Article:

Gilbert, Rebecca Anne, Hitch, Graham James and Hartley, Tom (2016) Temporal precision and the capacity of auditory-verbal short-term memory. The Quarterly Journal of Experimental Psychology. pp. 1-43. ISSN: 1747-0226

https://doi.org/10.1080/17470218.2016.1239749

Author accepted manuscript. In press: Quarterly Journal of Experimental Psychology

Temporal Precision and the Capacity of Auditory-Verbal Short-Term Memory

Rebecca A. Gilbert[1], Graham J. Hitch[1], and Tom Hartley[1]

[1]Department of Psychology, University of York, UK.

Note: Rebecca A. Gilbert is now at the Department of Experimental Psychology, University College London, UK.

Corresponding author
Dr Rebecca Gilbert
Department of Experimental Psychology
Psychology and Language Sciences
University College London
26 Bedford Way
London
WC1H 0AP
Phone: +44 7554 887 536
Email: becky.gilbert@ucl.ac.uk

Acknowledgement

We would like to thank Dr Stephen Darling for providing a computerised version of the Visual Patterns Test. The authors declare no conflicts of interest.

Abstract

The capacity of serially-ordered auditory-verbal short-term memory (AVSTM) is sensitive to the timing of the material to be stored, and both temporal processing and AVSTM capacity are implicated in the development of language. We developed a novel “rehearsal-probe” task to investigate the relationship between temporal precision and the capacity to remember serial order. Participants listened to a sub-span sequence of spoken digits and silently rehearsed the items and their timing during an unfilled retention interval. After an unpredictable delay, a tone prompted report of the item being rehearsed at that moment. An initial experiment showed cyclic distributions of item responses over time, with peaks preserving serial order and broad, overlapping tails. The spread of the response distributions increased with additional memory load and correlated negatively with participants’ auditory digit spans. A second study replicated the negative correlation and demonstrated its specificity to AVSTM by controlling for differences in visuo-spatial STM and nonverbal IQ. The results are consistent with the idea that a common resource underpins both the temporal precision and capacity of AVSTM. The rehearsal-probe task may provide a valuable tool for investigating links between temporal processing and AVSTM capacity in the context of speech and language abilities.

[Words: 198]

Key words: short-term memory; timing; subvocal rehearsal; serial order memory

Introduction

Auditory-verbal short-term memory (AVSTM) is related to a range of speech, language and cognitive skills including vocabulary development (Baddeley, Gathercole, & Papagno, 1998; Leclercq & Majerus, 2010), word learning (Majerus, Poncelet, Elsen, & van der Linden, 2006; Service, Maury, & Luotoniemi, 2007), reading (Anvari, Trainor, Woodside, & Levy, 2002; Martinez Perez, Majerus, & Poncelet, 2012), verbal reasoning (Kane et al., 2004) and verbal IQ (Cantor, Engle, & Hamilton, 1991). AVSTM is sensitive to phonemic similarity, word length and articulatory suppression in ways that suggest it operates as a "phonological loop", a serially organised speech input-output system comprising a phonological buffer store that loses information rapidly but can be refreshed by subvocal rehearsal (see Baddeley, 2007 for an overview). However, the mechanisms of representation within the phonological loop, and specifically representation of temporal and serial information, are much less clear. A deeper understanding of these mechanisms is vital because serial order is not just a byproduct of AVSTM, but a defining characteristic, essential to its function.

Accordingly the capacity of AVSTM is typically measured using immediate serial recall (ISR) tasks, where an arbitrary sequence of spoken items must be remembered in the correct order. Span is the limit on the number of items beyond which accurate serial recall is the exception rather than the rule. For familiar items such as digits, span is around 7 items and reflects principally memory for serial order. Thus, order errors predominate when span is exceeded (Aaronson, 1968) and their distribution reflects ordinal distance within the sequence, with migrations to adjacent positions the most frequent (Bjork & Healy, 1974). Order errors are highly sensitive to temporal characteristics of spoken sequences such as their rhythmic structure (Frankish, 1985; Hartley, Hurlstone, & Hitch, 2016; Hitch, Burgess, Towse, & Culpin, 1996; Ryan, 1969a, 1969b). These results indicate that the capacity-defining limits of memory for serial order should not be considered independently of timing. However, although links between AVSTM for serial order and temporal properties (speech rate, duration, rhythm) of verbal materials have been well studied, a limitation of previous empirical studies is that performance is typically measured in terms of memory for item *order*, not item *timing*. Thus it is not clear to what extent the precision of an underlying temporal representation is subjected to the same capacity limitations that determine accurate maintenance of serial order.

This issue has some significance for current debates surrounding the role of auditory temporal processing in more general speech and language abilities in children and adults, where one prominent claim is that variability in processing temporal information is predictive of phonological development and skills (Goswami, 2011; Tierney & Kraus, 2014). This view is supported by evidence showing that temporal variability in auditory input processing is linked to speech and language abilities in typical and atypical development (Benasich & Tallal, 2002; Grube, Cooper, & Griffiths, 2013; Thomson & Goswami, 2008; Wolff, Michel, & Ovrut, 1990; Wolff, 2002; Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014). In addition, both children and adults with atypical language development show characteristic deficits in AVSTM for serial order (Corkin, 1974; Martinez Perez, Majerus, Mahot, & Poncelet, 2012; Martinez Perez, Majerus, & Poncelet, 2013), and individual differences in the variability of rhythmic tapping are related to AVSTM (digit span) in typical adults (Saito, 2001). The relationships among variability in auditory temporal processing, serial order AVSTM and speech and language skills could be explained under the assumption that memory for serial order relies on the accurate perception and stable representation of auditory timing.

Indeed, some computational models of short-term memory propose mechanisms whereby serial order information is derived from aspects of timing (Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Hartley & Houghton, 1996; Hartley et al., 2016). In these models, each item is associated with the state of a temporally-sensitive context signal at encoding. When this signal is replayed at retrieval, a competitive queuing mechanism is proposed in which the items are reactivated by the signal in parallel, and the item with the highest activation is selected for retrieval and then supressed. Other models use the same competitive queuing system but propose a positional rather than time-based mechanism to encode item order, where the state of the context signal changes in response to each item (irrespective of the inter-item interval), and the sequence timing is therefore not retained (see Hurlstone, Hitch, & Baddeley, 2014 for a review of competitive queuing and Macken, Taylor, & Jones, 2015 for an example of an account that does not use this mechanism). In both types of competitive queuing models, patterns of recall and order errors are well explained by noise introduced in the item activation levels when the context signal is replayed. Items that have been encoded with more similar contexts are more likely to be confused with one another at retrieval, and items with fewer contextual competitors (e.g. the first and last items) are more likely to be retrieved correctly.

One reason for the continued existence of models with both temporal and positional coding systems is that, empirically, the extent to which ISR depends on accurate memory for sequence timing is not well understood. It could be that ISR depends only on a coarse-grained representation of sequence timing sufficient to discriminate serial order, with no need for greater precision. Alternatively, a fine-grained representation of sequence timing might impact the ability to recall items in the correct order – this type of account might provide a more parsimonious explanation for effects linking timing to memory capacity and to speech and language.

While direct evidence on the temporal precision of AVSTM is lacking, previous work has begun to examine the relationship between STM for sequence order and timing. For example, Farrell and McLaughlin (2007) presented irregularly-timed sequences and compared post-cued recognition memory for their serial order or temporal rhythm. Patterns of performance differed between the two recognition tasks, suggesting a possible dissociation between ordinal and temporal information in STM. Although Farrell (2008) found striking similarities when he went on to compare the recall of rhythm and order, he interpreted the differences as continuing to question the notion of a single timing signal driving memory for serial order. However, one problem in interpreting differences between tasks is that they may differ in unintended ways, highlighting the value of seeking independent converging evidence. In the present studies we examined an alternative approach that involves measuring temporal precision in the covert serial recall of isochronous sequences.

Recent discussions of the nature of capacity limitations in the domain of visuo-spatial STM raise issues that may be equally applicable to AVSTM. In visuo-spatial STM, the questions concern whether items are represented with fixed or variable precision, and whether or not limited resources can be dynamically allocated to precision. Broadly speaking, theoretical accounts of visual memory capacity can be categorised as fixed slot or dynamic resource models. Fixed slot models (e.g. Luck & Vogel, 1997) assume that a limited number of items can be remembered with a given precision, and that no information is retained for additional items beyond the number of slots. By contrast, resource models (e.g. Ma, Husain, & Bays, 2014) propose that a shared pool of resources can be used to represent items of any number or complexity, such that the precision of the item representations decreases as the overall cognitive load increases. Other models are more nuanced, involving combinations of fixed slots and dynamic resource allocation (see e.g. Zhang & Luck, 2008). The tasks used to investigate these issues measure the precision with which continuous features such as location, colour and orientation can be remembered as a function of variables such as cognitive load (Wilken & Ma, 2004).

Although little work has been done to determine whether there is an analogous issue in AVSTM, it has been suggested that the flexible resource allocation model applies to non-verbal auditory memory (Joseph, Teki, Kumar, Husain, & Griffiths, 2016). A recent study by Kumar et al. (2013) found that the precision of immediate memory for the pitch of auditory tones decreased with increasing memory load (number of tones in the sequence). Similarly, Teki and Griffiths found that precision in the reproduction of auditory temporal intervals decreased with increased memory load (number of durations in the sequence). However, it is not clear whether these results generalise to temporal precision in memory for sequences of spoken words. In the present study, our initial aim was to test whether a similar trade-off exists between memory load and the temporal precision of sequence representations in AVSTM. If so this would suggest flexible allocation of limited resources to temporal precision in AVSTM broadly analogous to precision in visuo-spatial STM (Ma et al., 2014).

To achieve this aim we first had to devise a method of measuring temporal precision for sequences in STM. The method we developed was somewhat analogous to auditory-motor synchronization-continuation tasks (Thomson & Goswami, 2008) where an auditory-motor ‘entrainment’ period is followed by unpaced tapping in which the participant attempts to maintain the same timing. Timing variability in these tasks is measured by the standard deviation of the inter-tap intervals. However, unlike tapping tasks, which typically assess synchronization to tones, our task was designed to assess the temporal precision of rehearsing novel sequences in AVSTM where the entrainment stimulus is a series of spoken words. Another important difference from the finger-tapping paradigm is that our task does not involve overt behaviour and requires instead subvocal rehearsal of the input sequence. Overt actions such as finger-tapping or spoken repetition would provide continuous sensorimotor feedback about timing that might influence participants' performance. Furthermore, overt speech may differ qualitatively from inner speech, for instance in the necessity for discrete item selection and/or the abstractness of item representations (Oppenheim & Dell, 2008, 2010).

In our novel "rehearsal-probe" task[1], a listener is presented with an auditory-verbal sequence and instructed to subvocally rehearse it exactly as presented, that is, with accurate item and inter-item timing in addition to item and order accuracy. The listener continues subvocalising the sequence in this manner until hearing a probe consisting of a brief tone and is instructed to respond with the item being rehearsed when the probe was presented. By presenting the probe after variable and unpredictable delays, the response data provide a picture of the time-course of the listener's sequence reproduction. In particular, by cumulating responses after the same probe delay across multiple trials, we are able to quantify the temporal variability of responses. We use short, subspan sequences presented twice over before the start of rehearsal in order to eliminate or at least reduce serial order errors and allow a focus on the accuracy of timing. However, we note that, in our view, a serial order error is necessarily an error of timing in which the temporal imprecision is big enough to change the item's serial position. Thus, rather than treating coarse-grain serial and fine-grain temporal errors as distinct phenomena with distinct causes, we investigate the more parsimonious possibility that they have a common origin. In Experiment 1 we show that temporal precision decreases when AVSTM is loaded with irrelevant items, and is positively correlated with AVSTM span. Experiment 2 goes on to show that the relationship between temporal precision and span replicates with a larger sample and remains when individual differences in non-verbal IQ and visuo-spatial STM are controlled. These findings are consistent with the idea that fine-grained sequence timing and serial order depend on common limited resources that are specific to AVSTM.

Experiment 1

Our first experiment had two aims, one of which was to establish the rehearsal-probe task as a useful method for assessing the temporal precision of regularly-paced subvocal rehearsal. If this task provides a valid measure of subvocal rehearsal timing, we would expect to see systematic variation in the frequency of responding with each item in the sequence as a function of the delay of the probe. Thus, the frequency of responding with any particular item would peak around the time-points when it would be expected to be rehearsed (based on the presented timing) and would decline with increasing temporal interval either side. The widths of these frequency distributions would provide an empirical index of the temporal precision of the rehearsal sequences.

The second aim was to explore the hypothesis that the temporal precision of subvocal rehearsal is constrained by the resources that underpin AVSTM storage capacity. To achieve this we made use of the STM preload paradigm (Baddeley & Hitch, 1974; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; FitzGerald & Broadbent, 1985; Halford, Maybery, O’Hare, & Grant, 1994; Morris, Gick, & Craik, 1988), and required participants to hold irrelevant letters in AVSTM while performing the rehearsal-probe task. The logic of the preload paradigm holds that, if the two tasks depend on the same set of limited resources, then increasing the load in the first (preload) task will result in fewer resources available, and consequently a functional increase in memory load, for the second task. In the present study, the temporal precision of rehearsal in the probe task is expected to decrease when the number of preload items is increased. If, on the other hand, temporal precision is invariant to memory load, temporal precision should remain the same despite a change in the size of the letter preload list.

We also took the opportunity to seek converging evidence for a dynamic trade-off between temporal precision and memory load by examining the relationship between individual differences in AVSTM span and temporal precision. We assume that individuals with lower AVSTM spans have fewer of the resources that underpin storage capacity than high-span individuals, and will therefore use a higher proportion of these resources to maintain and rehearse a sub-span sequence of any given length. Thus, the memory load experienced when performing the rehearsal-probe task with a fixed number of items will be higher for low-span than high-span individuals. If the temporal precision of rehearsing a sub-span sequence is limited by the same resources that limit span, then the temporal variability of rehearsal should correlate negatively with digit span.

Method

Participants

Participants were 25 undergraduate volunteers from the University of York (18 women, mean age 19.50 years, range 18 to 22 years). Participants were eligible if they spoke English as a first language and did not report a diagnosed hearing problem or developmental language disorder (e.g. dyslexia, SLI). A small cash reward or course credit hours was given as compensation.

Materials

The digits 0 through 9 were recorded monaurally at 44100 Hz in a sound-attenuated booth by a female native English speaker. Post-recording processing of the sound files was carried out using Audacity (Mazzoni & Dannenburg, 2000) and Praat (Boersma & Weenink, 1992) software. Each digit recording was adjusted to a duration of 400 ms, matched for maximum amplitude, and all pitch contours were aligned and reduced by 50%. Finally, the acoustic onsets of the digits were adjusted so that their p-centres would occur at regular intervals during sequence presentation (see Supplementary Material for more information about pitch modification and p-centre adjustment). The letter stimuli were recorded and processed in the same manner as the digit stimuli, with the exception that they did not undergo pitch contour manipulation or p-centre adjustment (see Supplementary Material for details). The probe stimulus was a 150 ms 440 Hz steady-state tone with 15 ms onset and offset ramps. Stimulus sequences were presented via an E-Prime program (Psychology Software Tools, Inc., Pittsburgh, PA).

Design

The experiment had a within-subjects design in which all participants completed alternating blocks of high and low memory load trials. A set of 8 probe delays was distributed randomly within each block. Probe times were selected such that, given perfect rehearsal timing, they would occur during the second rehearsal cycle, at either the beginning or end of each of the four 400 ms items (i.e. excluding the 100 ms inter-item silences). This resulted in the following set of times: 2067, 2333, 2567, 2833, 3067, 3333, 3567, and 3833 (in ms from the offset of the last digit in the presented sequence to the onset of the probe tone). The dependent variables were the temporal variability of rehearsal (computed as the circular standard deviation of response distributions, see Results section), and digit span.

Procedure

Participants were tested individually in a quiet room in a single session with digit span data collected prior to the rehearsal-probe task.

In the digit span task, participants heard random sequences of digits presented at one item per second, starting with a list length of 3 and increasing up to 12. The 3-10 digit sequences never contained a repeated item, while the 11- and 12-digit sequences contained 1 or 2 non-adjacent item repetitions, respectively. Stimuli were the same as in the rehearsal-probe task and spoken immediate serial recall was recorded by the experimenter. There were three trials at each list length, and the task continued until the participant responded incorrectly to all three sequences of a given length or completed all three at length 12. Digit span was scored as the average list length of the last three correct trials. For example, if a participant responded correctly to all three 5-digit lists, one 6-digit list and no 7-digit lists, then the score was calculated as (6 + 5 + 5)/3 = 5.33.

In the rehearsal-probe task participants were told that, in each trial, they would hear a sequence of 4 digits repeated twice in succession (this was in order to provide rhythmic continuity between the end and beginning digits). They were told to continue silently rehearsing the list in the same order and at the same pace as during presentation until they heard a brief tone, at which point they were to respond with the digit they were currently rehearsing by pressing the corresponding number key on the keyboard. They were asked not to speak aloud or move any part of their body while rehearsing and were monitored by the experimenter throughout.

Before starting the rehearsal-probe task, the experimenter demonstrated three trials with the rehearsal phase spoken aloud. After hearing the tone, the experimenter explained, e.g., “I was saying the number 7 when I heard the tone, so I’m responding with number 7” before pressing the ‘7’ key on the computer keyboard. Participants then completed six practice trials; three with overt rehearsal followed by three with silent rehearsal. The overt rehearsal trials allowed the experimenter to verify that the participant understood the task and was able to correctly identify the digit being articulated when they heard the probe. If the participant responded incorrectly during the overt practice trials the experimenter reviewed the instructions and the trials were repeated. No feedback was given about participants’ 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 after the first rehearsal list cycle. In these cases, the experimenter clarified the instructions and the practice trials were repeated.

After practicing the rehearsal-probe task alone, it was combined with a memory preload. Each trial began with a quasi-random sequence of 2 or 6 auditory letters and ended with their serial recall via corresponding key presses on the computer keyboard. The rehearsal-probe task was sandwiched in between, with a 1-second silent interval separating the end of the letter sequence presentation and start of the rehearsal task. The letters and digits were presented at the same isochronous rate of two items per second. Figure 1 illustrates the procedure. After a digit response was made to the rehearsal probe, letter recall began with a visual cue (“Letters?”) and ended with a key press (Enter). There was no time limit for responses to the rehearsal probe or the letter recall cue. Participants were told they had to remember both the letters and the numbers accurately in each trial, and were asked not to guess.

Figure 1 here

Participants were informed about the length of the letter sequences at the start of each block of trials and completed the low load condition in the first of the alternating blocks. There were 16 trials in each block (two for each of eight probe times) and three blocks per condition, resulting in six trials for each probe time per condition. Probe times were selected randomly within blocks and blocks were separated by self-paced breaks, with participants encouraged to rest between blocks.

Results

One participant responded with the item from the same serial position across all probe times, so their data were excluded from further analysis. For the remaining 24 participants, performance on the rehearsal-probe task was only examined on trials where correct recall of the preload (letter) list exceeded a threshold. This was to ensure the validity of the memory load manipulation. The criteria were recall of both letters in the low load condition and of at least three letters in the high load condition, in both cases irrespective of order. Applying these thresholds excluded 9.0% and 12.9% of trials in the low and high memory load conditions respectively, and ensured a minimum difference of at least one stored letter between them. After removing trials with item errors in response to the rehearsal probe (0.3% and 0.9% in the low and high load conditions respectively), the proportions of responses corresponding to the item in each serial position were calculated at each probe interval. Figure 2 shows the means and standard errors for the proportions of responses for each item and probe time. As expected, the frequency of responding with each of the 4 items peaked near the times corresponding to perfect performance and declined on either side. The considerable overlap among the curves shows that the consistency of responses within probe times was not perfect, but nor were the responses random. The distributions were cyclical, most clearly for the first and last items. Thus, as probe delay increased, the frequency of first item responses rose, then declined and finally began to rise again, while over the same period the frequency of final item responses declined and then rose.

Figure 2 here

Because of the cyclical nature of the response distributions, circular standard deviations (CSDs) were used to measure their temporal variability. CSD is similar to linear standard deviation, but measures dispersion around the mean in terms of phase. To calculate phases, the duration of the rehearsal cycle was represented by the circumference of a circle, with the beginning and end of the cycle represented by the same point (0 or 2π radians). The 8 probe delays were then mapped on to the circumference, converted to phase angles and the frequencies of responses at each phase angle used to calculate the CSD for each of the 4 distributions. The calculation was performed using MATLAB (The MathWorks Inc., 2010) and the MATLAB CircStats toolbox (Berens, 2009). First, a mean resultant vector for each distribution was computed by averaging unit vectors corresponding to each data point (i.e. the angular probe time corresponding to a single response). The circular variance was then computed as one minus the resultant vector length, which ranged from 0 (no variance) to 1 (maximum dispersion). Finally, the CSD was estimated as the square root of two times the circular variance[2].

Table 1 shows the CSDs for the 1st, 2nd, 3rd and 4th items in the sequence and the mean CSD over all items as a function of memory load. Overall, CSDs were larger in the high load condition and were similar for all 4 items in the sequence with the possible exception of a slight reduction for the last item (see also Figure 2). A two-way (load x serial position) repeated-measures ANOVA on CSDs confirmed a statistically significant effect of load, $F(1,21) = 10.28$, $MSE = 0.03$, $p = .004$, no effect of serial position, $F(3,63) = 1.04$, $MSE = 0.04$, $p = .383$, and no interaction, $F(3,63) = 2.13$, $MSE = 0.02$, $p = .105$.

Table 1 here

Examination of individual differences in performing the rehearsal-probe task revealed a statistically significant negative correlation between individual mean CSDs averaged over the two load conditions and auditory digit span scores, $r = -.59$, $p = .003$. Figure 3 shows the scatterplot.

Figure 3 here

Discussion

We begin by considering the utility of the rehearsal-probe task as a tool for analysing the time-course and temporal precision of subvocal rehearsal. In broad terms it is clear that performance exhibits a number of relevant features. Thus (a) the probability of responding with an item from each of the 4 positions varied cyclically with the delay of the probe, (b) the peaks of the distributions occurred close to (but consistently slightly after) the times corresponding to perfectly timed subvocal rehearsal, (c) the distributions had broadly similar shapes, and finally (d) the width of the distributions, as measured by the CSD, was sensitive to the experimental manipulation of memory load and showed systematic individual differences. Lags in the peaks of the distributions, while not relevant to the response dispersion measure, might reflect a separable source of temporal deviation in this task. For instance, these lags could indicate a static delay at the start of the rehearsal period or in the process of detecting the probe and selecting an item for response.

As regards theoretical interpretation, the twin observations that (a) the timing of rehearsal was significantly more variable with a high memory preload and (b) participants with lower auditory digit spans tended to rehearse with less temporal precision provide converging evidence for a resource-based account of AVSTM in which shared resources limit both temporal precision and storage capacity.

Thus, as an interim summary, the pattern of performance in the rehearsal-probe task is broadly consistent with covert serial recall, and CSD appears to be a robust measure of its temporal precision and gives results consistent with the hypothesis that common resources constrain both span and temporal precision in AVSTM.

Experiment 2

Although Experiment 1 provided evidence for a correlation between the temporal precision of subvocal rehearsal and digit span, the sample size and the absence of control variables limit the strength of the conclusion drawn. We therefore conducted a second study with the aim of replicating the correlation using a larger sample and ruling out the possibility that it is mediated by general ability or non-verbal STM. To this end we included measures of intelligence and visuo-spatial STM in addition to digit span.

Intelligence is highly correlated with executive attentional control in working memory (Conway, Kane, & Engle, 2003; Engle, Tuholski, Laughlin, & Conway, 1999) and has been shown to relate to variability in motor-timing tasks (Madison, Forsman, Blom, Karabanov, & Ullén, 2009), even when motivation is manipulated (Ullén, Söderlund, Kääriä, & Madison, 2012). Because we were interested in the specificity of the relationship between auditory temporal precision and memory for serial order, it was important to control for any aspect of this association that could be attributable to more general processes. By measuring individual differences in non-verbal IQ we aimed to control for shared variance between performance on the digit span and rehearsal probe tasks that can be attributed to more general individual differences factors such as attention, motivation, executive control and motor-timing abilities.

We also considered the possibility that domain-general STM abilities might account for some of the relationship between the digit span and rehearsal probe tasks, even after controlling for differences in non-verbal IQ. To this end we used a visuo-spatial STM task as it requires modality-independent STM processes shared with the digit span and rehearsal probe tasks, such as attentional control (Cowan et al., 2011; Majerus et al., 2010, 2016), and does not overlap with the critical elements of the two latter tasks, viz. STM for serial order and auditory temporal processing. Visuo-spatial STM reflects a separate system from AVSTM (see e.g. Baddeley, 1986; Cocchini et al., 2002; Smith & Jonides, 1997) but as there is some evidence for modality-independent serial ordering processes (Hurlstone et al., 2014; Majerus et al., 2010), it was important to ensure that individual differences in serial order were not captured by the domain-general STM control measure. For this reason we assessed visuo-spatial STM using the Visual Pattern Test (VPT; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Wilson, Scott, & Power, 1987) which measures span for simultaneously presented visuo-spatial patterns. Unlike auditory-verbal digit span, and alternative tests of visuo-spatial STM such as Corsi Blocks (Corsi, 1972), the VPT involves no element of serial ordering.

In contrast to Experiment 1, there was no within-participant load manipulation, as the aim was only to replicate and extend the between-participant results. We also used a longer list of 6 rather than 4 digits in the rehearsal-probe task in an attempt to increase its sensitivity to individual differences in AVSTM capacity, while still ensuring that participants were rehearsing sub-span sequences. We planned to use regression analysis to determine the extent to which variance shared between auditory digit span and the variability of rehearsal timing can be accounted for by individual differences in IQ and visuo-spatial STM. We expected to replicate the negative correlation between auditory digit span and timing variability during rehearsal. A further prediction was that regression analysis would show that temporal variability predicts digit span scores even after controlling for any variance accounted for by non-verbal IQ and visuo-spatial STM.

Method

Participants

Participants were 40 undergraduate and postgraduate volunteers (29 female, mean age 19.8 years, range 18 to 24 years) from the University of York. The compensation, eligibility criteria, ethical approval and informed consent details were the same as those described earlier.

Design

All participants completed the rehearsal-probe, digit span and visual pattern span tasks together with a measure of nonverbal IQ.

Materials

The stimuli for the rehearsal-probe and digit span tasks were the same as in Experiment 1. Non-verbal IQ was measured using the Matrix Reasoning subset of the WASI battery (Wechsler, 1999), and visual STM was measured using the Visual Patterns Test (Della Sala et al., 1999; Wilson et al., 1987).

Procedure

Digit span was administered first, followed by the rehearsal-probe task, WASI Matrix Reasoning, and the VPT. The entire session lasted between 50 and 60 minutes. Apart from the differences described here, the procedures for administering and scoring the digit span and rehearsal-probe tasks were the same as in Experiment 1.

Unlike Experiment 1, here we administered the rehearsal-probe task on its own, without any preload (i.e. letter recall) task. Sequences for the rehearsal-probe task consisted of 6 digits selected at random from the set 0-9 without replacement. Figure 4 shows the structure of a single trial. Probe times were selected randomly on each trial from a set of 14, 12 of which were such that, if the sequence were perfectly replicated during the rehearsal delay, probes would occur either during the beginning (75 ms after onset) or end (325 ms after onset) of one of the six digits during the second rehearsal list cycle. The two additional probe times were added to extend the probe period beyond the second rehearsal cycle in order to avoid participants' anticipation of the probes. These probe times were aligned to occur at the end of the first rehearsal cycle and the beginning of the third cycle. The resulting 14 probe times were evenly spaced, occurring every 250 ms from 2825 to 6075 ms after the end of sequence presentation (see Supplementary Material). Each probe time occurred 3 times per block of 42 trials and participants completed a total of 126 trials (9 trials x 14 probe times) after demonstration and practice as described in the Experiment 1 Procedure.

Figure 4 here

The WASI Matrix Reasoning task consists of a series of 35 abstract visual matrices with one blank (missing) square. Participants were asked to indicate which of five image options below the matrix would complete the pattern if it were inserted in the location of the missing square. Verbal responses were recorded by the experimenter. The 35 items were ordered in increasing difficulty. In accordance with the test manual, participants started with the 7th matrix and worked onwards until they met the stopping criterion (either four consecutive wrong answers, or four out of five consecutive wrong answers), or reached the end of the test. The dependent measure was the average of the index numbers of the three most complex matrices that were responded to correctly (i.e. if the last three correct responses were to matrices 20, 23, and 24, then the score was 22.33). The maximum possible score was 34.

In the VPT, participants were shown patterns of filled and empty squares within grids of increasing size, and were asked to recall the pattern by filling in a blank grid after a short delay (example stimuli and more details are provided in the Supplementary Material). Participants completed three trials per grid size, starting with the smallest (grid 1, 4 squares) and continuing until all three trials at a given grid size were incorrect or all trials for the largest size (grid 14, 30 squares) were completed. The dependent measure was the average grid number of the three largest grids to which the participant responded correctly. There were 14 grid sizes in total, so the maximum possible score was 14.

Results

Rehearsal-probe task. After removing the trials where no response was given or an extra-list response was given (0.73% and 0.97% of the data, respectively), proportions of responses for items in each serial position were calculated across probe times. Figure 5 summarises performance in the rehearsal-probe task in terms of the proportions of responses for each serial position across probe times, averaged over all participants. The main features are broadly similar to those shown previously for 4-item sequences, with the exception of a stronger tendency for the response distributions to become wider and flatter over time. As before, the cyclic nature of the distributions is clearly shown, with the frequency of first item responses rising during the final item and that of final item responses declining during the first item.

Figure 5 here

A one-way ANOVA on CSDs revealed that the serial position effect was statistically significant, $F(5,195) = 14.68$, $MSE = 0.12$, $p < .001$. Pairwise comparisons with Bonferroni correction showed that the mean CSD was significantly lower for position 1 compared to positions 2-4 (all *p*s < .01) and for position 6 compared to positions 2-5 (all *p*s < .05).

Table 2 here

Individual differences analyses. Descriptive data for all measures are presented in Table 3.

Table 3 here

Relationships between the measures were initially examined using two-tailed Pearson's *r* correlations (see Table 4). These revealed a statistically significant negative correlation between mean CSD and digit span (see Figure 6) and a statistically significant albeit weaker positive correlation between VPT and digit span. Non-verbal IQ correlated negatively and non-significantly with mean CSD and positively but non-significantly with VPT.

Table 4 here

Prior to regression analysis the data were examined and found to have met the assumptions of independence of errors (Durbin-Watson = 1.83), homoscedasticity and no multicollinearity (maximum VIF = 1.16, minimum tolerance = 0.86). A multiple linear regression analysis was conducted to determine the amount of variance in digit span scores uniquely accounted for by each variable. The results are shown in Table 4.

Entry into the regression was forced, with WASI MR entered first, VPT entered second and mean CSD entered third. The rationale was that non-verbal intelligence and non-verbal STM should be controlled for in order to make a more conservative estimate of the amount of variance in verbal 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). The outcome of this analysis was very clear in showing that only CSD emerged as a statistically significant predictor of auditory digit span.

Table 5 here

Discussion

We begin by noting the robustness of the distinctive pattern of cyclical response distributions over elapsed time in the rehearsal-probe task, with those for 6 items looking remarkably similar to those for 4 items in Experiment 1. While timing precision appears to be more variable here than in Experiment 1, this is to be expected given the longer sequence length (6 digits rather than 4) and the longer probe delays. Unlike in Experiment 1, here there was a significant serial position effect. This looks to be an effect of time rather than position, given that response variability for the 6th item is noticeably greater at later compared to earlier probe times (see Figure 5). We note also the robustness of individual differences in performance in the rehearsal-probe task in that the ability to recall auditory-verbal sequences in the correct serial order, as measured by digit span, was once again strongly related to the temporal precision of rehearsing sub-span sequences. The present results went further by showing that temporal precision was a statistically significant predictor of variation in digit span scores even after controlling for individual differences in nonverbal IQ and visuo-spatial STM. Thus it seems that there is not only a close but a highly specific relationship between the ability to time item rehearsal precisely when operating below span and the capacity to recall such items in the correct order.

General Discussion

We set out to establish the rehearsal-probe task as a method for studying the time-course of subvocal rehearsal in AVSTM, and to explore the hypothesis that temporal precision in AVSTM is constrained by the resources that also limit its storage capacity. In two experiments we observed lawful patterns of data; response proportions for each item rose and fell over the rehearsal delay times, and peaked in the expected serial order. The response patterns proved to vary meaningfully both between individuals and in response to experimental manipulation of STM load. In particular, the CSDs of responses for each item across probe times provide, we argue, a useful measure of the temporal variability of sequence representations during STM maintenance. Overall, these experiments demonstrate the utility of the rehearsal-probe task for investigating the temporal precision of auditory-verbal sequence representations without the use of overt motor execution to mark ongoing event timing. The responses from this task form a rich source of information about rehearsal timing, including not only precision but other characteristics such as the overall rate and static offset (i.e. phase shift). In future research, the rehearsal-probe method may prove to be a useful tool for addressing questions about the temporal properties of subvocal rehearsal, for example in different populations, in relation to individual differences measures, or in response to changes in sequence stimuli and timing.

In Experiment 1 a preload manipulation was used to increase the load on domain-specific AVSTM resources while keeping all aspects of the rehearsal task (number of items, sequence timing, probe timing) constant. We found that increasing the storage load in AVSTM resulted in a decrease in the temporal precision of rehearsing a sequence of only 4 digits. It appears that timing precision in AVSTM is not static within individuals, but varies in response to a change in load on the resources that determine the number of items that can be stored in the correct serial order. Our results are broadly consistent with the previous findings that increased memory load leads to poorer precision in recall for auditory temporal intervals (Teki & Griffiths, 2014) and for the pitch of tones (Kumar et al., 2013). These results have been used to support a dynamic resource allocation account of non-verbal auditory memory (Joseph et al., 2016).

We also found that individual differences in the temporal variability of rehearsing sub-span digit sequences were negatively correlated with the storage capacity of AVSTM as measured by auditory digit span. This correlation was found in two experiments using different list lengths, suggesting that it reflects a reliable, general relationship. It also seems to be specific to AVSTM as it persists when differences in nonverbal IQ and visuo-spatial STM are controlled for. The correlation is important given that digit span does not demand precise encoding and maintenance of temporal information beyond the coarse-grained relative timing of digits necessary to distinguish their order. A trivial explanation would be to argue that CSD simply reflects the tendency to make order errors rather than temporal variability of the sequence representation. However, given that the digit sequences were short (span or sub-span length) and presented twice over, any errors during recall would have been infrequent. Also, if the differences in response variability could be explained by order errors alone then there should be ceiling effects (i.e. perfect or near-perfect timing) for high-span participants, but this was not the case. It therefore seems unlikely that span- and load-related differences in the rehearsal-probe task response distributions merely reflect order errors during rehearsal.

We interpret our data on individual differences and memory load as converging evidence for the view that a common pool of resources limits the precision with which temporal information is represented in AVSTM and the number of items that can be stored in the correct order. Our working hypothesis is that, in contrast with the fixed time-based storage capacity in the early models of the phonological loop (Baddeley, Thomson, & Buchanan, 1975), AVSTM is a more flexible system in which resources are dynamically allocated to temporal precision with implications for amount stored. We note that this interpretation is analogous to accounts of visuo-spatial STM capacity in terms of the assignment of limited resources to the precision of representing dimensions such as location, colour and orientation (Ma et al., 2014), raising the interesting possibility that general principles may apply across modalities. The resources involved in AVSTM would nevertheless seem to be quite separate, given our evidence that they are independent of individual differences in visuo-spatial pattern span.

We note however other possible ways of interpreting the relationship between temporal processing and AVSTM load/capacity reported here. Rather than there being a dynamic trade-off in resource allocation, serial order storage capacity may directly constrain temporal precision or vice versa. For instance, increased load on AVSTM capacity may impact the speed with which items can be selected and retrieved, which in turn could impact the ability to reproduce a sequence of items at a consistent pace[3]. Indeed, in isolation, our data

(Experiment 1) shows that manipulating memory load affects the precision of timing. However, consideration of earlier evidence argues against a simple unidirectional relationship, in that manipulating the timing of items affects memory capacity where the load (i.e., the total number and duration of items) is held constant (Hartley et al., 2016; Ryan, 1969a). Thus at present it seems that the most parsimonious explanation of these reciprocal effects is that they depend on common resources. Further research will be needed to exclude more complex explanations and to further characterise both shared and dissociable mechanisms of timing and serial order in AVSTM.

The present evidence of a relationship between temporal precision and overall storage capacity in AVSTM is relevant to the possibility of a unifying theoretical account of correlations between AVSTM and temporal information processing in other tasks (Grube et al., 2013; Saito, 2001; Tierney & Kraus, 2013) and its sensitivity to temporal variables such as the rhythmic structure of a sequence (Frankish, 1985; Hitch et al., 1996; Ryan, 1969a, 1969b). One way this might be achieved is through shared mechanisms whereby serial order information is derived from temporal information, as in computational models in which order is represented by the activation of internal oscillators tuned to different frequencies in the input (Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016). Other competitive queuing models explain serial order recall using purely positional codes or gradients (e.g. Henson, 1998; Page & Norris, 1998; but see also e.g. Hughes, Chamberland, Tremblay, & Jones, 2016; Macken, Taylor, & Jones, 2015), where changes in the context signal are driven by the presentation of a new item, irrespective of the (absolute or relative) timing between items. Thus in models with positional coding systems, no information about sequence timing is retained beyond that which discriminates order. Without a temporally-sensitive mechanism for the representation of serial order, it is not clear how models in this latter category could explain the reciprocal patterns of effects among sequence timing, serial order STM and load/capacity (Hartley et al., 2016; Hitch et al., 1996).

More generally, our findings may shed light on the relationship between auditory-temporal processing and language development. In particular, some researchers have argued that the precision of auditory temporal processing across multiple timescales is critical for the development of phonological skills (Goswami, 2011; Tierney & Kraus, 2014). Indeed, phonological deficits have been found to co-occur with auditory timing imprecision (e.g. Flaugnacco et al., 2014; Grube et al., 2013; Thomson & Goswami, 2008; Wolff, 2002) and reduced AVSTM capacity for serial order (e.g. Brady, Shankweiler, & Mann, 1983; Martinez Perez, Majerus, Mahot, et al., 2012; Martinez Perez et al., 2013; Rapala & Brady, 1990). While this issue remains controversial due to conflicting findings regarding the prevalence of temporal processing deficits in dyslexia (Papadopoulos, Georgiou, & Parrila, 2012; Protopapas, 2014) and alternative explanations for the co-occurrence of phonological and AVSTM impairments (e.g. Marshall, Snowling, & Bailey, 2001; Ramus & Szenkovits, 2008), our findings suggest some scope for reconciliation. In providing empirical evidence for a relationship between timing precision and serial order in AVSTM, our results are consistent with the view that developmental links between auditory temporal processing, serial order STM and phonological skills reflect the coupling of auditory timing and serial order in phonological learning. An important task for future work will be to determine whether these results extend to temporal precision for non-isochronous and unpredictably-timed spoken word sequences, which more closely resemble the temporal characteristics of natural speech.

In conclusion, we have shown that the rehearsal-probe task provides a robust and sensitive method for sampling the time-course of maintaining items in AVSTM via paced subvocal rehearsal. Using this task to study the rehearsal of short sequences, we have found that the availability of items varies cyclically over time, with overlapping distributions that preserve serial order. We have demonstrated that the CSD of these distributions provides a measure of the temporal precision of maintenance. Two experiments have shown further that the CSD increases as a function of the storage load on AVSTM and correlates negatively and selectively with individual differences in AVSTM span. We conclude the time-course of subvocal rehearsal can be measured reliably using the probe technique, and interpret the results as converging evidence that a common pool of resources limits both the temporal precision of representations in AVSTM and its overall storage capacity. We suggest that these resources may be key to further understanding the role of temporal information in the encoding of serial order in AVSTM and, more generally, the part played by temporal and serial representation in speech and language.

[Words: 7,771]

References

Aaronson, D. (1968). Temporal Course of Perception in an Immediate Recall Task. *Journal of Experimental Psychology*, *76*(1), 129–140.

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.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

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. 10.1016/S0022-5371(75)80045-4

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behavioural Brain Research*, *136*(1), 31–49.

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10), 1–21.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*(1), 80–97.

Boersma, P., & Weenink, D. (1992). Praat: Doing Phonetics By Computer.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, *35*(2), 345–367.

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. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. 10.1037//0033-295X.106.3.551

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.

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. 10.3758/BF03194326

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. 10.1016/j.tics.2003.10.005

Corkin, S. (1974). Serial-Ordering Deficits in Inferior Readers. *Neuropsychologia*, *12*(3), 347–354. 10.1016/0028-3932(74)90050-5

Corsi, P. M. (1972). Human memory and the medial temoral region of the brain. *Dissertation Abstracts International*, *34*(819B).

Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., & Christ, S. E. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*, *23*(10), 2852–63. 10.1162/jocn.2011.21625

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.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

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. 10.1037/0278-7393.34.1.128

Farrell, S., & McLaughlin, K. (2007). Short-term recognition memory for serial order and timing. *Memory & Cognition*, *35*(7), 1724–1734. 10.3758/BF03193505

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*. 10.3389/fnhum.2014.00392

Frankish, C. R. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*(2), 200–209. 10.1016/0749-596X(85)90024-5

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. 10.1016/j.tics.2010.10.001

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. 10.1080/17588928.2013.825236

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. 10.1111/j.1467-8624.1994.tb00820.x

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. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135–178. 10.1016/j.cogpsych.2016.05.001

Henson, R. N. (1998). Short-term memory for serial order: the Start-End Model. *Cognitive Psychology*, *36*(2), 73–137. 10.1006/cogp.1998.0685

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 A, *49*(1), 116–139.

Hughes, R. W., Chamberland, C., Tremblay, S., & Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory and Language*, *90*, 126–146. 10.1016/j.jml.2016.04.006

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), 339–73. 10.1037/a0034221

Joseph, S., Teki, S., Kumar, S., Husain, M., & Griffiths, T. D. (2016). Resource allocation models of auditory working memory. *Brain Research*, *1640*, 183–192. 10.1016/j.brainres.2016.01.044

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. 10.1037/0096-3445.133.2.189

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. 10.1080/17588928.2012.716416

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.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–81. 10.1038/36846

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

Macken, B., Taylor, J., & Jones, D. M. (2015). Limitless capacity: a dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, *6*, 293. 10.3389/fpsyg.2015.00293

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. 10.1016/j.intell.2008.07.006

Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. *Cerebral Cortex*, *26*(1), 166–179. 10.1093/cercor/bhu189

Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., … Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience*, *22*(11), 2570–93. 10.1162/jocn.2009.21378

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. 10.1080/09541440500446476

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid Auditory Processing and

Phonological Ability in Normal Readers and Readers With Dyslexia. *Journal of Speech, Language, and Hearing Research*, *44*(4), 925–940. 10.1044/1092-4388(2001/073)

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. 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. 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. 10.1016/j.ridd.2013.04.005

Mazzoni, D., & Dannenburg, R. (2000). Audacity.

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. 10.3758/BF03197047

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. 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. 10.3758/MC.38.8.1147

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781.

Papadopoulos, T. C., Georgiou, G. K., & Parrila, R. K. (2012). Low-level deficits in beat perception: neither necessary nor sufficient for explaining developmental dyslexia in a consistent orthography. *Research in Developmental Disabilities*, *33*(6), 1841–56. 10.1016/j.ridd.2012.04.009

Protopapas, A. (2014). From temporal processing to developmental language disorders: mind the gap. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1634), 20130090. 10.1098/rstb.2013.0090

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, *61*(1), 129–41. 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. 10.1007/BF00383371

Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*(2), 137–147. 10.1080/14640746908400206

Ryan, J. (1969b). Temporal Grouping, Rehearsal and Short-Term Memory. *The Quarterly Journal of Experimental Psychology*, *21*(2), 148–155. 10.1080/14640746908400207

Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory*, *9*(4), 313–322. 10.1080/09658210143000164

Service, E., Maury, S., & Luotoniemi, E. (2007). Individual differences in phonological learning and verbal STM span. *Memory & Cognition*, *35*(5), 1122–1135.

10.3758/BF03193483

Smith, E. E., & Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42. 10.1006/cogp.1997.0658

Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Frontiers in Psychology*, *5*, 1329. 10.3389/fpsyg.2014.01329

The MathWorks Inc. (2010). MATLAB - The Language of Technical Computing. Natick, Massachusetts: The MathWorks, Inc.

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. 10.1016/j.jphysparis.2008.03.007

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language*, *124*(3), 225–231. 10.1016/j.bandl.2012.12.014

Tierney, A. T., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, *8*, 949. 10.3389/fnhum.2014.00949

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. 10.1016/j.intell.2012.01.012

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–35. 10:1167/4.12.11

Wilson, J. T. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, *5*(3), 249–255. 10.1111/j.2044-835X.1987.tb01060.x

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing*, *15*(1-2), 179–206. 10.1023/A:1013880723925

Wolff, P. H., Michel, G. F., & Ovrut, M. (1990). The Timing of Syllable Repetitions in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research*, *33*(2), 281–289. 10.1044/jshr.3302.281

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–. 10.1073/pnas.1406219111

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–5. 10.1038/nature06860

Table 1. Mean CSDs (SEM in parentheses) for the 4 serial positions and the two memory load conditions in Experiment 1; 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.

	Serial Position				
Load condition	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)	

Table 2. Mean CSDs (SEM in parentheses) for the 6 serial positions in Experiment 2. The maximum possible CSD value is 1.41, and the minimum possible CSD is 0.26 in this experiment.

Serial Position						
1	2	3	4	5	6	Mean
0.61 (0.03)	0.72 (0.04)	0.73 (0.04)	0.73 (0.03)	0.66 (0.03)	0.58 (0.03)	0.67 (0.03)

Table 3. Means, standard deviations, and range of possible scores on all measures in Experiment 2.

Measure	Mean	SD	Range of possible scores
Mean circular standard deviation	0.67	0.19	0.26 - 1.41
Digit span average	6.94	1.06	5.00 - 12.00†
WASI Matrix Reasoning average	29.65	2.84	0.00 - 34.00
Visual Patterns Test average	10.06	1.72	0.00 - 14.00

Note: WASI = Weschler Abbreviated Scale of Intelligence; 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 at list lengths of 4, 5 and 6 (and no correct responses for 7-digit lists) results in an average digit span score of 5.

Table 4. Pearson's *r* correlations between measures in Experiment 2.

Measure	Mean CSD	Digit span	VPT
Mean CSD	1.00		
Digit span	-.59***	1.00	
VPT	-.13	.32*	1.00
WASI MR	-.26	.07	.30

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

Table 5. Results of multiple linear regression analysis of factors predicting the variance in digit span scores in Experiment 2.

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
WASI MR	.03	.06	.07	-.01	.06	.02	-.06	.05	-.17
VPT				.20	.10	.32*	.18	.08	.29*
Mean CSD							-3.37	.74	-.59***
R		.07			.32			.65	
R^2		<.01			.10			.43	
F for ΔR^2		.20			3.93			20.55***	

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

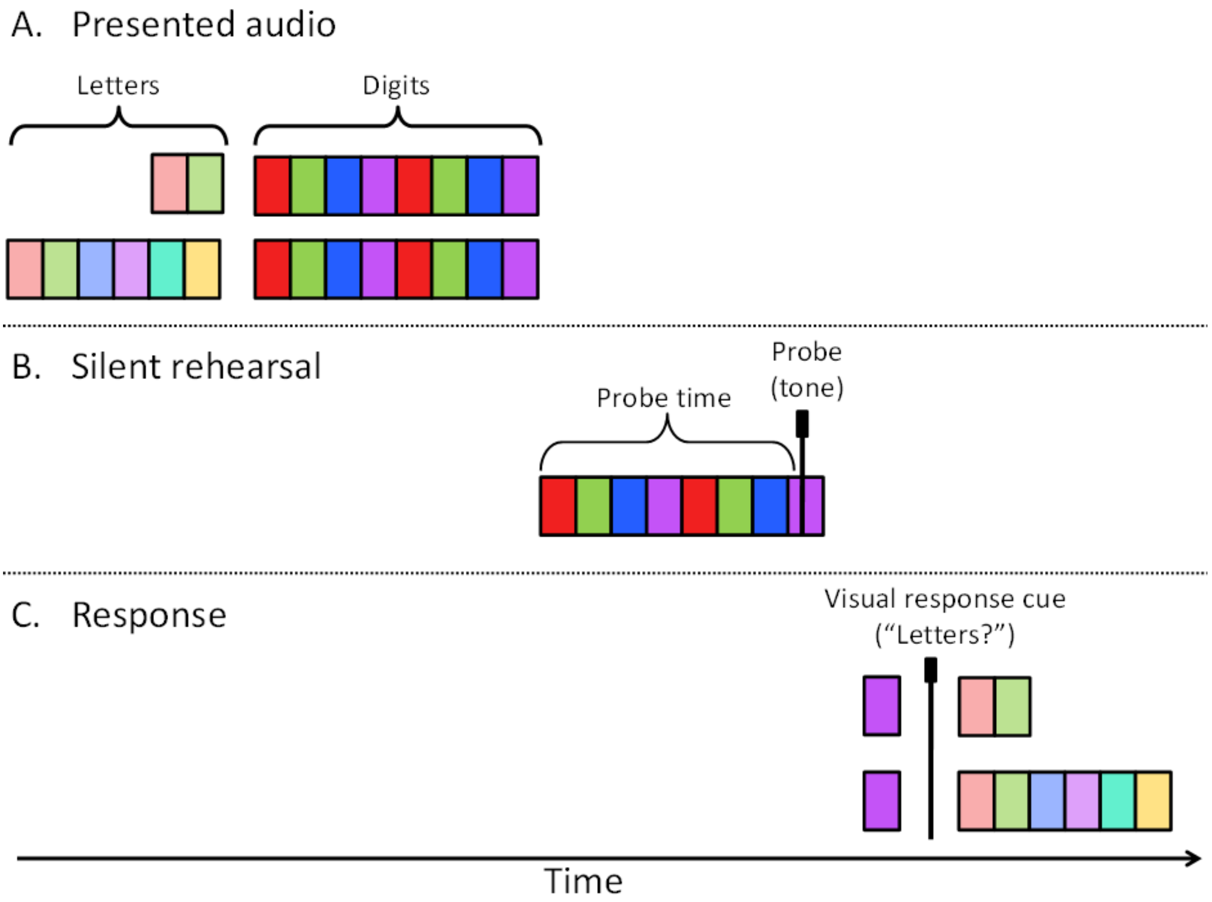

Figure 1. Representation of a single trial in the low preload (2-letter lists, top row in A and C panels) and high preload (6-letter lists, bottom row in A and C panels) conditions in Experiment 1. Boxes represent the 500 ms items (400 ms sounds and 100 ms inter-stimulus intervals, adjusted to align item-specific p-centres). Colours denote the serial positions within the letter and digit lists. 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. After the rehearsal-probe portion of the trial is finished, a visual cue appears to signal the start of the response period for the 2- or 6-letter preload list.

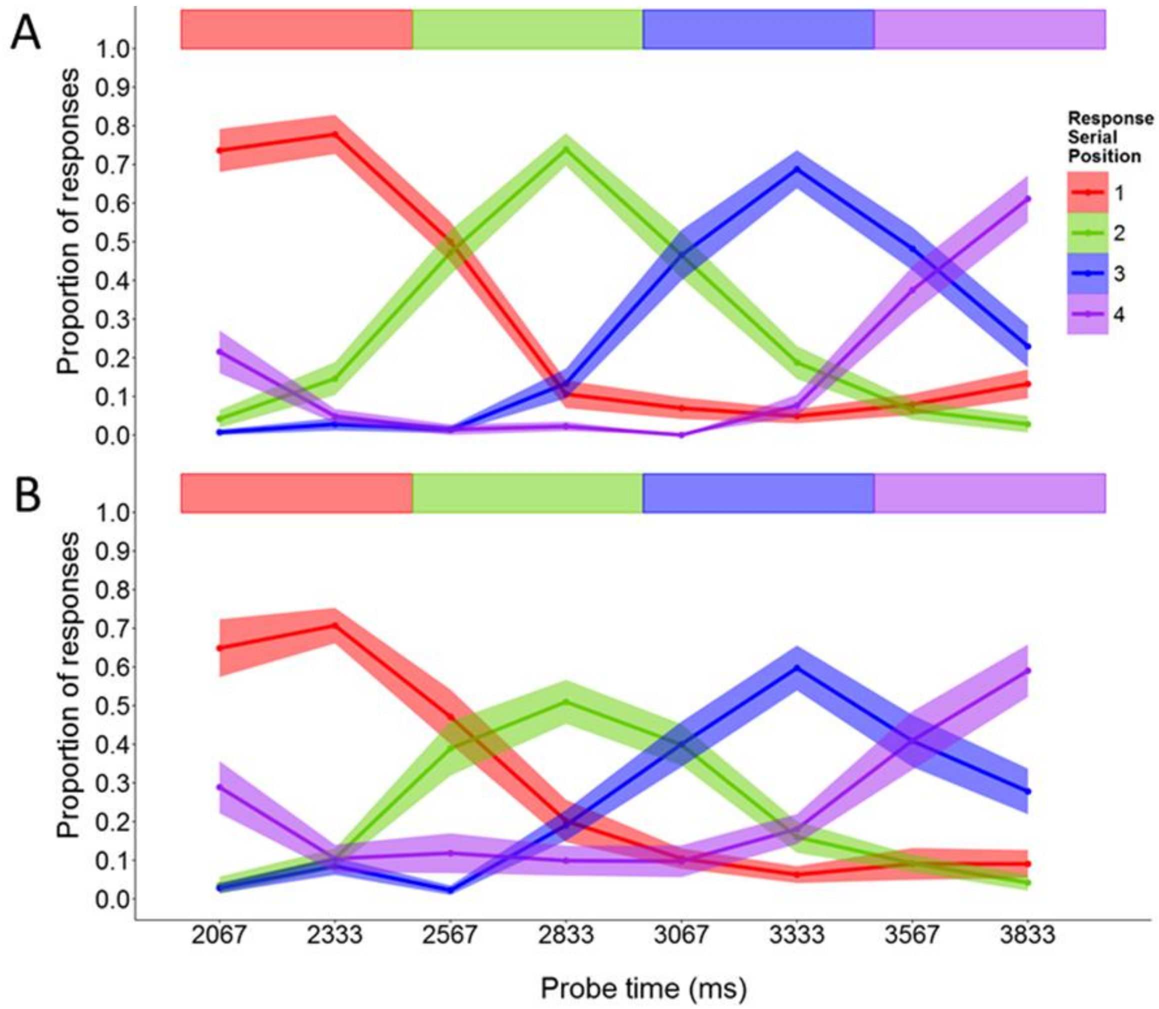

Figure 2. Mean proportions of responses for each serial position across probe times in the low (2-letter preload; A) and high (6-letter preload; B) load conditions from Experiment 1. The widths of the ribbons at each probe time reflect the SEMs. Rectangles above the plot show the item timing if the sequence were rehearsed exactly as presented.

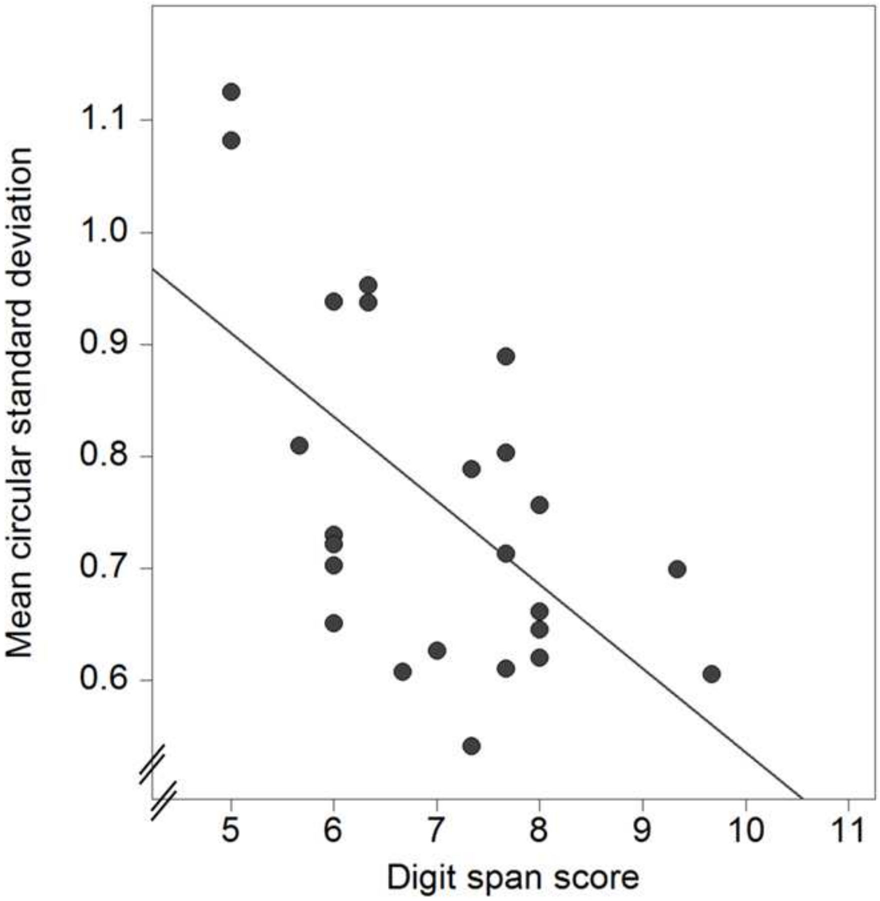

Figure 3. Scatterplot of the relationship between digit span scores and mean CSDs, averaged over the two load conditions from Experiment 1. The solid line shows the linear regression model for mean CSD predicted by digit span. The y-axis minimum is the mean CSD value that would be obtained from perfectly consistent responses in this task (0.52). The maximum possible mean CSD value is 1.41.

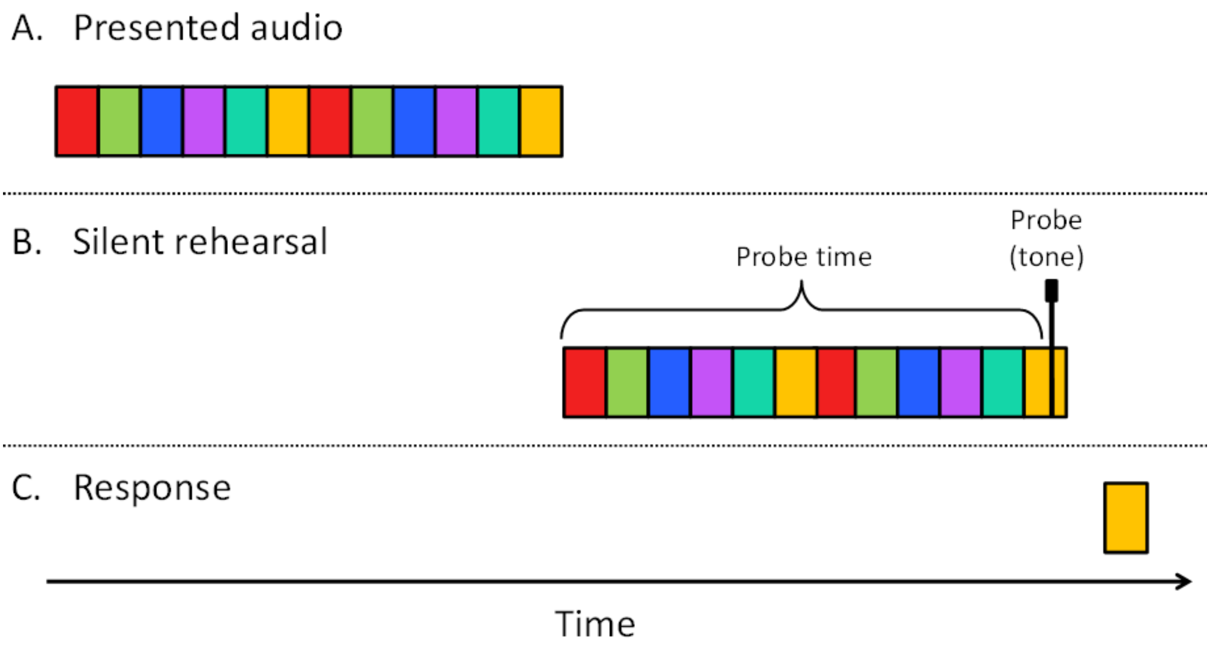

Figure 4. Representation of a single trial in the rehearsal-probe task in Experiment 2. The boxes represent the 400 ms sequence items and 100 ms inter-item silence (with item-specific adjustments to align p-centres). The colours of the boxes denote the serial positions of the list items. Each trial begins with the perceptually isochronous presentation of a 6-digit sequence, repeated twice (A). Immediately after the end of the second sequence presentation, the participant begins rehearsing the list exactly as it was presented, until s/he hears a tone (B). Upon hearing the tone, the participant stops rehearsing and responds with the digit that s/he was currently rehearsing at the same time that the tone was presented by pressing the corresponding digit on the keyboard (C). After a response is made, the trial ends and there is a short inter-trial interval followed by fixation cross signalling the start of the next trial.

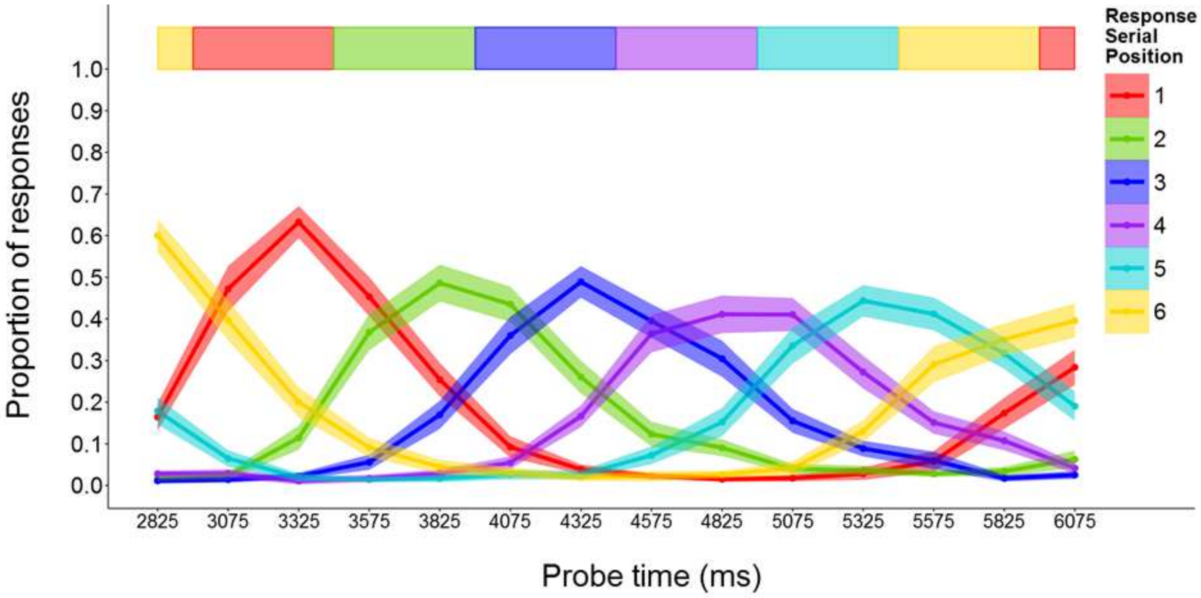

Figure 5. Response proportions for the six serial positions across probe times in Experiment 2. The response proportions are averaged over all participants, and the widths of the ribbons at each probe time reflect the SEMs. Rectangles above the plot show the item timing if the sequence were rehearsed exactly as presented.

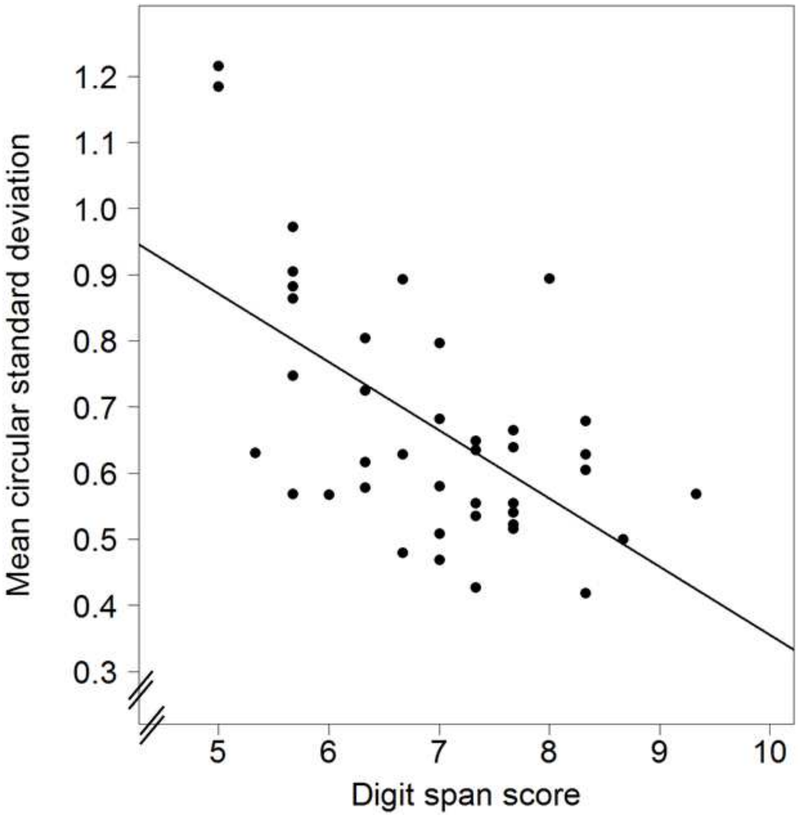

Figure 6. Scatterplot showing the negative correlation between digit span scores and temporal variability (mean CSD) for 6-digit rehearsal lists from Experiment 2. The line shows the simple linear regression for mean CSDs as predicted by digit span scores. The y-axis minimum is the mean CSD value that would be obtained from perfectly consistent responses in this task (0.26). The maximum possible mean CSD is 1.41.

[1] We refer to this task as the rehearsal-probe task because of the similarities with rehearsal in STM, namely the repetition of a memory sequence via inner speech. However, we are not suggesting that the subvocal repetition in this task is equivalent to natural rehearsal as it is used to recall sequences over short intervals, where there are no constraints on the timing or order of item repetition.

[2] This equation produces a range of possible CSD values from 0 to the square root of 2. However, in the rehearsal-probe task, the minimum possible CSD is a function of the number of probes per item and the number of items in the sequence, and is greater than zero. This minimum value is stated in the Results section of each experiment.

[3] We would like to thank Steve Majerus for suggesting this interpretation.