-300 ms, i.e. P2). The widths of the temporal windows were selected based on previous research and on the range of peak latencies observed in the data, where the latencies of later components (e.g. Tb/P2) are known to be relatively more variable across participants (Simpson & Prendergast, 2013).

7.1.7 Frequency data processing and analysis

The continuous EEG data sets were first high-pass filtered using a non-causal FIR filter with a 0.01 Hz cut-off frequency. The high-pass filtered continuous data sets were then epoched into arbitrary one-second windows for ease of data segment rejection prior to independent components analysis (ICA). At this stage, any noisy or flat channels were also rejected to facilitate artifact detection and improve the ICA results. The epoched (segmented) data was then subjected to an artifact detection procedure for non-stereotyped artifacts only, such as non-ocular muscle activity or static noise, using an amplitude threshold of +/- 75 μV for epoch marking followed by visual inspection. After these epochs were rejected, the remaining epochs were subjected to an ICA[22] (Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). The ICA sphering matrix was then exported to the complete high-pass filtered continuous data set, and the independent components reflecting any clear constant (e.g. EKG,

---

[22] The ICA algorithm was infomax (Bell & Sejnowski, 1995) as implemented as the 'binica' type in the 'runica' EEGLAB function, using the 'extended' option for to extract sub-Gaussian sources (Lee, Girolami, & Sejnowski, 1999).

50 Hz) or stereotyped (e.g. vertical eye blinks and horizontal saccades) noise sources were rejected. Identification of artifactual independent components was based on the topography and frequency spectrum of the component, as well as the qualitative characteristics of the amplitude over time (Groppe, Makeig, & Kutas, 2009; Onton & Makeig, 2006). The channels removed prior to ICA were spherically interpolated. Then, the continuous data sets were epoched into three segments corresponding to the three segments of each trial; sequence presentation period (4.5 seconds), rehearsal segment (1.5 seconds for Time A or 2.5 seconds for Time B), and inter-trial interval (1.5 seconds). Finally, artifact detection procedures were applied to the ICA-cleaned trial segments using a +/- 75 μV amplitude threshold and visual inspection.

Four subjects had to be excluded from the frequency analysis due to a very high number of artifacts in the data, which resulted in poor ICA results and left too few artifact-free epochs in these data sets at the analysis stage. For the remaining 16 participants, the number of epochs retained after the final (post-ICA) artifact-rejection process ranged from 103 to 177 (M = 158.31, SD = 20.15) sequence presentation sections out of the 180 trials. For rehearsal sections of the trials, the number of remaining epochs ranged from 104 to 176 (M = 160.63, SD = 19.23) out of 180. Finally, out of the 120 ITI sections that were not preceded by a behavioural response[23], the number of epochs included in the analyses ranged from 62 to 117 (M = 103.50, SD = 15.39).

Because the power at low frequencies is particularly affected by differences in the time window when this window is close to the minimum length required for a reliable power estimate (3 times the 2 Hz cycle length, or 1.5 seconds), the epoch durations were matched across the three trial segments. The 1.5-second rehearsal interval of the Time A trials and the first 1.5 seconds of the Time B rehearsal interval were combined. As a result, the epochs used for comparison were: the last 1.5 seconds of the sequence presentation period, the first 1.5 seconds of the rehearsal period, and the first 1.5 seconds of the baseline period. To determine whether there were any differences in the 2 Hz evoked (phase-locked) power among the three trial segments, the epochs were averaged within individuals and trial segments. Discrete Fourier transforms (DFTs) of the averaged signals were computed, where the number of points was based on the next power of 2 after the number of samples in the data. This resulted in power values for frequencies from 0 (DC) to 500 Hz (Nyquist frequency) with a resolution of 0.49 Hz. The resulting evoked frequency-power distributions were normalised by taking the log of the power in order to account for the $1/f$ distribution.

---

[23] Sixty of the 180 trials were catch (response) trials, meaning that a behavioural response was required after the presentation of the probe at the end of the rehearsal interval. While the ITI period did not begin until after the participant entered a response, the 60 ITIs that followed the catch trials were nonetheless excluded from the analysis due to potential contamination from movement artifacts.

7.2 Results

7.2.1 *Behavioural results*

Participants' responses to catch trials during the EEG rehearsal task were scored according to two criteria; accuracy and consistency. For the accuracy measure, the serial positions of the responses were scored as correct or incorrect over all trials, and this was converted to a percentage correct. For Time A trials, the probe would occur during the 4th item in the sequence if the sequence were rehearsed with the exact presentation timing, therefore responses to Time A trials were scored as correct if they corresponded to the 4th serial position of the rehearsal sequence and were scored as incorrect for any other serial position. This same scoring procedure was used with Time B trials, where the item in the 2nd serial position was scored as correct and all other responses were incorrect. The percentage of correct responses within each Time category was then calculated for each participant and overall (see Figure 65 for distributions of catch trial response percentages at Time A and Time B over all participants). Overall, there were 45.83% correct responses across both probe times, with 42.83% correct catch trial responses to Time A probes and 48.83% correct responses to Time B probes. Importantly, the response accuracy was well above the 25% chance response rate. Also, the tendency to respond with the immediately preceding item (i.e. 3rd item at Time A, 1st item at Time B) was consistent with the results from previous experiments showing a tendency for responses to lag slightly behind the expected timing.

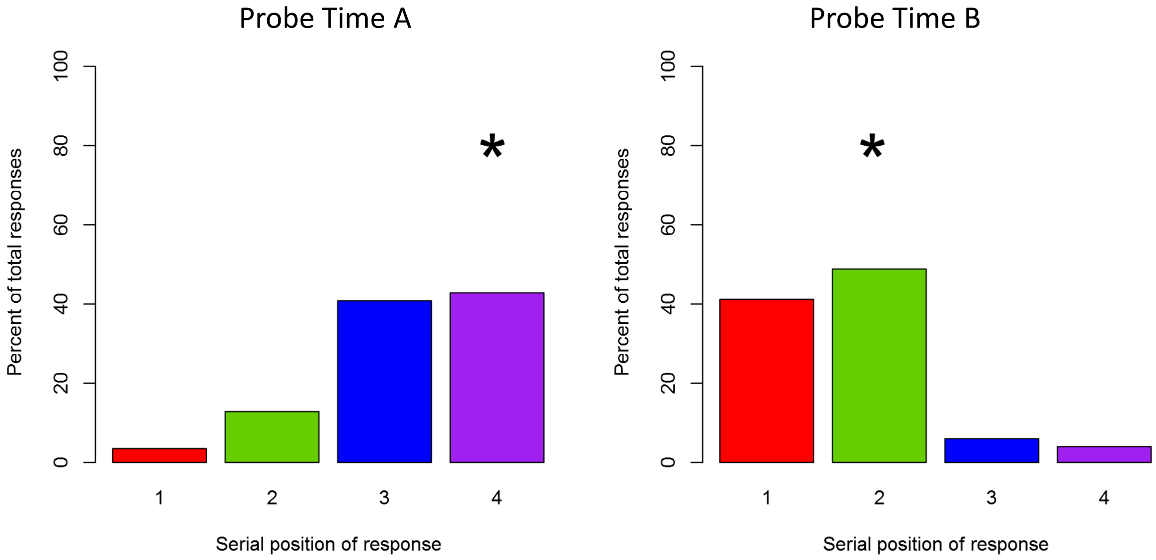

Figure 65. Percentages of total responses to catch trials corresponding to items in each serial position. Percentages are reported within the responses to probes presented at Time A (left) and Time B (right). The correct response at each probe time is denoted with a * above the serial position.

As a measure of response consistency, CSDs were calculated for each individual's distribution of serial positions of the responses to all 30 catch trials at each probe time (A and B). The computation of CSDs in this case was different from that described in Chapter 2 and used in previous experiments because of the difference in task design used here. In the previous experiments, there were many probe times with only a few responses per probe time, so CSDs were calculated using the distributions of probe times (in radians) for each serial position. In the present experiment, there are only two probe times, but many responses per probe, so the CSDs were calculated using the distributions of serial position responses within each probe time. To do this, the four serial positions were converted to equidistant locations around a circle, in radians. Then, the CSD of these serial position radian distributions was calculated within each probe time. Highly consistent responses to probes would result in the same radian values, and thus low CSDs, whereas more varied responses to probes would result in a greater dispersion of radians, and therefore higher CSDs. The average of these two CSD measures provided an overall CSD score for each participant. Table 28 shows the means, standard deviations, minima and maxima for all behavioural measures. It is important to note that, because these CSD values were computed differently, they are not directly comparable to those in previous experiments.

Table 28. Means, standard deviations, minima and maxima of digit span scores, rehearsal catch trial accuracy (percent correct), and CSDs.

Measure	Mean	SD	Minimum	Maximum
Digit span	7.12	1.14	5.00	9.33
Accuracy (% correct)	45.83	29.53	1.67	93.33
CSD - probe time A †	0.77	0.31	0.00	1.26
CSD - probe time B †	0.68	0.24	0.26	1.27
Mean CSD †	0.72	0.25	0.22	1.27

† The range of possible CSD scores in this experiment is 0.00 to 1.41, where a score of 0 indicates perfect consistency in response serial positions at a given probe time, and CSD scores increase as the set of responses at a given probe time approach a uniform distribution across all four serial positions (1.41).

Correlations between behavioural measures were conducted to investigate the relationships between digit span, response accuracy (percent correct) and response consistency (CSD). There was a significant negative correlation between digit span and probe time B CSD scores, *r* = -.45, *p* = .045 (see Figure 66). The relationships between digit span and both Time A CSD and mean CSD scores were also negative but did not reach significance (*r* = -.21, *p* = .370 and *r* = -.36, *p* = .122 respectively).

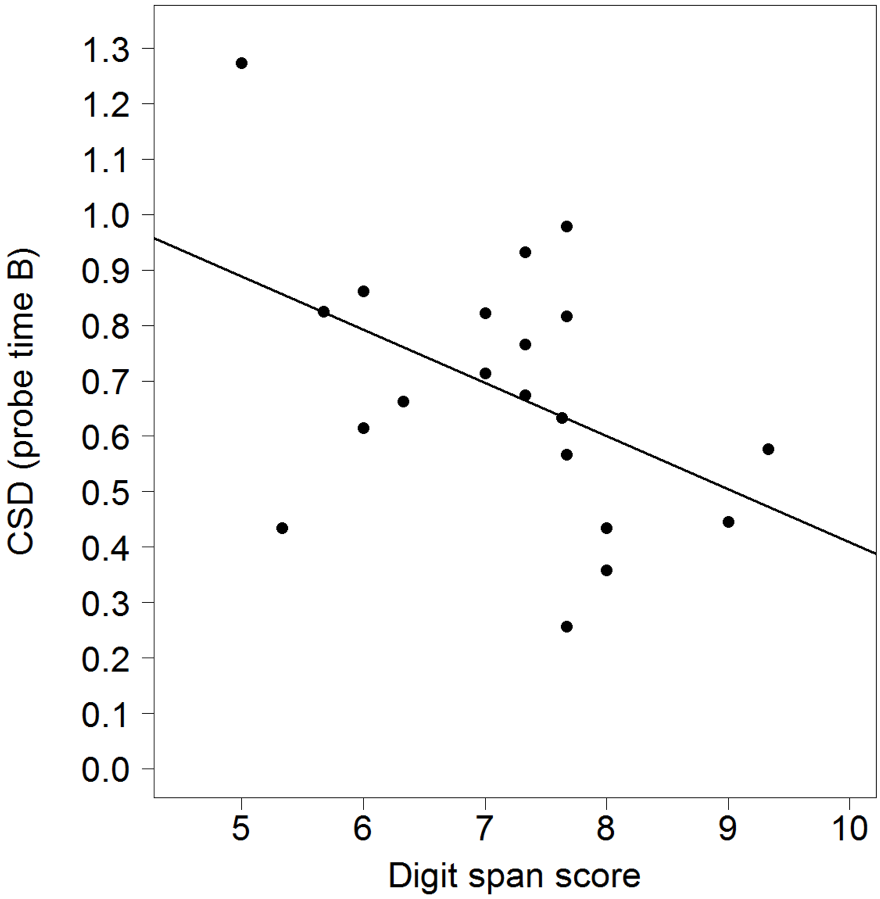

Figure 66. Relationship between CSDs for Time B probes and digit span. Solid line shows the least squares regression line. The minimum and maximum CSD scores were 0 and 1.41 in this experiment. CSD values in this experiment were based on distributions of serial position responses within a single probe time (Time B), rather than distributions of probe times within serial position response categories, and thus are not directly comparable to the CSD values reported in previous experiments.

Response accuracy was not significantly related to either digit span, $r = -.16$, $p = .504$, or CSD measures (all *r*s < -.29, all *p*s > .05). The lack of a significant linear relationship between response accuracy and CSD is not surprising given the non-linear relationship between the two variables (see Figure 67). This is due to the fact that, in order to achieve high response consistency (low mean CSD), the individual is likely be on one of the two extremes of the response accuracy spectrum. On the other hand, individuals with low response consistency (high mean CSD) will have a more even spread of serial position responses at a given time, some of which will be correct and others incorrect, resulting in mid-to-low response accuracy (random responding = 25% accuracy). There is a general pattern to the data that places individuals roughly into three categories: high accuracy and high consistency, low accuracy and high consistency, and low consistency with mid-to-low accuracy. This pattern is visible in the inverted U-shaped curve that characterises the relationship between CSD and response accuracy, shown in Figure 67.

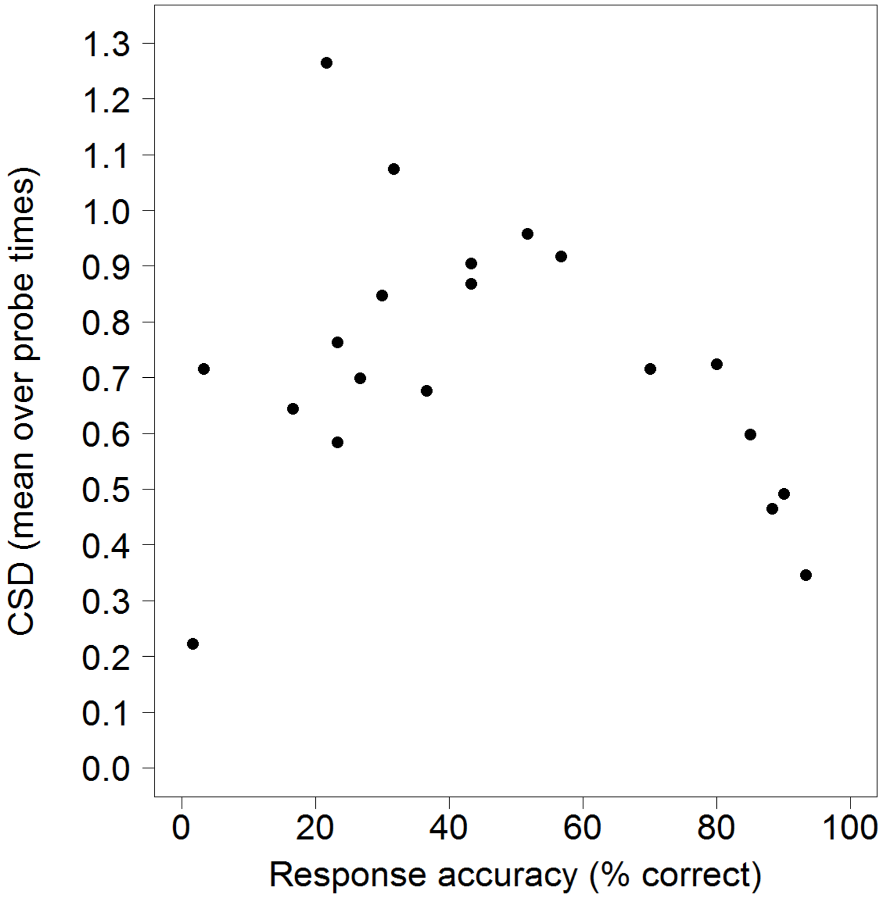

Figure 67. Relationship between response accuracy (percent of correct responses to catch trials, x-axis) and the mean CSD over the two probe times (y-axis).

7.2.2 ERP results

The grand average ERPs for match and mismatch conditions (irrespective of probe time) are shown in Figure 68. The results show characteristic auditory ERP waveform, as easily identified at the vertex (Cz). The averaged waveforms in response to match and mismatch probes were highly similar across channels, with the exception of left temporal (T7, TP7) and left posterior (P7, PO7, O1) sites. There were more obvious differences between grand average ERPs to Time A and Time B probes, as shown in Figure 69. The later, Time B probes tended to elicit a more negative N1 deflection at the vertex, as well as more positive Ta peaks in the right temporal (T8, TP8) and right posterior (P8, PO8, O2) areas.

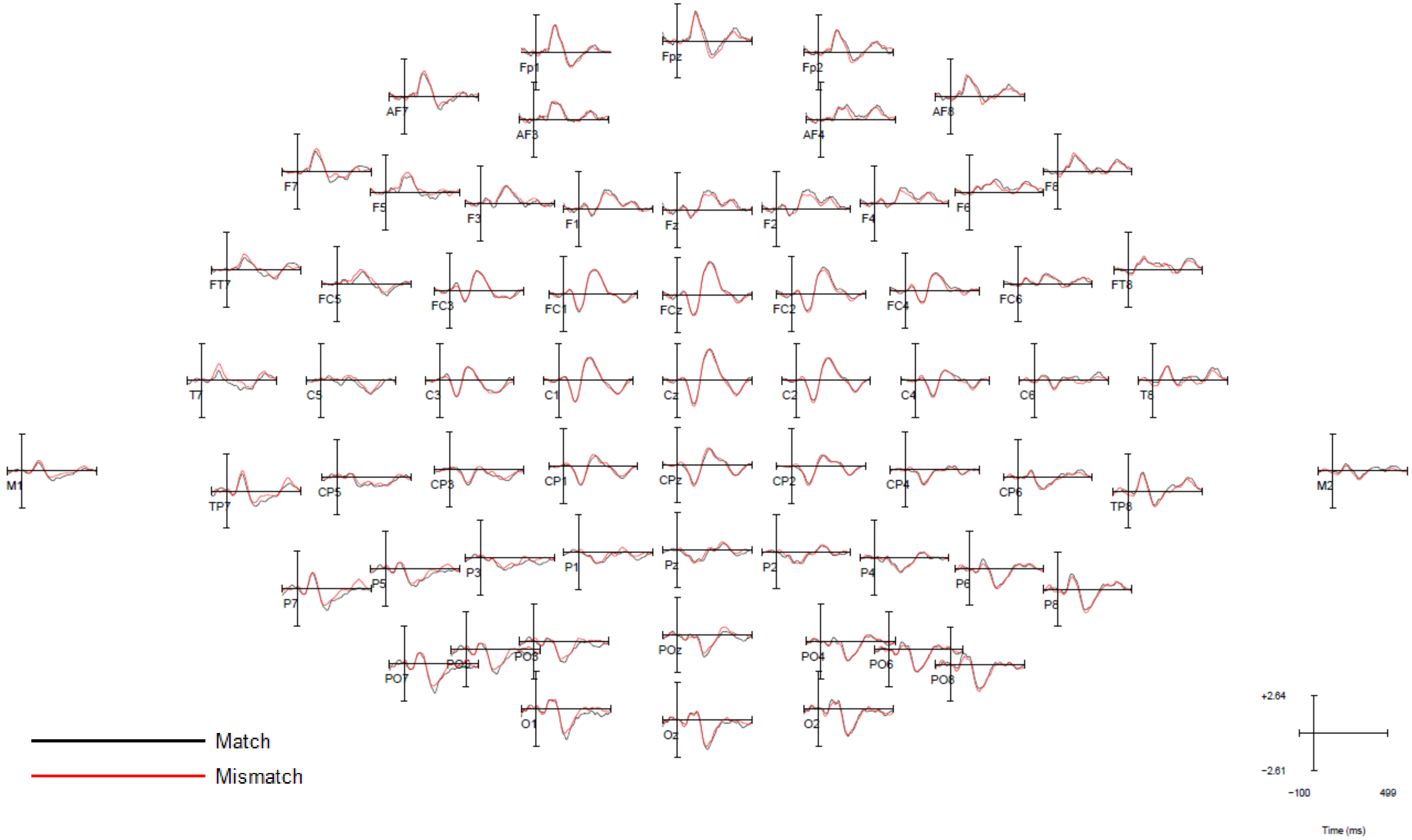

Figure 68. Grand average ERPs (N = 19) from Match (black) and Mismatch (red) trials, averaged over the two probe times.

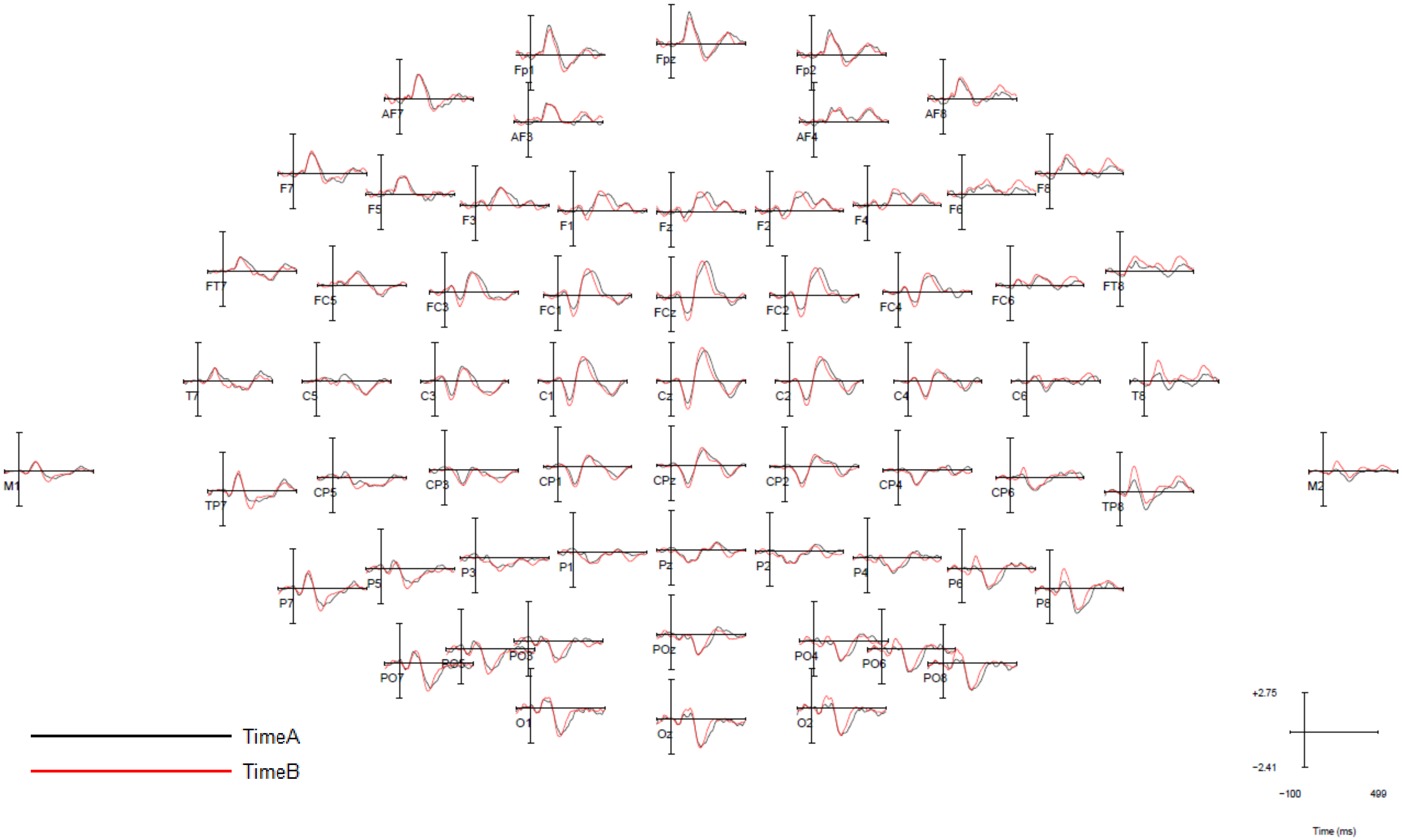

Figure 69. Grand average ERPs (N = 19) from Time A (black) and Time B (red) trials, averaged over the two probe types.

As previously mentioned, the T-complex was identified at the right and left temporal lobe, and occurred most prominently in sites T8, TP8, T7 and TP7 (see Figure 70). Peak amplitude of the Ta component was defined as the highest positive amplitude in the 75-125 ms time window. The aim of this analysis was to determine whether reliable ERP amplitude differences exist as a function of the probe type. In order to account for the possibility that these differences might be only be present at one of the two probe times and in one of the two brain hemispheres and channels, these additional factors were included in the analysis. Thus a repeated measures ANOVA was performed with peak Ta amplitude as the dependent measure and the within-subject factors of probe type (Match vs. Mismatch), probe time (Time A vs. Time B), hemisphere (left vs. right), and site (central temporal vs. temporal-parietal). Results showed a significant main effect of time, $F(1,18) = 12.61$, $p = .002$, reflecting a larger positive peak amplitude for Time B probes compared to Time A probes ($M_B = 1.84$, $M_A = 1.40$, $M_{\text{diff}} = -0.44$, $SE_{\text{diff}} = 0.12$, see Figure 70). There were no main effects of type, $F(1,18) = 1.28$, $p = .273$, hemisphere, $F(1,18) = .02$, $p = .878$, or channel, $F(1,18) = 4.01$, $p = .060$.

The time x hemisphere interaction was significant, $F(1,18) = 21.81$, $p < .001$, with a larger difference in Ta peak amplitude between Time A and Time B in the right hemisphere ($M_A = 1.13$, $M_B = 2.06$) compared to the left hemisphere ($M_A = 1.67$, $M_B = 1.61$, see Figure 70). The type x time interaction was not significant, $F(1,18) = 0.23$, $p = .637$, nor was the type x hemisphere interaction, $F(1,18) = 2.08$, $p = .167$. There were no significant interactions between channel and type ($F(1,18) = .01$, $p = .916$), time ($F(1,18) = 1.65$, $p = .215$) or hemisphere ($F(1,18) = 2.00$, $p = .174$).

For the three-way interactions, there was a significant interaction between type, hemisphere and channel on peak Ta amplitude, $F(1,18) = 4.51$, $p = .048$ (see Figure 71). This interaction was followed up with a simple effects analysis, which revealed that the interaction was driven by a significant effect of probe type in the left hemisphere and temporal channel (T7). The grand average Ta peak amplitude at T7 was significantly greater in the mismatch

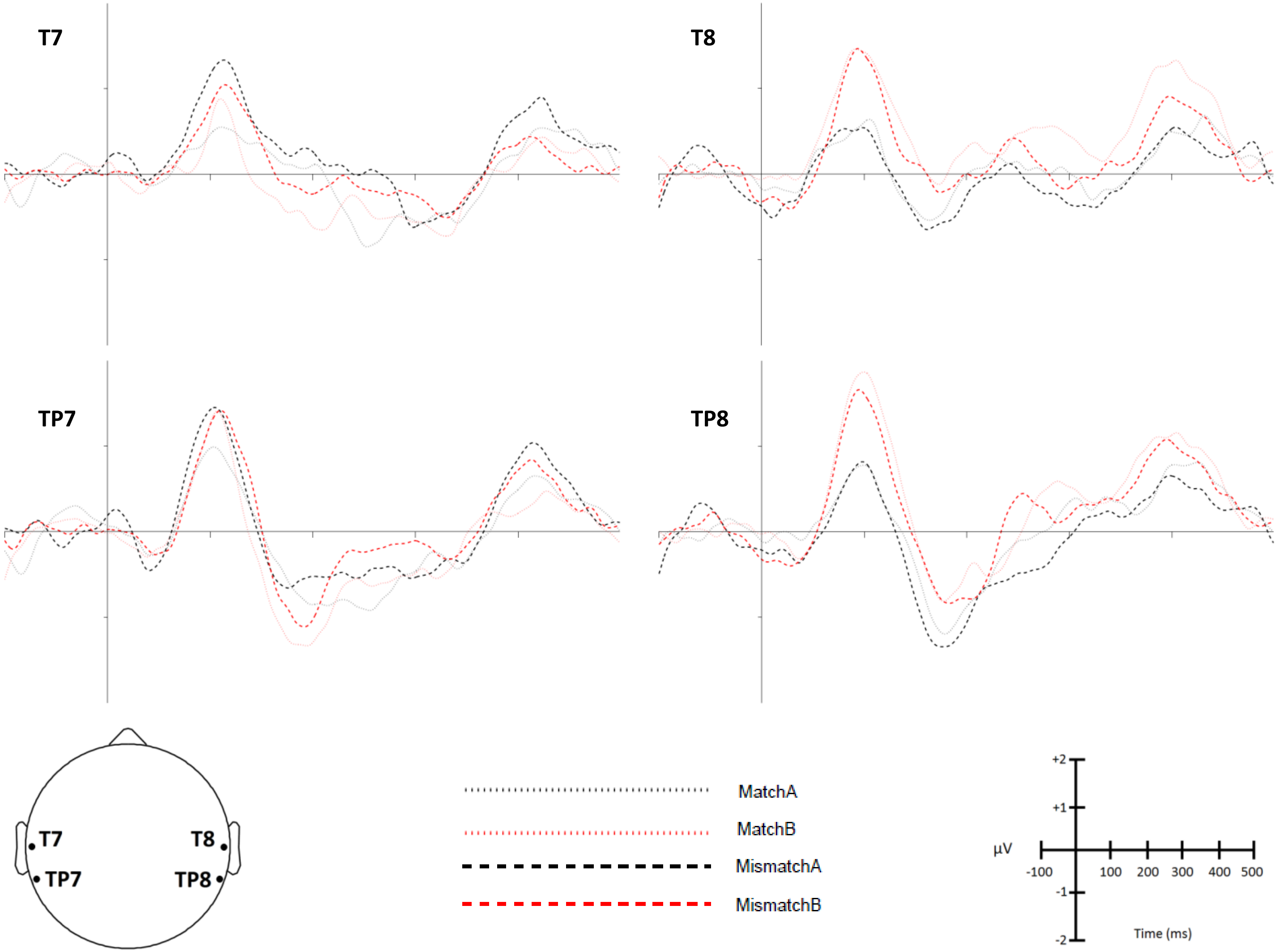

Figure 70. Grand average ERPs at left (T7, TP7) and right (T8, TP8) temporal sites. Line types and colours denote the four conditions.

condition compared to the match condition, (M_{match} = 1.22, $M_{mismatch}$ = 1.65, M_{diff} = -0.43, 95% $C.I._{diff}$ = [-0.82, -0.04], *p* = .034). The interaction between type, time and channel was not significant, *F*(1,18) = 0.02, *p* = .894, nor were the interactions between type, time and hemisphere, *F*(1,18) = 1.04, *p* = .322, and time, hemisphere and channel, *F*(1,18) = 0.33, *p* = .575.

The four-way interaction between type, time, hemisphere and channel was significant, *F*(1,18) = 5.00, *p* = .038. A simple effects analysis revealed that there were significant effects of probe type on mean peak amplitudes at Time A in both left hemisphere channels (T7: M_{diff} = -0.68, 95% $C.I._{diff}$ [-1.14, -0.22], *p* = .006; TP7: M_{diff} = -0.29, 95% $C.I._{diff}$ [-0.57, -0.02], *p* = .041). The interaction with the channel factor was due to the significantly larger peak amplitude at TP7 compared to T7 in the Match A condition only (M_{diff} = -0.52, 95% $C.I._{diff}$ = [-0.87, -0.18], *p* = .005). The effects of probe type on Ta peak amplitudes in the left hemisphere channels at Time A were not significant for the Time B probes. Figure 71 shows the significantly larger Ta peak amplitudes for Mismatch A compared to Match A conditions in the left hemisphere T7 and TP7 sites.

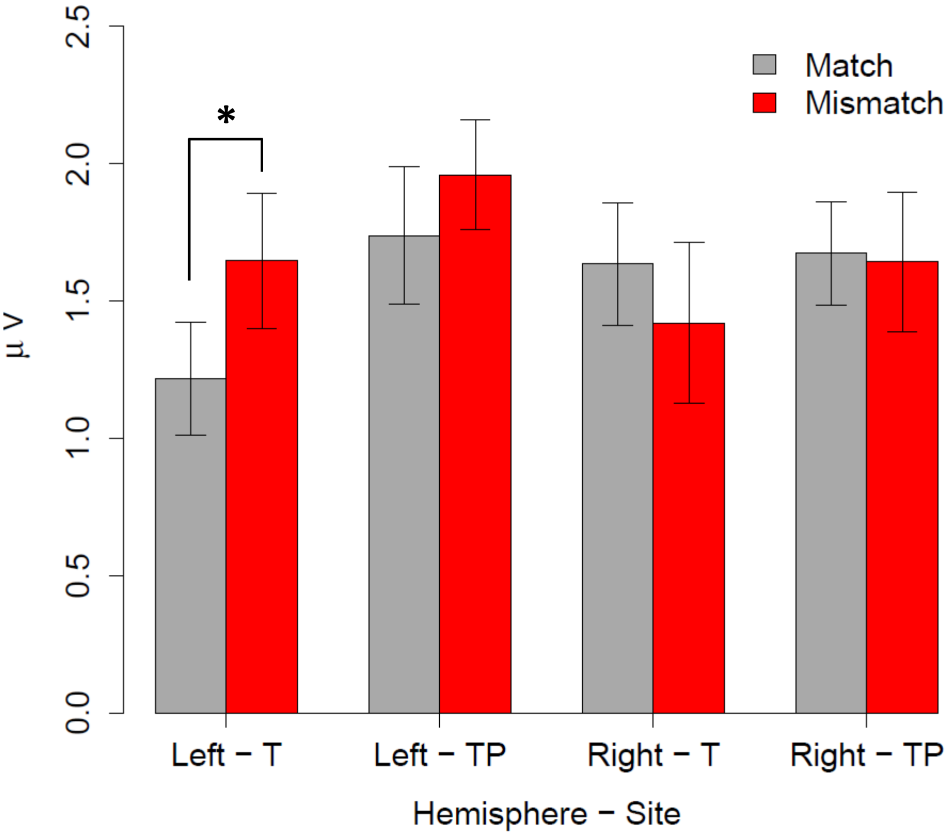

Figure 71. Positive peak amplitudes between 75-125 ms (Ta component of the T-complex) in ERPs to match (gray) and mismatch (red) probes in left temporal sites (T7, TP7) and right temporal sites (T8, TP8). Error bars represent 95% confidence intervals for between-participant variance. * denotes significance in within-participant comparisons.

To identify differences related to the probe type in the Tb component (negative deflection at temporal sites corresponding to the P2 at the vertex), a repeated-measures ANOVA was run on the negative peak amplitudes between 150 and 300 ms. There four factors: type (Match vs. Mismatch), time (Time A vs. Time B), hemisphere (left vs. right), and channel (central temporal vs. temporal-parietal). There were no main effects of time, $F(1,18) = 0.01$, $p = .919$, type, $F(1,18) = 1.34$, $p = .262$, or hemisphere, $F(1,18) = 1.21$, $p = .287$. There was a main effect of channel, $F(1,18) = 10.21$, $p = .005$, with more negative peak amplitudes in the temporal-parietal channels compared to the central-temporal channels ($M_{\text{temporal-parietal}} = -1.85$, $M_{\text{central-temporal}} = -1.21$, $M_{\text{diff}} = -0.64$, 95% $\text{C.I.}_{\text{diff}}$ [-1.06, -0.22]). There was also a significant interaction between type and hemisphere, $F(1,18) = 9.01$, $p = .008$. The mean negative peak amplitude was more negative in the Match condition compared to the Mismatch condition in the left hemisphere, but not in the right hemisphere ($M_{\text{match-left}} = -1.88$, $M_{\text{mismatch-left}} = -1.44$, $M_{\text{diff}} = 0.44$, 95% $\text{C.I.}_{\text{diff}}$ [0.80, 0.07], see Figure 72). No other interactions were significant.

Finally, in order to determine whether the effect of probe type was related to catch trial response accuracy, a supplementary analysis was conducted on ERPs averaged across participants who were split according to the median response accuracy value (36.7%). There were 9 participants in both the higher accuracy group (Mean accuracy = 68.5%, SD = 20.3%, minimum = 43.3%, maximum = 93.3%) and the lower accuracy group (Mean accuracy = 19.8%, SD = 10.8%, minimum = 1.67%, maximum = 31.67%), with the participant at the median excluded. This analysis showed that the ERPs differences between probe types in the higher accuracy group were amplified compared to the grand average ERPs, whereas the effect of probe type was reduced or eliminated in the lower accuracy group.

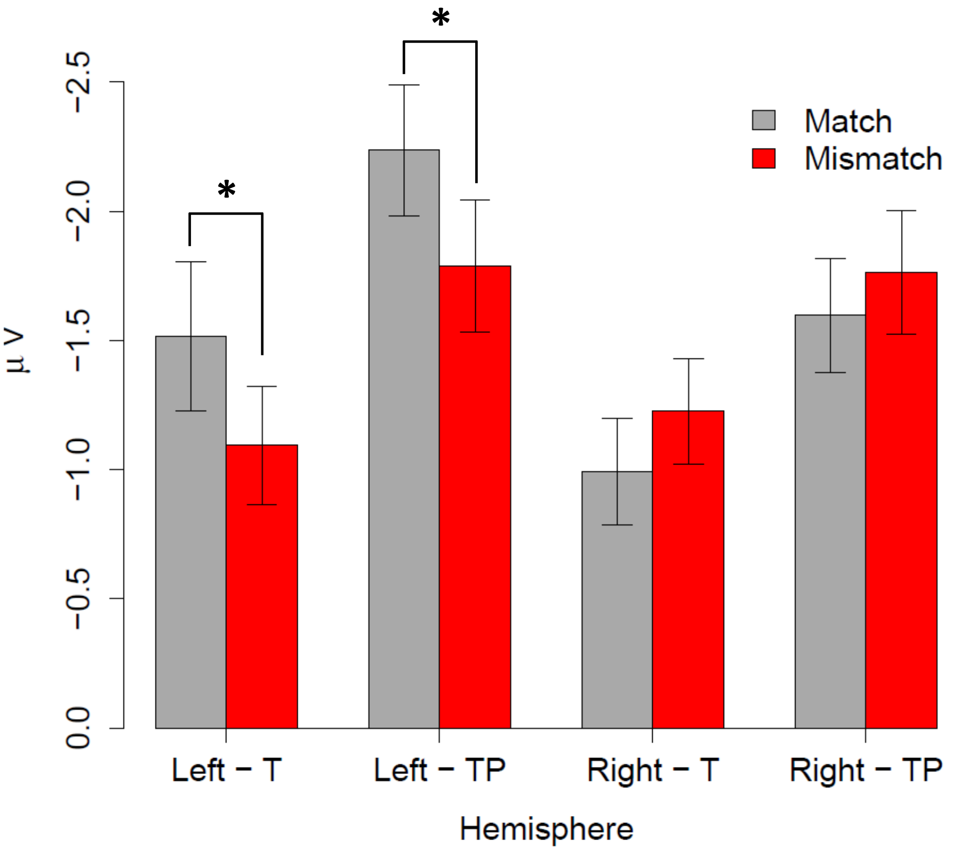

Figure 72. Negative peak amplitudes between 150-300 ms in ERPs to match (gray) and mismatch (red) probes in left temporal sites (T7, TP7) and right temporal sites (T8, TP8). Error bars represent 95% confidence intervals for between-participant variance. * denotes significance in within-participant comparisons.

To summarise the results of the ERP analysis, the Ta positive peak amplitudes (75-125 ms) to Match probes were significantly reduced compared to Mismatch probes in the T7 left hemisphere temporal site. In addition, the Tb negative peak amplitudes (150-300 ms) in response to Match probes were more negative compared to Mismatch probes in the left hemisphere. As to the effects of probe time, the later, Time B probes elicited larger Ta positive peak amplitudes compared to Time A probes, and this was most evident in the right hemisphere.

7.2.3 Frequency results

Figure 73 shows the log evoked power up to 15 Hz for the sequence presentation, rehearsal and baseline segments. The gray lines show the average over channels for each participant and the red lines show the grand average within each segment type. The frequency-power distributions reveal a few immediately noticeable differences among the trial segments. First, there are local peaks in power around 2, 4 and 6 Hz in the sequence presentation segment, and to a lesser extent, in the rehearsal segment. These local peaks appear most reliably in the individual data at 2 Hz, with more between-participant differences at the 4 and 6 Hz peaks. However, these three local peaks are not immediately apparent in the baseline segment. A second noticeable difference is that there are local increases in low alpha (8-10 Hz) power in the sequence presentation and rehearsal segments, but not in the baseline segment. Finally, there is greater power in the baseline condition compared to the sequence presentation and rehearsal segments, particularly at low frequencies, which is attributable to more broadband noise.

Because of the relative increase in log evoked power at low frequencies in the baseline segment, a direct comparison of log evoked power at 2 Hz would obscure the differences between segments. Moreover, the aim of the analysis was to determine whether there were significantly larger local peaks in 2 Hz log power in the sequence presentation and rehearsal segments compared to baseline. For that reason, the ratio of 2 Hz log power to the log power in the two neighbouring frequency bins (i.e. 1.47 and 2.45 Hz) was calculated as the dependent measure. Larger ratios reflect higher 2 Hz log power relative to that in neighbouring bins, and thus a larger local peak. On average, this ratio was highest for the sequence presentation segment (M = 0.58, SD = 0.03, 95% C.I. [0.56, 0.59]), followed by the rehearsal segment (M = 0.54, SD = 0.03, 95% C.I. [0.53, 0.56]), and then the baseline segment (M = 0.48, SD = 0.03, 95% C.I. [0.46, 0.50]). The distributions of these ratios across participants are shown in Figure 74.

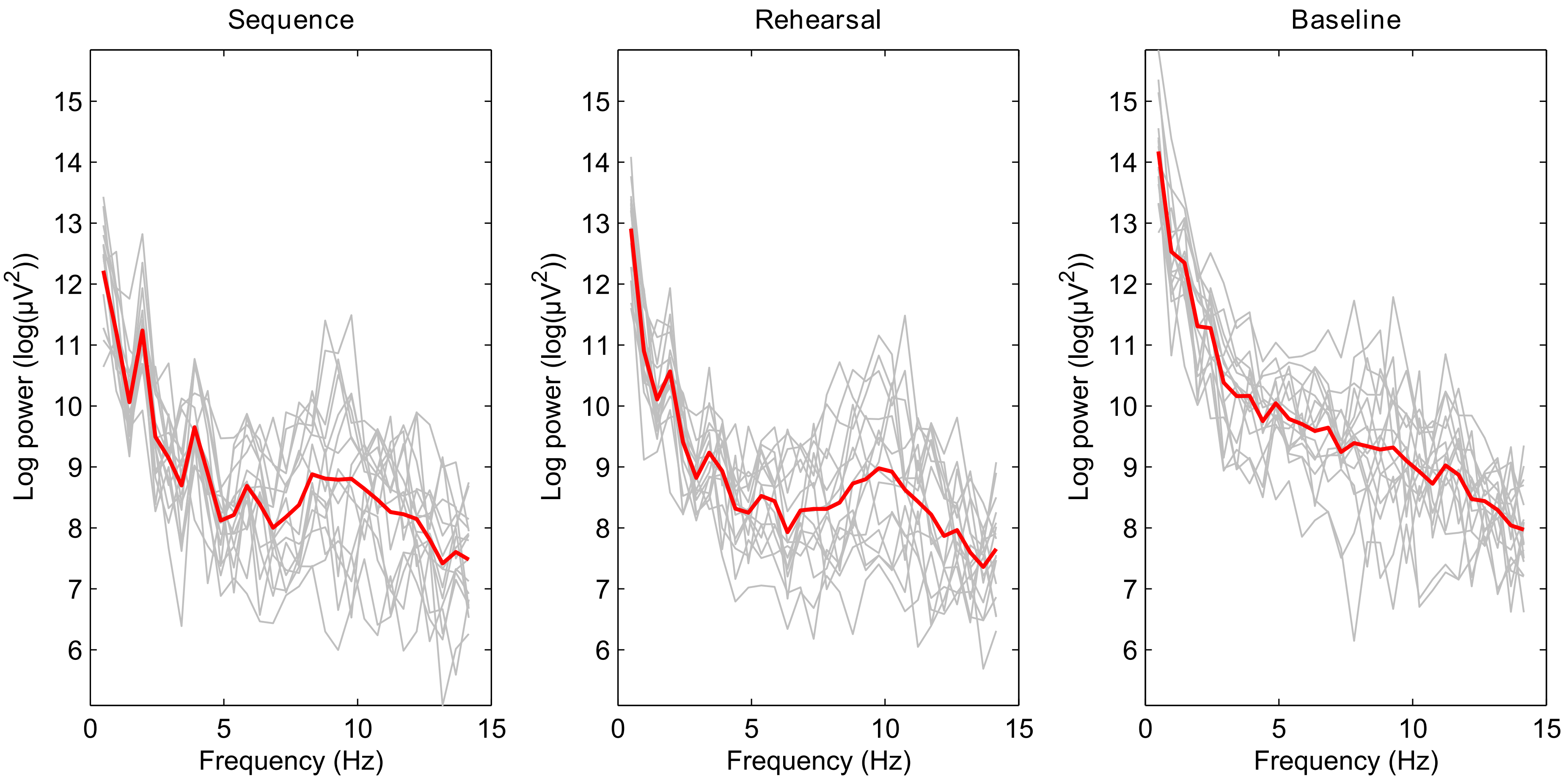

Figure 73. Log evoked power in the sequence presentation (left), rehearsal (middle) and baseline (right) trial segments. The gray lines show the individual participants' power averaged over all channels, and the red lines show the grand averages within each segment type.

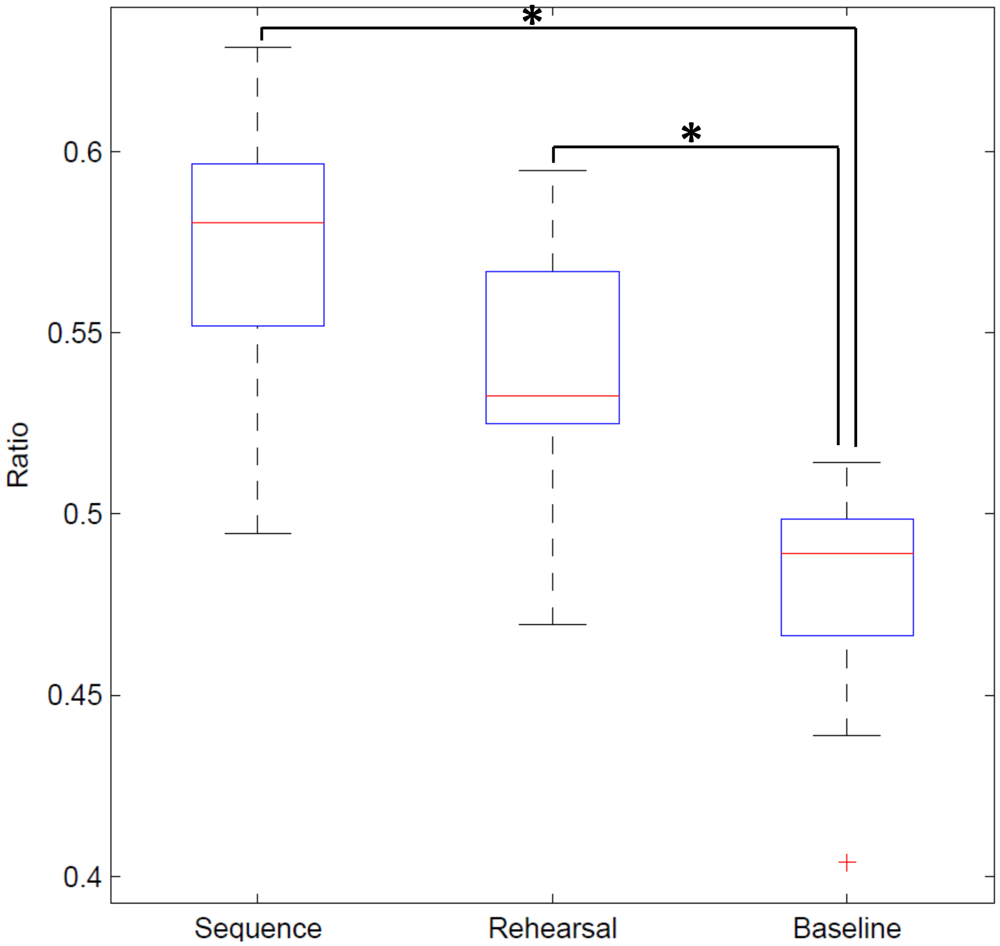

Figure 74. Distributions of ratios of 2 Hz log evoked power to log evoked power in the two neighbouring frequency bins (1.47 and 2.44 Hz) for 1.5 second segments of the sequence presentation period (left), subvocal rehearsal period (middle) and baseline period (ITIs, right). * denotes significant difference within participants.

The ratios of 2 Hz log evoked power to the log power in neighbouring frequency bins met the assumption of sphericity (Mauchly's test $p = .635$) for a repeated measures one-way ANOVA. The factor was trial segment with three levels; sequence presentation, rehearsal, and baseline. There was a significant effect of trial segment on the 2 Hz ratio, $F(2,30) = 31.77$, $p < .001$. Planned comparisons revealed that the 2 Hz ratio in the sequence presentation segment was significantly higher than that in the baseline segment, $F(1,15) = 59.82$, $p < .001$. The 2 Hz ratio in the rehearsal segment was also significantly higher than that in the baseline segment, $F(1,15) = 34.01$, $p < .001$.

To determine whether individual differences in the 2 Hz log evoked power ratios were related to any behavioural outcomes, correlations were calculated between the ratio data from the sequence presentation and rehearsal segments with the measures of response accuracy (percent correct) and response consistency (CSD). This analysis revealed that greater 2 Hz log evoked power ratios during the sequence presentation period were positively correlated with response accuracy, $r = .63$, $p = .008$. No other correlations were significant.

7.3 Discussion

Overall, this experiment demonstrated that the rehearsal-probe task can be used with EEG to detect neural markers for the contents of rehearsal, as well as local peaks in phase-locked oscillations at the rate of item presentation. These results show that the effect of probe match/mismatch type on ERPs exists even in the context of a rehearsal paradigm, where multiple words are held in memory and subvocalised cyclically. The difference in ~200 ms negative peak amplitudes to match and mismatch probes within each probe time is striking given that the only difference between these conditions was the serial position of the probe item. The use of a paced rehearsal paradigm also showed that, not only does phase-locking occur to the exact digit presentation rate, but a relative increase in 2 Hz phase-locked power can also be detected in the rehearsal periods.

While differences in ERPs as a function of the contents of inner speech have been previously reported, this experiment is the first to show that these differences can be detected in the context of a self-paced rehearsal paradigm, where multiple verbal tokens are held in memory. Consistent with the hypothesis, the ERP results showed small but significant within-participant differences as a function of the probe type. These differences emerge as early as ~100 ms after probe presentation in the left temporal Ta peak amplitude, and more reliably at ~200-250 ms in the left temporal Tb component. In particular, positive peak amplitudes in the left temporal Ta component at ~100 ms were suppressed for the earlier (Time A) Match compared to Mismatch probes, while these peak amplitudes were not significantly different between conditions in the later (Time B) probes. This earlier N1 suppression effect might have implications for our current understanding of the level(s) at which feedback expectations are generated from motor efference copies, given that the match process in this experiment occurred at the abstract phonological, rather than acoustic/feature level. It appears that temporal overlap is sufficient for eliciting this earlier effect, even without feature overlap. More speculatively, it might be possible that participants were 'subvocalising' the fine-grained features of the presented sequence, rather than speaking the same digits with their own inner 'voice'. In this sense, it is possible that some participants were engaging in a rehearsal process that was more like imagined hearing than imagined speech. In addition to the earlier ERP difference, negative peak amplitudes in the left temporal Tb component at ~ 200 ms were enhanced (more negative) for Match compared to Mismatch probes, and this effect was significant for both probe times. The results are consistent with previous results showing N1/N1m (i.e. Ta) suppression

(Numminen & Curio, 1999) and P2/P2m (i.e. Tb) enhancement (Tian & Poeppel, 2012) for match compared to mismatch probes.

One possible reason for the interaction between probe type and time for the earlier Ta component is that the subvocalised item representations were stronger or more precise earlier in the rehearsal interval, and became weaker or less precise as the rehearsal duration increased. The ERP data support this hypothesis in that the left temporal peak amplitudes for Match probes at Time A were significantly reduced relative to both the Mismatch probes at Time A, and to both probe types in left temporal channels at Time B. That is, under the hypothesis that the effect of probe type was not clearly observed at Time B due to a weakening of the Match effect for the later probes, then the Match B responses should be similar to the Mismatch responses, and indeed this was the case (see positive peaks in left hemisphere sites in Figure 70).

The effect of probe type on ERPs was no doubt weakened by inaccurate rehearsal timing in some participants. While some participants responded to catch trials with the correct item, others tended to respond incorrectly (usually with the immediately preceding item). The degree of response consistency was also variable between participants, and together these two measures characterised most of the response data; highly consistent and highly accurate, highly consistent but highly inaccurate, and inconsistent and somewhat inaccurate. The probe type manipulation should have worked as intended for participants with consistently accurate responses, so the effects of probe type on ERPs were probably driven by these individuals. Participants who were consistently inaccurate in their responses to the catch trials were likely not rehearsing the presented probe item on the Match trials, meaning that the probe type manipulation may have been invalid. Inconsistent responses to catch trials likely reflect more variable timing, where the participant was sometimes rehearsing the correct item on Match trials and sometimes not. In this case, any effect of the probe type manipulation on ERPs may have been present but dampened by the inclusion of trials in the Match condition that were, in fact, mismatch. This interpretation was supported by the results of a supplementary analysis involving a comparison between participant groups split on response accuracy. The higher accuracy group showed an increased difference compared to the grand average, and the effect of probe type was reduced or eliminated in the lower accuracy group. This supports the interpretation that the effect of probe type on ERPs was driven by the contents of rehearsal, and suggests that the true effect size is likely to be greater than that observed in the present study due to the inclusion of ERPs to 'Match' probes that were not actually aligned with the subvocal matching item as intended.

Because the effect of probe type on ERPs is small, and because precise alignment between rehearsal and probe timing is crucial for a Match condition, future experiments using the rehearsal-probe task to detect changes in ERPs as a function of probe type would benefit from measuring individual differences in rehearsal timing properties behaviourally, before the EEG task. This would allow the use of behavioural timing measures to increase sensitivity, reliability and validity of the probe type manipulation. For instance, after analysing the behavioural data, the probe times could be tailored to match predicted rehearsal timing at the individual level.

In addition to the effect of probe type, there was also a clear effect of probe time on the ERP data. The main effect of probe time was predominantly over the right hemisphere temporal and posterior sites, indicating that the neural generator of this difference is separate from that responsible for the effect of probe type. The effect of probe time was most likely a result of the difference in predictability, where the onsets of the later (Time B) probes were more predicable than the earlier (Time A) probes. This is because a probe occurred on every trial and they were restricted to a known period, which means that, as rehearsal time elapsed past the point of the Time A probes (1.5 seconds), the expectation for probe onset became higher. Moreover, some participants may have more explicitly learned that there were only two possible probe times, and for these individuals the effect would be even greater. While these results were not explicitly hypothesised, they are consistent with previous ERP data on temporal predictability (Poeppel, 2003). In future work, the effect of probe time on ERPs could be reduced or eliminated by 1) adding trials where no probe is presented (though there would have to be another cue to signal the end of the trial), 2) introducing a range of probe times in order to reduce the predictability. However, the predictability of probe times is not a major issue for the rehearsal-probe match/mismatch paradigm because the effects of probe type and probe time appear to be separable.

The frequency analysis showed that, in both the sequence presentation and rehearsal intervals, there was an overall reduction in evoked (phase-locked) log power at lower frequencies with peaks at the presented sequence rate relative to log power in the two neighbouring frequency bins. Moreover, these peaks were detected at the level of the whole brain by averaging over all channels, so these results may be a conservative measure of the true effect. In contrast to the sequence presentation and rehearsal periods, a local peak at 2 Hz was not present in the baseline interval. These results are consistent with previous work showing neural entrainment to low frequency rhythms in speech perception, and this experiment is the first to show that phase-locked rhythmic activity can be detected during a paced subvocal rehearsal task.

In future work, it would be possible to modify the sequence presentation rate to determine whether the local peaks in evoked log power track the rate change. This would help to determine whether the local peaks are a result of general cognitive factors related to both sequence encoding and rehearsal (e.g. memory, attention) or artifacts of other spectral differences (e.g. increased power in the alpha band), rather than reflections of neural activity related to the prominent rhythms in auditory sequence stimuli as well as the pace of subvocal rehearsal. Furthermore, the frequency design could be improved by extending the rehearsal and baseline periods in order to get better estimates of low frequency power. The use of source localisation techniques and connectivity analyses would also be useful for determining the neural generators of the oscillatory activity at the rate of sequence presentation and subvocalisation.

In addition to the group analysis, this experiment also revealed interesting results at the individual differences level. The relative log power at the sequence presentation rate was significantly positively correlated with response accuracy. This result shows that the degree of neural entrainment to low frequency auditory-verbal timing is related to the accurate reproduction of sequence timing during rehearsal, which suggests that the perception of timing during sequence encoding may be particularly important for temporal accuracy during rehearsal. Under this hypothesis, it would be predicted that factors that have been previously shown to affect rehearsal timing (e.g. memory pre-loading, number of items) might also exert an influence on the strength of neural entrainment during sequence encoding.

Finally, response variability at Time B probes was significantly negatively correlated with individual differences in auditory-verbal STM capacity for serial order. This finding is worth noting as it is consistent with the negative correlations observed in previous experiments between CSD and digit span. This result also shows that meaningful individual differences in response variability are observed with only two probe times and multiple trials for each time (as opposed to many probe times and fewer trials per time, as in the previous experiments).

7.4 Chapter Summary

The use of the rehearsal-probe task with EEG recording demonstrated the ability to detect neural markers of both the content and timing of inner speech in the context of self-paced sequence rehearsal. This was accomplished via the timed presentation of match/mismatch item probes and the analysis of the phase-locked low frequency spectra of duration-matched sequence presentation, subvocal rehearsal and baseline (ITI) intervals. ERPs to match compared to mismatch item probes revealed an early (~100 ms) suppression in a left temporal site after shorter rehearsal delays, as well as a later (~200 ms) enhancement in left temporal channels after both shorter and longer rehearsal delays. Evoked log power in the sequence presentation period showed a relative peak at the 2 Hz item presentation rates, and importantly, this peak was evident in the rehearsal period but absent in the baseline interval. Further, the magnitude of the 2 Hz peak during sequence presentation was related to rehearsal timing accuracy, showing that neural entrainment to slow speech rhythms during STM encoding may be an important contributor to timing accuracy in STM, and by extension to serial order recall. Overall, the experiment has shown the promising potential of the use of the rehearsal-probe paradigm with EEG for detecting neural correlates of subvocal rehearsal in memory experiments, as well as contributing toward an understanding of the brain basis of auditory-verbal STM for order and timing.

CHAPTER 8

GENERAL DISCUSSION

The rehearsal-probe task provides the first experimental paradigm for the detection of subvocal rehearsal and measurement of its temporal properties. Performance in the novel task was established across multiple experiments with reliable and consistent results. Temporal precision in particular was responsive to changes in item-order familiarity, memory load, and rhythmic grouping (Chapter 3). In addition, timing variability for subspan sequences was found to be closely coupled to auditory digit span performance, exhibiting a negative correlation which was replicated in multiple experiments. As hypothesised, rehearsal timing was more variable in groups of adults and children with developmental language disorders compared to age- and IQ-matched controls, and this group difference was closely linked to individual and group differences in auditory STM span. Computational modelling of the data revealed that the evolution of response distributions over time can be described using four theoretically motivated temporal parameters. An analysis of the distributions of these parameters across experiments provided new information about the timing of paced rehearsal in typical adults, and suggested new avenues for behavioural and computational research. Finally, the use of the task with EEG shows that paced rehearsal can be detected at the level of brain activity, for instance through the increased prominence of the rehearsal rate in the frequency spectrum during the maintenance interval, and through evoked responses to presented matched or mismatched verbal probes.

The primary finding that emerged through the use of this task is that changes in the temporal properties of subvocal rehearsal are consistent with the notion that STM for timing is resource-limited. A negative association between individual differences in auditory-verbal STM span and timing precision for subspan sequences was firmly established across several experiments, including one study designed specifically to test this hypothesis with a larger sample and control tasks (Expeirment 5, Chapter 4). The relationship between timing precision and STM span can be explained by assuming that the precision of sequence timing is a function of the available, limited cognitive resources, where individuals with lower STM spans will have less available resources compared to individuals with higher spans given the same STM load. That timing precision changes dynamically with the available resources was also supported by the increase in within-participant timing variability in response to increased memory load in the most convincing test of this question, Experiment 3 (Chapter 3), where memory load was manipulated without confounding alterations to the subspan rehearsal sequences.

Three other experiments in Chapter 3 were designed to manipulate memory load but instead indicated sensitivities of the dependent measure to confounding factors. In retrospect, the use of different presentation rates (Experiment1), over-learned sequences (Experiment 2) and temporal grouping patterns (Experiment 4) may have introduced additional factors that biased the results. Nonetheless, all of the experiments designed to test memory load demonstrate that timing precision is not invariant to changes in the content and within-sequence timings of sequences matched for duration and number of items. These results would not be predicted under the hypothesis that timing precision during subvocal rehearsal is a) unrelated to the STM content, or b) fixed when STM load is below span, only deteriorating once a load threshold has been reached. This latter view is analogous to the fixed-slot models of visual WM, which propose that a certain number of visuospatial objects can be held in memory with fixed precision, and that errors occur when the to-be-remembered material exceeds the number of slots (see e.g. Zhang & Luck, 2008). Similarly, the finding that timing precision is responsive to memory load may also pose a problem for models of STM for serial order that postulate a ‘positional’, discrete coding system (as discussed in more detail in the next section).

While the set of seven behavioural experiments (Chapters 3-5) generally showed that the rehearsal-probe task and CSD measure provide a useful method for examining rehearsal timing, some of these experiments also served to highlight additional considerations for future research with the task paradigm. Experiments 1, 2 and 4 in particular revealed certain limitations to the flexibility of the task, namely the difficulty manipulating load separately from other important sequence factors such as list duration, number of items and item rate. The ideal design for experiments using the rehearsal-probe task is one that compares timing for sequences that are matched for the following characteristics: list duration, number of items, presentation rate, and item-order predictability. In addition, the probe times should be matched across the tasks to be compared in order to avoid biasing the mean CSD dependent measure. While this presents a challenge to experimental design, the successful matching of these factors across conditions in Experiment 3 shows that it is not an insurmountable issue. Finally, careful consideration should be taken when using auditory sequences with different rhythmic patterns, whether within the rehearsal-probe task or in other STM paradigms, with particular attention paid to the potential benefit for metrical structure and detrimental effect of irregular or unstructured timing.

The application of the rehearsal-probe task toward understanding developmental language disorders was addressed in two experiments (Chapter 5) comparing timing precision in groups of adults with dyslexia (Experiment 6) and children with SLI (Experiment 7) to age- and IQ-matched controls. As predicted, the groups with developmental language disorders showed increased timing variability for sub-span sequences. These findings have implications for current debates about the primary cognitive deficits in developmental language disorders, as well as their causal/risk factors. Namely, the results are difficult to reconcile with the view that phonological impairments and verbal STM deficits characteristic of individuals with developmental language disorders can be entirely explained by primary impairments in the quality or phonological representations (Snowling et al., 1991; Snowling, 1998), or in access to these representations (Ramus, 2003). The results from Experiments 6 and 7 are instead consistent with a growing body of evidence suggesting abnormal auditory temporal processing in these disorders (Goswami, 2011; Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014), which may be causally related to the co-occurring deficits in STM for serial order.

Preliminary computational modelling work (Chapter 6) has demonstrated that the response proportion data across experiments can be described using four separable temporal properties with minimal assumptions about the underlying mechanisms. While the modelling work is still at an early stage, it has shown an ability to capture the main attributes of the data and quantify the rehearsal timing properties in individual and group data sets. Continued work on modelling this data is a worthwhile endeavour given the potential for determining the specific temporal properties that might vary meaningfully between individuals and groups, and for generating testable hypotheses about response proportions at any given probe time. Also, the development of a robust model could lead to use as a test bed for evaluating alternative experimental designs. For instance, specific experimental designs could be simulated with realistic parameter values (e.g. those estimated by the meta-analysis in Chapter 6), and the simulations could inform decisions about the set of probe times that would be most sensitive to hypothesised effects. In future work, the use of improved experimental designs and modelling to estimate population-level temporal properties of rehearsal could prove useful for identifying how certain groups might differ in one or more aspect of rehearsal timing.

Finally, the rehearsal-probe task's potential in contributing toward the identification of neural correlates of subvocal speech was demonstrated in Chapter 7, where the behavioural task was combined with EEG recording. This experiment showed that neural markers for both the contents and timing of inner speech can be identified in the context of a paced cyclic rehearsal task. Differences in event-related responses to match versus mismatch item probes were identified in the left temporal lobe sites. Also, local peaks in phase-locked power spectra at the item presentation rate were seen in sequence presentation and rehearsal periods, but these were absent in the baseline period. This result is suggestive of neural oscillatory signatures that are directly related to the paced, cyclical subvocal rehearsal process. In addition, the magnitude of local spectral peaks during sequence presentation was significantly related to response accuracy, meaning that individual differences in the fidelity of timing information during STM maintenance might originate in the encoding process. While there was no direct link observed between the EEG measures and timing variability, the experiment was not designed to detect individual differences. Thus an important next step is to replicate the effects in future experiments and determine which aspects of the EEG, if any, relate to individual differences in timing variability and/or STM span. The neural correlates might prove useful for detecting subvocal rehearsal in less constrained STM tasks, and potentially serve as markers of dysfunction relevant to disorders of language development.

8.1 Implications for Models of STM for Serial Order

Given these empirical findings, a critical question that follows is what drives the effect? That is, what are the mechanisms that underlie changes in temporal precision for subspan sequences as a function of available resources? This question can be addressed within the framework of contemporary models of STM for serial order, the dominant computational accounts of which employ some form of parallel item activation[24] with a competitive queuing (CQ) process (e.g. Brown et al., 2000; Burgess & Hitch, 1999; Hartley & Houghton, 1996; Henson, 1998; Page & Norris, 1998, see Hurlstone et al., 2014 for a recent review). The general principle of this class of models works as follows. First, when auditory-verbal items are presented they activate their own representations (i.e. item nodes). At the same time, these item activations are associated with a changing time- or event-based context signal. During this encoding process, adjacent sequence items will be associated with similar states of the context signal. First and last sequence items will have fewer neighbouring items and thus they will be more uniquely associated with the beginning and end of the signal. Emergent properties of the mechanism account for the frequency of transposition errors (particularly for mid-list items), primacy and recency effects.

At the retrieval stage, the context signal is replayed and the items are re-activated with their associated strengths. Items compete for selection at the recall via the strengths of their activations, where the item with the strongest activation is selected with the highest probability, and then suppressed to prevent continued selection (see Figure 75). The item with the second strongest activation is selected with the second highest probability, and so on. There are a few steps in this process where noise can be introduced to account for imperfect recall, which occasionally occurs for span- or subspan-length sequences.

---

[24] See Lashley (1951), Houghton and Hartley (1996) and Henson et al. (1996) for arguments related to parallel item activation in serially-ordered recall.

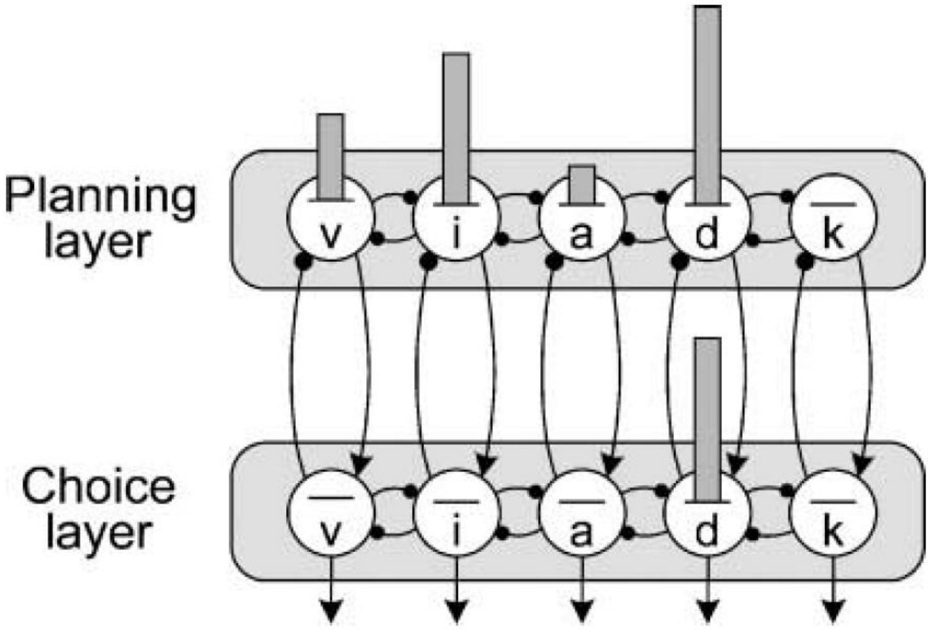

Figure 75. Representation of the competitive queuing process from Bohland et al. (2010). Item nodes are activated in parallel ("Planning layer") by a context signal (not shown), and these activations are fed into a competitive "choice layer". In the competitive choice layer, item activations are subjected to a winner-takes-all process, where only the item with the strongest relative activation remains active and selected for output. Arrows show excitatory connections and filled black circles show inhibitory connections. This example shows the representation of the letter sequence "diva", where the items are activated according to their relative associations with an early state of the context signal. Thus, the first letter ("d") is most active in the planning layer, followed by "i", "v" and "a". The letter "d" is selected for output and its item node is suppressed via the inhibitory feedback from the choice layer. This process continues whereby changes in the context signal cause changes to item activations until all items have been recalled.

Parallel activation models of serial order are supported by both behavioural and neuroscientific evidence. For instance, the presence of anticipation errors (i.e. responding with a sequence item at an earlier serial position) in ISR suggests that STM items are activated in parallel in the first stage of recall, followed by a determination of their order (Lashley, 1951). This is in contrast to serial item activation, as posited by earlier simple associative chaining models, where each item serves as a cue for the next item. Complementing theoretical arguments for the existence of parallel item activation, a study using intra-cranial recording of neuronal electrical activity in the prefrontal cortices of monkeys showed that movement sequences were associated with distinct patterns of neural activity (Averbeck, Chafee, Crowe, & Georgopoulos, 2002). As shown in Figure 76, the relative strengths of these ensembles of activity revealed dynamic changes in prefrontal representations corresponding to each movement over time when the monkeys drew serially-ordered line segments. Note that the activations of each movement representation, shown as coloured lines in Figure 76, were activated in parallel before the execution of the sequence at time 0, where the relative strengths of activations specified the position within the sequence. Also, these activations are non-zero before and after reaching peak strength, and there is considerable overlap among the activations. This is in contrast to a discrete onset of item representations that would be predicted from a serial activation system, such as simple chaining. Also worth noting is the striking resemblance between these single-neuron recordings and the serial position response proportion functions generated by the rehearsal-probe task.

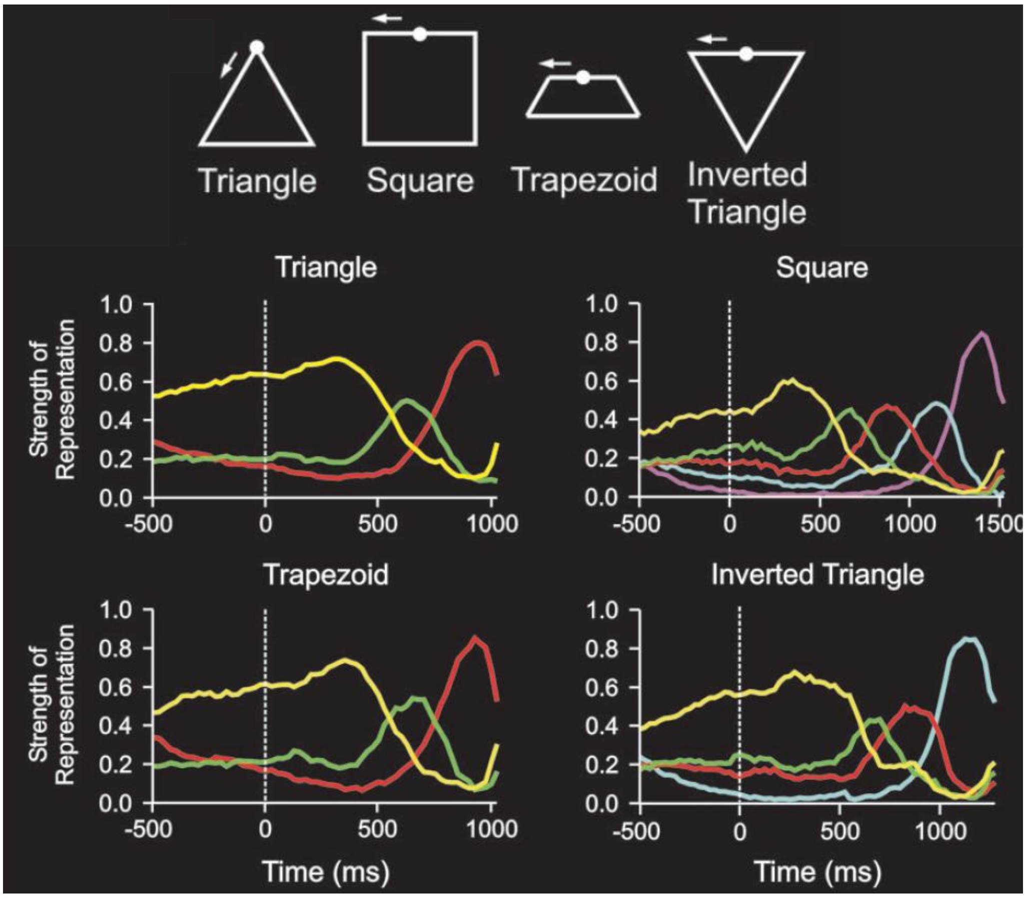

Figure 76. Strengths of line segment representations over time derived from single-neuron recordings in monkey prefrontal cortex from Averbeck et al. (2002). Top panel shows the shapes drawn by monkeys starting at time 0; circle indicates the starting point and arrow shows the direction of copying. Bottom plots show the relative strengths of serially-ordered line segment representations before and during the motor sequence execution for each shape.

Within this framework of STM for serial order, it is possible to consider more specifically how limitations on available resources might lead to decreased temporal precision. Resource limitations could reduce the precision of the context signal itself, or increase the amount of noise present during the formation of associations between items and the context signal and/or the reactivation of items when the context signal is replayed. That is, when memory load increases within an individual, or when memory load is effectively higher in individuals with lower STM capacities compared to those with higher capacities, then there may be fewer resources available to represent precision in the context signal. Alternatively, the relative decrease in available resources might introduce more noise into item activations either at encoding or during retrieval. When the context signal is replayed at retrieval, the coarser context signal and/or increased noise would result in more overlap among item activations. Consequently, this would mean less distinction among items activations at any given time, and therefore more variability in responses around the mean times associated with each item. Given enough imprecision in the signal or noise in the item activations, this model would predict an increase in serial order transposition errors.

This attempt to understand how the rehearsal-probe data relates to current models of STM for serial order reveals two critical issues. First, it is not clear how rehearsal should be instantiated within parallel activation plus CQ models. The ability to reconcile the rehearsal-probe data with individual and classes of models depends on exactly how the rehearsal process is implemented. Unlike overt ISR, which is the outcome these models seek to explain, rehearsal is covert and therefore is not necessarily equivalent to overt retrieval in all respects. One possibility is that subvocal rehearsal involves the cyclical reinstatement of the context signal, without the discrete item selection process (via CQ) that occurs during overt retrieval. If this is the case, then the rehearsal timing data could be understood as a proxy for the underlying dynamically-changing item activations driven by the context signal.

Another possibility is that cyclic rehearsal should be represented as repetitions of the complete retrieval process, where each item is selected for (subvocal) recall. In this case, the relative strengths of item activations would be fed into the competitive choice layer, where discrete items are selected serially throughout the rehearsal period. By this account, it would also be possible to make serial order errors during rehearsal, and (according to some models) each item would be suppressed after selection. It seems as though this implementation of rehearsal would predict less overlap among serial position response probabilities over time, especially after the item has been ‘recalled’ (selected) during rehearsal because its activation should then be immediately suppressed and the response probability should drop to 0. Instead, the observed response probabilities clearly show considerable overlap among positions and a gradual decline in response probability for each position after its peak. Therefore, at present, the data appear to be more consistent with the view that rehearsal does not involve a discrete item selection process, and that the response data are a proxy for the dynamic parallel activations of items over time. More generally, the consideration of this issue reveals the need for more explicit instantiations of subvocal rehearsal in these models. One exception to the general failure to address rehearsal is the Burgess and Hitch (1992) model, which treats rehearsal as repeated retrieval via CQ with item suppression.

Computational modelling of these alternative rehearsal processes would produce concrete predictions that could then be compared in terms of their relative fit to the rehearsal-probe data.

When considering how the rehearsal-probe data can be accounted for by models of serial order, a second issue that arises is related to the nature of the context signal. Models can be generally classified according to the use of position-based versus time-based coding signals, where the former refers to a discrete advancement of the signal linked to items/events, while in the latter, the signal change is continuous. Examples of these types of signals for positional coding are shown in Figure 77. Models that assume position-based coding include, for example, C-SOB (Lewandowsky & Farrell, 2008), the Primacy model (Page & Norris, 1998), the Start-End model (Henson, 1998), Houghton (1990), and Burgess & Hitch (1992). Models that use time-based coding include OSCAR (Brown et al., 2000), BUMP (Hartley, Hurlstone, & Hitch, unpublished), and later versions of the Burgess & Hitch model (1999, 2006). In models with position-based coding, the presented items drive the changes in either the state of the context signal (e.g. Burgess & Hitch 1992, the Start-End model) or in the relative strength of item activations (e.g. the Primacy model). Consequently, precise temporal information about the sequence is not retained with a position-based coding system. If a subspan sequence is encoded according to discrete positional steps, it is not clear how load-dependent differences in temporal precision could emerge, or indeed how precise timing could be represented at all.

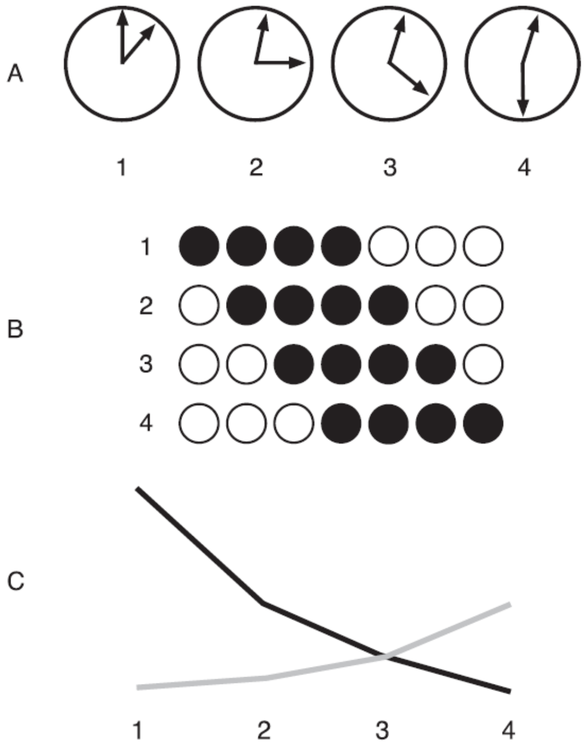

Figure 77. Examples of time- and position-based context signals from Hurlstone et al. (2014). The OSCAR model (Brown et al., 2000) employs a temporal coding system, where a sequence is encoded according to the phases of a set of endogenous oscillators. This system is represented with a clock analogy (A). Burgess and Hitch conceptualised the context signal as a 'moving window' of partially-overlapping sets of activations, as shown in (B). In the 1992 model this signal was event-driven (i.e. position-based), meaning that the window moved forward with each item, whereas in the more recent versions (1999, 2006) the signal moves forward with the passage of time. Positional coding can also be achieved with gradients, such as those used in the Start-End model (Henson, 1998), where items are encoded according to their *relative* distance from the start and end of a sequence.

By contrast, the models using a continuous time-based signal to encode item order would more easily account for the rehearsal-probe results because a gradually-changing signal allows scope for varying levels of precision. With a time-based signal, differences in temporal precision might emerge, for example, due to varying levels of resolution in the signal during encoding and/or recall. When an individual has relatively fewer resources available, either compared to other individuals with greater capacity or to other tasks with reduced memory load, then the resolution of the timing signal might be represented with lower resolution. This is unlike position-based coding signals, which only predict increased serial order errors with increased noise in item activations, and which do not provide a mechanism for the encoding of fine-grained item timng. Therefore a time-based coding system appears to be necessary so that sequence timing can be encoded and recalled in a continuous and dynamic fashion, where temporal precision can vary as a function of the available resources.

To conclude, the most straightforward mechanistic explanation of the response proportion data seems to be that they are a proxy for the underlying parallel item activations in memory, where activations are driven by a gradually-changing context signal. This interpretation is certainly tentative given that the rehearsal-probe responses are clearly not a direct measure of the purported item activations, and because a discrete response is required in response to the probe. Considerations about the implications of the rehearsal-probe data for models of STM for serial order have highlighted a need for more explicit implementations of the subvocal rehearsal process. This is important because the abilities of current models to account for the rehearsal-probe data is likely to vary depending on 1) the precise assumptions about rehearsal (e.g. with a discrete item selection process versus without), and 2) how the context signal is represented (e.g. via position- versus time-based signals). Future work with computational modelling could clarify these issues by making more concrete predictions about the rehearsal-probe data under specific assumptions about subvocal rehearsal and serial ordering mechanisms.

8.2 Future Research

The rehearsal-probe task is a promising method that, thus far, has only been used with a restricted set of items (digits) and with predictable timings. One exception to the non-isochronous sequence timing was in Experiment 4, but this grouping pattern was nonetheless predictable as participants practiced rehearsing the 3-3 grouping pattern before the task, and the same pattern was used on each trial. Therefore an obvious extension of the rehearsal-probe paradigm would be with different auditory-verbal or non-verbal phonological stimuli, and with unpredictable item presentation timings. These two avenues for future work would determine the extent to which the results reported here generalise to different types of verbaliseable stimuli, and the relative influence of bottom-up (stimulus driven) versus top-down (attention driven) factors on STM for order and timing.

For example, timing precision could be measured using lists of non-words. Non-words are of particular interest because they simulate a naturalistic word learning situation, where novel sequences of sublexical units (phonemes/syllables) must be remembered in the correct order without recourse to lexical representations in LTM and without semantic associations to support retrieval. Word learning is thought to be heavily influenced by auditory STM for serial order (Baddeley et al., 1998; Gupta & MacWhinney, 1997; Majerus & Boukebza, 2013) and auditory rhythmic grouping (Boucher, 2006), so the study of covert rehearsal for non-word sequences of verbaliseable material (e.g. syllables) has the potential to reveal even greater differences in rehearsal timing related to memory load, individual STM capacity, and language abilities. Furthermore, the application of the rehearsal model could help determine which temporal properties might be particularly affected by the lack of support from lexical and semantic networks, and by extension, what aspects of auditory STM for timing are most important for vocabulary acquisition.

Unpredictable item timing would present a more ecologically valid test of auditory-verbal STM for order and timing as it relates to speech and language development. This is because speech rhythms tend to be only quasi-rhythmic (Peelle & Davis, 2012), and timing for the same string of words can vary widely according to, for instance, the speaker and semantic intent. The experiments presented in this thesis cannot determine whether timing precision during rehearsal reflects the ability to use prior knowledge about sequence timing to direct the focus of attention and to guide rehearsal timing, or whether it reflects lower-level processing of the stimuli timing on each trial. Future experiments using unpredictable timing patterns could contribute to differentiating among these possibilities by determining whether 1) rehearsal timing precision suffers when the presented sequence timing is less predictable, and 2) rehearsal timing precision still relates to individual differences in STM span, memory load manipulations, and language abilities. If performance on the rehearsal-probe task is largely a result of top-down control, then performance should generally suffer when sequence timing is unpredictable. Likewise, if the observed differences in timing precision are primarily a result of differences in top-down control, then the differences observed across Experiments 1-7 should be reduced or eliminated when sequence timing is unpredictable.

A broader issue related to timing is whether strictly-timed, self-paced subvocal rehearsal is representative of more naturally occurring, ‘free-running’ rehearsal. It is not clear at this point what the answer is to this question, but it is an important topic that merits investigation. When the rehearsal-probe task was conceptualised, regular sequence timing and instructions to exactly replicate this timing were considered the strongest approach toward attempting to measure covert speech, with the reasoning that this design had the best chance of providing a reliable measure. Given the success of the task thus far, it seems quite possible that the task could be used with untimed forward-order rehearsal instructions. That is, participants could be instructed to simply continue rehearsing the items in the correct (forward) serial order, without explicit instructions about replicating sequence timing. Provided that individuals tend to rehearse sets of similar sequences with broadly the same temporal attributes, then the task should work with unpaced rehearsal as well. Indeed, the modelling results estimated that participants generally rehearsed at a slightly slower pace, which might be indicative of a tendency to deviate away from the presented rate toward a more optimal or natural rehearsal rhythm. An extension of the rehearsal-probe task to untimed rehearsal in future work would be useful for examining how rehearsal occurs in a less restrictive and more ecologically valid situation, and it could reveal more pronounced differences between individuals and developmental groups.

The unanticipated results in particular were useful in generating questions for future research. For instance, in Experiment 2, timing variability within participants increased for digits in a familiar counting sequence compared to equal-length randomly ordered digit sequences. Future investigations could aim to investigate the influence of over-learned sequences on the ability to reproduce stimulus-specific timing during rehearsal, for example using other familiar verbal sequences of arbitrary items (alphabet, days, months, etc.). The notion that STM for timing of these lists might differ from unpredictably ordered verbal sequences is supported by evidence that 1) overlearned sequences might be stored and processed differently from non-ordinal stimuli (Pariyadath, Plitt, Churchill, & Eagleman, 2012), and 2) timing information appears to be linked to item-order associations in Hebbian learning tasks (Hitch et al., 2009).

In Experiment 4, contrary to predictions, timing variability was greater for temporally-grouped compared to isochronous sequences. As discussed in Chapter 3, this may have been due to the lack of hierarchical metrical structure linking the within- and between-subgroup timings, which has been shown to affect performance thresholds in auditory change detection tasks (Grube & Griffiths, 2009). To clarify this remaining question, Experiment 4 could be repeated with parametric manipulations of sequence timing, both regular and irregular, and with both metrically structured and unstructured sequences. Under the present hypothesis, timing should be more precise for sequences with regular and hierarchically-structured temporal grouping patterns. If, on the other hand, temporal grouping is found to be universally detrimental to timing precision, this would be at odds with predictions and it would suggest that there is not a direct relationship between timing precision and STM for serial order, as regular temporal grouping patterns are known to improve recall for supraspan lists (Ryan, 1969a).

The use of EEG with the rehearsal-probe task provides another promising avenue for future research. As discussed in Chapter 7, an obvious next step given the results of the EEG experiment is to extend the design to include different presentation rates. If the relative increase in phase-locked power reflects periodic neural activity directly related to the rate of sequence perception and rehearsal, then this peak in the frequency spectra should track changes in the rate of item presentation and rehearsal. In future work with the ERP paradigm, behavioural rehearsal-probe data could be collected before the EEG task in order to account for individual differences in rehearsal timing, and to maximise the probability of aligning the presented probe with the onsets of specific items during rehearsal. More generally, continued efforts to establish the EEG signatures of rehearsal contents and timing could result in the ability to detect these markers with less constrained memory tasks. Such tasks might involve, for instance, variations on sequence stimuli and timing, full serial recall after the rehearsal period, or a preload task. Also, as previously mentioned, it would be useful to be able to measure unpaced rehearsal activity. The EEG frequency spectra might provide a measure of spontaneous rehearsal rates within individuals, provided that they are similar across multiple trials. The location and strength of any low-frequency local peaks in the frequency spectra could then be analysed with respect to experimental manipulations and individual or group differences in order to reveal possible relationships with auditory-verbal STM for serial order and timing, as well as potential neural markers for developmental language disorders.

8.3 Conclusion

In this thesis, I have argued that there is a need for a direct measure of rehearsal, and I addressed this problem by developing and testing a novel task for this purpose. I also put forward the claim that timing precision for auditory-verbal sequences may be determined, at least in part, by the availability of shared limited resources. In support of this hypothesis, the results of experiments with the rehearsal-probe task have convincingly demonstrated that temporal variability for subspan sequences is inversely related to auditory-verbal STM span. There is scope for improvement in the task methods and in further development of computational modelling, but overall, the rehearsal-probe paradigm is a reliable method for capturing the temporal dynamics of subvocal rehearsal, and performance in the task is related to important individual and group differences in auditory memory and language development. Future studies could determine the extent to which these findings generalise to other types of auditory stimuli, and whether the neural correlates of rehearsal timing can be detected in typical memory tasks without constraints on rehearsal timing. Finally, the dynamic resource account of STM for sequence timing has implications for different classes of models of STM for serial order. The findings presented here appear to suggest an important role for fine-grained timing information in the encoding and retrieval of serial order, in auditory-verbal STM capacity, and in language development.

APPENDIX A: CIRCULAR STANDARD DEVIATION

The function used for the calculation of circular standard deviation (CSD) was written by Philipp Berens as part of the Matlab CircStats toolbox (Berens, 2009). In the present research, the angles ('alpha') corresponded to positions of the probe times around a circle.

```
function [s s0] = circ_std(alpha, w, d, dim)
% s = circ_std(alpha, w, d, dim)
%   Computes circular standard deviation for circular data
%   (equ. 26.20, Zar).
%
%   Input:
%     alpha sample of angles in radians
%     [w        weightings in case of binned angle data]
%     [d    spacing of bin centers for binned data, if supplied
%           correction factor is used to correct for bias in
%           estimation of r]
%     [dim  compute along this dimension, default is 1]
%
%     If dim argument is specified, all other optional arguments can be
%     left empty: circ_std(alpha, [], [], dim)
%
%   Output:
%     s     angular deviation
%     s0    circular standard deviation
%
% PHB 6/7/2008
%
% References:
%   Biostatistical Analysis, J. H. Zar
%
% Circular Statistics Toolbox for Matlab

% By Philipp Berens, 2009
% berens@tuebingen.mpg.de - www.kyb.mpg.de/~berens/circStat.html

if nargin < 4
  dim = 1;
end

if nargin < 3 || isempty(d)
  % per default do not apply correct for binned data
  d = 0;
end

if nargin < 2 || isempty(w)
  % if no specific weighting has been specified
  % assume no binning has taken place
    w = ones(size(alpha));
else
  if size(w,2) ~= size(alpha,2) || size(w,1) ~= size(alpha,1)
    error('Input dimensions do not match');
  end
end

% compute mean resultant vector length
r = circ_r(alpha,w,d,dim);

s = sqrt(2*(1-r));      % 26.20
```

```
s0 = sqrt(-2*log(r));     % 26.21
```

The ‘circ_std’ function is dependent upon the ‘circ_r’ function, also written by Berens as part of the CircStats toolbox. The ‘circ_r’ function calculates the mean resultant vector length given a set of angles (‘alpha’), and optionally, weightings for binned data (‘w’), spacing for binned data (‘d’), and a dimension argument (‘dim’).

```
function r = circ_r(alpha, w, d, dim)
% r = circ_r(alpha, w, d)
%   Computes mean resultant vector length for circular data.
%
%   Input:
%     alpha sample of angles in radians
%     [w        number of incidences in case of binned angle data]
%     [d    spacing of bin centers for binned data, if supplied
%           correction factor is used to correct for bias in
%           estimation of r, in radians (!)]
%     [dim  compute along this dimension, default is 1]
%
%     If dim argument is specified, all other optional arguments can be
%     left empty: circ_r(alpha, [], [], dim)
%
%   Output:
%     r     mean resultant length
%
% PHB 7/6/2008
%
% References:
%   Statistical analysis of circular data, N.I. Fisher
%   Topics in circular statistics, S.R. Jammalamadaka et al.
%   Biostatistical Analysis, J. H. Zar
%
% Circular Statistics Toolbox for Matlab

% By Philipp Berens, 2009
% berens@tuebingen.mpg.de - www.kyb.mpg.de/~berens/circStat.html

if nargin < 4
  dim = 1;
end

if nargin < 2 || isempty(w)
  % if no specific weighting has been specified
  % assume no binning has taken place
    w = ones(size(alpha));
else
  if size(w,2) ~= size(alpha,2) || size(w,1) ~= size(alpha,1)
    error('Input dimensions do not match');
  end
end

if nargin < 3 || isempty(d)
  % per default do not apply correct for binned data
  d = 0;
end

% compute weighted sum of cos and sin of angles
r = sum(w.*exp(1i*alpha),dim);
```

```
% obtain length
r = abs(r)./sum(w,dim);

% for data with known spacing, apply correction factor to correct for
bias
% in the estimation of r (see Zar, p. 601, equ. 26.16)
if d ~= 0
  c = d/2/sin(d/2);
  r = c*r;
end
```

APPENDIX B: MODELLING SUPPLEMENTARY MATERIAL

8.4 Matlab functions used in the model

The computational model presented in Chapter 6 was implemented in Matlab using both standard and custom functions. Figure 78 shows the key functions used in the model. Each box represents a separate Matlab function, where the boxes with white backgrounds indicate custom functions and the boxes with gray backgrounds refer to standard Matlab functions. The arrows show where data is passed between functions.

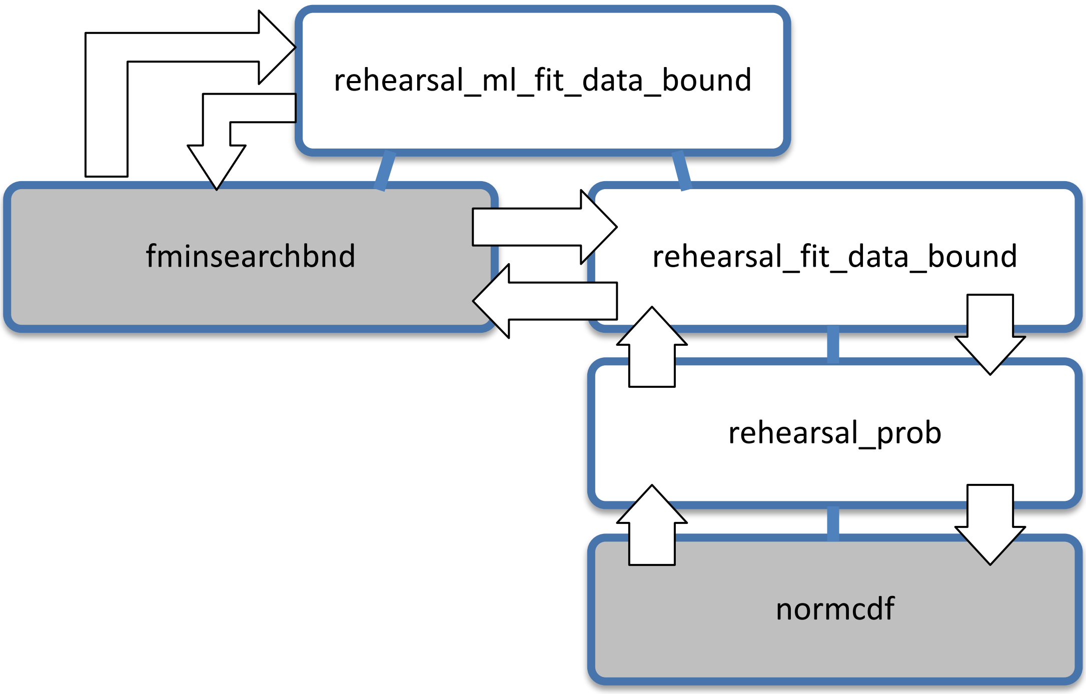

Figure 78. The relationships among functions used in the model. Each box represents a function, and the arrows show how data is passed between functions. Boxes with gray backgrounds indicate standard Matlab functions, while those with white backgrounds are custom built.

The roles of these functions will now be presented, starting with the 'rehearsal_ml_fit_data_bound' function, which is called directly by the user.

8.4.1 rehearsal_ml_fit_data_bound

The 'rehearsal_ml_fit_data' function serves as a wrapper for the minimisation algorithm 'fminsearchbnd'. The main role of this function is to call 'fminsearchbnd' with the 'rehearsal_fit_data_bound' function, where the input parameters are also passed to this function. The role of the 'fminsearchbnd' function is described in later in this section. This function takes several inputs that describe the rehearsal-probe task conditions (probe times, number of items, item onset times, item durations, and list duration). Additionally, the function is given some starting values (*X0*) for the four parameters, as well as the actual data (*observed*). This function returns ‘fitted’, which is a list of the values of the 4 parameters after the model has been fitted to the data.

```
function [fitted,fval] = rehearsal_ml_fit_data_bound(probe_times, nitems,
onsets, durations, list_duration, X0, observed)

% Set the fminsearch tolerance and maximum iteration values
options = optimset('TolFun', 0.0000001, 'MaxIter', 600, 'MaxFunEvals',
          1e10);

% Set the lower and upper boundaries for all parameters
LB = [0.05, 0.00001, 0.00001, Inf];
UB = [1.95, Inf, Inf, Inf];

[fitted, fval, exitflag, output] = fminsearchbnd(@(x)
rehearsal_fit_data_bound(probe_times, nitems, onsets, durations,
list_duration, x(1), x(2), x(3), x(4), observed),...
    X0, LB, UB, options)

% Take the modulus of the offset parameter and half the list duration
if mod(abs(fitted(4)), list_duration/2) ~= abs(fitted(4))
    if fitted(4) < 0
        fitted(4) = mod(fitted(4), list_duration/2);
    else
        fitted(4) = mod(fitted(4), -list_duration/2);
    end
end

% Print result to the screen
fprintf(1,'m_r = %1.4f\ns_cum = %1.4f\ns_con = %1.4f\nm_off =
%1.4f\nf_val = %1.4f\n', fitted(1), fitted(2), fitted(3), fitted(4),
fval)
```

To start with, the ‘options’ variable is set to the specified optimisation parameters via the ‘optimset’ function, and this is passed to the ‘fminsearchbnd’ function. ‘Optimset’ is a standard Matlab function which replaces default optimisation values in ‘fminsearchbnd’. Next, the lower and upper boundaries of the parameter values are specified by *LB* and *UB* in the function. The ‘*Inf*’ value refers to infinite, meaning that the parameter can take any value. The first parameter, rate, is bounded by 0.05 and 1.95. This means that the fitted rehearsal rate cannot take any value outside of this range. The two error parameters, cumulative and constant error, cannot take values lower than 0.00001, but there is no upper limit to these parameters. The offset parameter can take any value in the ‘fminsearchbnd’ algorithm.

8.4.2 fminsearchbnd

The ‘fminsearchbnd’ function is a bounded version of the standard ‘fminsearch’ Matlab function, and it is called by the ‘rehearsal_ml_fit_data_bound’ function. The purpose of ‘fminsearchbnd’ is to find the set of parameters (input values) that minimise the output value of a function given the function handle and starting parameter values. In this case, the function to be minimised is 'rehearsal_fit_data_bound', and the parameters are represented by *x*. The 'fminsearchbnd' function uses the Nelder-Mead Simplex algorithm for minimization of a multidimensional error surface. Listed below are the input and output arguments for 'fminsearchbnd', as well as the author and version information.

```
function [x,fval,exitflag,output] =
fminsearchbnd(fun,x0,LB,UB,options,varargin)
% Author: John D'Errico
% E-mail: woodchips@rochester.rr.com
% Release: 4
% Release date: 7/23/06
```

The 'fminsearchbnd' function returns the value(s) of the parameter(s) in *x* that minimise the input function *f(x)*. In the case of the rehearsal model, *x* (which is assigned to the variable *fitted* in the 'rehearsal_ml_fit_data_bound' function) is a vector containing the four parameters in the order in which they are entered into the 'rehearsal_data_fit_bound' function (rate, cumulative error, constant error, offset). LB and UB refer to the lower and upper bounds of the parameter value(s). The output variable *fval* is the value of the specified function evaluated using the parameters in *x.* The output variable *exitflag* indicates whether 'fminsearchbnd' terminated successfully or unsuccessfully. The 'fminsearchbnd' function terminates successfully when a minimum value has been located in the error space in which the difference between the current and surrounding x-coordinates is less than or equal to a tolerance value *TolX,* and when the difference between the current function value and corresponding y-values is less than or equal to a tolerance value *TolFun*. This function terminates unsuccessfully when it reaches a specified maximum number of iterations or function executions without reaching the tolerance threshold. The *TolX* and *TolFun* thresholds can be set using 'optimset' in the 'rehearsal_ml_fit_data_bound' function, as in the 'rehearsal_ml_fit_data_bound' function. Finally, the output variable *output* contains additional information about the 'fminsearchbnd' operation, such as the number of iterations completed by 'fminsearchbnd' before termination.

8.4.3 rehearsal_fit_data_bound

This function is called indirectly within the 'rehearsal_ml_fit_data_bound' function, and is passed the same input arguments. The purpose of this function is to calculate and output the value of the discrepancy function between the data and the model fit using the current parameter values. Because this function is called in 'fminsearchbnd', the 4 input parameters that are adjusted by fminsearchbnd(*m_r, s_cum, s_con, m_off*) change with each adjusted call.

```
function L = rehearsal_fit_data_bound(probe_times, nitems, onsets,
durations, list_duration, m_r, s_cum, s_con, m_off, observed)
```

```
[resp_prob] = rehearsal_prob(probe_times, nitems, onsets, durations,
list_duration, m_r, s_cum, s_con, m_off);

L = zeros(1,length(probe_times));
N = zeros(nitems,length(probe_times));
for i = 1:nitems
    for j = 1:length(probe_times)
        % find all of the responses to trials with probe j
        allRespJ = observed(observed(:,1) == j,2);
        % count the number of item i responses to those trials
        N(i,j) = sum(allRespJ == i);
    end
    % log-likelihood is the response count (scalar N) times the log of
    % probability for the item at each probe time (vector resp_prob)
    l = N(i,:).*log(resp_prob(i,:));
    L = L + l;
end
L = -sum(L);
```

This function starts by calculating the modelled response probabilities based on the current set of parameters. This is done via the ‘rehearsal_prob’ function (described below). Then, for each serial position, the number of observed responses for that item is calculated across probe times. The number of observed responses for each item across probe times is multiplied by the log of the modelled response probability for that item across probe times. This serves as the likelihood measure, and the result from each serial position is added to produce a single likelihood value.

The final output of this function is *L*, which is the negative-signed value of the maximum log likelihood of the data given the current 4 input parameters. This value is first calculated as the maximum log likelihood, meaning that a larger value indicates a better fit to the data. Finally, the *L* value is converted to a negative number because the ‘fminsearchbnd’ function minimizes the output value of any function it is given. The *L* value is returned to ‘fminsearchbnd’ for minimisation.

8.4.4 rehearsal_prob

The role of the ‘rehearsal_prob’ function is to compute the modelled response probabilities across the probe times for a given set of parameter values. This function takes all of the same arguments as the previous functions (probe times, number of items, item onset times, item durations, list duration, and the 4 model parameters). Note that this function does not take the observed data as an input, because there is no comparison made here between the modelled response probabilities and the observed data. This function is called by the ‘rehearsal_fit_data’ function above, which uses the output from the ‘rehearsal_prob’ function to calculate the maximum likelihood values. In the case of ‘rehearsal_prob’, the 4 model parameters (*m_r, s_cum, s_con, m_off*) are passed from each run of ‘fminsearchbnd’ to ‘rehearsal_fit_data’, and then to ‘rehearsal_prob’ in order to calculate the probability that a response is a given serial position at a given probe time.

```
function p = rehearsal_prob(probe_times, nitems, onsets, durations,
list_duration, m_r, s_cum, s_con, m_off)

t = probe_times;

for i = 1:nitems
    istart = onsets(i);
    iend = onsets(i) + durations(i);

    mmu = t*m_r + m_off;
    mstd = sqrt((s_cum*t).^2 + s_con.^2);

    for q = -2:100
        pstart = normcdf(istart + list_duration*q, mmu, mstd);
        pend = normcdf(iend + list_duration*q, mmu, mstd);
        pitem(q+3,:) = pend - pstart;
    end
    psitem(i,:) = sum(pitem);
end
p = psitem;
```

The output of this function is a set of probabilities ‘*p*’. For each serial position *i*, cumulative density functions are computed over multiple rehearsal cycles to estimate the probability that the item in the given position is being rehearsed at a given probe time. The cumulative probabilities are added up over multiple rehearsal cycles to account for the fact that, at any point in time, the number of list cycles already rehearsed cannot be assumed. Also, there are usually non-zero probabilities of responding with items from adjacent rehearsal cycles, and these probabilities must also be considered.

The cumulative probability density curves are produced using the ‘normcdf’ function. These distributions are centred on a mean *mmu* with some standard deviation *mstd*. The mean of the distribution is calculated using the list of probe times and the two parameters which affect the centres of the response probability distributions for each serial position during rehearsal: the rate (*m_r*) and the initial offset (*m_off*). If the rate is the same as the item presentation rate (*m_r* = 1.0) and there is no initial offset or delay at the start of the rehearsal period (*m_off* = 0.0) then the distributions *pstart* and *pend* will be centred at each probe time (*mmu* = *t*).

For each probe time *t* there is a corresponding computed mean *mmu* and standard deviation *mstd*, and for each *mmu* and *mstd* a normal cumulative density function is computed and the cumulative probability is evaluated at the item start time *istart* and the item end time *iend*. The ‘normcdf’ function, part of the Matlab Statistics toolbox, produces a normal cumulative density function given 3 input arguments: X, mu, and sigma (with optional input arguments pcov and alpha). 'normcdf' returns the cumulative distribution function of the normal distribution with mean *mu* and standard deviation *sigma* evaluated at the values of *X.*

The probability of responding with a particular item at a given time is calculated by first finding the cumulative probability of the serial position response based on the modelled mental onset time of that item in a particular rehearsal cycle. The mental onset time (*pstart*) is a normal distribution, where the mean is determined by the rate and offset parameters, and the standard deviation is determined by the constant and cumulative error parameters. Next, the cumulative probability of the serial position response is also calculated for the mental offset time (*pend*) for that item in a given rehearsal cycle. Finally, the cumulative probability for the mental onset time is subtracted from that for the offset time, and the result is the probability of responding with that item (in a particular rehearsal cycle) at that time.

This process is repeated over all probe times and for multiple rehearsal cycles. More precisely, the probabilities of responding with each item are calculated between the mental onsets and offsets of each item for -2 to 100 times the list duration. This means that the probabilities are calculated for two cycles before the expected rehearsal cycle (to account for the fact that individuals may rehearse at a much slower rate) and up to 100 rehearsal cycles after the sequence is presented, which is an arbitrarily large number of iterations. In actuality, the probability of rehearsing a given item is only likely to be non-zero at neighbouring list cycle. Thus, for most data sets, the probability of rehearsing a given item quickly becomes 0 as the number of rehearsal list cycles increases past the expected list cycle toward 100.

REFERENCES

Aaronson, D. (1968). Temporal Course of Perception in an Immediate Recall Task. *Journal of Experimental Psychology*, *76*(1), 129–140.

Adams, H. F. (1915). A note on the effect of rhythm on memory. *Psychological Review*, *22*, 289–299.

Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *NeuroImage*, *73*, 191–199.

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–64. doi:10.1038/nn1800

Alt, M., & Spaulding, T. J. (2011). The effect of time on word learning: an examination of decay of the memory trace and vocal rehearsal in children with and without specific language impairment. *Journal of Communication Disorders*, *44*(6), 640–654. doi:10.1016/j.jcomdis.2011.07.001

Andreou, L.-V., Kashino, M., & Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hearing Research*, *280*(1-2), 228–35. doi:10.1016/j.heares.2011.06.001

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*(2), 111–130.

Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, *14*(5), 919–924. doi:10.3758/BF03194122

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation, Vol. 2* (pp. 89–195). London: Academic Press.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 13172–7. doi:10.1073/pnas.162485599

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–8. doi:10.1111/j.1467-9280.2007.01949.x

Baciu, M. V., Rubin, C., Décorps, M. A., & Segebarth, C. M. (1999). fMRI assessment of hemispheric language dominance using a simple inner speech paradigm. *NMR in Biomedicine*, *12*, 293–298.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory ? *Quarterly Journal of Experimental Psychology*, *20*(3), 249–264.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

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

Baddeley, A. D. (2007). *Working memory, thought, and action. Oxford Psychology Series* (Vol. 45). New York, NY: Oxford University Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–73.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 47–89). New York, NY: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. doi:10.1016/S0022-5371(75)80045-4

Baddeley, A. D., & Wilson, B. (1985). Phonological Coding and Short-Term in Patients without Speech. *Journal of Memory and Language*, *24*(4), 490–502.

Balasubramaniam, R., Wing, A. M., & Daffertshofer, A. (2004). Keeping with the beat: movement trajectories contribute to movement timing. *Experimental Brain Research*, *159*(1), 129–34. doi:10.1007/s00221-004-2066-z

Banai, K., & Ahissar, M. (2013). Musical Experience, Auditory Perception and Reading-Related Skills in Children. *PLoS ONE*, *8*(9), e75876. doi:10.1371/journal.pone.0075876

Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1359–1367. doi:10.1037/a0015792.Discrete

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), 1–11. doi:10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(10), 1–15. doi:10.1167/11.10.6

Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, *231*(5890), 851–854. doi:10.1126/science.1158023

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*, 54. doi:10.1186/1471-2202-12-54

Bell, A. J., & Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, *7*(6), 1129–1159. doi:10.1162/neco.1995.7.6.1129

Beneventi, H., Tønnessen, F. E., & Ersland, L. (2009). Dyslexic children show short-term memory deficits in phonological storage and serial rehearsal: an fMRI study. *The International Journal of Neuroscience*, *119*(11), 2017–43. doi:10.1080/00207450903139671

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10), 1–21.

Betts, B. J., & Jorgensen, C. (2006). Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment. *Journal of Human-Computer Interaction*, *18*, 1242–1259.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin*, *130*(6), 858–886.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*(1), 80–97.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173–189.

Blankenship, A. B. (1938). Memory span: A review of the literature. *Psychological Bulletin*, *35*(1), 1–25. doi:10.1037/h0061086

Boersma, P., & Weenink, D. (1992). Praat: Doing Phonetics By Computer. Retrieved from http://www.praat.org/

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–29. doi:10.1162/jocn.2009.21306

Boucher, V. J. (2006). On the Function of Stress Rhythms in Speech: Evidence of a Link with Grouping Effects on Serial Memory. *Language and Speech*, *49*(4), 495–519. doi:10.1177/00238309060490040301

Bower, G. H., & Winzenz, D. (1969). Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, *80*(2), 1–17.

Bowey, J. A. (1997). What Does Nonword Repetition Measure? A Reply to Gathercole and Baddeley. *Journal of Experimental Child Psychology*, *67*, 295–301.

Brady, S. (1986). Short-term memory, phonological processing, and reading ability. *Annals of Dyslexia*, *36*(1), 138–153. doi:10.1007/BF02648026

Brodie, D. A., & Prytulak, L. S. (1975). Free recall curves: Nothing but rehearsing some items more or recalling them sooner? *Journal of Verbal Learning and Verbal Behavior*, *14*(5), 549–563.

Brown, G. D. A., & Hulme, C. (1995). Modeling Item Length Effects in Memory Span: No Rehearsal Needed? *Journal of Memory and Language*, *34*, 594–621.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-Based Memory for Serial Order. *Psychological Review*, *107*(1), 127–181.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, *31*(4), 429–460. doi:10.1016/0749-596X(92)90022-P

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. doi:10.1037//0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. doi:10.1016/j.jml.2006.08.005

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*(2), 229–246.

Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., … Burani, C. (2007). Non-Word Repetition in Children with Specific Language Impairment: A Deficit in Phonological Working Memory or in Long-Term Verbal Knowledge? *Cortex*, *43*(6), 769–776. doi:10.1016/S0010-9452(08)70505-7

Catts, H. W., Adlof, S. M., Hogan, T., & Weismer, S. E. (2005). Are Specific Language Impairment and Dyslexia Distinct Disorders? *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378–1396. doi:10.1044/1092-4388(2005/096)

Chen, C., & Stevenson, H. W. (1988). Cross-linguistic differences in digit span of preschool children. *Journal of Experimental Child Psychology*, *46*(1), 150–158.

Chincotta, D., & Hoosain, R. (1995). Reading rate, articulatory suppression and bilingual digit span. *European Journal of Cognitive Psychology*, *7*(2), 201–211. doi:10.1080/09541449508403100

Coady, J. A., Mainela-Arnold, E., & Evans, J. L. (2013). Phonological and lexical effects in verbal recall by children with specific language impairments. *International Journal of Language & Communication Disorders*, *48*(2), 144–59. doi:10.1111/1460-6984.12005

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. doi:10.3758/BF03194326

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, *28*(4), 529–538. doi:10.3758/BF03201243

Collier, G. L., & Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 602–627. doi:10.1037/0096-1523.21.3.602

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences*, *39*(5), 1005–1014. doi:10.1016/j.paid.2005.03.020

Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory & Cognition*, *26*(2), 330–342. doi:10.3758/BF03201144

Conrad, R. (1963). Acoustic Confusions and Memory Span for Words. *Nature*, *197*(4871), 1029–1030. doi:10.1038/1971029a0

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, *4*(3), 161–169. doi:10.1016/S0022-5371(65)80015-9

Conway, A. R. ., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. . (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. doi:10.1016/S0160-2896(01)00096-4

Corkin, S. (1974). Serial-Ordering Deficits in Inferior Readers. *Neuropsychologia*, *12*(3), 347–354. doi:10.1016/0028-3932(74)90050-5

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error Biases in Inner and Overt Speech: Evidence from Tongue Twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 162–175.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex*, *45*(1), 119–130. doi:10.1016/j.cortex.2007.09.008

Corriveau, K. H., Pasquini, E., & Goswami, U. (2007). Basic Auditory Processing Skills and Specific Language Impairment: A New Look at an Old Hypothesis. *Journal of Speech, Language, and Hearing Research*, *50*(3), 647–667. doi:10.1044/1092-4388(2007/046)

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*(1), 1–17. doi:10.1016/0749-596X(92)90002-F

Cowan, N., Keller, T. A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal Memory Span in Children: Speech Timing Clues to the Mechanisms Underlying Age and Word Length Effects. *Journal of Memory and Language*, *33*(2), 234–250. doi:10.1006/jmla.1994.1012

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The Role of Attention in the Development of Short-Term Memory: Age Differences in the Verbal Span of Apprehension. *Child Development*, *70*(5), 1082–1097. doi:10.1111/1467-8624.00080

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 599–607.

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*(4), 183–91.

Cuvo, A. J. (1975). Developmental differences in rehearsal and free recall. *Journal of Experimental Child Psychology*, *19*(2), 265–278. doi:10.1016/0022-0965(75)90090-9

Da Costa Pinto, A. (1991). Reading Rates and Digit Span in Bilinguals: The superiority of Mother Tongue. *International Journal of Psychology*, *26*(4), 471–483. doi:10.1080/00207599108247135

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

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, *89*(1), 63–100. doi:10.1037//0033-2909.89.1.63

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85 Pt 2*, 761–8. doi:10.1016/j.neuroimage.2013.06.035

Einstein, G. O., Pellegrino, J. W., Mondani, M. S., & Battig, W. F. (1974). Free-recall performance as a function of overt rehearsal frequency. *Journal of Experimental Psychology*, *103*(3), 440–449. doi:10.1037/h0037165

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, *71*(1), 43–51. doi:10.1111/j.2044-8295.1980.tb02728.x

Ericcson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science , 208* (4448), 1181–1182. doi:10.1126/science.7375930

Ester, E. F., Fukuda, K., May, L. M., Vogel, E. K., & Awh, E. (2014). Evidence for a fixed capacity limit in attending multiple locations. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 62–77. doi:10.3758/s13415-013-0222-2

Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological Similarity and Trace Degradation in the Serial Recall Task: When CAT helps RAT, but not MAN. *International Journal of Psychology*, *34*(5-6), 301–307. doi:10.1080/002075999399602

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 128–145. doi:10.1037/0278-7393.34.1.128

Farrell, S., Wise, V., & Lelièvre, A. (2011). Relations between timing, position, and grouping in short-term memory. *Memory & Cognition*, *39*(4), 573–87. doi:10.3758/s13421-010-0053-0

Ferguson, M. A., Hall, R. L., Riley, A., & Moore, D. R. (2011). Communication, listening, cognitive and speech perception skills in children with auditory processing disorder (APD) or Specific Language Impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *54*(1), 211–27. doi:10.1044/1092-4388(2010/09-0167)

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. (M. Carmichael, Ed.) (4th ed., p. 915). London: SAGE Publications.

Fischler, I., Rundus, D., & Atkinson, R. C. (1970). Effects of overt rehearsal procedures on free recall. *Psychonomic Science*, *19*(4), 249–250.

FitzGerald, P., & Broadbent, D. E. (1985). Order of report and the structure of temporary memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 217–228.

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., … Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00392

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Child Development*, *37*(2), 283–299.

Flavell, J. H., Green, F. L., & Flavell, E. R. (2000). Development of Children's Awareness of Their Own Thoughts. *Journal of Cognition and Development*, *1*(1), 97–112. doi:10.1207/S15327647JCD0101N_10

Flavell, J. H., Green, F. L., Flavell, E. R., & Grossman, J. B. (1997). The Development of Children's Knowledge about Inner Speech. *Child Development*, *68*(1), 39–47. doi:10.1111/j.1467-8624.1997.tb01923.x

Ford, J. M., Roach, B. J., & Mathalon, D. H. (2010). Assessing corollary discharge in humans using noninvasive neurophysiological methods. *Nature Protocols*, *5*(6), 1160–8. doi:10.1038/nprot.2010.67

Fortin, C., & Massé, N. (1999). Order information in short-term memory and time estimation. *Memory & Cognition*, *27*(1), 54–62. doi:10.3758/BF03201213

Fortin, C., Rousseau, R., Bourque, P., & Kirouac, E. (1993). Time estimation and concurrent nontemporal processing: Specific interference from short-term-memory demands. *Perception & Psychophysics*, *53*(5), 536–548. doi:10.3758/BF03205202

Frankish, C. R. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*(2), 200–209. doi:10.1016/0749-596X(85)90024-5

Frankish, C. R. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 469–479.

Frankish, C. R. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, *9*(7), S5–S22. doi:10.1002/acp.2350090703

Fraser, J., Goswami, U., & Conti-Ramsden, G. (2010). Dyslexia and Specific Language Impairment: The Role of Phonology and Auditory Processing. *Scientific Studies of Reading*, *14*(1), 8–29. doi:10.1080/10888430903242068

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*(2), 177–182. doi:10.1016/j.conb.2010.03.005

Galton, F. (1887). Supplementary Notes on "Prehension" in Idiots. *Mind*, *12*, 79–82.

Garrity, L. I. (1975). An electromyographical study of subvocal speech and recall in preschool children. *Developmental Psychology*, *11*(3), 274–281.

Garrity, L. I. (1977). A Review of Short-Term Memory Studies of Covert Speech in Young Children. *The Journal of Psychology*, *95*(2), 249–261. doi:10.1080/00223980.1977.9915887

Garrity, L. I., & Donoghue, J. T. (1977). A Follow-up Electromyographical Study of Subvocal Speech and Recall in Preschool Children. *The Journal of Psychology*, *95*(2), 275–283. doi:10.1080/00223980.1977.9915890

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83–94. doi:10.3758/BF03210559

Gathercole, S. E., & Adams, A.-M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, *33*, 672–688.

Gathercole, S. E., Adams, A.-M. M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, *22*(2), 201–207.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological Memory Deficits in Language Disordered Children: Is There a Causal Connection? *Journal of Memory and Language*, *29*, 336–360.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966–979.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *54*(1), 1–30. doi:10.1080/02724980042000002

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological Short-term Memory and Vocabulary Development: Further Evidence on the Nature of the Relationship. *Applied Cognitive Psychology*, *13*, 65–77.

Gathercole, S. E., Willis, C., & Baddeley, A. D. (1991). Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology*, *82*(3), 387–406. doi:10.1111/j.2044-8295.1991.tb02407.x

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*, 349–367. doi:10.1017/S0142716400009267

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory*, *2*(2), 103–27. doi:10.1080/09658219408258940

Geva, S., Bennett, S., Warburton, E. A., & Patterson, K. (2011). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology*, *25*(3), 323–343. doi:10.1080/02687038.2010.511236

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*, *134*, 3071–3082. doi:10.1093/brain/awr232

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*(June), 130. doi:10.3389/fpsyg.2011.00130

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, *3*, 238. doi:10.3389/fpsyg.2012.00238

Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*(March), 1–5. doi:10.3389/fpsyg.2013.00138

Ghitza, O., Giraud, A.-L., & Poeppel, D. (2013). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, *6*, 1–4. doi:10.3389/fnhum.2012.00340

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1-2), 113–26. doi:10.1159/000208934

Gill, C. B., Klecan-Aker, J., Roberts, T., & Fredenburg, K. A. (2003). Following directions: Rehearsal and visualization strategies for children with specific language impairment. *Child Language Teaching and Therapy*, *19*(1), 85–103. doi:10.1191/0265659003ct245oa

Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. *Neuron*, *56*(6), 1127–1134.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–7. doi:10.1038/nn.3063

Glanzer, M., & Meinzer, A. (1967). The effects of intralist activity on free recall. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 928–935.

Glenberg, A., & Adams, F. (1978). Type I rehearsal and recognition. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 455–463.

Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior*, *16*(3), 339–352.

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*, *31*(23), 8502–11. doi:10.1523/JNEUROSCI.0208-11.2011

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. doi:10.1016/j.tics.2010.10.001

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szucs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: the Ba-Wa distinction in developmental dyslexia. *Developmental Science*, *14*(1), 34–43. doi:10.1111/j.1467-7687.2010.00955.x

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10911–6. doi:10.1073/pnas.122368599

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage*, *45*(4), 1199–211. doi:10.1016/j.neuroimage.2008.12.038

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. doi:10.1371/journal.pbio.1001752

Grube, M., Cooper, F. E., & Griffiths, T. D. (2013). Auditory temporal-regularity processing correlates with language and literacy skill in early adulthood. *Cognitive Neuroscience*, *4*(3-4), 225–230. doi:10.1080/17588928.2013.825236

Grube, M., & Griffiths, T. D. (2009). Metricality-enhanced temporal encoding and the subjective perception of rhythmic sequences. *Cortex*, *45*(1), 72–79. doi:10.1016/j.cortex.2008.01.006

Grube, M., Kumar, S., Cooper, F. E., Turton, S., & Griffiths, T. D. (2012). Auditory sequence analysis and phonological skill. *Proceedings of the Royal Society B*, *279*(1746), 4496–4504. doi:10.1098/rspb.2012.1817

Grüsser, O.-J. (1995). On the History of the Ideas of Efference Copy and Reafference. In C. Debru (Ed.), *Essays in the History of the Physiological Sciences: Proceedings of a Network Symposium of the European Association for the History of Medicine and Health* (pp. 35–56). Strasbourg: Rodopi.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*(7), 1213–1236. doi:10.1080/02724980343000071

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language*, *59*(2), 267–333. doi:10.1006/brln.1997.1819

Halford, G. S., Maybery, M. T., O'Hare, A. W., & Grant, P. (1994). The Development of Memory and Processing Capacity. *Child Development*, *65*(5), 1338–1356. doi:10.1111/j.1467-8624.1994.tb00820.x

Hall, D., & Gathercole, S. E. (2011). Serial recall of rhythms and verbal sequences: Impacts of concurrent tasks and irrelevant sound. *Quarterly Journal of Experimental Psychology*, *64*(8), 1580–1592. doi:10.1080/17470218.2011.564636

Hämäläinen, J. A., Fosker, T., Szücs, D., & Goswami, U. (2011). N1, P2 and T-complex of the auditory brain event-related potentials to tones with varying rise times in adults with and without dyslexia. *International Journal of Psychophysiology*, *81*(1), 51–9. doi:10.1016/j.ijpsycho.2011.04.005

Hämäläinen, J. A., Leppänen, P. H. T., Torppa, M., Müller, K., & Lyytinen, H. J. (2005). Detection of sound rise time by adults with dyslexia. *Brain and Language*, *94*(1), 32–42. doi:10.1016/j.bandl.2004.11.005

Hämäläinen, J. A., Rupp, A., Soltész, F., Szücs, D., & Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: an MEG study. *NeuroImage*, *59*(3), 2952–61. doi:10.1016/j.neuroimage.2011.09.075

Hanley, J. R. (1997). Does Articulatory Suppression Remove the Irrelevant Speech Effect? *Memory*, *5*(3), 423–431. doi:10.1080/741941394

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review*, *10*(2), 435–444. doi:10.3758/BF03196503

Hansen, J., & Bowey, J. A. (1994). Phonological Analysis Skills, Verbal Working Memory, and Reading Ability in Second-Grade Children. *Child Development*, *65*(3), 938–950. doi:10.1111/j.1467-8624.1994.tb00794.x

Harmony, T., Fernández, T., Silva, J., Bosch, J., Valdés, P., Fernández-Bouzas, A., … Rodríguez, D. (1999). *Do specific EEG frequencies indicate different processes during mental calculation? Neuroscience Letters* (Vol. 266, pp. 25–28).

Hartley, T. (2001). Syllabic phase: a bottom-up representation of the temporal structure of speech. In J. A. Bullinaria & W. Lowe (Eds.), *Connectionist Models of Cognition and Perception: Proceedings of the Seventh Neural Computation and Psychology Workshop* (pp. 227–289). Brighton, England: World Scientific.

Hartley, T., & Houghton, G. (1996). A Linguistically Constrained Model of Short-Term Memory for Nonwords. *Journal of Memory and Language*, *31*(35), 1–31.

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (n.d.). *Effects of Rhythm on Memory for Spoken Sequences: A Model and Tests of its Stimulius-Driven Mechanism.*

Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology*, *72*(1), 119–133. doi:10.1348/000709902158801

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 644–655.

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*(2), 180–90. doi:10.1111/j.1469-8986.2005.00272.x

Helenius, P., Uutela, K., & Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain*, *122*(5), 907–913. doi:10.1093/brain/122.5.907

Henry, L. A., & Millar, S. (1991). Memory span increase with age: A test of two hypotheses. *Journal of Experimental Child Psychology*, *51*(3), 459–484.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(49), 20095–100. doi:10.1073/pnas.1213390109

Henson, R. N. (1998). Short-term memory for serial order: the Start-End Model. *Cognitive Psychology*, *36*(2), 73–137. doi:10.1006/cogp.1998.0685

Henson, R. N., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, *38*(4), 426–40.

Henson, R. N., Hartley, T., Burgess, N., Hitch, G. J., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*(8), 1307–1334. doi:10.1080/02724980244000747

Henson, R. N., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained Memory: Error Patterns Rule out Chaining Models of Immediate Serial Recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 80–115. doi:10.1080/027249896392810

Herman, A. B., Houde, J. F., Vinogradov, S., & Nagarajan, S. S. (2013). Parsing the Phonological Loop: Activation Timing in the Dorsal Speech Stream Determines Accuracy in Speech Reproduction. *Journal of Neuroscience*, *33*(13), 5439–5453. doi:10.1523/JNEUROSCI.1472-12.2013

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. doi:10.1038/nrn2113

Hiscock, M., Cheesman, J., Inch, R., Chipuer, H. M., & Graff, L. A. (1989). Rate and variability of finger tapping as measures of lateralized concurrent task effects. *Brain and Cognition*, *10*(1), 87–104. doi:10.1016/0278-2626(89)90077-8

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal Grouping Effects in Immediate Recall: A Working Memory Analysis. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 116–139. doi:10.1080/713755609

Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language*, *61*(1), 97–111. doi:10.1016/j.jml.2009.02.004

Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, *7*(4), 347–362. doi:10.1111/j.2044-835X.1989.tb00811.x

Hitch, G. J., Halliday, M. S., & Littler, J. E. (1989). Item Identification Time and Rehearsal Rate as Predictors of Memory Span in Children. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *41*(2), 321–337.

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Heffernan, T. M. (1991). Speech, "Inner Speech", and the Development of Short-Term Memory: Effects of Picture-Labeling on Recall. *Journal of Experimental Child Psychology*, *51*(2), 220–234.

Holm, L., Ullén, F., & Madison, G. S. (2011). Intelligence and temporal accuracy of behaviour: unique and shared associations with reaction time and motor timing. *Experimental Brain Research*, *214*(2), 175–183. doi:10.1007/s00221-011-2817-6

Holm, L., Ullén, F., & Madison, G. S. (2013). Motor and executive control in repetitive timing of brief intervals. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 365–380.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–15. doi:10.1111/j.1467-7687.2009.00848.x

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–38. doi:10.1162/089892902760807140

Houghton, G. (1990). The problem of serial order: a neural network model of sequence learning and recall. In R. Dale, M. Mellish, & C. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.

Houghton, G., & Hartley, T. (1996). Parallel Models of Serial Behaviour: Lashley Revisited. *Psyche*, *2*(25), 1–25.

Huang, L. (2010). Visual working memory is better characterized as a distributed resource rather than discrete slots. *Journal of Vision*, *10*(14), 8. doi:10.1167/10.14.8

Hulme, C., Roodenrys, S., Brown, G. D. A., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, *86*(4), 527–536. doi:10.1111/j.2044-8295.1995.tb02570.x

Hulme, C., Silvester, J., Smith, S., & Muir, C. (1986). The effects of word length on memory for pictures: Evidence for speech coding in young children. *Journal of Experimental Child Psychology*, *41*(1), 61–75.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech Rate and the Development of Short-Term Memory Span. *Journal of Experimental Child Psychology*, *38*(2), 241–253. doi:10.1016/0022-0965(84)90124-3

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, *47*(1), 72–87.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373.

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*(6), 674–689. doi:10.1016/j.cortex.2010.07.010

Hwang, G., Jacobs, J., Geller, A., Danker, J., Sekuler, R., & Kahana, M. J. (2005). EEG correlates of verbal and nonverbal working memory. *Behavioral and Brain Functions*, *1*, 1–13. doi:10.1186/1744-9081-1-20

Jacobs, J. (1887). Experiments on “prehension.” *Mind*, *12*, 75–79.

Jacoby, L. L. (1973). Encoding processes, rehearsal, and recall requirements. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 302–310.

Jacoby, L. L., & Bartz, W. H. (1972). Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, *11*(5), 561–565.

Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A.-C. (2011). Is the word-length effect linked to subvocal rehearsal? *Cortex*, *47*(4), 484–493. doi:10.1016/j.cortex.2010.07.007

Jantz, T. K., Tomory, J. J., Merrick, C., Cooper, S., Gazzaley, A., & Morsella, E. (2014). Subjective aspects of working memory performance: Memoranda-related imagery. *Consciousness and Cognition*, *25C*, 88–100. doi:10.1016/j.concog.2014.01.013

Jarrold, C., & Hall, D. (2013). The Development of Rehearsal in Verbal Short-Term Memory. *Child Development Perspectives*, *7*(3), 182–186. doi:10.1111/cdep.12034

Jarrold, C., Thorn, A. S. C., & Stephens, E. (2009). The relationships among verbal short-term memory, phonological awareness, and new word learning: evidence from typical development and Down syndrome. *Journal of Experimental Child Psychology*, *102*(2), 196–218. doi:10.1016/j.jecp.2008.07.001

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006). Lexical and semantic influences on item and order memory in immediate serial recognition: evidence from a novel task. *Quarterly Journal of Experimental Psychology*, *59*(5), 949–64. doi:10.1080/02724980543000141

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–55.

Jorgensen, C., & Binsted, K. (2005). Web Browser Control Using EMG Based Sub Vocal Speech Recognition. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS)* (pp. 294c.1–294c.8.). IEEE. doi:10.1109/HICSS.2005.683

Jorgensen, C., & Dusan, S. (2010). Speech interfaces based upon surface electromyography. *Speech Communication*, *52*(4), 354–366. doi:10.1016/j.specom.2009.11.003

Kail, R. (1997). Phonological Skill and Articulation Time Independently Contribute to the Development of Memory Span. *Journal of Experimental Child Psychology*, *67*(1), 57–68. doi:10.1006/jecp.1997.2393

Kail, R., & Park, Y.-S. (1994). Processing Time, Articulation Time, and Memory Span. *Journal of Experimental Child Psychology*, *57*, 281–291.

Kalm, K., Davis, M. H., & Norris, D. (2012). Neural mechanisms underlying the grouping effect in short-term memory. *Human Brain Mapping*, *33*(7), 1634–47. doi:10.1002/hbm.21308

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. doi:10.1037/0096-3445.133.2.189

Kauramäki, J., Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., & Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *The Journal of Neuroscience*, *30*(4), 1314–21. doi:10.1523/JNEUROSCI.1950-09.2010

Kee, D. W., Morris, K., Bathurst, K., & Hellige, J. B. (1986). Lateralized interference in finger tapping: Comparisons of rate and variability measures under speed and consistency tapping instructions. *Brain and Cognition*, *5*(3), 268–279. doi:10.1016/0278-2626(86)90031-X

Keeney, T. J., Cannizzo, S. R., & Flavell, J. H. (1967). Spontaneous and Induced Verbal Rehearsal in a Recall Task. *Child Development*, *38*(4), 953–966.

Kirchner, D. M., & Klatzky, R. L. (1985). Verbal Rehearsal and Memory in Language-Disordered Children. *Journal of Speech and Hearing Research*, *28*(4), 556–565.

Kliegl, R., Smith, J., Heckhausen, J., & Baltes, P. B. (2009). Mnemonic Training for the Acquisition of Skilled Digit Memory. doi:10.1207/s1532690xci0404_1

Kopp, F., Schröger, E., & Lipka, S. (2004). Neural networks engaged in short-term memory rehearsal are disrupted by irrelevant speech in human subjects. *Neuroscience Letters*, *354*(1), 42–45.

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*, *61*(2), 188–203. doi:10.1016/j.ijpsycho.2005.10.001

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399. doi:10.1016/j.tics.2010.06.005

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

Laasonen, M., Service, E., & Virsu, V. (2001). Temporal order and processing acuity of visual, auditory, and tactile perception in developmentally dyslexic young adults. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(4), 394–410. doi:10.3758/CABN.1.4.394

Laasonen, M., Virsu, V., Oinonen, S., Sandbacka, M., Salakari, A., & Service, E. (2012). Phonological and sensory short-term memory are correlates and both affected in developmental dyslexia. *Reading and Writing*, *25*(9), 2247–2273. doi:10.1007/s11145-011-9356-1

Large, E. W., & Jones, M. R. (1999). The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review*, *106*(1), 119–159.

Larsen, J. D., & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: do they have a common source? *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*(8), 1249–68. doi:10.1080/02724980244000765

Lashley, K. S. (1951). The Problem of Serial Order in Behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York: Wiley.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133*(3), 339–354.

Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00482

Leclercq, A.-L., Maillart, C., & Majerus, S. (2013). Nonword Repetition Problems in Children With Specific Language Impairment: A Deficit in Accessing Long-Term Linguistic Representations? *Topics in Language Disorders*, *33*(3), 238–254.

Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology*, *46*(2), 417–427.

Lee, T.-W., Girolami, M., & Sejnowski, T. J. (1999). Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. *Neural Computation*, *11*(2), 417–441. doi:10.1162/089976699300016719

Lehmann, M., & Hasselhorn, M. (2010). The dynamics of free recall and their relation to rehearsal between 8 and 10 years of age. *Child Development*, *81*(3), 1006–1020. doi:10.1111/j.1467-8624.2010.01448.x

Lehmann, M., & Hasselhorn, M. (2012). Rehearsal dynamics in elementary school children. *Journal of Experimental Child Psychology*, *111*(3), 552–560.

Lehongre, K., Morillon, B., Giraud, A.-L., & Ramus, F. (2013). Impaired auditory sampling in dyslexia: further evidence from combined fMRI and EEG. *Frontiers in Human Neuroscience*, *7*(August), 1–8. doi:10.3389/fnhum.2013.00454

Leong, V., & Goswami, U. (2014a). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, *308*, 141–161. doi:10.1016/j.heares.2013.07.015

Leong, V., & Goswami, U. (2014b). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00096

Leong, V., Hämäläinen, J. A., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, *64*(1), 59–73.

Lewandowsky, S., & Farrell, S. (2008). Short-Term Memory: New Data and A Model. *Psychology of Learning and Motivation*, *49*, 1–48. doi:10.1016/S0079-7421(08)00001-7

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition* (p. 359). London: SAGE Publications.

Lian, A., Karlse, P. J., & Eriksen, T. B. (2004). Opposing effects of phonological similarity on item and order memory of words and nonwords in the serial recall task. *Memory*, *12*(3), 314–337. doi:10.1080/09658210344000026

Locke, J. L., & Fehr, F. S. (1970a). Subvocal Rehearsal as a Form of Speech. *Journal of Verbal Learning and Verbal Behavior*, *9*, 495–498.

Locke, J. L., & Fehr, F. S. (1970b). Young children's use of the speech code in a recall task. *Journal of Experimental Child Psychology*, *10*(3), 367–373.

Lorås, H., Stensdotter, A.-K., Ohberg, F., & Sigmundsson, H. (2013). Individual differences in motor timing and its relation to cognitive and fine motor skills. *PloS One*, *8*(7), e69353. doi:10.1371/journal.pone.0069353

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2013.06.006

Luck, S. J., & Zhang, W. W. (2004). Fixed resolution, slot-like representations in visual working memory. *Journal of Vision*, *4*(8), 149–149. doi:10.1167/4.8.149

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. doi:10.1016/j.neuron.2007.06.004

Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology*, *3*(170), 1–10. doi:10.3389/fpsyg.2012.00170

Lyon, D. R. (1977). Individual Differences in Immediate Serial Recall: A Matter of Mnemonics? *Cognitive Psychology*, *9*, 403–411. doi:10.1016/0010-0285(77)90014-7

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

Madison, G. S. (2001). Variability in isochronous tapping: Higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: HUman Perception and Performance*, *27*(2), 411–422.

Madison, G. S. (2014). Sensori-motor synchronisation variability decreases as the number of metrical levels in the stimulus signal increases. *Acta Psychologica*, *147*, 10–16. doi:10.1016/j.actpsy.2013.10.002

Madison, G. S., Forsman, L., Blom, Ö., Karabanov, A., & Ullén, F. (2009). Correlations between intelligence and components of serial timing variability. *Intelligence*, *37*(1), 68–75. doi:10.1016/j.intell.2008.07.006

Madison, G. S., Karampela, O., Ullén, F., & Holm, L. (2013). Effects of practice on variability in an isochronous serial interval production task: Asymptotical levels of tapping variability after training are similar to those of musicians. *Acta Psychologica*, *143*(1), 119–128. doi:10.1016/j.actpsy.2013.02.010

Mainela-Arnold, E., & Evans, J. L. (2005). Beyond Capacity Limitations: Determinants of Word Recall Performance on Verbal Working Memory Span Tasks in Children with SLI. *Journal of Speech, Language, and Hearing Research*, *48*(4), 897–909. doi:10.1044/1092-4388(2005/062)

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: New evidence from a novel word learning paradigm. *Journal of Experimental Child Psychology*, *116*(4), 811–828.

Majerus, S., Leclercq, A.-L., Grossmann, A., Billard, C., Touzin, M., Van der Linden, M., & Poncelet, M. (2009). Serial order short-term memory capacities and specific language impairment: no evidence for a causal association. *Cortex*, *45*(6), 708–20. doi:10.1016/j.cortex.2008.10.006

Majerus, S., Linden, M. Van Der, Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*, *51*(2), 297–306. doi:10.1016/j.jml.2004.05.002

Majerus, S., Poncelet, M., Elsen, B., & van der Linden, M. (2006). Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. *European Journal of Cognitive Psychology*, *18*(6), 848–873. doi:10.1080/09541440500446476

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, *93*(2), 95–119. doi:10.1016/j.jecp.2005.07.005

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, *107*(2), 395–419. doi:10.1016/j.cognition.2007.10.003

Majerus, S., Vrancken, G., & Van der Linden, M. (2003). Perception and short-term memory for verbal information in children with specific language impairment: further evidence for impaired short-term memory capacities. *Brain and Language*, *87*(1), 160–161.

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, *94*(20), 10979–10984. doi:10.1073/pnas.94.20.10979

Maki, R. H., & Schuler, J. (1980). Effects of rehearsal duration and level of processing on memory for words. *Journal of Verbal Learning and Verbal Behavior*, *19*(1), 36–45.

Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, *30*(3), 247–256. doi:10.3758/BF03214280

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language & Communication Disorders*, *44*(4), 466–88. doi:10.1080/13682820802591643

Marshall, P. H., & Werder, P. R. (1972). The effects of the elimination of rehearsal on primacy and recency. *Journal of Verbal Learning and Verbal Behavior*, *11*(5), 649–653.

Martinez Perez, T., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a specific impairment of serial order short-term memory in dyslexic children. *Dyslexia*, *18*(2), 94–109. doi:10.1002/dys.1438

Martinez Perez, T., Majerus, S., & Poncelet, M. (2012). The contribution of short-term memory for serial order to early reading acquisition: evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *111*(4), 708–23. doi:10.1016/j.jecp.2011.11.007

Martinez Perez, T., Majerus, S., & Poncelet, M. (2013). Impaired short-term memory for order in adults with dyslexia. *Research in Developmental Disabilities*, *34*(7), 2211–2223. doi:10.1016/j.ridd.2013.04.005

Maybery, M. T., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: implications for models of serial verbal memory. *Journal of Memory and Language*, *47*(3), 360–385.

Mazzoni, D., & Dannenburg, R. (2000). Audacity. Retrieved from audacity.sourceforge.net

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “Specifics” of Specific Reading Disability and Specific Language Impairment. *Journal of Child Psychology and Psychiatry*, *41*(7), 869–874. doi:10.1111/1469-7610.00674

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348–367.

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to Read: The Role of Short-Term Memory and Phonological Skills. *Journal of Experimental Child Psychology*, *58*(1), 112–133. doi:10.1006/jecp.1994.1028

Mcguire, P. K., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, *26*(1), 29–38.

Meltzner, G. S., Sroka, J., Heaton, J. T., Gilmore, L. D., Colby, G., Roy, S., … De Luca, C. J. (2008). Speech Recognition for Vocalized and Subvocal Modes of Production using Surface EMG Signals from the Neck and Face. In *Interspeech* (pp. 2667–2670).

Morris, R. G., Gick, M. L., & Craik, F. I. M. (1988). Processing resources and age differences in working memory. *Memory & Cognition*, *16*(4), 362–366. doi:10.3758/BF03197047

Morton, J., Marcus, S. M., & Frankish, C. R. (1976). Perceptual Centers (P-centers). *Psychological Review*, *83*(5), 405–408.

Muneaux, M., Ziegler, J. C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: evidence from French. *NeuroReport*, *15*(8), 1255–1259. doi:10.1097/01.wnr.0000127459.31232.c4

Murray, D. J. (1967). The Role of Speech Responses in Short-Term Memory. *Canadian Journal of Psychology*, *21*(3), 263–276.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, *78*(4), 679–684.

Muter, V., & Snowling, M. J. (1998). Concurrent and Longitudinal Predictors of Reading: The Role of Metalinguistic and Short-Term Memory Skills. *Reading Research Quarterly*, *33*(3), 320–337. doi:10.1598/RRQ.33.3.4

Nairne, J. S. (2002). Remembering Over the Short-Term: The Case Against the Standard Model. *Annual Review of Psychology*, *53*, 53–81.

Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language*, *50*(2), 113–133. doi:10.1016/j.jml.2003.09.005

Naveh-Benjamin, M., & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *38*, 739–751.

Nicolson, R. I. (1981). The Relationship Between Memory Span and Processing Speed. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and Learning* (pp. 179–183). Boston, MA: Springer US. doi:10.1007/978-1-4684-1083-9

Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading Disabilities in SLI and Dyslexia Result From Distinct Phonological Impairments. *Developmental Neuropsychology*, *34*(3), 296–311. doi:10.1080/87565640902801841

Nithart, C., Demont, E., Metz-Lutz, M.-N., Majerus, S., Poncelet, M., & Leybaert, J. (2011). Early contribution of phonological awareness and later influence of phonological memory throughout reading acquisition. *Journal of Research in Reading*, *34*(3), 346–363. doi:10.1111/j.1467-9817.2009.01427.x

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What Does Motor Efference Copy Represent? Evidence from Speech Production. *Journal of Neuroscience*, *33*(41), 16110–16116. doi:10.1523/JNEUROSCI.2137-13.2013

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, *272*(1), 29–32.

Oganian, Y., & Ahissar, M. (2012). Poor anchoring limits dyslexics' perceptual, memory, and reading skills. *Neuropsychologia*, *50*(8), 1895–905. doi:10.1016/j.neuropsychologia.2012.04.014

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. In *Progress in brain research* (Vol. 159, pp. 99–120). Elsevier. doi:10.1016/S0079-6123(06)59007-7

Oppenheim, G. M. (2012). The case for subphonemic attenuation in inner speech: comment on Corley, Brocklehurst, and Moat (2011). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 502–12. doi:10.1037/a0025257

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. doi:10.1016/j.cognition.2007.02.006

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: the flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–60. doi:10.3758/MC.38.8.1147

Ornstein, P. A., Naus, M. J., & Liberty, C. (1975). Rehearsal and Organizational Processes in Children 's Memory. *Child Development*, *46*(4), 818–830.

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: measuring musical timing skills. *Dyslexia*, *9*(1), 18–36. doi:10.1002/dys.233

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781.

Palmer, S. E., & Ornstein, P. A. (1971). Role of rehearsal strategy in serial probed recall. *Journal of Experimental Psychology*, *88*(1), 60–66. doi:10.1037/h0030658

Papagno, C., & Vallar, G. (1995). Verbal Short-term Memory and Vocabulary Learning in Polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 98–107. doi:10.1080/14640749508401378

Pariyadath, V., Plitt, M. H., Churchill, S. J., & Eagleman, D. M. (2012). Why overlearned sequences are special: distinct neural networks for ordinal sequences. *Frontiers in Human Neuroscience*, *6*, 1–9. doi:10.3389/fnhum.2012.00328

Parkinson, A. L., Korzyukov, O., Larson, C. R., Litvak, V., & Robin, D. A. (2013). Modulation of effective connectivity during vocalization with perturbed auditory feedback. *Neuropsychologia*, *51*(8), 1471–1480. doi:10.1016/j.neuropsychologia.2013.05.002

Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent Effects of Grouping by Time, Voice, and Location on Response Timing in Verbal Serial Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1349–1355.

Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory Processing of Amplitude Envelope Rise Time in Adults Diagnosed With Developmental Dyslexia. *Scientific Studies of Reading*, *11*(3), 259–286. doi:10.1080/10888430701344280

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Letters to Nature*, *362*, 342–345.

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*, 1–17. doi:10.3389/fpsyg.2012.00320

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. doi:10.1093/cercor/bhs118

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, *36*, 179–190.

Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquière, P., & Wouters, J. (2011). Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. *Research in Developmental Disabilities*, *32*(6), 2810–9. doi:10.1016/j.ridd.2011.05.025

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, *41*(1), 245–255. doi:10.1016/S0167-6393(02)00107-3

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1071–86. doi:10.1098/rstb.2007.2160

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie*, *50*(4), 408–412.

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*, 777. doi:10.3389/fnhum.2013.00777

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212–218. doi:10.1016/S0959-4388(03)00035-7

Ramus, F. (2014). Neuroimaging sheds new light on the phonological deficit in dyslexia. *Trends in Cognitive Sciences*, *18*(6), 274–5. doi:10.1016/j.tics.2014.01.009

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, *61*(1), 129–41. doi:10.1080/17470210701508822

Rapala, M. M., & Brady, S. (1990). Reading ability and short-term memory: The role of phonological processing. *Reading and Writing*, *2*(1), 1–25. doi:10.1007/BF00383371

Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping*, *29*(11), 1231–42. doi:10.1002/hbm.20460

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1638–1654.

Remez, R. E., Thomas, E. F., Dubowski, K. R., Koinis, S. M., Porter, N. A. C., Paddu, N. U., … Grossman, Y. S. (2013). Modulation sensitivity in the perceptual organization of speech. *Attention, Perception & Psychophysics*. doi:10.3758/s13414-013-0542-x

Rosen, S. (1999). Language disorders: A problem with auditory processing? *Current Biology*, *9*(18), R698–R700. doi:10.1016/S0960-9822(99)80443-6

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, *31*(3-4), 509–527. doi:10.1016/S0095-4470(03)00046-9

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*(1), 63–77.

Rundus, D. (1974). Output order and rehearsal in multi-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 656–663.

Rundus, D. (1980). Maintenance rehearsal and long-term recency. *Memory & Cognition*, *8*(3), 226–230. doi:10.3758/BF03197610

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 99–105.

Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*(2), 137–147. doi:10.1080/14640746908400206

Ryan, J. (1969b). Temporal Grouping, Rehearsal and Short-Term Memory. *The Quarterly Journal of Experimental Psychology*, *21*(2), 148–155. doi:10.1080/14640746908400207

Saint-Aubin, J., & Poirier, M. (1999a). Semantic similarity and immediate serial recall: is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *52*(2), 367–94. doi:10.1080/713755814

Saint-Aubin, J., & Poirier, M. (1999b). The Influence of Long-term Memory Factors on Immediate Serial Recall: An Item and Order Analysis. *International Journal of Psychology*, *34*(5-6), 347–352. doi:10.1080/002075999399675

Saito, S. (1997). When articulatory suppression does not suppress the activity of the phonological loop. *British Journal of Psychology*, *88*(4), 565–578. doi:10.1111/j.2044-8295.1997.tb02658.x

Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory*, *9*(4), 313–322. doi:10.1080/09658210143000164

Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research*, *40*(1), 10–18. doi:10.1111/1468-5884.00070

Salamé, P., & Baddeley, A. D. (1982). Disruption of Short-Term Memory by Unattended Speech: Implications for the Structure of Working Memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150–164.

Saoud, H., Josse, G., Bertasi, E., Truy, E., Chait, M., & Giraud, A.-L. (2012). Brain-speech alignment enhances auditory cortical responses and speech perception. *The Journal of Neuroscience*, *32*(1), 275–281. doi:10.1523/JNEUROSCI.3970-11.2012

Sato, M., Troille, E., Ménard, L., Cathiard, M.-A., & Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*. doi:10.1007/s00221-013-3510-8

Schulkind, M. D. (1999). Long-term memory for temporal structure: Evidence from the indentification of well-known and novel songs. *Memory & Cognition*, *27*(5), 896–906. doi:10.3758/BF03198542

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural Brain Research*, *216*(2), 685–691. doi:10.1016/j.bbr.2010.09.015

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biological Psychology*, *87*(1), 146–151. doi:10.1016/j.biopsycho.2011.02.021

Schwartze, M., Tavano, A., Schröger, E., & Kotz, S. A. (2012). Temporal aspects of prediction in audition: cortical and subcortical neural mechanisms. *International Journal of Psychophysiology*, *83*(2), 200–207. doi:10.1016/j.ijpsycho.2011.11.003

Scott, M. (2013). Corollary Discharge Provides the Sensory Content of Inner Speech. *Psychological Science*, (July), 0956797613478614–. doi:10.1177/0956797613478614

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, *133*(4), 286–292. doi:10.1121/1.4794932

Service, E., Maury, S., & Luotoniemi, E. (2007). Individual differences in phonological learning and verbal STM span. *Memory & Cognition*, *35*(5), 1122–1135. doi:10.3758/BF03193483

Severin, F. T., & Rigby, M. K. (1963). Influence of digit grouping on memory for telephone numbers. *Journal of Applied Psychology*, *47*(2), 117–119.

Shafer, V. L., Schwartz, R. G., & Martin, B. (2011). Evidence of deficient central speech processing in children with specific language impairment: The T-complex. *Clinical Neurophysiology*, *122*(6), 1137–1155.

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, *16*(4), 219–27. doi:10.1002/hbm.10046

Shergill, S. S., Tracy, D. K., Seal, M., Rubia, K., & McGuire, P. K. (2006). Timing of covert articulation: an fMRI study. *Neuropsychologia*, *44*(12), 2573–7. doi:10.1016/j.neuropsychologia.2006.04.005

Simpson, M. I. G., & Prendergast, G. (2013). Auditory magnetic evoked responses. In G. G. Celesia (Ed.), *Disorders of Peripheral and Central Auditory Processing Handbook of Clinical Neurophysiology* (Vol 10., pp. 253–270).

Snowling, M. J. (1998). Dyslexia as a Phonological Deficit: Evidence and Implications. *Child and Adolescent Mental Health*, *3*(1), 4–11. doi:10.1111/1475-3588.00201

Snowling, M. J. (2000). *Dyslexia*. Oxford: Blackwell.

Snowling, M. J., Bishop, D. V. M., & Stothard, S. E. (2000). Is Preschool Language Impairment a Risk Factor for Dyslexia in Adolescence? *Journal of Child Psychology and Psychiatry*, *41*(5), 587–600. doi:10.1111/1469-7610.00651

Snowling, M. J., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*(03), 369–373. doi:10.1017/S0142716400009279

Snowling, M. J., Nation, K., Moxham, P., Gallagher, A., & Frith, U. (1997). Phonological Processing Skills of Dyslexic Students in Higher Education: A Preliminary Report. *Journal of Research in Reading*, *20*(1), 31–41. doi:10.1111/1467-9817.00018

Soltész, F., Szűcs, D., Leong, V., White, S., & Goswami, U. (2013). Differential Entrainment of Neuroelectric Delta Oscillations in Developmental Dyslexia. *PLoS ONE*, *8*(10), e76608. doi:10.1371/journal.pone.0076608

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489.

Standing, L., Bond, B., Smith, P., & Isely, C. (1980). Is the immediate memory span determined by subvocalization rate? *British Journal of Psychology*, *71*(4), 525–539. doi:10.1111/j.2044-8295.1980.tb01764.x

Stigler, J. W., Lee, S.-Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, *23*(1), 1–20.

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*, 1–9. doi:10.3758/s13414-014-0690-7

Surprenant, A. M., Lecompte, D. C., & Neath, I. (2000). Manipulations of irrelevant information: Suffix effects with articulatory suppression and irrelevant speech. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(2), 325–348. doi:10.1080/713755892

Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology*, *107*(3), 306–324.

Tehan, G., & Lalor, D. M. (2000). Individual differences in memory span: The contribution of rehearsal, access to lexical memory, and output speed. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(4), 1012–1038.

Teki, S., Chait, M., Kumar, S., Shamma, S. A., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife*, *2*(e00699), 1–16. doi:10.7554/eLife.00699

The MathWorks Inc. (2010). MATLAB - The Language of Technical Computing. Natick, Massachusetts: The MathWorks, Inc.

Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, *29*(3), 334–348. doi:10.1111/j.1467-9817.2006.00312.x

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology - Paris*, *102*(1), 120–129. doi:10.1016/j.jphysparis.2008.03.007

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *52*(2), 303–24. doi:10.1080/713755823

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*(314). doi:10.3389/fnhum.2012.00314

Tian, X., & Poeppel, D. (2013). The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing. *Journal of Cognitive Neuroscience*, *25*(7), 1020–1036. doi:10.1162/jocn_a_00381

Tierney, A. T., & Kraus, N. (2013a). The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. *Journal of Neuroscience*, *33*(38), 14981–14988. doi:10.1523/JNEUROSCI.0612-13.2013

Tierney, A. T., & Kraus, N. (2013b). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language*, *124*(3), 225–231. doi:10.1016/j.bandl.2012.12.014

Timm, J., Sanmiguel, I., Keil, J., Schröger, E., & Schönwiesner, M. (2014). Motor Intention Determines Sensory Attenuation of Brain Responses to Self-initiated Sounds. *Journal of Cognitive Neuroscience*, *26*(7), 1481–1489. doi:10.1162/jocn_a_00552

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). Test of word reading efficiency. *Austin, TX: Pro-Ed*.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition*, *18*(6), 559–567. doi:10.3758/BF03197098

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *5*(2), 193–197.

Tzeng, O. J. (1973). Stimulus meaningfulness, encoding variability, and the spacing effect. *Journal of Experimental Psychology*, *99*(2), 162–166.

Ullén, F., Forsman, L., Blom, O., Karabanov, A., & Madison, G. S. (2008). Intelligence and variability in a simple timing task share neural substrates in the prefrontal white matter. *The Journal of Neuroscience*, *28*(16), 4238–43. doi:10.1523/JNEUROSCI.0825-08.2008

Ullén, F., Söderlund, T., Kääriä, L., & Madison, G. S. (2012). Bottom–up mechanisms are involved in the relation between accuracy in timing tasks and intelligence — Further evidence using manipulations of state motivation. *Intelligence*, *40*(2), 100–106. doi:10.1016/j.intell.2012.01.012

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), 8780–5. doi:10.1073/pnas.1117465109

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. doi:10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Von Holst, V. E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Die Naturwissenschaften*, *37*(20), 464–476. doi:10.1007/BF00622503

Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Waber, D. P., Weiler, M. D., Bellinger, D. C., Marcus, D. J., Forbes, P. W., Wypij, D., & Wolff, P. H. (2000). Diminished Motor Timing Control in Children Referred for Diagnosis of Learning Problems. *Developmental Neuropsychology*, *17*(2), 181–197. doi:10.1207/S15326942DN1702_03

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212. doi:10.1037//0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). Comprehensive test of phonological processing: CTOPP.

Wang, S., & Gathercole, S. E. (2013). Working memory deficits in children with reading difficulties: Memory span and dual task coordination. *Journal of Experimental Child Psychology*, 188–97. doi:10.1016/j.jecp.2012.11.015

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 186–210.

Warmington, M., Stothard, S. E., & Snowling, M. J. (2013). Assessing dyslexia in higher education: the York adult assessment battery-revised. *Journal of Research in Special Educational Needs*, *13*(1), 48–56. doi:10.1111/j.1471-3802.2012.01264.x

Waters, G. S., Rochon, E., & Caplan, D. (1992). The Role of High-Level Speech Planning in Rehearsal: Evidence from Patients with Apraxia of Speech. *Journal of Memory and Language*, *31*, 54–73. doi:10.1016/0749-596X(92)90005-I

Watkins, M. J., LeCompte, D. C., Elliott, M. N., & Fish, S. B. (1992). Short-Term Memory for the Timing of Auditory and Visual Signals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 931–937.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*(2), 89–104.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Weist, R. M. (1972). The role of rehearsal: Recopy or reconstruct. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 440–450.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the Time Course of Phonological Encoding. *Journal of Memory and Language*, *34*(3), 311–334. doi:10.1006/jmla.1995.1014

Wickelgren, W. A. (1964). Size of Rehearsal Group and Short-Term Memory. *Journal of Experimental Psychology*, *68*(4), 413–9.

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. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, *19*(2), 97–102. doi:10.1080/14640746708400077

Wing, A. M. (1977). Effects of type of movement on the temporal precision of response sequences. *British Journal of Mathematical and Statistical Psychology*, *30*(1), 60–72. doi:10.1111/j.2044-8317.1977.tb00724.x

Wing, A. M., & Kristofferson, A. B. (1973a). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, *14*(1), 5–12. doi:10.3758/BF03198607

Wing, A. M., & Kristofferson, A. B. (1973b). The timing of interresponse intervals. *Perception & Psychophysics*, *13*(3), 455–460. doi:10.3758/BF03205802

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing*, *15*(1-2), 179–206. doi:10.1023/A:1013880723925

Wolff, P. H., Cohen, C., & Drake, C. (1984). Impaired motor timing control in specific reading retardation. *Neuropsychologia*, *22*(5), 587–600. doi:10.1016/0028-3932(84)90023-X

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, 1406219111–. doi:10.1073/pnas.1406219111

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., … Service, E. (2014). Two Distinct Auditory-Motor Circuits for Monitoring Speech Production as Revealed by Content-Specific Suppression of Auditory Cortex. *Cerebral Cortex First Published Online January 9, 2014*. doi:10.1093/cercor/bht351

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–5. doi:10.1038/nature06860

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434–41. doi:10.1177/0956797611417006

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *The British Journal of Mathematical and Statistical Psychology*, *57*(Pt 1), 173–81. doi:10.1348/000711004849222