Published in final edited form as:

J Neurolinguistics. 2018 November ; 48: 176–189. doi:10.1016/j.jneuroling.2018.04.014.

Short-term memory span in aphasia: Insights from speech-timing measures

Christos Salis¹, Nadine Martin², Sarah V. Meehan², and Kevin McCaffery²

¹Newcastle University, Speech & Language Sciences, King George VI building, Queen Victoria Road, Newcastle upon Tyne, NE1 7RU, UK, +44 191 208 8875, christos.salis@ncl.ac.uk

²Temple University, Department of Communication Sciences & Disorders, Ritter Annex, 1301 Cecil B Moore Avenue, Philadelphia, PA 19122, USA, +1 215 204 1870

Abstract

Auditory-verbal short-term memory impairments are part and parcel of aphasia and interfere with linguistic processing. To date, the science about short-term memory impairments in aphasia has been generated and dominated by studying measures of accuracy, that is, span length. Because accuracy is expressed through speech, examining the speech-timing characteristics of persons with aphasia as they engage in spoken recall could reveal insights about the manner in which accuracy is achieved.

Six speech-timing measures (e.g., response durations, pause durations) were elicited from the speech waveform of word span tasks from twelve people with aphasia. Speech-timing measures were compared to neuro-typical control participants. Speech-timing performance between erroneous and correct responses in the aphasia group was also examined. Across all measures, people with aphasia produced considerably longer speech-timing patterns in comparison to control participants. Memory load affected some measures in people with aphasia and control participants. Speech-timing in correct response trials was shorter than responses in erroneous trials. Memory span correlated only with one measure, namely, speech time (defined as the sum of each individual word duration in a response). Speech time also correlated with the following measures: Aphasia severity (Aphasia Quotient of the Western Aphasia Battery), spontaneous speech, and language comprehension (also measured by the Western Aphasia Battery).

Some protracted speech-timing patterns in the aphasia group may be explained by a deregulation of activation-decay patterns. However, in the absence of further evidence from people with aphasia, possible issues around the sensitivity of some speech-timing measures limit firmer conclusions. Speech-timing measures are response-time measures, which have not been systematically studied in studies of short-term or working memory in aphasia and as such, can push the current boundaries of knowledge of short-term and working memory impairments in aphasia, not only in stroke related aphasia but also other neurological conditions.

Correspondence to: Christos Salis.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Short-term memory; aphasia; speech-timing

1. Introduction

The links between the integrity of verbal short-term and working memory (STM and WM respectively) and language abilities in aphasia have been attracting considerable interest in both theoretical (Howard & Nickels, 2005; Martin & Allen, 2008; Martin & Ayala, 2004; Martin & Reilly, 2012; Murray, 2012; Wright & Fergadiotis, 2012) and clinical aphasiology (DeDe, Ricca, Knilans, & Trubl, 2014; Minkina, Rosenberg, Kalinyak-Fliszar, & Martin, 2017; Paek & Murray, 2015; Salis, Hwang, Howard, & Lallini, 2017; Zakariás, Keresztes, Marton, & Wartenburger, 2018). STM and WM impairments co-occur with aphasia and interfere with core linguistic processing at multiple levels, from phonology and word to sentence and discourse (Martin, 2009; Sung, McNeil, Pratt, Dickey, Hula, Szuminsky & Doyle, 2009). Various theoretical positions about the precise nature of STM and WM impairments and their links to aphasia have been proposed to account for the diverse associative or dissociative patterns of STM and WM functioning (e.g., Majerus, Attout, Artielle, & van der Kaa, 2015; Martin & Allen, 2008; Verhaegen, Piertot, & Poncelet, 2013). For example, Majerus et al. (2015) reported that people with aphasia may present with a range of deficits, affecting particular sub-processes of STM functioning: Selective item deficits, selective order deficits, generalized item deficits, and serial order deficits. Another theoretical position distinguishes phonological vs. semantic STM deficits (Martin & Ayala, 2004; Martin & Allen, 2008).

As a construct, STM pertains to the temporary storage of a limited amount of information in a relatively unprocessed state (Cowan, 2008). Unlike the related construct of WM, STM does not involve manipulation of information, which is the distinguishing feature of WM. In contrast to STM, WM draws heavily upon attention related processes (e.g., updating, inhibiting) responsible for the temporary storage of information while it is being manipulated (Cowan, 2008). Another distinguishing feature in the two constructs is how they are measured. STM measures are considered simple (e.g., serial recall of words), whereas WM measures are considered complex (e.g., alphabet span in which a series of presented words must be rearranged to recall in alphabetical order) (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002). A recent review of standardized STM/WM measures in aphasia research since 1980 has shown that both simple and complex measures are increasing in popularity (Murray, Salis, Martin, & Dralle, 2018).

The present study investigates several measures of STM performance elicited from word span, a simple measure of STM. The method we used involves the temporal analysis of the speech of persons with aphasia and neuro-typical control participants as they recall lists of words serially. Because STM by its very nature is time-limited, speech-timing measures afford precision that more typical measures of accuracy do not. In the remainder of the introduction, we draw from the wider, non-aphasia literature on speech-timing measures of STM and WM as our framework. The consensus of that literature is that speech-timing

measures reflect covert processes in STM architecture. We then discuss the relevance of speech-timing measures in terms of levels of activation-decay of STM representations and processing speed.

1.1 Measures of accuracy and speech timing in STM

Measures (standardized or not) are quantified records, which are gathered from testing procedures and enable researchers to describe, quantify and consequently understand constructs, that is, real phenomena of theoretical and clinical interest (Edwards & Bagozzi, 2000). Consequently, measures form the basis of understanding a phenomenon. In STM and WM research in aphasia, accuracy is the prototypical measure that has been used for over a century (Eling, 2015). In STM, accuracy is often defined as the maximum number of words a person can recall correctly from a list, usually in serial manner. This measure of accuracy is often referred to as memory span and is usually demonstrated through speech production. Speech, therefore, is the physical medium through which the abstract, verbal representations held in STM are realised. To date, the science about STM impairments in aphasia has been generated and dominated by studying measures of accuracy (i.e., span size) elicited from spoken production with some evaluation of variation in accuracy, either as a function of items that are being recalled (e.g., number of syllables, frequency, concreteness), or accuracy of recall at various serial positions of a list. Yet, the temporal characteristics of speech output have seldom been scrutinised.

Nevertheless, in literature domains beyond aphasia, researchers have broadened the knowledge of STM (and to a lesser extent WM) by examining directly the temporal characteristics of the speech output as it is produced in real time, when participants recall words. Cowan (1992), who carried out a comprehensive study of speech-timing measures in developmental STM research, provides a list of measures and how they are elicited from a speech sound waveform (see Figure 1). The most popular measures that have attracted repeated attention across literature domains are word duration, inter-word pauses (pauses hereafter), and preparation time. The literature domains and studies that used speech-timing measures are as follows: *Developmental psychology* (Cowan, 1992, 1999; Cowan, Keller, Hulme, Roodenrys, McDougall, & Rack, 1994; Cowan, Wood, Keller, Nugent, & Keller, 1998; Cowan, et al., 2003), *mainstream cognitive psychology* (Haberlandt, Lawrence, Krohn, Bower, & Thomas, 2005; Hulme, Newtown, Cowan, Stuart, & Brown, 1999; Tehan & Lalor, 2000; Towse, Cowan, Hitch, & Horton, 2008), and *aging research* (Kynette, Kemper, Norman, & Cheung, 1990). We examine some of these studies in more detail in the next two paragraphs to highlight: (a) Selected reported correlations between accuracy and speech-timing measures, as well as other key findings; and (b) theoretical explanations of memory performance that have been based on speech-timing measures. We should note that there is some variation in the terminology of speech-timing measures among authors. In the present paper, we adhere to the terminology used by Cowan (1992).

Some, though not all, speech-timing measures have been shown to correlate with accuracy of span in STM tasks but this varies across studies. Examining four to five year old children, Cowan (1992) reported statistically significant positive correlations between accuracy in word span and response time ($r = .59$), pronunciation time ($r = .60$), and speech time ($r = .$

82). Greater accuracy was associated with longer durations in these speech-timing measures. None of the other measures depicted in Figure 1 correlated with accuracy. Tehan and Lalor (2000) examined relationships of some timing measures with digit span accuracy in university students and found that only speech time correlated significantly, albeit weakly, with digit span ($r = .36$). Pause or preparation time did not correlate with digit span.

Much of the early research on speech-timing measures in STM and WM focused on the study of pauses and what they reveal about covert processes of STM and WM, especially as memory load increases (i.e., the number of items a person is asked to recall). Chiefly due to the number and length of pauses in the longest of list lengths (i.e., four words), Cowan (1992) argued that the decay rate of information in STM does not decrease steadily, from the first word in the list onwards. Instead, there is fluctuation of decay and activation levels (or reactivation) of the word items as the person engages in recall. Hulme et al. (1999) argued that during pauses two covert processes take place: (a) Identification of a word to be recalled for a given position in the list (that is, item and order information); and (b) restoration (or redintegration) of the decayed trace before being used as a response. Towse et al. (2008) proposed a similar hypothesis for WM, the recall reconstruction hypothesis, positing that longer pauses stem from longer memory searches. In the preparation time measure (shown in Figure 1), three processes may be taking place, namely, rehearsal of the list before recall, memory search, and motor speech programming. Interpreting variation in word length, Haberlandt et al. (2005) argued that the similar patterns of spoken durations and pauses may indicate that search processes coincide with articulatory planning processes. In the present study, we examine the value of some of these measures in determining the nature of verbal STM impairment in aphasia.

In some studies, memory load was a factor that modified some speech-timing measures but not others. For example, Cowan et al. (1994, 2003) found that memory load affected pauses to a greater extent than word durations. In other studies with university students, memory load was evident in word and pause durations, with words and pauses in list-initial and list-final positions having shorter lengths than words and pauses in list-medial positions, thus demonstrating inverse recall curve shapes (Haberlandt et al., 2005). This pattern has also been observed in pauses (Haberlandt et al., 2005).

1.2 The present study

Speech-timing measures belong to the wider category of response-timing measures that log time between stimulus and response. Table 1 summarizes aphasia studies of STM and WM that reported data from response-timing measures, including two studies of speech-timing measures (Kinsbourne, 1972; Martin, Saffran, & Dell, 1996). With the exception of Mayer and Murray (2012), most other studies shown in Table 1 used recognition memory paradigms, based on variants of Sternberg's task (Sternberg, 1966, 1969). In this task, trials comprising a list of words or pictures are presented sequentially, one item at a time. The list is followed by a probe word (or two words as in the case of Attout, van der Kaa, George, & Majerus, 2012) that may have either occurred in the trial (target), or not (distractor). The person needs to decide (i.e., recognize) if the probe word was present among the list of words that was presented. Both accuracy and response-time are logged.

The response-time data available in the studies that used Sternberg's task show mixed findings between persons with aphasia and neuro-typical controls. For example, the person reported by Hamilton and Martin (2005, 2007) was slower as well as less accurate than neuro-typical controls. However, in Attout and colleagues (2012) only one of the two persons was slower than neuro-typical controls in one of the two STM tasks. To our knowledge, only two studies reporting a total of three persons with aphasia studied speech-timing measures to investigate STM to date (Kinsbourne, 1972; Martin et al., 1996). Both studies elicited the same speech-timing measure, which comprised the duration of the last stimulus word in addition to the response time (see Figure 1). Data from control participants were not reported in either study.

From a theoretical stance, investigating speech-timing measures in STM tasks in persons with aphasia is important for two reasons. The first reason has to do with the concept of maintenance of activation and rates of decay of information held in STM (Cowan, 1992). The second reason relates to the construct of processing speed and memory (Cowan & Kail, 1996; Kail & Salthouse, 1994). We sketch these issues next and revisit them in the discussion. Each of these constructs represents a potential processing impairment underlying the language and verbal STM impairment in aphasia.

1.2.1 Activation-decay patterns—Activation-decay limitation is one mechanism that gives rise to STM impairment in aphasia and that has been related to speech-timing patterns in aphasia. Martin and colleagues (Dell, Martin, & Schwartz, 1997; Martin, 2009, 2012; Schwartz, Dell, Martin, & Saffran, 1994) posit that the nature of STM difficulties in aphasia stems from an inability to consistently maintain activation of phonological, semantic and conceptual information of words that people with aphasia are asked to recall. Recall of words earlier in a list is supported by greater semantic activation that accumulates via feedforward and feedback activation over the time course of word retrieval, compared to words at the end of a list, which are supported more by recent phonological activation relative to semantic activation. This differential distribution of semantic and phonological support contributes to the primacy and recency effects observed in verbal recall in neuro-typical speakers. The evidence supporting the activation-decay view comes from connectionist modelling studies (e.g., Dell et al., 1997) as well as behavioral studies examining correlations between language measures and STM span (e.g., Martin & Ayala, 2004), serial position effects (Martin & Saffran, 1997), and verbal learning (e.g., Martin & Saffran, 1999). As shown in Table 1, Martin et al. (1996) utilised only one speech-timing measure that relates to but is different from preparation time (as defined in this study) to elucidate the role of activation-decay. Their participant was tested on two-word lists. Preparation time for words in the second position in the list was found to be greater (i.e., slower) than that for words in the first position in the list. Kinsbourne (1972) reported a similar pattern in the two patients he studied. In both studies, the maximum number of words that participants were tested on was two. In both studies, preparation time was slower for two words than one word. This pattern could be interpreted as evidence for slowed activation or reduction in processing speed as memory load increases. These limited findings suggest that speech-timing is sensitive to memory load.

The pattern of activation in Martin's account, suggests less accumulation of semantic support of the final items in the list compared to earlier items. Using a broader range of speech-timing measures in typically developing children, Cowan (1992) found that activation patterns in STM recall do not follow a steady, monotonic pattern. Cowan's results suggested that decay during spoken recall is partly counteracted by reactivation of items and this is evident in fluctuations in pause durations. Martin and Saffran (1997) and Martin and Ayala (2004), among other authors (e.g., Warrington & Shallice, 1969; Wilshire, Keall, & O'Donnell, 2010), reported differential patterns of performance accuracy in some persons with aphasia. Some persons show primacy effects (i.e., better recall for items presented at the beginning of a list), while others show recency effects (i.e., better recall for items presented at the end of a list). The only thorough analysis of speech-timing data that demonstrated clear primacy-recency patterns as evinced by speech-timing measures are the inverse curve patterns reported by Haberlandt et al. (2005), with faster response times at the beginning and end of lists. The question that arises in relation to aphasia is whether speech-timing measures are sensitive to memory load and what they reveal about the architecture of STM in terms of activation-decay patterns. We address these issues in the present study.

1.2.2 Processing speed, speech and STM—The concept of processing speed as modifying STM ability has existed for some time in developmental psychology and aging research, among other areas, in an attempt to understand development of STM skills in children (e.g., Cowan & Kail, 1996) and decline of STM as people become older (Kail & Salthouse, 1994). A model, initially proposed by Kail and Salthouse (1994) and subsequently adopted and adapted by Cowan and Kail (1996), relates STM span (i.e., accuracy) to two types of processing speed: Global, or general, processing speed and task-specific processing speed. What distinguishes the two types of processing speed are the measures used to investigate them (Edwards & Bagozzi, 2000). With reference to STM as measured through spoken production, task-specific processing speed relates specifically to the speed at which people can articulate words within a given time limit in a particular task (speeded element). On the other hand, global processing speed is measured with tasks unrelated to speech, but nonetheless have a speeded element. In reviewing relevant studies, Cowan and Kail (1996) showed that STM is modified by both global and task-specific processing speed measures.

Global processing speed is invariably affected in brain damage (Posner & Rueda, 2002) and aphasia, even in minimal aphasia (Neto & Santos, 2012). In relation to speech production skills, Bose and van Lieshout (2008) reported abnormal and slow articulatory patterns in speeded speech tasks in people with mild aphasia. A motoric impairment was unlikely to be an explanation for the slow articulatory patterns because the participants did not present with apraxia of speech or dysarthria. In addition to aphasia, apraxia of speech is considered a motor speech disorder that often co-occurs with aphasia. McNeil and Kent (1990) compared speech-timing patterns (in non-memory tasks) in persons with apraxia of speech and conduction aphasia. Both groups were similar in their ability to adjust their speech rate. McNeil and Kent argued for a phonetic-motoric component contributing to the speech patterns in both groups. Similar findings were also reported in more recent studies with

similar comparisons (e.g., Bose, van Lieshout, & Square, 2007). However, other studies suggest otherwise (e.g., Maas, Mailend, & Guenther, 2015).

To summarize, the overarching motivation for the present study hinges on the potential of speech-timing measures to expand the current knowledge boundaries of STM deficits in aphasia. We carried out speech-timing analyses of the spoken responses of persons with aphasia and control participants while they engaged in spoken repetition of word lists of increasing length. To our knowledge, this is the first systematic study of speech-timing measures of STM in aphasia. In this study, we relate the speech-timing patterns of persons with aphasia to activation-decay patterns and processing speed theory.

The following three main research questions and related hypotheses guided the present study:

1. What are the similarities and differences between people with aphasia and neuro-typical control participants in speech-timing STM measures? Given that persons with aphasia, often have slower speech production skills, when it comes to speech skills as utilized in STM tasks, the following working hypotheses can be drawn: (a) Speech-timing measures will be slower than those of neuro-typical control participants; and (b) based on evidence from unimpaired speakers, and to some extent persons with aphasia (as discussed in previous sections), speech-timing measures will be sensitive to memory load increments.
2. How do speech-timing measures in people with aphasia differ between correct and erroneous trials in STM span? Comparisons between correct and erroneous trials in studies that employed speech-timing measures in STM tasks have not been studied previously. For a long time, studying erroneous responses in aphasia research has been a fruitful way of refining our understanding of cognitive-linguistic processes and advancing theoretical positions (e.g., Ellis & Young, 1996). A comparison between correct and erroneous trials could help us understand if activation-decay patterns are similar in the two types of responses. We hypothesized differences in speech-timing measures in correct and erroneous trials, and also that effects of memory load would be more evident in erroneous than correct trials.
3. To what extent do speech-timing measures correlate with accuracy in STM and language measures in people with aphasia? The rationale for asking this question has to do with the construct validity of speech-timing measures of STM. If the two types of measures of STM (accuracy and speech-timing) measure the same underlying cognitive construct, there would be a correlation between STM accuracy and speech-timing measures. Relatedly, previous studies that used measures of STM accuracy showed that STM correlates with language measures (e.g., Martin & Saffran, 1997; Murray, 2012). The language measure we had at our disposal was language ability as measured by the Western Aphasia Battery. Consequently, the hypothesis we tested was that speech-timing measures would correlate with this language ability measure.

2. Method

2.1 Participants

Twelve people (8 males, 4 females) who presented with aphasia as a result of stroke and seven neuro-typical control participants (4 males, 3 females) were included. Hearing and vision were within normal limits in both groups. The mean age of the aphasia group was 54.8 years ($SD = 9.6$, range = 31–68). The mean age of the control group was 61 years ($SD = 10$, range = 48–72). The mean level of education of the aphasia group was 13.5 years ($SD = 2.4$, range = 10–18). The mean level of education of the control group was 13.2 years ($SD = 2.9$, range = 12–16). The aphasia group had a word span of 2.4 words ($SD = .79$, range = 1–4). The control group had a word span of 5.4 words ($SD = .49$, range = 4–7).

The people with aphasia were in the chronic stage and were medically stable. The mean duration between aphasia onset and time of testing was 61.1 months ($SD = 52.7$, range = 12–203). All participants were right handed. Detailed biographical information, language abilities as measured by the Western Aphasia Battery (Kertesz, 1982) and STM abilities are shown in Table 2. Eleven people had suffered a single symptomatic left hemisphere stroke. One person had suffered a bilateral stroke (person 5). Additionally, one person (person 6) had suffered an infarcted stroke and had undergone a craniotomy associated with cerebral hemorrhage. One person (person 4) had suffered a hemorrhagic stroke. All other participants had suffered infarcted strokes. Apraxia of speech assessment (Dabul, 2000) revealed possible apraxia of speech of mild severity in two participants (persons 2 and 6).

The project received ethical clearance from the Internal Review Board of Temple University. Individuals in both groups gave voluntary informed consent before their involvement in the project.

2.2 Data elicitation and measurement

The speech material used to generate the speech-timing measures came from retrospective audio recordings of the word repetition span subtest of the *Temple Assessment of Language and Short-term Memory in Aphasia* (TALSA; Martin, Minkina, Kohen, & Kalinyak-Fliszar, 2018; Martin, 2012). The word span subtest includes high-imageability, high-frequency nouns in memory loads from two to five words, presented in ascending order, starting with a memory load of two words. There were 10 trials for each memory load.

Testing took place in a quiet room. Participants listened to each trial, which was presented auditorily through a computer, at a rate of one word per second. The inter-word interval in each trial was one second. The delivery of stimuli was controlled by E-Prime 2.0 Professional experiment presentation software running on Dell OptiPlex 7010 with external speakers (Logitech R-10). At the end of each trial, a visual prompt was presented on the computer screen (500 ms after the end of the last word in the trial), which prompted participants to begin recalling the words serially. The participant's spoken response was recorded on a digital voice recorder (Tascam DR-40 Linear PCM Recorder). The test was discontinued at the memory load at which a participant made errors in six or more trials. Errors were phonemic deviations from the target, omissions, substitutions of another word (related or unrelated to the target), or word order errors. No participant with aphasia was

tested in memory loads of more than four words. The control participants were tested at memory loads of two to seven words but only trials of two to four words were included in the analyses.

The speech-timing measures were obtained with Praat speech analysis software (Boersma & Weenink, 2015), using the semi-automated function of silent pause identification of Praat. Silent segments were defined as those that were greater than 200 ms (Peach, 2013; Peach & Coelho, 2016). Praat generated an acoustic spectrograph of each recording and segmented the spectrographs of the audio-recorded TALSA administrations between silent and speech segments. The segments were checked and adjusted manually for accuracy by listening to each segment and viewing the acoustic spectrograph. Praat generated the durations of segments of interest.

Six measures were calculated (as shown in Figure 1) from the duration figures generated by Praat. The measures were as follows: (1) Response time (i.e., the duration of the whole response); (2) preparation time (i.e., the silent period between the word end measurement of the final word in a trial and the participant's first word recall); (3) pronunciation time (i.e., the duration of the whole response minus the preparation time); (4) speech time (i.e., the sum of individual word durations in a response); (5) word length (i.e., the mean duration of words in a response); and (6) inter-word pause (i.e., the mean duration of silent pauses that occurred between words in a response). If a response did not contain a pause (as defined earlier), a notional pause of zero was logged. In such cases, the duration of individual words in the response (i.e., word length) was derived by dividing the duration of the speech segment by the number of words. For example, if the trial "brush, lamp" was identified by Praat as one speech segment (i.e., without a pause between the two words) with a total duration of 1012 ms. A pause of zero ms was logged. The duration of each word was logged as 506 ms (i.e., 1012 divided by two). When participants produced speech disfluencies such as filled pauses (e.g., "hm", "ehm") or mazes (e.g., "something", "can't remember"), the boundaries and durations of these segments were logged in Praat. However, these durations were not included in the measurements. Responses that were incomplete in terms of the number of words a participant was required to produce at a specified memory load were also excluded from all analyses.

Reliability—The first, third and fourth authors as well as a speech-language pathology student at the final year of her studies carried out the analyses. All had received training and practice on the data extraction protocol. To ascertain the accuracy of the analyses, a second rater (from within the research team) analyzed speech material from four persons with aphasia (33% of the sample). Point-to-point reliability in the Praat measurements was 94%. A disagreement was defined as a discrepancy of 50 ms or more between two raters in the measurement of each segment.

2.3 Data analyses

Between group analyses—These analyses relate to the first research question. In both groups, only correct trials were considered for these analyses, that is, trials where all words were produced serially in terms of accuracy (Cowan, 1992; Hulme et al., 1999). Responses

that contained phonological deviations from the target words were considered errors and were excluded from these analyses. For each of the six measures described above, descriptive statistics (estimated marginal means, standards errors) were calculated for each group for each memory load (two, three, four words). The inferential statistical analyses involved a series of mixed-effects models of estimated marginal means (one model per measure). The models included the following fixed-effects: Group (aphasia, control), memory load (two, three, four words) and group by memory load interaction. Pairwise comparisons were Bonferroni corrected.

Aphasia group – Responses type analysis (i.e., error vs. correct)—These analyses related to the second research question. In this analysis, speech-timing patterns between correct and erroneous responses in the aphasia group were included. The erroneous responses did not include trials that were incomplete or were abandoned. The number of data points that were utilised in these analyses by memory load were as follows: Two words = 15, three words = 32, four words = 21. The statistical analyses were similar to those in the between group analyses. The fixed-effects were response type (error, correct), memory load (two, three, four words), and response type by memory load interaction. Pairwise comparisons were Bonferroni corrected.

Additionally, we compared the number of syllables in the words that participants actually produced in error responses to the number of syllables they should have produced, had their responses been correct. This analysis¹ is particularly relevant for understanding further the word length measure because differences between correct and error trials in word length could be attributed to the greater (or fewer) number of syllables participants may produce in error as compared to correct responses. For example, participant 3, in response to the stimulus trial *gun, scissors* (three syllables), said *queen, and feather*, (four syllables, counting the additional word *and*). In the following example, the same participant, in response to the stimulus *shoe, girl, ball, camel* (five syllables), said *shoe, ball, camel, earth* (five syllables). In this example, the number of syllables in the erroneous trial is the same as the stimulus trial. For the statistical comparison between numbers of syllables in correct vs. error responses, we carried out a two-tailed Mann-Whitney test.

Correlational analyses—There were two correlational analyses, which related to the third research question. In the first analysis, Spearman *rho* correlations (uncorrected) were carried out between mean performance on each speech-timing measure and STM span (i.e., accuracy) (shown in Table 2). In these analyses mean performance across all memory loads was used. This involved using only correct responses. In the second analysis (again Spearman *rho*, uncorrected), the choice of speech-timing measures was influenced by the results of the first set of correlations. The scores of each participant on each of the Western Aphasia Battery subtests (Table 2) were correlated with speech time, which was the only measure that resulted in a significant correlation in the first analysis.

¹We are grateful to the anonymous reviewer who prompted us to consider the issue of word length and the wider implications of the correct-erroneous comparisons and their interpretations.

3. Results

3.1 Between group results

Descriptive statistics for the correct responses in the two groups are shown in Table 3.

For response time, there was an effect of group, $F(1, 42) = 26.263$, $p < .001$, an effect of memory load, $F(2, 42) = 8.858$, $p < .001$, and no interaction between group and memory load, $F(2, 42) = 1.682$, $p = .19$. In the aphasia group, pairwise comparisons across memory loads were significant between two and three words, $p < .001$, and two and four words, $p < .001$. The difference between three and four words was not significant, $p = .85$. In the control group, none of the pairwise comparisons between different memory loads was significant, $p = .22$ or higher.

For preparation time, there was an effect of group, $F(1, 42) = 10.726$, $p < .001$, no effect of memory load, $F(2, 42) = .400$, $p = .67$, and no interaction between group and memory load, $F(2, 42) = .343$, $p = .71$.

For pronunciation time, there was an effect of group, $F(1, 42) = 30.968$, $p < .001$, an effect of memory load, $F(2, 42) = 15.989$, $p < .001$, and an interaction between group and memory load, $F(2, 42) = 3.708$, $p = .03$. In the aphasia group, pairwise comparisons across memory loads were significant between two and three words, $p < .001$, two and four words, $p < .001$, but not between three and four words, $p = .17$. In the control group, none of the pairwise comparisons were significant, $p = .09$ or higher.

For speech time, there was an effect of group, $F(1, 42) = 10.890$, $p < .001$, an effect of memory load, $F(2, 42) = 32.226$, $p < .001$, and no interaction between group and memory load, $F(2, 42) = .761$, $p = .47$. In the aphasia group, pairwise comparisons at each memory load were significant between two and three words, $p < .001$, two and four words, $p < .001$, but not between three and four words, $p = .16$. In the control group, all pairwise comparisons across memory loads were significant between two and three words, $p = .03$, two and four words, $p < .001$, but not between three and four words, $p = .06$.

For word length, there was an effect of group, $F(1, 42) = 12.274$, $p < .001$, no effect of memory load, $F(2, 42) = .729$, $p = .48$, and no interaction between group and memory load, $F(2, 42) = .444$, $p = .64$.

For pauses, there was an effect of group, $F(1, 42) = 21.812$, $p < .001$, no effect of memory load, $F(2, 42) = .207$, $p = .81$, and no interaction between group and memory load, $F(2, 42) = .812$, $p = .45$.

3.2 Aphasia group – Response type analysis (i.e., correct vs. error)

Table 4 presents the descriptive statistics for erroneous responses. Data from correct responses (shown in Table 3) have also been included for ease of comparison. We should note that in these analyses the pairwise comparisons were between correct and error responses at the same memory load condition (e.g., differences between correct and error responses at memory load of two words). The median number of syllables participants

produced in correct vs. error responses was 4 and 4 respectively. A Mann-Whitney test showed no difference in number of syllables between correct and error responses, $U = 2072.5$, $p = .61$.

For response time, there was an effect of response type, $F(1, 43) = 7.239$, $p = .01$, an effect of memory load, $F(2, 43) = 3.592$, $p = .03$, and no interaction between response type and memory load, $F(2, 43) = .632$, $p = .53$. Pairwise comparisons showed significant differences between error and correct responses in two words, $p = .01$. No other significant differences were found in three or four words, $p = .25$ or higher.

For preparation time, there was an effect of response type, $F(1, 43) = 5.715$, $p = .02$, no effect of memory load, $F(2, 43) = 2.683$, $p = .08$, and an interaction between response type and memory load, $F(2, 43) = 3.632$, $p = .03$. Pairwise comparisons showed a significant difference between error and correct responses of two words, $p < .001$. No other significant difference was found in three or four words, $p = .70$ or higher.

For pronunciation time, there was an effect of response type, $F(1, 43) = 4.707$, $p = .03$, an effect of memory load, $F(2, 43) = 26.803$, $p < .001$, and an interaction between response type and memory load, $F(2, 43) = 14.525$, $p < .001$. Pairwise comparisons did not show a significant difference between error and correct responses in two words, $p = .20$. There was an unexpected significant difference in three words, $p < .001$, with erroneous responses being longer than correct ones. Finally, there was a significant difference in four words, $p < .001$.

For speech time, there was an effect of response type, $F(1, 43) = 18.803$, $p < .001$, an effect of memory load, $F(2, 43) = 27.664$, $p < .001$, and an interaction between response type and memory load, $F(2, 43) = 13.861$, $p < .001$. Pairwise comparisons showed no significant difference between error and correct responses in two words, $p = .93$, a significant difference in three words, $p < .001$, and no significant difference in four words, $p = .61$.

For word length, there was an effect of response type, $F(1, 43) = 37.353$, $p < .001$, an effect of memory load $F(2, 43) = 36.341$, $p < .001$, and an interaction between response type and memory load, $F(2, 43) = 31.984$, $p < .001$. Pairwise comparisons showed no significant difference between error and correct responses in two words, $p = .42$, a significant difference in three words, $p < .001$, and no significant difference in four words, $p = .75$.

For pauses, there was an effect of response type, $F(1, 43) = 5.147$, $p = .02$, no effect of memory load, $F(2, 43) = 1.113$, $p = .33$, and no interaction between response type and memory load, $F(2, 43) = 2.092$, $p = .13$. Pairwise comparisons showed a significant difference between error and correct responses in two words, $p = .03$, no significant difference in three words, $p = .84$, and no significant difference in four words, $p = .08$.

Correlational analyses—The results of the first analysis, the bivariate correlations between word span (Table 2) and speech-timing measures, are shown in Table 5. The only statistically significant correlation was between STM word span and speech time, $\rho = .69$, $p = .01$. None of the other correlations was significant. Table 6 shows the results of the second analysis, correlations between speech time and performance in the language tasks of

the Western Aphasia Battery. Speech time correlated with Aphasia Quotient, spontaneous speech, and spoken language comprehension.

4. Discussion

The primary motivation of this study was to examine a range of speech-timing measures obtained from a serial word recall task in order to improve our understanding of the processes that contribute to the success and failure of STM in aphasia with reference to activation-decay and processing speed. Three main research questions guided this study, which we address in the next sections, together with their corresponding hypotheses. We also compare our findings to related findings in other literature domains where it was possible to draw direct comparisons from available published data. Given the relative sparseness of speech-timing analyses in STM and WM research more widely, such comparisons would help future researchers understand if speech-timing measures have generic or population-specific properties.

4.1 What are the similarities and differences between people with aphasia and neuro-typical participants in speech-timing STM measures?

In all six measures there were significant differences between the two groups such that people with aphasia were slower than control participants (main effects). This finding confirms the first hypothesis (i.e., speech-timing measures will be slower in persons with aphasia than neuro-typical participants). When memory load was considered, there were differences in response time, pronunciation time, and speech time, with three-word lists produced more slowly than lists of two words by people with aphasia. Counter-intuitively, there were no differences in these measures (or in the remaining three measures) between three- and four-word lists. These findings suggest that response time, pronunciation time, and speech time may be more sensitive measures of memory load than the other three measures (i.e., preparation time, word length, pauses). This means that the hypothesis (i.e., speech-timing measures would be sensitive to memory load increments) may be upheld only partially. Another explanation for the absence of a difference between three and four words could be due to statistical sensitivity. There were fewer trials in the comparisons between three and four words. Six of the 12 participants had a span of two words, three participants had a span of three words and only one participant had a span of four words (see Table 2). Furthermore, we examined the words in the experimental stimuli in terms of number of syllables (i.e., monosyllabic vs. bisyllabic words) across memory loads, in case there were more bisyllabic words in three-word lists (in comparison to two and four word lists) that could have affected the speech-timing measures. However, in a post-hoc analysis, the distribution of monosyllabic and bisyllabic words in the three memory loads was similar, $\chi^2(3) = 2.50, p = .32$. It is worth pointing out that speech time was the only measure able to discern differences not only in the aphasia group but also in the control group. In the control group, there was also a significant difference between two and three words and a non-significant trend ($p = .06$), suggesting progressively slower speech time as memory load increased.

If response time, pronunciation time and speech time are conceptualized as processing speed measures (task-specific, rather than global), in accordance with Kail and Salthouse's (1994) hypothesis, then processing speed may modify STM capacity in aphasia. While it is intuitive to suggest that these three measures could be conceptualized as task-specific processing speed measures, in the absence of more general processing speed measures in the present study, firmer conclusions cannot be drawn. Clearly, this is an avenue for further research that could examine if global processing speed (elicited from non-STM measures) and task-specific processing speed measures (i.e., STM speech-timing measures) co-vary.

In the present data, the faster response times in the control group in comparison to the aphasia group suggest that processing speed was fast across memory loads. Although there were main effects of memory load in the response and speech time measures, the pairwise comparisons that were carried out in the control group were not significant, reflecting the small memory demands in that group. However, in the aphasia group, memory load did exert a small effect in response times between two and three words but not between three and four words. Pronunciation time (i.e., the duration of the whole response minus the preparation time) was the only measure that revealed an interaction between group and memory load. The pairwise comparisons revealed that the interaction was driven by fluctuations in pronunciation time only in the aphasia group. Similar to response time, pronunciation time differed in memory loads of two and three words (as well as two and four words), but not in three and four words. Limited statistical sensitivity (as explained previously) could be the reason behind these findings.

Memory load did not exert any influence on preparation time or word and pause length. One possible explanation for these findings could be the relative lack of sensitivity of these measures in discerning differences in terms of memory load. Lack of sensitivity was also raised by Cowan (1992) in developmental psychology. Preparation time, word, and pause length measures are derived from the descriptively "larger" measures of response time, pronunciation time and speech time because they involve particular segments of the response utterances (see Figure 1). Lack of sensitivity is a plausible explanation for the absence of memory load effects, but it does not account for the lack of difference between memory loads of three and four words. However, Cowan et al. (1994) proposed that greater memory load (in this case of four words) elicits responses more quickly because participants may be concerned about forgetting and, consequently, produce the words in the four-word lists as fast as words in the three-word lists. This suggests that participants may engage a strategy as to how they choose to respond when they sense the possibility of being unsuccessful. If this is the case, clearly persons with aphasia have a meta-metacognitive awareness as to how to modify the speed of delivery in order to be successful.

Another possible explanation for the lack of memory load effect in preparation time, word, and pause length measures may be that levels of activation did not fluctuate as memory load increased. This is plausible because the measures were elicited only from correct trials where levels of activation were optimum (Martin et al., 1996). However, if this were the case, why then did memory load become evident in the other measures (response time, pronunciation time and speech time)? It is possible that the relative superior sensitivity of these measures may have revealed the memory load differences. As mentioned earlier,

speech time in particular revealed statistical differences within the neuro-typical participants (whereas other measures did not) as well as in persons with aphasia. This suggests that as a measure, speech time may be more sensitive than the other measures used in this study.

The responses of the aphasia group were more fragmented than the responses in the control group as the pause duration data suggest. However, neither the aphasia group nor the control group exhibited longer pauses as memory load increased. From all six speech-timing measures that have been studied in other literature domains, pauses attracted the greatest attention and have spurred interest in determining what kind of cognitive processing is occurring during these pauses. The presence and length of pauses have been interpreted as periods of memory searches by researchers in the developmental (e.g., Cowan et al., 1999; Cowan et al., 1998) and adult literature (e.g., Hulme et al., 1999; Tehan & Lalor, 2000). Hulme et al. (1999) linked pauses to identification of a word to be recalled for a given position in a trial. This interpretation fits well with the word retrieval deficit, which is a defining feature of aphasia, and a deficit that also relates to STM impairment in aphasia (Martin, 2009; Martin, 2012). Hulme and colleagues also suggested that during pauses, restoration (or redintegration) of the decayed trace takes place before being used as a response. This hypothesis is not supported by the present findings (i.e., absence of a memory load effect on pause time).

In the word length measure, the main effect of group was also evident, with the aphasia group taking longer to produce each word. Kynette and colleagues (1990) also found word length differences between younger and older adults. These authors only used a fixed memory load of five words. In the present study, word length was not affected by memory load in either group. In young adult participants, Haberlandt et al. (2005) found that word durations differed as a function of memory load (four, five, six words), a finding similar to Cowan et al. (1998) for children. However, in other studies with young adults the findings are mixed. Tehan and Lalor (2000) and Towse et al. (2008) did not find differences in word length as memory load increased. In the developmental literature, Cowan (1992) also reported mixed findings for the word length measure, such that in some conditions (lists with phonologically dissimilar words, e.g., cat, pen) word length varied by memory load, while in a task with phonologically similar words (e.g., cat, bat), word length was not affected by memory load. While there have been attempts to relate word length to STM in non-aphasia studies, aphasia studies that examined word durations in tasks with minimal STM demands, found that word durations in speakers with aphasia are longer than durations of neuro-typical speakers (e.g., Ryalls, 1986). We return to this issue below when we discuss the error patterns.

4.2 How do speech-timing measures in people with aphasia differ between correct and erroneous trials in STM span?

In all six measures, erroneous responses were considerably longer than correct ones. Based on this finding, the hypothesis that there would be a difference between correct and erroneous trials is upheld. There was also considerable variability in all measures. For example, in preparation time, although there was no overall effect of memory load, a difference emerged in lists of two words. The longer preparation time in erroneous trials

may suggest that participants may have had an awareness that the trial to be produced was going to be erroneous. Word length was longer only in three words. The difference in the number of syllables participants produced in error responses as compared to the number of syllables they should have produced was not significant. This was also the case in a related post-hoc comparison that involved number of syllables only in three-word error responses, Mann-Whitney $U = 252$, $p = .59$. As discussed earlier, people with aphasia are slower in producing words in general (e.g., Baum et al., 1990). The present finding of slower word production in erroneous three-word lists cannot be readily explained. A difficulty in explaining speech-timing measures in erroneous responses lies in the unpredictability of word choices in these cases. Consequently, the words that are produced could differ from words in correct responses in terms of lexical variables (e.g., frequency, imageability). Such differences are likely to contribute to longer response times in the speech-timing measurements of erroneous responses than their correct counterparts. Although we compared the number of syllables in error trials against the number of syllables participants should have produced, we did not compare the number of syllables in the actual correct trials participants did produce. This constitutes an inherent limitation of this analysis. However, as reported earlier (section 4.1), the distribution of monosyllabic and bisyllabic words was similar. The alternative analysis, that is, counting the actual number of syllables in correct trials and comparing them against the number of syllables in error trials, would have resulted in very uneven sample sizes and made such a comparison unreliable (Zimmerman, 2003). In summary, the mixed findings show that the hypothesis that memory load would be greater in erroneous than correct trials cannot be upheld.

The question that arises is why there was a difference between correct and erroneous trials, mindful of the important methodological limitation that such a comparison entails. It is not possible to discuss the findings in the erroneous trials in the context of studies involving neuro-typical children and adults because error responses were not analyzed in these studies. The only exception is Haberlandt et al. (2005) who scrutinized pauses in erroneous lists of six words in neuro-typical young adults. Pauses were shorter in list-initial and list-final positions in comparison to pauses in list-medial positions. However, in the study by Haberlandt and colleagues participants were asked to say “pass” in place of any words they could not recall, whereas in our study this was not the case. This may explain the discrepancy between the two studies in terms of pause patterns in erroneous trials. In our data, one reason for these highly variable and irregular patterns between erroneous and correct responses could stem from difficulties in regulating activation levels (e.g., Martin et al., 1994; Schwartz et al., 1994). This model of word processing includes two parameters of spreading activation that regulate efficient and enduring access to word representations in all language tasks: Connection strength (impairment leading to slowing of activation) and activation decay (impairment leading to too-fast decay of activation). Although the latter parameter plays a key role in maintaining activation of representations in STM, impairment of each parameter can contribute to word processing and verbal STM impairment. Future studies that include many more participants and complementary methodological paradigms (e.g., connectionist modeling or response time tasks as shown in Table 1) might be able to distinguish contributions of each of these processing characteristics (connection strength and activation decay) to accuracy. For example, future studies could examine patterns of

performance in individuals whose word retrieval impairments are defined as being rooted in a connection strength impairment (slowed activation), activation maintenance impairments (too fast decay of activation) or some combination of both activation parameters.

4.3 To what extent do speech-timing measures correlate with accuracy in STM and language measures in people with aphasia?

We posed this question because we were interested in investigating the construct validity of speech-timing measures and standard measures of STM accuracy, and more broadly, language abilities as measured by the Western Aphasia Battery. The first set of correlations showed that STM span in people with aphasia correlated only with the speech time measure (i.e., the sum of individual word durations in correct trials). The correlation was moderate ($\rho = .69$). None of the other measures correlated significantly with STM span. The hypothesis that speech-timing measures will correlate with accuracy in STM can only be partially upheld. Two studies compared speech time and span in the literature and both reported similar findings. For word span in children, Cowan (1992) also reported a strong correlation between speech time and word span ($r = .82$, $r = .77$, for phonologically dissimilar and similar word lists respectively). Cowan also found only two other measures that correlated with STM span: Pronunciation and response time. Tehan and Lalor (2000) who studied neuro-typical young adults, also found that speech time correlated with digit span ($r = .36$). Preparation time and pauses did not correlate with digit span in that study.

The second set of correlations was carried out between the speech time measure and the language subtest scores of the Western Aphasia Battery. The measures that yielded significant, fairly moderate correlations were the Aphasia Quotient (a composite measure and a crude index of severity of aphasia), spontaneous speech and auditory comprehension. Repetition of words and sentences, which has inherent STM demands, did not correlate with speech time, which was elicited in this experiment through repetition of word lists. However, the repetition subtest of the Western Aphasia Battery involves phrases and sentences and as such differs from the word span task. One might expect a correlation between speech time and the repetition subtest if speech time was elicited from a sentence repetition task. Naming ability did not correlate with speech time. Mindful of the small sample size and the issue of multiple comparisons, firm conclusions cannot be drawn. In relation to the hypothesis (i.e., speech-timing measures would correlate with language ability), the patterns of correlations suggest that the hypothesis is partially upheld.

4.4 Summary, implications and conclusions

The present study is the first systematic and comprehensive investigation of speech-timing measures of STM in aphasia. As such, it was exploratory. Where possible, we drew comparisons between our data and data reported in other literature domains, notably developmental psychology. Some similarities emerged between our and previous studies. The speech-timing measures revealed differences between persons with aphasia and neuro-typical control participants. Memory load was evident in some measures but not others. Similar findings were also reported in other studies. Previous findings, for example, those that showed differences in speech-timing measures between younger and older children, were also found in the present study. Each measure offered a novel glimpse into the latent

mechanisms of STM in aphasia. The patterns we found show time-protracted and fragmented speech, particularly evident in erroneous responses. This is consistent with what researchers and clinicians observe when they assess STM in persons with aphasia, especially when STM errors arise, as the results from erroneous trials suggest. We have offered preliminary interpretations for some of these patterns in relation to processing speed and activation-decay patterns. Processing speed theory as utilised in STM research in other populations has not been considered seriously in STM and WM research in aphasia. This is an avenue of research that could prove theoretically as well as clinically useful. For example, speech-timing measures in span tasks or other samples of connected speech, may help discern mild/minimal aphasia and the potential for full recovery after stroke or other medical conditions. To date, theoretical support for the role of activation-decay patterns in STM span, and indeed other theoretical positions of STM impairment, comes predominantly from studies of accuracy.

Another wider implication of the present study has to do with whether speech-timing measures would relate to other response-time measures (discussed in section 1.2), elicited from input tasks. The issue of modality (i.e., input vs. output) in STM is a long-standing theme in aphasia STM research (e.g., Howard & Franklin, 1990; Romani, 1992; Vallar & Papagno, 2002). Studies on this theme have sought to understand whether STM comprises input and output buffers that can be differentially impaired in some persons with aphasia (e.g., Romani, 1992). Furthermore, it is worth noting that these studies have used only measures of accuracy to advance theoretical accounts of a modular architecture of STM that consists of input and output buffers. Response-time measures that yield information about input processes have been used in aphasia STM research (see Table 1). To date, however, no study has compared response-time measures for input and output (i.e., using speech-timing measures), complemented by measures of accuracy and other indices of cognitive-linguistic processing. It would be important to carry out such a study for at least two theoretical reasons: First, to investigate the issue of a modular architecture of STM in terms of input and output buffers, and second, to investigate similarities and differences between response timing in input STM tasks and speech timing in output tasks. Such a comparison would help refine the activation-decay account as well as help us understand if processing speed theory has a place in STM research in aphasia.

Finally, we should acknowledge the limitations of the current study. The small sample size, relatively large number of measures and lack of a priori analysis of statistical power may diminish the external validity of our findings. Like many other studies in aphasia, there could be have been inherent variability of performance (e.g., Tyler, 1992). Another limitation is that the 200ms duration that defined pauses may not have been sensitive to detect differences in the correct vs. error analyses. Future studies should address these limitations.

To conclude, in this study we explored the potential of speech-timing measures for STM research in aphasia. While the use of speech-timing (and also response-time) measures in STM has existed for some time, such measures have not been adopted widely in aphasiology. This is an uncharted territory, which can only enrich our knowledge of the

factors that affect STM and WM deficits in aphasia. We hope our study inspires researchers to examine response-time measures comprehensively in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this paper was supported by National Institute on Deafness and other Communication Disorders Center of the National Institutes of Health under award number R21DC 008782 and R01DC013196. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are grateful to all participants who assisted us in this project. We would also like to thank Kelsey Homan who carried out some of the speech-timing analysis.

References

1. Allen MC, Martin RC, Martin N. Relations between short-term memory deficits, semantic processing, and executive function. *Aphasiology*. 2012; 26:428–461. [PubMed: 22736889]
2. Attout L, van der Kaa M, George M, Majerus S. Dissociating short-term memory and language impairment: The importance of item and serial order information. *Aphasiology*. 2012; 26:355–382.
3. Barde LHF, Schwartz MF, Chrysikou EG, Thompson-Schill SL. Reduced short-term memory span in aphasia and susceptibility to interference: Contribution of material-specific maintenance deficits. *Neuropsychologia*. 2010; 48:909–920. [PubMed: 19925813]
4. Baum S, Blumstein S, Naeser M, Palumbo C. Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language*. 1990; 39:33–56. [PubMed: 2207620]
5. Boersma P, Weenink D. Praat: Doing phonetics by computer [Computer programme]. Version 5.4.19. 2015
6. Bose A, van Lieshout P. Effects of utterance length on lip kinematics in aphasia. *Brain and Language*. 2008; 106:4–14. [PubMed: 18440061]
7. Bose A, van Lieshout P, Square PA. Word frequency and bigram frequency effects on linguistic processing and speech motor performance in individuals with aphasia and normal speakers. *Journal of Neurolinguistics*. 2007; 20:65–88.
8. Conway ARA, Cowan N, Bunting MF, Theriault DJ, Minkoff SRB. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*. 2002; 30:163–183.
9. Cowan N. Verbal memory span and timing of spoken recall. *Journal of Memory and Language*. 1992; 31:668–684.
10. Cowan N. The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology*. 1999; 72:193–209. [PubMed: 10047439]
11. Cowan N. What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*. 2008; 169:323–338. [PubMed: 18394484]
12. Cowan N, Kail R. Covert processes and their development in short-term memory. In: Gathercole S, editor *Models of Short-term Memory*. Hove, UK: Psychology Press; 1996. 29–50.
13. Cowan N, Keller TA, Hulme C, Roodenrys S, McDougall S, Rack J. Verbal memory span in children: Speech timing cues to the mechanism underlying age and word length effects. *Journal of Memory and Language*. 1994; 33:234–250.
14. Cowan N, Wood NL, Wood PK, Keller TA, Nugent LD, Keller CV. Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General*. 1998; 127:141–160. [PubMed: 9622911]
15. Cowan N, Towse JN, Hamilton Z, Sauls JS, Elliott EM, Lacey JF, Hitch G. Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*. 2003; 132:113–132. [PubMed: 12656300]

16. Dabul B. Apraxia Battery for Adults. 2. Austin, TX: Pro-Ed; 2000.
17. DeDe G, Ricca M, Knilans J, Trubl B. Construct validity and reliability of working memory tasks for people with aphasia. *Aphasiology*. 2014; 28:692–712.
18. Dell G, Martin M, Schwarz MF. A case-series test of the interactive two-step model of lexical access. Predicting word repetition from picture naming. *Journal of Memory and Language*. 1997; 56:490–520.
19. Doshier BA, Ma J. Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1998; 24:316–335.
20. Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. *Psychological Methods*. 2000; 5:155–174. [PubMed: 10937327]
21. Eling P. Kurt Goldstein's test battery. *Cortex*. 2015; 63:16–26. [PubMed: 25241395]
22. Ellis AW, Young AW. *Human Cognitive Neuropsychology: A textbook with readings*. Hove, UK: Psychology Press; 1996.
23. E-Prime Professional [Computer software]. Pittsburgh, PA: Psychology Software Tools;
24. Haberlandt K, Lawrence H, Krohn T, Bowe K, Thomas JG. Pauses and durations exhibit a serial position effect. *Psychonomic Bulletin & Review*. 2005; 12:152–158. [PubMed: 15945208]
25. Hamilton AC, Martin RC. Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience*. 2005; 5:1–13.
26. Hamilton AC, Martin RC. Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex*. 2007; 43:112–123. [PubMed: 17334211]
27. Howard D, Franklin S. Memory without rehearsal. In: Vallar G, Shallice T, editors *Neuropsychological Impairments of Short-term Memory*. Cambridge, UK: Cambridge University Press; 1990. 287–318.
28. Howard D, Nickels L. Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology*. 2005; 22:42–77. [PubMed: 21038240]
29. Hulme C, Newton P, Cowan N, Stuart G, Brown G. Think before you speak: Pauses, memory search rate, and trace reintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1999; 25:447–463.
30. Kail R, Salthouse TA. Processing speed as a mental capacity. *Acta Psychologica*. 1994; 86:199–225. [PubMed: 7976467]
31. Kertesz A. *Western Aphasia Battery*. New York: Grune & Stratton; 1982.
32. Kinsbourne M. Behavioral analysis of the repetition deficit in conduction aphasia. *Neurology*. 1972; 22:1126–1132. [PubMed: 4673556]
33. Kynette D, Kemper S, Norman S, Cheung H. Adults' word recall and word repetition. *Experimental Aging Research*. 1990; 16:117–121. [PubMed: 2090462]
34. Maas E, Mailend ML, Guenther FH. Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production. *Journal of Speech, Language and Hearing Research*. 2015; 58:185–200.
35. Majerus S, Attout L, Artielle M, van der Kaa M. The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia*. 2015; 77:165–176. [PubMed: 26275964]
36. Martin N. The role of semantic processing in short-term memory and learning: Evidence from aphasia. In: Thorn A, Page M, editors *Interactions between Short-term and Long-term Memory in the Verbal Domain*. New York: Psychology Press; 2009. 220–243.
37. Martin N. Management of communication deficits associated with memory disorders. In: Peach RK, Shapiro LP, editors *Cognition and Acquired Language Disorders*. St Louis, MI: Mosby; 2012. 275–297.
38. Martin N, Ayala J. Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. *Brain and Language*. 2004; 89:464–483. [PubMed: 15120538]
39. Martin N, Reilly J. Short-term/working memory impairments in aphasia: Data, models, and their application to aphasia rehabilitation. *Aphasiology*. 2012; 26:253–257.

40. Martin N, Minkina I, Kohen FP, Kalinyak-Fliszar M. Assessment of linguistic and verbal short-term memory components of aphasia. *Journal of Neurolinguistics*. 2018
41. Martin N, Saffran EM. Language and auditory-verbal short-term memory deficits on verbal learning: Evidence from aphasia. *Cognitive Neuropsychology*. 1997; 14:641–682.
42. Martin N, Saffran EN. Effects of word processing and short-term memory deficits on verbal learning: Evidence from aphasia. *International Journal of Psychology*. 1999; 34:330–346.
43. Martin N, Saffran EM, Dell GS. Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*. 1996; 52:83–113. [PubMed: 8741977]
44. Martin RC, Allen CM. A disorder of executive function and its role in language processing. *Seminars in Speech and Language*. 2008; 29:201–210. [PubMed: 18720317]
45. Maybery MT, Parmentier FBR, Jones DM. Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language*. 2002; 47:360–385.
46. Mayer JF, Murray LL. Measuring working memory deficits in aphasia. *Journal of Communication Disorders*. 2012; 45:325–339. [PubMed: 22771135]
47. Mills RH, Knox AW, Juola JF, Salmon SJ. Cognitive loci of impairments in picture naming by aphasia subjects. *Journal of Speech and Hearing Research*. 1979; 22:73–87. [PubMed: 502502]
48. Minkina I, Rosenberg S, Kalinyak-Fliszar M, Martin N. Short-term memory and aphasia: From theory to treatment. *Seminars in Speech and Language*. 2017; 38:17–28. [PubMed: 28201834]
49. McNeil MR, Kent RD. Motoric characteristics of adult aphasic and apraxic speakers. In: Hammond GR, editor *Advances in Psychology: Cerebral control of speech and limb movements*. New York: Elsevier; 1990. 349–386.
50. Murray LL. Direct and indirect treatment approaches for addressing short-term or working memory deficits in aphasia. *Aphasiology*. 2012; 26:317–337.
51. Murray L, Salis C, Martin N, Dralle J. The use of standardized short-term and working memory tests in aphasia research: A systematic review. *Neuropsychological Rehabilitation*. 2018; 28:309–351. [PubMed: 27143500]
52. Neto B, Santos ME. Language after aphasia: Only a matter of speed processing? *Aphasiology*. 2012; 26:1352–1361.
53. Novick JM, Kan IP, Trueswell JC, Thompson-Schill SL. A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*. 2009; 26:527–567. [PubMed: 20183014]
54. Paek EJ, Murray LL. Working memory approach for language and memory deficits in an individual with chronic aphasia: A case study. *eHearsay*. 2015; 5:86–98.
55. Peach RK. The cognitive basis for sentence planning difficulties in discourse after traumatic brain injury. *American Journal of Speech-Language Pathology*. 2013; 22:S285–S297. [PubMed: 23695905]
56. Peach RK, Coelho CA. Linking inter- and intra-sentential processes for narrative production following traumatic brain injury: Implications for a model of discourse processing. *Neuropsychologia*. 2016; 80:157–164. [PubMed: 26593882]
57. Posner MI, Rueda MR. Mental chronometry in the study of individual and group differences. *Journal of Clinical and Experimental Neuropsychology*. 2002; 24:968–976. [PubMed: 12647772]
58. Romani C. Are there distinct input and output buffers? Evidence from an aphasic patient with an impaired output buffer. *Language and Cognitive Processes*. 1992; 7:131–162.
59. Ryalls JH. An acoustic study of vowel production in aphasia. *Brain and Language*. 1986; 29:48–67. [PubMed: 3756461]
60. Salis C, Hwang F, Howard D, Lallini N. Short-term and working memory treatments for improving sentence comprehension in aphasia: A review and a replication study. *Seminars in Speech and Language*. 2017; 38:29–39. [PubMed: 28201835]
61. Schwartz MF, Dell G, Martin N, Saffran EM. Normal and aphasic naming in an interactive spreading activation model. *Brain and Language*. 1994; 47:391–394.

62. Schweickert R, Boruff B. Short-term memory capacity: Magic number of magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1986; 12:419–425.
63. Sternberg S. High-speed scanning in human memory. *Science*. 1966; 153:652–654. [PubMed: 5939936]
64. Sternberg S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*. 1969; 57:421–457. [PubMed: 5360276]
65. Sung JE, McNeil MR, Pratt SR, Dickey MW, Hula WD, Szuminsky NJ, Doyle PJ. Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology*. 2009; 23:1040–1052.
66. Swinney DA, Taylor OL. Short-term memory recognition search in aphasics. *Journal of Speech and Hearing Research*. 1971; 14:578–588. [PubMed: 5163892]
67. Tehan G, Lalor DM. Individual differences in memory span: The contribution of rehearsal, access to lexical memory, and output speed. *The Quarterly Journal of Experimental Psychology*. 2000; 53A:1012–1038.
68. Towse JN, Cowan N, Hitch G, Horton NJ. The recall of information from working memory: Insights from behavioural and chronometric perspectives. *Experimental Psychology*. 2008; 55:371–383. [PubMed: 19130763]
69. Tyler LK. *Spoken Language Comprehension: An experimental approach to disordered and normal processing*. Cambridge, MA: MIT Press; 1992.
70. Vallar G, Papagno C. Neuropsychological impairments of verbal short-term memory. In: Baddeley AD, Kopelman MD, Wilson BA, editors *The Handbook of Memory Disorders*. 2. Chichester, UK: Wiley; 2002. 249–270.
71. Verhaegen C, Pierrat F, Poncelet M. Dissociable components of phonological and lexical-semantic short-term memory and their relation to impaired word production in aphasia. *Cognitive Neuropsychology*. 2013; 30:544–563. [PubMed: 24547971]
72. Warrington EK, Shallice T. The selective impairment of auditory-verbal short-term memory. *Brain*. 1969; 92:885–896. [PubMed: 5364015]
73. Wilshire CE, Keall LM, O' Donnell DJ. Semantic contributions to immediate serial recall: Evidence from two contrasting aphasic individuals. *Neurocase*. 2010; 16:331–351. [PubMed: 20552532]
74. Wright HH, Fergadiotis G. Conceptualising and measuring working memory and its relationship to aphasia. *Aphasiology*. 2012; 26:258–278. [PubMed: 22639480]
75. Zakariás L, Keresztes A, Marton K, Wartenburger I. Positive effects of a computerised working memory and executive function training on sentence comprehension in aphasia. *Neuropsychological Rehabilitation*. 2018; 28:369–386. [PubMed: 26999324]
76. Zimmerman DW. A warning about the large-sample Wilcoxon-Mann-Whitney test. *Understanding Statistics*. 2003; 2:267–280.

Highlights

- The first systematic study of speech-timing measures of short-term memory span in stroke aphasia.
- Speech-timing measures differentiate performance between persons with aphasia and neuro-typical control participants.
- Speech-timing measures reveal information about levels of processing speed and activation of memory representations
- Speech-timing measures (and response-time measures in general) elicited from short-term memory and working memory tasks need to be studied more systematically in aphasiology.

Stimulus list:
broom, fish, needle

Participant's response:
broom ... fish ... needle

Response time:	$G - A$
Preparation time:	$B - A$
Pronunciation time:	$G - B$
Speech time:	$(C - B) + (E - D) + (G - F)$
Word length:	$\text{Speech time} / 3$
Inter-word pause:	$[(D - C) + (F - E)] / 2$

Figure 1.
Illustration of a stimulus list, with an accuracy response (adapted from Cowan, 1992)

Table 1

Summary of studies that used response-timing measures to index STM/WM

STM/WM measures	Brief description of response-timing measure	Studies
<i>word span</i>	spectrographic analysis of time lapse between onset of target word and onset of person's spoken response to each target word	Martin et al. (1996), N = 1
<i>auditory-verbal digit span</i>	spoken repetition; timing measured from onset of stimulus to onset of person's spoken response	Kinsbourne (1972), N = 2
<i>visual-verbal digit span</i>	spoken repetition of visually presented digits; timing measured from onset of stimulus to onset of person's spoken response	Kinsbourne (1972), N = 2
<i>auditory-m anual digit span</i>	pointing to auditorily presented digits written on card; timing measured from audible tap in tape recording	Kinsbourne (1972), N = 2
<i>item and order probe recognition</i>	response times elicited through computer key presses ("yes/no") for recognition tasks of word pairs (item and order), following presentation of single words (e.g., flower, pen, ruler) with a correct order probe flower-ruler	Attout et al. (2012), N = 2
<i>recent negatives</i>	response times elicited through computer key presses ("yes/no") for recognition tasks, following individual presentation of triplets of words or letters	Allen, Martin, & Martin (2012), N = 20; Barde, Schwartz, Chrysikou, & Thompson-Schill (2010), N = 20; Hamilton & Martin (2005, 2007), N = 1; Novick, Kan, Trueswell, & Thompson-Schill (2009), N = 3
<i>verbal and visual n-back</i>	response times elicited through computer key presses from person's correct responses	Mayer & Murray (2012), N = 14
<i>digit recognition</i>	response times elicited through computer key presses ("yes/no") in recognition task following a list of two, four or six digits	Swinney & Taylor (1971), N = 8
<i>pictorial recognition</i>	responses times elicited through computer key presses ("yes/no") in recognition task, following list of two or four pictures or shapes	Mills, Knox, Juola, & Salmon (1979), N = 10

Table 2

Aphasia group information: Biographical, language and word span measures

	1	2	3	4	5	6	7	8	9	10	11	12
	<i>Biographical information</i>											
Age	68	52	53	53	31	52	52	68	63	57	54	56
Gender	F	M	M	M	M	F	M	M	F	M	M	F
Education (years)	14	14	12	12	16	10	12	18	16	14	10	14
Time post onset (months)	36	69	86	19	12	24	203	343	87	37	91	35
	<i>Western Aphasia Battery: Severity of aphasia & specific language abilities</i>											
Aphasia Quotient	90.5	79.8	57.7	70.6	79	76.7	81.4	62.6	80.3	93.5	90.3	89.1
Spontaneous speech	18	15	9	15	7.9	12	14	12	14	19	18	17
Comprehension	9.75	8.2	5.95	8.1	7	8.95	9.7	6.9	9.65	9.75	9.45	9.35
Repetition	9.7	8.8	7.7	6.8	5.5	8.6	9	6.6	7.8	9.1	8	8.4
Naming	7.8	7.9	6.2	8.3	9	8.8	8	5.8	8.7	8.9	9.7	9.8
	<i>TALSA: STM span</i>											
Word span	4	3	2	2	1	3	2	2	3	3	2	2

Table 3

Correct responses in both groups: Estimated marginal means (SEs)

		Two words	Three words	Four words
Response time	<i>aphasia</i>	3635 (.383)	5667 (.400)	6504 (.664)
	<i>controls</i>	2542 (.502)	3146 (.502)	3834 (.502)
Preparation time	<i>aphasia</i>	1660 (.141)	1854 (.147)	1621 (.244)
	<i>controls</i>	1121 (.185)	1224 (.185)	1310 (.185)
Pronunciation time	<i>aphasia</i>	1974 (.270)	3817 (.282)	4882 (.467)
	<i>controls</i>	1420 (.353)	1916 (.353)	2524 (.353)
Speech time	<i>aphasia</i>	1360 (.91)	2058 (.95)	2421 (.158)
	<i>controls</i>	1192 (.119)	1639 (.119)	2044 (.119)
Word length	<i>aphasia</i>	659 (.32)	689 (.33)	605 (.56)
	<i>controls</i>	554 (.42)	523 (.42)	511 (.42)
Inter-word pause	<i>aphasia</i>	614 (.121)	877 (.126)	820 (.210)
	<i>controls</i>	215 (.158)	124 (.158)	159 (.158)

Note. Figures represent ms

Table 4

Error and correct responses: Estimated marginal means (SEs)

		Two words	Three words	Four words
Response time	<i>error</i>	6372 (841)	6777 (741)	8163 (908)
	<i>correct</i>	3635 (642)	5667 (671)	6504 (1112)
Preparation time	<i>error</i>	3263 (336)	1987 (296)	1841 (363)
	<i>correct</i>	1660 (256)	1854 (268)	1621 (444)
Pronunciation time	<i>error</i>	2814 (1445)	1182 (974)	10482 (5162)
	<i>correct</i>	1974 (524)	3817 (1364)	4882 (2082)
Speech time	<i>error</i>	1573 (304)	4789 (268)	2683 (328)
	<i>correct</i>	1360 (232)	2058 (242)	2421 (402)
Word length	<i>error</i>	782 (121)	2312 (107)	670 (131)
	<i>correct</i>	659 (93)	689 (97)	605 (161)
Inter-word pause	<i>error</i>	1556 (337)	771 (297)	1848 (364)
	<i>correct</i>	614 (258)	877 (269)	820 (446)

Note. Figures represent ms

Table 5

Correlations between STM word span and speech-timing measures

	<i>rho</i>	<i>p</i> values
STM word span		
Response time	.08	.81
Preparation time	-.18	.57
Pronunciation time	.29	.36
Speech time	.69	.01
Word length	.33	.29
Inter-word pause	-.24	.43

Table 6

Correlations between speech time and language abilities

	<i>rho</i>	<i>p</i> values
Speech time		
Aphasia Quotient	.59	.04
Spontaneous speech	.69	.01
Comprehension	.59	.04
Repetition	.48	.11
Naming	-.15	.64