Abstract

The aim of this study was to examine the validity and feasibility of sentence transcription testing (STT) for the purpose of examining the interplay between verbal working memory and central processing. The general area of interest is to understand working memory as a dynamic system that involves the management and integration of information from several temporal distances. Due to the world-wide conditions at the time this study was conducted (2020), the testing was online and computerized, which severely limited the controllability of the procedure leading to a high amount of exclusions and dubious results. The testing of 17 subjects, 9 females, 8 males with the average age of 30.5 ($SD = 9.5)$ yielded mixed results, excluding gender, impulsivity and age as likely factors for the variance. Following these results, a post hoc analysis was added to interpret if transcription data has validity as a tool for observing effects of interference on memory recall and the task at hand. This analysis did reveal patterns that reinforce the view of language processing as a multimodal task. The type of errors seems to follow tendencies of primacy, recency and availability, as well as proactive and retroactive interference. These tendencies of memory recall seem to work in unison or is a manifestation of syntactic, lexical, and presumably semantic processing and can be used to measure individual differences in language processing and the tendency to linguistically “fill in the gaps”. The variation seen within the sample does make transcription testing appealing for further studies. The main variance within the sample can be described as replacing words with other previously attended to information, and or forgetting words during transcription. These tendencies, might reveal properties about the interaction between executive function (EF) and verbal working memory (V-WM) as a source of individual difference. However, more validation studies are proposed for weaving out factors that might skew the results in this type of testing and modelling.

Title

Using sentence Transcription testing – As a way to test the interference effects and dynamics of verbal-working memory

Author

Sinal Bou Aram

Keywords

Sentence transcription testing verbal working memory central control processes agency memory bias availability recency primacy language processing interference theory syntactic processing
Using Sentence Transcription Testing –
As a way to test the interference effects and dynamics of verbal-working memory

Sinal Bou Aram

Supervisor: Ulf Träff
Examinator: Kenny Skagerlund
Abstract

The aim of this study was to examine the validity and feasibility of sentence transcription testing (STT) for the purpose of examining the interplay between verbal working memory and central processing. The general area of interest is to understand working memory as a dynamic system that involves the management and integration of information from several temporal distances. Due to the worldwide conditions at the time this study was conducted (2020), the testing was online and computerized, which severely limited the controllability of the procedure leading to a high amount of exclusions and dubious results. The testing of 17 subjects, 9 females, 8 males with the average age of 30.5 ($SD = 9.5$) yielded mixed results, excluding gender, impulsivity and age as likely factors for the variance. Following these results, a post hoc analysis was added to interpret if transcription data has validity as a tool for observing effects of interference on memory recall and the task at hand. This analysis did reveal patterns that reinforce the view of language processing as a multimodal task. The type of errors seems to follow tendencies of primacy, recency and availability, as well as proactive and retroactive interference. These tendencies of memory recall seem to work in unison or is a manifestation of syntactic, lexical, and presumably semantic processing and can be used to measure individual differences in language processing and the tendency to linguistically “fill in the gaps”. The variation seen within the sample does make transcription testing appealing for further studies. The main variance within the sample can be described as replacing words with other previously attended to information, and or forgetting words during transcription. These tendencies, might reveal properties about the interaction between executive function (EF) and verbal working memory (V-WM) as a source of individual difference. However, more validation studies are proposed for weaving out factors that might skew the results in this type of testing and modelling.
Keywords: Sentence transcription testing verbal working memory central control processes agency memory bias availability recency primacy language processing interference theory

Table of contents

**Introduction** 1

*Theoretical models for working memory* 2

**The hypothesis and research question** 11

**Method** 12

*Participants* 12

*Procedure* 14

*Test description and data variables* 15

**Constructing the sentences** 17

*Sentence composition* 18

*Sentence complexity based on symbolic cognitive models* 18

**Data processing** 23

*Inhibition performance* 23

*RIT-score calculation* 23

*Model of interference* 27

**Results** 30

*Within-group ANOVA of performance between tests* 30

*Age-differences* 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex-differences</td>
<td>33</td>
</tr>
<tr>
<td>Response inhibition testing results</td>
<td>33</td>
</tr>
<tr>
<td>Further analysis</td>
<td>34</td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td>36</td>
</tr>
<tr>
<td><strong>Conclusion</strong></td>
<td>44</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>45</td>
</tr>
<tr>
<td>Digital links</td>
<td>53</td>
</tr>
<tr>
<td>Appendix A</td>
<td>54</td>
</tr>
<tr>
<td>Appendix B</td>
<td>56</td>
</tr>
<tr>
<td>Appendix C</td>
<td>57</td>
</tr>
<tr>
<td>Appendix D</td>
<td>58</td>
</tr>
<tr>
<td>Appendix E</td>
<td>60</td>
</tr>
<tr>
<td>Appendix F</td>
<td>61</td>
</tr>
</tbody>
</table>
Using Sentence Transcription Testing -

As a way to test the interference effects and dynamics of verbal-working memory

There are today several testing techniques devoted to measure aspects of working memory (WM). Many of them assess working memory from the perspective of capacity (for example digit span, or sequencing tests) and retention strategy. While others focus on the working part of WM, for example by disrupting the rehearsal loop by recounting items in reverse (also known as the Brown-Peterson technique), which tests the ability to manipulate recently acquired information in a flexible, *on-the-fly*, way.

The sentence transcription test (STT) is meant to measure how interference can change the dynamics of *verbal working memory*. This is done by performing the task of transcribing (verbatim) a sentence under several conditions. The first condition involves listening to a sentence with no other stimuli present and writing a transcript of it as accurately as possible. The second condition is the same in all aspects but with the sentence followed by sound similar to human speech but is unintelligible. The third is a sentence but with a distracting sentence immediately following the, to be transcribed, sentence. And the fourth, distracting words following the, to be transcribed, sentence.

The purpose of this type of test is to pinpoint some specific characteristics of information processing of verbal items, as it relates to limitations of working memory and executive function. This type of testing might help us to understand if there is a difference within the general population when it comes to the interaction between long- and short term memory, and the implications on performance this might entail. To answer these questions, the STT is accompanied by an additional preceding test called the *response inhibition test* or RIT, which is meant to test the ability to inhibit a *prepotent response*, also called *response inhibition* or *impulse control*. This was done using a Go/NoGo procedure as will
be described in the methods section. This test is motivated because of claims that response inhibition is related to interference control (Friedman & Miyake, 2004; Anderson, 2003) and other aspects that pertains to the subject of cognitive adaptability (Congdon et. al., 2012). The connection to inhibition has notably been posed by Anderson (2003) to be related to how we control interference from information stored in Long term memory, LTM.

Another purpose is to evaluate if it is feasible to use this type of testing in general, and to discuss the challenges that emerge when interpreting this type of data.

Theoretical models for working memory

Verbal WM is a sub-category of WM that specifically deals with information that can be expressed verbally such as words, digits, and letters. It encompasses both conscious and unconscious treatment of perceived written and spoken verbal information, along with the ability to produce speech (Caplan & Waters, 1998). Our WM in general, is defined by the fact that it is a limited capacity system (APA, 2020), meaning that the amount and quality of information it can hold is limited (Miller, 1956) and fragile (Reisberg, 2019). The modern and heavily influential model of WM can be attributed to the seminal works of Baddeley and Hitch (1974; 1994), and Baddeley (1992; 2000). In this model of working memory, we circumvent the limitations of WM by the use of central executive control processes (Atkinson & Shiffrin, 1968; se. Daneman & Carpenter, 1980) to help us deal with real-life situations of retaining and manipulating information in mind. The term “central executive” is in Baddeley and Hiches WM model not dissimilar from the general idea of executive control and functions, the difference being that the “central executive” is specific to WM whereas executive functions/control refers to most aspects of cognition involved in planning, cognitive flexibility, problem solving, attentional control, inhibition (APA, 2019) etc. The central executive, following Baddeley and Hitches model is a system of
control that utilizes information that is being handled by slave- or sub systems. To put this into perspective, the task of reading, for example, usually engage subvocalization, which in turn is involved in establishing an articulatory rehearsal loop. The central executive is then in charge of engaging strategies to maintain (keep information fresh) this information so it’s not displaced by other information. The more this information is rehearsed the more stable and memorable it becomes. In general, all auditory information that is attended to enters the phonological buffer. Which is a passive representation (an “internal echo” Reisberg, 2019) used for holding information while its being processed (Baddeley & Hitch, 1994). Baddeley’s (1974) model posit that WM will exhibit different functions depending on the load. Smaller loads tend to be handled by a phonologcal loop (PL), while larger loads tend to engage central executive functions. What constitutes small, contrary to large, loads is both a question of quantity but also association. For example, digits are commonly understood as verbal items that are handled by our verbal working memory, but sequences of digits usually require several control processes to be retained (chunking and rehearsal in particular), due to the sequence usually not being handled semantically (Jones & Macken, 2015) and approximately (since they often need to be remembered explicitly). And since we are accustomed to retaining random sequences of digits or letters, we have reinforced cognitive mnemonics for this purpose (Jones & Macken, 2015). The same is not necessarily the case when it comes to retention of sentences. In many cases we rely on comprehending written and spoken material based on the associations we have drawn through experience. We also need to understand the sentence configuration (the syntax) of the words in order to correctly understand the meaning of a sentence (Hagoort et. al., 1999). This also requires holding on to the information, since early words in the sentence is of direct importance due to their implications on latter words. To illustrate the difference between these types of processing demands, let’s say you are told to remember the sequence 1, 7, 8, 9, you would probably do so by means of rehearsal
and chunking. However, if you are told to remember 1789 because that was the
year the French revolution begun, you would most likely associate this number
sequence to many other things (stored in LTM). In the case of sentence
comprehension, when we hear a sentence the phonological buffer activates and
keep the contents in the beginning of the sentence intact until the entire sentence
is finished, we are then able to dedicate higher order processing to understand the
sentence.

Historically, not all models of language processing have relied heavily on the
concept of phonological short-term memory (PSTM) (Jacquemot & Scott, 2006).
Baddeley’s original short-term memory (STM) concept has since its conception
been revised and expanded to include an additional component involved in speech
processing in particular, the episodic buffer (EB) (Baddeley, 2000). The EB
integrates the information that is cycled through the PL and the visuospatial
sketchpad (a visuospatial modelling feature for information) with information
retained in LTM, and infuses it with the sense of time (Baddeley, 2000). Which
also helps facilitate the chunking of information (Baddeley, 2000) as proposed by
Miller (Miller, 1956). In the context of PL, Caplan and Walters (1990) has
described that the PL does not seem to be an important dependency for the initial
understanding of a sentence, or what they call the first-pass parsing and
interpretation. But is important for maintaining the sentence in memory to
accomplish the task if delayed. In other words, the PL by itself does not contribute
to increased proficiency in parsing and interpreting, but can compensate for a low
capacity WM by facilitating the rehearsal processes (Caplan & Walters, 1990).

In this case, the participants are transcribing sentences, and thus, additional
components are added in comparison to more isolated WM tests like the digit span
test. These components are: syntax (grammatical structure), semantics (the
meaning derived from a sentence), lexical (word composition, length, syllable
count etc.) and properties of the sentences. These components also interact and require some sort of interference control, as will be described in a later passage.

It has been noted in discussions regarding the implications of syntax and memory that syntax and STM is intertwined (Caplan & Waters, 2013). And some models have integrated syntax with short term working memory (ST-WM), and conceptualized syntactic comprehension into two main aspects: parsing (analysis of sentence composition); and interpretation of the sentence (Caplan & Waters, 2013). However, following Caplan and Waters (2013) argument, some models that fit parsing and interpretation into ST-WM does not fully account for the aspects of skill accumulation, or the interaction between LTM memory and WM.

Endorsed by many philosophers and psycholinguistics is the view that language stands as a fundamental tool for higher order cognition (Carruthers, 2002). It has been concluded in several studies that language impairment is connected to overall cognitive function using IQ as a measurement (Gallinat & Spaulding, 2014). Furthermore, the relationship between language and WM has been proposed to be an important predictor for cognitive function (Baddeley, 2003). And in comparative studies across different language systems, the various ways language is structured has been indicated to influence cognitive function (Brown & Lennenberg, 1954).

Following studies about the relationship between intelligence and specific language impairment, SLI, in children, it has been difficult to pinpoint and generalize what aspect of intelligence that is likely to be affected (Botting, 2005). Following Bottings (2005) discussion regarding this, it is argued that SLI can affect several cognitive functions in a dynamic and perhaps unpredictable way. Based on this finding, the development of new testing tools of language can be useful for breaking down language into several processing systems, to further investigate how differences in language processing can interact with other aspects of cognition.
Another way to look at language processing as it pertains to the testing in this case, is through the theoretical viewpoint of *echoic memory* (EM). EM was first described by Niesser (1967), and refers to the ability of withholding auditory information, as a direct and accurate representation of recently heard information or stimuli. Its capacity range between 250ms-4s, and its most vivid report occurs at about 1s after auditory stimulation (Niesser, 1967). The notion of echoic memory is a part of what called *perceptual span*, meaning, that perceptible information from sensory systems are not processed as a whole but rather processed through several *modal-specific* buffers (Solso et. al., 2005). These buffers are proposed to filter but also facilitate higher order cognitive tasks. To understand a “word” in context, regardless if it is read or listened to, requires relevant information to be held and then combined with other relevant information. Some studies of echoic memory have indicated it to be particularly sensitive to subsequent auditory information, and can be erased altogether by interfering processing of verbal information (Bryden, 1971). This aspect will conceivably be tested using the STT.

In general, the relationship between language and cognitive performance is not fully understood but has been implicated to be relevant for understanding how different pathologies (for example, *temporal lobe epilepsy*, Carpentier, et al., 2001) and language systems (Carroll, 1964) affect cognitive development (Carpentier, et al., 2001). One example of how language processing is applicable to cognitive functions in general, is that syntactic processing is indicated to rely on WM, and compete for resources that are shared with other cognitive processes (Gordon et al., 2002).

Since this study is dealing with spoken material and the ability to rehearse information. It’s of importance to note that it has been observed in *neuroimaging* (Positron emission topography, PET, in this case) studies (Jonides et. al. 2005), that the same regions of the brain are activated during the internal rehearsal of
information as when this type of information is produced (i.e. speaking out loud and to yourself is similar on a neural level). As is relevant in this case, we generally find it difficult to internally rehearse a sentence and listening to another sentence and keep a record of both, indicating that our mental resources are limited (if not trained). This has been studied using concurrent articulation tests (trying to remember a series of numbers while reading something unrelated) and when put under these circumstances our working memory capacity (WMC) can be reduced with as much as a third of its usual storage capacity (Chincotta & Underwood, 1997). There is also research that indicate that two different processing demands (storage and processing) interfere with each other and similarly impair memory (Oberauer & Göthe, 2006). The subject of language processing has also been researched from the perspective of attention. Neuroimaging studies conducted by Shaywitz et. al. (2001), used fMRI to isolate different neural correlates, responsible for the processing of at least three associated attentional functions: selective attention, divided attention and executive function. And found that the activation that arise during language processing is system wide and encompasses widely distributed cortical regions that are distinct but at the same time integrated. For the STT tests, it is perhaps the aspect of selective attention, and executive function that is mostly relevant, since selective attention is the process of focusing on one stimuli while ignoring another and executive functions is reciprocally related to both attention and language processing (Bialystok, 2015).

A common stance within the psycholinguistic literature is that verbal information is processed by different areas of the brain working in parallel. Many more areas seem to engage when attending to verbal information in comparison to strict perceptual processing. This has been indicated by some fMRI studies (Carpenter et. al., 1999; Gabrieli et. al. 1996). Language processing does also seem to adhere and at the same time transcend input modality. When it comes to
the difference between the processing of listened to, as opposed to read verbal information. Studies have been able to locate the specific loci that seems to differ between processing auditory words (heard words) in contrast to reading printed words (Constable et. al., 2004). Auditory verbal information seems to engage the primary auditory cortex and bilaterally across the superior temporal gyrus (Constable et. al., 2004). While printed words activate other parts of the brain. However, significant overlap has been observed in the inferior frontal gyrus regardless of modal input type (Constable et. al., 2004). When it comes to processing the sentence post-perceptually it has been observed that increased activation across several sites of the left hemisphere occurs depending on if the sentence has an *object-relative clause* structure compared to *subject-relative clause* structure. The way different clause types increase activity, seems to be modal general which does further indicate that language processing, in part, transcend modal specificity. By introducing different kinds of distractions the STT can therefore test what type of stimuli interfere with the task of transcription. And if this affect the ability of the central executive or mnemonic strategies to function. The distractions during the STTs are also meant to produce automatic, on-the-fly, judgement that presumably require some sort of executive control over attention and inhibition to withhold the information to accomplish the task of transcription. Also, if information is processed with perceptual or lower level cognition. Does this pose a different dynamic, as opposed to the same information being processed semantically? For example by associating words and meaning with previous information in LTM. Studies has indicated that only relying on phonologically based shallow processing hinder performance in memory based tasks as compared to more semantically based *deep processing* (Fujii et. al., 2002). By using PET and MEG this has been hypothesized to be due to the latter engaging various parts including the hippocampal formation (Henke et. al., 1999) of the brain in comparison to the former (Fujii et. al., 2002).
To better understand the interaction between parallel processing demands as proposed to be engaged by language processing. It might be relevant to understand how information is retained or decay due to the effects of time, repetition, and interference. Time dependent effects affecting the immediate recall are known as *serial position effects*. One of the earlier observations of serial positioning effects on recall was described by Murdock (1962), where the retention of information follows a U-shaped curve, where information presented earlier in the sequence is subject of *primacy* and information presented last in sequence is subject of *recency*, both effects contributing to higher retention of items placed near the beginning and end in a presentation sequence. The proposal by Murdock (1962), was that the effects of primacy is due to the items having more opportunity to be rehearsed and encoded into LTM, while the last items has less time to decay. Work by Glanzer and Cunitz (1966) later described the recency effects as only being temporary, its only if the items in a list is to be recalled directly after presentation that the effects of recency is potent. When the information goes without recall for 30 seconds or more, recency is no longer a potent effect (Glanzer & Cunitz, 1966). Another finding is that under certain conditions the middle position can be retained due to the *von Restorff effect*, which describes the effect of higher retention of information that is novel in an otherwise more homogeneous series of information (Restorff, 1933). On the subject of interference, there are dualistic models of STM – LTM interaction, in particular the one developed by Waugh and Norman (1965). This model would not describe unrehearsed information as decaying, but rather, are displaced by other information. However, the effects as described above is thought to be tendencies of neural information storage (Atkinson & Shiffrin, 1968), not control mechanisms. In this study, the STT data might reveal a pattern of differences in controlling the effects of serial positioning, which might be useful for understanding if there is a difference in the utilization of control processes within the population. Particularly *retrieval control* and *comparison*, as proposed to be central processes by Atkinson and Shiffrin (1968).
One aspect of executive control as proposed by Anderson (2003) is inhibitory control. This prospect sees that we need to inhibit the response to a memory much the same way as we need to inhibit a prepotent response. To put this in the context of this case, when perceptual storage diminishes, we need to remember or retrieve information. But in order to differentiate between the relevant information and previously stored information we have to engage inhibition to block some information while allowing other information to pass. What further complicates how this inhibitive function operates is the added notion of associative strength to the retrieval cue. In essence, if an association to a cue has been repeated or rehearsed, when switching the association but not the cue, one tends to inherit the strength of the previous association, which contribute to the interference effect on retrieval. This notion was discredited following Andersons (2003) testing of different cue competitions. However, Andersons (2003) method primarily dealt with associations on a semantic basis, like categorical grouping (two red things, as compared to two vegetables etc.). In the case of the STT, items and cues are part of a sequence that conceivably implement yet another type of association of syntactic and lexical similarity and comparison. Therefore, an argument of this type of test is that it tests the proposed mechanism of inhibition on multiple channels of processing.

Another central processing technique is rehearsal of the information, in this case, the distractor stimuli presented during transcription, might give insight into what type of distractor might disrupt the ability to rehearse the information, and interfere both retroactively and proactively. The Brown-Peterson technique (disruption of rehearsal due to task of recalling items in reverse) has been used to show the destructive effects of rehearsal tampering (Chincotta & Underwood, 1997). The STT might yield similar results by disrupting the process of rehearsal.
In summation, following the theories of how we retain and process verbal information. The STT is meant to produce a real-time pattern of these aspects which can propel future modeling of neural information processes in general.

**The hypothesis and research question**

The STT is meant to give insight into the nature of information processing and memory retention. To understand the tests validity for this purpose, the test and data is to be critically examined to evaluate if transcription data can be useful for future modelling of memory and related control processes. To drive insight into this inquiry, the analysis is aided by a post hoc “proof of concept” *model of interference* that is based on established principles of memory retention, interference and serial positioning effects. Formally the research questions are posed as:

1. *Are the differences in distractor stimuli significant compared to control?*

   The variables of comparison are the grand means of the performance (the degree of error) for STT2, STT3, STT4 compared to STT1(control). How this variable is calculated will be described in the methods section.

2. *Are the differences in distractor stimuli significant compared to each other*?*

   *For example: Are STT2 (distracting phonemes) significantly different then STT3(distracting sentence).*

3. *Which distractor stimuli is the most potent distractor?*

4. *Is the variance within the samples attributable to age or sex?*

5. *Is STT performance related to RIT performance?*

6.[Post hoc]. *Does the STT test results reveal patterns that pertains to established principles of memory retention?*
Method

The testing used a within-group repeated measures design, and analyzed using one-way and two-way factorial ANOVA. This design is beneficial since the individuals pose as their own control. Therefore, the variability could more accurately reflect if the manipulation between measures has an effect on the individual. Additional analysis for examining correlation and difference with regards to age or sex, was done using *t*-tests, and *Pearson correlation*. The performance of the RIT-test was calculated from the probabilities derived from the *z*-score distribution for the response times (confirmed to be normally distributed using the *Shapiro-Wilk test*). Since the amount of correct “inhibitions” by the participants did not follow a normal distribution, a second method of calculating RIT performance was used (this will be described in the appropriate section). All tests were conducted using SPSS 26.

A complimentary model of interference was constructed to compare where bias is likely to influence the performance on the STT. In an endeavor to aid the analysis and interpretation of the types of response patterns observed. Furthermore, the model of interference is meant to be a *proof of concept* for a way to evaluate through modeling if the errors are attributable to known tendencies of serial positioning and interference effects (this will be described in a later section).

Participants

This test was meant to elucidate the dynamics of verbal working memory as it relates to the general population. And was therefore not directed towards a specific population. The test was distributed through several channels based on convenience, snowballing and the websites: accindi.se and facebook.se. The testing period started at the 8th of December 2020, after a period of pilot testing. The goal for sample size was 10-30 participants. At the end of the testing period (31th of December 2020), 17 (9 females, 8 males) *valid* (based on exclusion criteria as described in a later segment in the methods section) participants with
the average age of 30.50 ($SD = 9.50$, range 14-49) were selected for further analysis. Using the information page at the start of testing (Appendix C), all participants were informed about the purpose and subject matter of this study. And to participate, informed consent had to be agreed on prior to participation. The participants were informed that no personal information would be gathered except for age and gender. Verbal consent was acquired for the two minors (aged 14 and 16), from their caregiver.

No particular measures were taken to ensure participant homogeneity with regards to aspects such as neurological abnormalities, left-right handedness, education, proficiency in using and communicating/typing using the QWERTY keyboard layout (which was used as a reference for setting the time-limit of the answering). With regards to the latter, typing speed is an important factor to consider. Studies have demonstrated that factors that affect typing speed are keyboard size (Sears, et al., 1993). This does imply that typing on a cellular phone or tablet does decrease typing speed. In an attempt to circumvent this variable, the link to the study specified that the participant was required to use a personal computer (PC) or laptop, that both use a larger keyboard interface. If the participants followed this however, was not controlled by any other measure. Another variable with regards to typing speed is age, while there is evidence to suggest that the motor aspect of typing increases during the teens and later slows down as we age, the overall typing speed does not slow down due to older typists being able to more accurately approximate letter presses in advance (Salthouse, 1984; Bosman, 1993). But a reminder is that typing is a skill that develops through training. From the perspective of the sociological concept of digital natives, one could assume that the probability of one acquiring a high level of typing skill, should be greater for generations that grew up during 80s and afterwards, due to the digital revolution. However, in the absence of controlling for typing speed, no inference has been drawn based on age and typing speed. For reference, when
compared to handwriting, using a keyboard to type is quicker on average (Horne et al., 2011), but when it comes to cognitive load and memorization, handwriting might be advantageous (Shibata & Omura, 2018). Therefore, it is difficult to assess which of these techniques should be considered to be more appropriate. But given that this test was conducted online, typing was considered to be the more feasible option.

No extra measures were taken to control the level of language ability. Usually verbal ability is controlled for by ensuring an equal level of verbal ability as measured by tests like the WAIS-IV verbal test (2008). In this case, such a test was not conducted due to the importance of brevity. To compensate for this however, the sentences was comprised in such a way, that variance based on differences like vocabulary should not be a determining factor. This was done by using very common words.

Performing several tests like these in a row can (conceivably) cause fatigue, increasing the drop-out rate, therefore, brevity, as mentioned, was valued. And the test was constructed to take no more than 10 minutes (which ended up taking between 7-8 minutes for most participants).

Procedure

Due to the world-wide conditions during December 2020, this test was conducted online. It was written in html (javascript & Php) and took the form of a webpage. There are few advantages other than practicality for using online-based testing of cognitive function. Its most notable disadvantage being the lack of control over the testing environment and procedure as well as not being able to ask the participants in person about the testing.

The participants were first greeted with an information page, describing the study, its subject matter and purpose (see Appendix A, for reference). The tests were given in two different orders. In place was a randomizer, which
automatically gave each participant a 50/50 chance of getting one of the two different tests. In the end, 7 participants from test sequence 1 (T1) and 10 participants from test sequence 2 (T2) was selected for further analysis. Both T1 and T2 started with the Response inhibition test (RIT), followed by four training transcription tests, which was then followed by the actual tests in the order, STT1; STT2; STT4; STT3, for T1, and STT1; STT3; STT4; STT2, for T2. Each STT pose a different distractor stimulus, and only one sentence per distractor had to be transcribed. In both T1 and T2, a question was posed after the testing, asking: Which test did you find the most difficult? After which, the participants were sent to a thank you screen. On which, the participant could comment on the test and give their email, if they wanted a copy of the report after its completion. Session-logging was collected in an attempt to check if the responder “cheats” during the test. The session-logging records the amount of times the page has been viewed by the user, if a session number is higher than 1, the participant has refreshed the page.

Test description and data variables

Response inhibition test (RIT) – This test required the participants to complete a 50 letter slideshow, with the instruction to press on all letters except the letter “X” as quickly as possible. All slides were uniform black frames with a white capitalized letter centralized in the frame (Figure 1.). The participants used the mouse to press on all letters.

*The participants followed the instruction to press on all letters that appears except the letter “X”*

![M](image)

Figure 1. The slideshow showing a letter centered in the screen as it would have appeared during testing
Two variables were collected; the first, response time (RT in milliseconds), was a measure of the time interval between each keypress. These intervals were then added together to produce a mean RT, which was later used to calculate a score; the second variable was the amount of errors, measured as how many times the participant pressed on X. In total, 7 Xs was presented in all tests. The placement of these was the same for all tests. And was spaced with both small and large distances from each other. Of the total of 50 letters, X appeared on pages: 13, 16, 20, 27, 36, 41, and 46.

*(All of the following STTs consisted of one transcription task each)*

**STT1** – Hearing a nine-word sentence with no distractor present. That activated when the participant pressed the “*Tryck för att lyssna*” (press to listen)-button (Figure 2). After the sentence was spoken, a text input field emerged where the participant wrote the transcription. After a set interval (25 +/- 3 seconds depending on the length of the sentence) after the text field emerged, the answer was sent to the server, and the participant was directed to the next page. All STTs was preceded with a written instruction.

![STT1](image)

**Figure 2. The picture shows the page for STT1 as the participants would have seen it**

**STT2** – Same as STT1 but with “speech-like” distractor sounds after the first sentence. Disturbing stimuli was created by slicing words into phonemes and then scramble them. Which produced nine sound bursts with the same rhythmical pattern as the spoken sentences. The phonemes were identified using a visualization of the audio spectrum (in audio-editing software). Studies
have indicated that nonsensical sound compared to intelligible speech have different effects on recall, which has been hypothesized to be due to the former being encoded and processed as sounds and not integrated verbal units (Rowe & Rowe, 1976).

**STT3** – Same as STT1 but with a distracting second sentence following the to-be-transcribed sentence. Speech sounds has been indicated to be more destructive for recall of verbal information (Rowe & Rowe, 1976).

**STT4** – Same as STT1 but with distractor words following the to-be-transcribed sentence.

Constructing the sentences - Sentence composition

The sentences were all spoken in Swedish since that is the participants’ primary language. Only one announcer was recorded to use in all of the sentences. The voice was that of a 34-year-old male (it seems not to be of significant importance to have varying presentation voices from the point of view of EM, Watkins & Watkins, 1980), with a clearly articulated and fluent pronunciation. The rhythmical flow of words was the same in all sentences, consisting of four words, pause, two words, small pause, three words. Totaling 9 words across all of the tests. To control for semantic bias, the sentences were generated using a syntactic structure algorithm developed by Chomsky (2002; based on earlier work by Chomsky). And consists of two parts: a noun-phrase and a verb-phrase, which in this case was expanded with two added adjectives, a preposition, adverb, and a second noun. Based on this algorithm 20 sentences (see Appendix B for reference) was randomly generated (using the Python programming language) from a subset of words, ensuring the same grammatical (syntactic) composition in every sentence.

The argument for why this method was used, was to try to eliminate or reduce the influence imposed by semantic reasoning. To exemplify such a reasoning
ability, think about the sentence: “It’s important to drink at least two litres of ____ every day”. Most people would without great effort infer that the missing word is water. So by introducing semantically coherent sentences, there is reason to think that the WM dynamics is aided by experience and intuition. And since those are not variables incorporated in the test, they need to be eliminated as far as feasibly possible. Another reason is to remove the human error in judgement, by the author, of trying to pick sentences of equal transcription difficulty.

Sentence complexity

Sentence complexity is a multifaceted concept. From the point of view of sentence comprehension, complexity can be defined by the number and type of clauses it contains (Norman et. al., 1992). Multiclause sentences are a type of complexity arising from combining many clauses into one sentence. In addition, some clause types are more difficult than others, particularly from the perspective of WM. For example, clauses that are embedded in the sentence that are important to understand in detail, in order to understand the relationship between the items in the sentence (Norman et al. 1992). Which requires multiple iterations of analysis before the sentence is correctly understood. In this case, comprehension is secondary to the superficial withholding of information. This is because the participant is directed to give an accurate transcript, not an accurate understanding of the information.

Another type of complexity is sentence length. Since we know that our WM is limited, all sentences were limited to 9 words. In total, 35 words were used to generate the sentences. These were picked from a vocabulary consisting of the 1000 most common Swedish words (Stam, 2013) in their respective grammatical grouping. One reason for using common words is based on the research finding that elaborate and exotic words might stick more than others (Craik & Tulving, 1972; 1975).
The vocabulary consisted of:


Shared lexical elements*:

“Utanför”, “Framför” (2 of 3)

*Criteria for lexical similarity; the word element has to be at least 3 or more adjacent letters.


Shared lexical elements:

“husets”, “landets” (2 of 7)

2b. Adjective: 7 items: “sista”(last), ”första”(first), ”äldsta”(oldest), ”yngsta”(youngest), ”minsta”(smallest), ”långa”(tall, long), ”korta”(short).

Shared lexical elements:

“sista”, ”första”, ”äldsta”, ”yngsta”, ”minsta” (5 of 7)


[3]. Verb phrase: 3a. Verb ”sågs”(seen), ”skrevs”(written), ”kördes”(driven), ”lagades” (repaired, fixed), ”öppnades”(opened), ”arbetades” (worked), ”tvättades”(washed), ”lånades”(borrowed), ”hämtades”(picked up, gathered).
Shared lexical elements:

"kördes", "lagades", "öppnades", "arbetades", "tvättades", "lånades", "hämtades". (7 of 9)


Example of a sentence generated from this vocabulary and used in one of the training STTs:

“Inuti deras minsta tåg hämtades långsamt skolans korta balkong”

<table>
<thead>
<tr>
<th>Inuti</th>
<th>Preposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>deras</td>
<td>Article</td>
</tr>
<tr>
<td>minsta</td>
<td>Adjective</td>
</tr>
<tr>
<td>tåg</td>
<td>Noun</td>
</tr>
<tr>
<td>hämtades</td>
<td>Verb</td>
</tr>
<tr>
<td>långsamt</td>
<td>Adverb</td>
</tr>
<tr>
<td>skolans</td>
<td>Article</td>
</tr>
<tr>
<td>korta</td>
<td>Adjective</td>
</tr>
<tr>
<td>balkong</td>
<td>Noun</td>
</tr>
</tbody>
</table>

The sentences used in all the testing

STT\textsubscript{1} Training

“Framför stadens långa maträtt”

STT\textsubscript{2} Training

“Inuti deras minsta tåg hämtades långsamt skolans korta balkong”

STT\textsubscript{3} Training

(Syntactically incoherent)

“Långa framför deras snabb bil störst bil hennes öppnad”
STTPRELIM

“Inuti husets minsta katt tvättades långsamt hennes äldsta hund”

STT1

“Framför hennes minsta tåg tvättades långsamt hennes yngsta maträtt”

STT2

“Inuti stadens sista bil tvättades snabbt husets sista pojke”

STT3

“Utanför landets äldsta hund kördes snabbt deras första flicka”

And

“Framför hennes korta adress hämtades snabbt landets långa pojke”

STT4

“Utanför hennes sista pojke träffades snabbt stadens minsta katt”

And

“Antalet, program, samhället, poäng, rad, sådant, höll, kallade, företaget”

Sentence complexity based on symbolic cognitive models

Semantic variation was excluded per design with the use of identical syntactic structure. The semantics, although nonsensical, still follow the same associative complexity across all sentences. This was ensured by adapting them into a theoretical model for verbal representation of knowledge. In this case the sentences were analyzed using the symbolic theoretical model: adaptive control of thought (ACT-R), developed by Anderson et. al. (2004).
Example of a randomly generated sentence:

“Inuti deras minsta tåg hämtades långsamt skolans korta balkong”

Translated to English:

“Inside their smallest train the balcony was picked up slowly”

![Diagram of sentence structure]

Figure 3. An illustration of sentence complexity based on ACT-R (Anderson et. al., 2004)

Semantic complexity is in this case based on amount of, and branching from node structures. In the example above the first layer of complexity is “the train picked up the balcony”. The second degree of complexity is added through attributes: “inside the smallest train the balcony was picked up slowly”. The third degree of complexity is adding ownership. When analyzed in this manner, all sentences are equally complex, semantically. The current stance within the scientific literature is that there is a delineation between symbolically based models and connectionist models, like the parallel distributed processing (PDP) model. But since connectionist models are difficult to adapt to word level associations, a symbolic model was deemed more appropriate.
Data processing

Inhibition performance

The way to calculate performance in response inhibition is by extracting the metric of stop-signal reaction time (SSRT) (Congdon et. al., 2012). The scientifically common way to test this type of inhibition is through using the stop-signal or Go/NoGo tasks (Congdon et. al., 2012). There is however no unity on how to calculate the SSRT, particularly because outliers might skew the data, leaving a less reliable average. In a discussion on the topic of calculating response inhibition, Congdon et. al. (2012) recommends the use of standards derived from the currently collected testing data. Which can then be used to evaluate the performance.

The inhibition test used in this study is a Go/NoGo-type test, meaning that it requires the participant to respond to one type of stimuli and not to respond or inhibit that response for another type of stimuli. The theory that dictates what actually constitute performance is the horse-race model (Boucher et. al., 2007), which in simple terms explains the process of inhibiting as response as a competition between two signals. Where one signal carrying the response action (go-action), competes with another signal that inhibits action (nogo-action). There are factors that can modulate this type of performance, perhaps due to desynchronization of circadian rhythm (May & Hasher, 1998; has been contested by Bratzke et. al., 2012), lack of sleep (in particular for younger participants) (Wilckens et. al., 2014), and the use of CNS stimulants (Tannock, et. al., 1995). These potential variables of disturbance were not controlled for in this study.

RIT-score calculation

The Shapiro-Wilk test was used to test if the variables: response time, RT, and the number of correct inhibitions, was normally distributed. For the RT variable the significance was .942 ($\alpha = .05$), which indicates that it is normally distributed.
Whereas the correct inhibition variable had a significance of .024 (α = .05), which indicates that it is not normally distributed. From figure 4, the correct inhibitions are skewed to the right, there is however no way to estimate where the top tapers of.

The first factor RT is calculated using the z-score distribution. Which in gives us the probability of scoring a higher RT (lower is better) based on the performance data distribution. As for participant 1, which had an approximate RT of 800ms, was about 50ms quicker than the sample mean of 852.94ms, this places him or her -0.34 standard deviations or z-values (z) from the midpoint of the z-distribution. Referring to the z-table, the probability of anyone responding slower than participant 1 was 63.2% (P(x>Z) = 0.63202). Participant 2, with an RT of 880ms was placed closer to the mean with P(x>Z) = 0.43158, which indicates a probability of a higher RT of 43.16%. These probabilities were then added to the second factor: the amount of errors. Participant 1 with 6 out of 7 correct response inhibitions scored P_{RI}(x<Z) = 0.6554, which is the probability of anyone making more errors. When adding these two probabilities together P_{RT}(x>Z) + P_{RI}(x<Z) equals the RITSCORE (0.63202+0.6554 = 1.28742). Participant 2 with 7 (maximum amount of correct responses) out of 7 inhibitions, but with a slower RT, scored: 0.43158+0.84566 = 1.27724. To mention is that the RT or response time, is not actually a metric of response latency nor is it stop-signal reaction time (SSRT), but rather the interval between clicks as determined by dividing the total amount of time the test took with the amount of slides. The time it takes for the webpage to display the slide and send the information to the server will add additional time to this number. If this has changed the RT in any substantial way, is difficult to evaluate. This added variable was assumed to be constant across the testing.

A second type of calculation was done based on multiplying the Z-score probability of RT with the amount of correct inhibitions.
As can be seen in figure 4, a significant proportion got 6 or 7 out of 7 in the inhibition test indicating a higher probability that the top of the scale is not the likely top in performance. Usually this type of Go/NoGo testing involves many more slides then the 50 used in this case, together with a dynamic modulation of the presentation latency based on the response time of the participants’ previous performance (Congdon et. al., 2012). The argument for using only 50 slides was to minimize dropout rates.

Figure 4.  Diagram showing skewed distribution in the RIT-test

Transcription performance and coding

The performance was determined using three types of metrics. One based on the severity of the error, measured in total amount of letters being correct divided by the maximum amount of letters that can be correct.

Example:

Transcription of the sentence (STT3):

_Utanför landets äldsta hund kördes snabbt deras första flicka_

Contains 53 letters in total. And was transcribed by a participant as:

_Utanför stadens största hund hämtades snabbt landets långa flicka_

Whereas, 23 of 53 letters were correctly transcribed. Equaling 43.40% correctly transcribed letters. These percentages were then used in an ANOVA analysis, to
measure the significance of the variance between tests and sample groups (age, gender, RIT-performance). This metric however does not tell us what type of error was prevalent, it only tells us the degree of error as compared to a correct verbatim transcript. To estimate the type of error we first coded the data based on five different tendencies, as follows: DNUI = did-not-understand-instruction; DNFT = did-not-finish-typing; LW = lost words; MW = missed-words; RW = replaced-words; MWRW = missed-words and replaced-words; NE = no-errors. It’s important to note that missed words could be due to reasons that have different implications and causes. The data reflects that lost words are either due to the participant not having finished typing or that the participant was not able to finish the sentence due to them not retaining the last words in the sentence. The way to determine this was by checking for unfinished words, for example:

The correctly transcribed sentence:

“Inuti deras minsta tåg hämtades långsamt skolans korta balkong”

Was transcribed by participant 1 as:

“Inuti deras minsta tåg hämtades långsamt skolans kort”

The abrupt end of the last words indicates that the participant did not finish typing, and was therefore given the code: DNFT (did-not-finish-typing).

Whereas participant 4 transcribed the sentence as:

“Inuti deras minsta tåg”

In where it is reasonable to assume that if the participant had (confidently) known the next word he or she would have begun typing it. Another clue is the shortness of the answer, if the lost words would be due to unfinished typing, it would most probably be nearer to the end of the sentence.
The difference between missed words (MW) and lost words (LW) is that missed words is if a word is missing in the sentence, while LW is the words in the end of the sentence that are absent and not subject of DNFT.

Not surprisingly the code: DNUI (*did-not-understand-instruction*) was more common for the earlier tasks. But did not extend throughout the testing.

Exclusion criteria was based on DNFT across all tests with no prevalence of other errors.

**Model of interference**

The next stage of data processing was to create a model of interference, which is meant to compare and estimate the effects of interference on the observed data to aid in interpreting if transcription data is sensitive to the effects of interference and serial position effects.

The process begun by placing the presented information on an \([ m \times n ]\) matrix, comprised of nine \(n\)-vectors (columns), with each presented word \([a_{m,n}]\) placed in its own cell, from left to right. Then, each sentence was placed in its own \(m\)-vector (row), in the order they were presented to the participants.

For example: \(a_{4,5} = \text{tvättades}, a_{7,5} = \text{snabbt},\) and so on.

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
<th>N9</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Framför</td>
<td>stadens</td>
<td>långa</td>
<td>matriätt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>Inuti</td>
<td>deras</td>
<td>minsta</td>
<td>tåg</td>
<td>hämtades</td>
<td>långsamt</td>
<td>skolans</td>
<td>korta</td>
<td>balkong</td>
</tr>
<tr>
<td>M3</td>
<td>Inuti</td>
<td>husets</td>
<td>minsta</td>
<td>katt</td>
<td>tvättades</td>
<td>långsamt</td>
<td>hennes</td>
<td>äldsta</td>
<td>hund</td>
</tr>
<tr>
<td>M4</td>
<td>Framför</td>
<td>hennes</td>
<td>minsta</td>
<td>tåg</td>
<td>tvättades</td>
<td>långsamt</td>
<td>hennes</td>
<td>yngsta</td>
<td>matriätt</td>
</tr>
<tr>
<td>M5</td>
<td>Inuti</td>
<td>stadens</td>
<td>sista</td>
<td>bil</td>
<td>tvättades</td>
<td>snabbt</td>
<td>husets</td>
<td>sista</td>
<td>pojke</td>
</tr>
<tr>
<td>M6</td>
<td>Utanför</td>
<td>hennes</td>
<td>sista</td>
<td>pojke</td>
<td>träffades</td>
<td>snabbt</td>
<td>stadens</td>
<td>minsta</td>
<td>katt</td>
</tr>
<tr>
<td>M7</td>
<td>Utanför</td>
<td>landets</td>
<td>äldsta</td>
<td>hund</td>
<td>kördes</td>
<td>snabbt</td>
<td>deras</td>
<td>första</td>
<td>flicka</td>
</tr>
<tr>
<td>M8</td>
<td>Framför</td>
<td>hennes</td>
<td>korta</td>
<td>adress</td>
<td>hämtades</td>
<td>snabbt</td>
<td>landets</td>
<td>långa</td>
<td>pojke</td>
</tr>
</tbody>
</table>

Figure 5. *Presentation sequence matrix for T1 at the end of testing.*

Then, each of these cells was assigned a weight. The weight is meant to predict what words are more likely to displace other words of the same type. In other
words, it is meant to give an estimation of where interference might influence the task of transcription. The weight was determined by combining five different variables:

**In-sequence bias** (quotient): The proportion of the number of places in which the word can coexist based on the syntactic structure of the sentence. Or from the point of view of the matrix above, the number of columns in where a word can fit. For example, in vector N1, all words can only occur in one place. Therefore, all words in N1 has a sequence bias of 1 out of 9. Except for its first occurrence in where it is 1 of 4. For every sentence that is being presented, the placement of the word in the sequence is reinforced (by adding 1/9), thus creating a larger impact the more sentences that is being presented. If the syntactic structure is reinforced i.e. that we know that two nouns is part of every sentence, there is reason to think that blanks or memory lapses will more likely be filled with guesses. Also, if a word-type can coexist in two places, the bigger proportion (2/9) of the total amount of words can be displaced. Therefore, the probability of the word being a source of bias increases due to it fitting in two places instead of one.

**Proximity bias** (integer): Is the placement in where each particular word appears in time (given by the row number). For example, the word: Framför, appears in row 1, 5, and 10. These numbers constitute the bias given by time dependant recency. In case of the word Framför, at depth M1 the proximity bias equals 1, at M4 it equals 1+4, at M8 it equals 1+4+8 and so forth. If a word is mentioned on several places in the same sentence, the proximity value is added two times for that row thus increasing the bias. Therefore, it reflects both time-proximity as well as availability.

**Lexical bias** (quotient of integer): Is the number of times the word shares a morpheme with another word. In row 1, we have three different words: Framför, Inuti, and Utanför. Giving us three words in total, but Framför and Utanför, shares the morpheme för, making them more similar to each other on a lexical basis then
does *inuti*. Therefore, *Framför* and *Utanför*, both receive 2 out of 3 in lexical bias while *inuti* receives 1 out of 3. This however is not the case before the M6 vector, and the first appearance of the word *utanför*.

*Size bias* (quotient or integer): Is the closeness the word has to the average word length in its respective category. In this case *Framför* = 7; *Inuti* = 5; *Utanför* = 7: giving us 6.3 (19/3). Both *Framför* and *Utanför*, receives 1 / (7/6.3) (9/10 or 0,90) (all values above the mean has to be divided by one in the numerator, or flipped 6.3/7, to represent how close the number is to the mean which is a maximum of 1), whereas *Inuti*, receives 5/6.3 (50/63 or 0.79). If the same word-type can occur in two places, with the existence of a new unique within-group word. The new word is counted to form the mean. For example, N2 and N7 represents two places in which, the word-type can coexist. At depth M3 (in sequence) and N2 a total of 3*[stadens : deras : husets]* and at M3 and N7, 2*[skolans : hennes]* coexist. Giving the mean average of (7+5+6+7+6) / 5 = 6.2.

*In-line bias* (integer): Is the distance of separation between the spoken word and the start of the transcription. Measured as an inverse of the horizontal axis of the matrix. The first words all receive 9, since they are temporally closer to the onset of the writing.

*Semantic bias* (unknown): Would be the associative closeness as determined by prior experience and intuition. This number has to be inferred by the actual data, since it’s difficult to reliably quantize in the absence of a larger corpus of data.

All factors are then multiplied to give each item its weight.
In summation, this modelling is based on combining the effects of: [1] Time, [2] Similarity, and [3] Repetition. All of these effects have been calculated by multiplying the proportions as they relate to each other. Both N vectors and M vectors are representations of time. The M vectors represent sequence or time on the temporal dimension of L1 (immediate proximity of output). And the N-vectors represent time on L2 (intermediate effects between testing). L1 is modelled to account for serial position effects. While L2 follow the availability of information and, less prominently, structure reinforcement. The model does predict that the effect of interference will accelerate the further into testing one goes.

To note is that vectors N4 and N9, both have very low weight. This is due to a combination of two factors; they are numerous; they do not share any lexical morphemes. They represent unique lexical entities, and are not variations of a common archetype like *framför* and *utanför* or *minsta* and *sista*.

### Results

Within-group ANOVA of performance between tests

A linear within-group ANOVA based on the performance metrics of the different STT tests indicated a significant difference between the different measures across all participants. Wilks’ Lambda = 0.191, F (4, 13) = 13.76, p = <.0001, $\eta^2_p = 0.809$. The between measures Tukey post-hoc test showed a significant difference between STT1 and STT2 ($p = .003$); STT1 and STT3 ($p <$
.001); STT1 and STT4 ($p = <.001$). However, no significant mean difference outside of the STT1 comparison was found. Indicating that no particular distractor is significantly more potent then another across the entire sample (all participants). However, as figures 5 shows, the mean follows different curves independent of testing order. This variance is most likely due to variation and not fatigue or proficiency. Even though the difference is not great, STT3 (distracting sentence) seems to be the most consistently difficult test across both groups. This is echoed by the answer by the participants to the question: “which sentence was the hardest one to transcribe”, in where 6 out of 17 answered STT3, (7 of 17 answered that they were all equally difficult, and only a handful was distributed over the remaining tests)

Mean of test performance from lowest to highest ($N = 17$)
STT1 – STT2 – STT3 – STT4

Mean of tests based on testing order 1 ($n = 7$)
STT1 – STT2 – STT4 – STT3

Mean of tests based on testing order 2 ($n = 10$)
STT1 – STT3 – STT4 – STT2

Figures 7. Diagrams showing grand mean performance for each STT test

Age-differences

Based on a one-way ANOVA, the only significant differences between the age groups; 14-21 ($n = 3$); 24-28($n = 6$); 34-35($n = 5$); 42-49($n = 3$), was in the
performance of STT4, $F_{obs} = 3.763$, $F_{crit}(3, 13) = 3.41$ ($p = .038$). Tukey-Kramer Post-hoc testing found that the difference between age groups was attributable to the difference between the group 14-21 ($M = 0$, $SD = 0$) and 24-28 ($M = 58.18$, $SD = 32.34$). And between 24-28 and 34-35 ($M = 18.19$, $SD = 19.86$). Additional tests were conducted to test the correlation between age and total STT-performance. For this test the *Pearson* yielded .20, which is not significant enough to reject the null hypothesis in any direction. Since this tests WM, and WM capacity is altogether known to decline with age, a negative correlation was expected. Further bivariate correlations were conducted with two-tailed significance hypothesis testing (since correlation in either direction is relevant). The only significant correlation was seen for STT3 (distracting sentence test) with .51 ($p = .045$). STT3 also correlated significantly with response time (RT) with -.58 ($p = .019$). These correlations however, have no value since they can better be explained by three factors: 1. Response time can variate due to the way the test was conducted, if it was conducted on a PC or laptop as opposed to a mobile phone, the participant would reasonably have used both a faster connection (mobile phones usually have higher latency) and a keyboard. And 2. Mobile phone use (which the participants were told not to use) reasonably result in lower STT performance due to the slower typing speed. And 3. The STT3 was the most difficult test in general which potentiates factors 1 and 2. To circumvent the issue of typing speed and latency, correlation was further investigated with the use of coding, in which three main codes were applied: Missing words (MW), lost words (LW), and replaced words (RW). Using these codes, both MW and LW was grouped together, and RW was alone. The correlation between age and MW&LW was .13 (not significant), and RW: -.27 (not significant). All correlations can be viewed in Appendix E. In summation, the data does not reflect any substantial age related effects. The only tentatively relevant correlation, although non-significant is the correlation between age and prevalence of replaced words, this will be discussed in the discussions section.
Sex-differences

To test the difference between sex groups on performance a t-test was conducted. Significant differences was only found for STT3, in where the female group \((n = 9, M = 42.14, SD = 8.75)\) on average performed better than the male group \((n = 8, M = 27.59, SD = 16.10)\) with a t-statistic of \(t(15) = -2.353, p = .033, d = 1.14\). This is clearly an artifact of random chance and sample size. And further t-tests on all other variables yielded no significant results.

Based on these results the conclusion is that neither age nor gender seems to be responsible for the observed variance. Although there might in fact be a difference within a general population, no such difference has been shown to have a significant bearing on the data in this case.

Response inhibition testing results

A one-way ANOVA comparing age-group and RIT-score yielded no significant results. Additional ANOVA tests on response time and error rate yielded no significant results. The variance in the RIT cannot be explained by age. Similarly, the ANOVA test results for sex did not yield any significant difference. To test the relationship to RIT-performance and STT-performance Pearson correlation tests was conducted on both RIT-score calculations which produced coefficients of 0.23 for the first method, and 0.24 for the second method, which is not significant enough to reject the null hypothesis. One further test was conducted by establishing two categories based on the top 3 and bottom 3 RIT-score results. These groups were then used in a 2x4 factorial ANOVA [RIT-high/low [2] x Performance [4]]. Which yielded no significant results based on performance of all STT tests. Therefore, the null hypothesis for RIT- and STT performance is retained, the RIT performance and STT does not seem to be related to the data in this case. All RIT correlation data with regards to both the type and severity of STT-performance can be viewed in Appendix E.
Further analysis

To interpret the transcription data from the perspective of interference, the testing results are presented alongside weighed vectors based on the interference effects as described in the methods section. A visual representation (heat map) of the predicted bias can be seen in Appendix F.

![Table](image)

Figure 8. Performance for the last test (STT3) for group T1 compared to bias.

* Here it becomes evident that the N2 vector imposes a great deal of bias on the N7 vector.

Figure 9. Performance for the last test (STT2) for group T2 compared to bias.
Figure 10. Performance for the STT2 for group T1 compared to bias.

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
<th>N9</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>8.67</td>
<td>7.25</td>
<td>6.17</td>
<td>5.14</td>
<td>4.08</td>
<td>3.36</td>
<td>2.85</td>
<td>2.70</td>
<td>2.53</td>
</tr>
<tr>
<td>M2</td>
<td>7.93</td>
<td>6.63</td>
<td>5.67</td>
<td>4.82</td>
<td>3.92</td>
<td>3.18</td>
<td>2.68</td>
<td>2.30</td>
<td>2.06</td>
</tr>
<tr>
<td>M3</td>
<td>6.51</td>
<td>5.43</td>
<td>4.67</td>
<td>3.84</td>
<td>3.10</td>
<td>2.48</td>
<td>2.01</td>
<td>1.74</td>
<td>1.55</td>
</tr>
<tr>
<td>M4</td>
<td>5.53</td>
<td>4.67</td>
<td>4.03</td>
<td>3.34</td>
<td>2.72</td>
<td>2.13</td>
<td>1.70</td>
<td>1.42</td>
<td>1.23</td>
</tr>
<tr>
<td>M5</td>
<td>4.71</td>
<td>3.95</td>
<td>3.43</td>
<td>2.88</td>
<td>2.38</td>
<td>1.90</td>
<td>1.54</td>
<td>1.27</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Correct | Innut | stades | sista | bil | trāvītes | snabb | sistes | stīsa | piekļ.

Figure 11. Performance for STT4 for group T1 compared to bias.

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
<th>N9</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>8.67</td>
<td>7.25</td>
<td>6.17</td>
<td>5.14</td>
<td>4.08</td>
<td>3.36</td>
<td>2.85</td>
<td>2.70</td>
<td>2.53</td>
</tr>
<tr>
<td>M2</td>
<td>7.93</td>
<td>6.63</td>
<td>5.67</td>
<td>4.82</td>
<td>3.92</td>
<td>3.18</td>
<td>2.68</td>
<td>2.30</td>
<td>2.06</td>
</tr>
<tr>
<td>M3</td>
<td>6.51</td>
<td>5.43</td>
<td>4.67</td>
<td>3.84</td>
<td>3.10</td>
<td>2.48</td>
<td>2.01</td>
<td>1.74</td>
<td>1.55</td>
</tr>
<tr>
<td>M4</td>
<td>5.53</td>
<td>4.67</td>
<td>4.03</td>
<td>3.34</td>
<td>2.72</td>
<td>2.13</td>
<td>1.70</td>
<td>1.42</td>
<td>1.23</td>
</tr>
<tr>
<td>M5</td>
<td>4.71</td>
<td>3.95</td>
<td>3.43</td>
<td>2.88</td>
<td>2.38</td>
<td>1.90</td>
<td>1.54</td>
<td>1.27</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Correct | Umanef | henses | sista | polje | trāfīdes | snabb | stādae | minste | kant.

Figure 12. Performance for the STT4 for group T2 compared to bias.

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
<th>N9</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>8.67</td>
<td>7.25</td>
<td>6.17</td>
<td>5.14</td>
<td>4.08</td>
<td>3.36</td>
<td>2.85</td>
<td>2.70</td>
<td>2.53</td>
</tr>
<tr>
<td>M2</td>
<td>7.93</td>
<td>6.63</td>
<td>5.67</td>
<td>4.82</td>
<td>3.92</td>
<td>3.18</td>
<td>2.68</td>
<td>2.30</td>
<td>2.06</td>
</tr>
<tr>
<td>M3</td>
<td>6.51</td>
<td>5.43</td>
<td>4.67</td>
<td>3.84</td>
<td>3.10</td>
<td>2.48</td>
<td>2.01</td>
<td>1.74</td>
<td>1.55</td>
</tr>
<tr>
<td>M4</td>
<td>5.53</td>
<td>4.67</td>
<td>4.03</td>
<td>3.34</td>
<td>2.72</td>
<td>2.13</td>
<td>1.70</td>
<td>1.42</td>
<td>1.23</td>
</tr>
<tr>
<td>M5</td>
<td>4.71</td>
<td>3.95</td>
<td>3.43</td>
<td>2.88</td>
<td>2.38</td>
<td>1.90</td>
<td>1.54</td>
<td>1.27</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Correct | Umanef | henses | minste | sadu | trāfīdes | snabb | minste | kant.

Figure 13. Performance for the STT3 for group T2 compared to bias.
Discussion

The goal of this study was to test if transcription testing can be used to generate useful data for analyzing patterns of memory retention. To aid in answering this fundamental question, several questions was posed which will be discussed hereafter. The first question: Are the differences in distractor stimuli significant compared to control? Was meant to elucidate if distractions in some way lowers the performance for the task of transcription. Based on the data, this was concluded to be the case; distractions seem to significantly lower performance for all participants across all age- and sex groups when compared to control (STT1).

When it comes to answering the question: Which distractor stimuli is the most potent distractor? The impact that the different distractors had on performance was not clear. The biggest difference however, was observed for STT3 (distracting sentence). This result was expected. Following the notion that a distraction has to be inhibited, a distracting sentence engages processing by several systems that perhaps all, to some degree, has to be inhibited. As well as the notion of selective attention and executive function as described in the theory section. This is conceivably more taxing then inhibiting random sounds or words such as the STT2&4. Compared to STT2 in which the distraction is that of a jumbled mess of phonemes, a phonological loop (Baddeley et. al., 1984) perhaps is not established and therefore not competing for resources. When treating the question: Is the variance within the samples attributable to age or sex? The data does not reflect that the variance is dependent on age or sex. One explanation could be that as we age we develop strategies that help us compensate for limitations of WM, this would however be a broad jump when the sample is of this size. Overall the research on this subject is mixed, one study compared two groups with mean ages of 23.9 and 73.9 years, and found the older group to be “non-strategic” and less flexible in spontaneously engaging memory rehearsal strategies, in free-recall word lists (Sanders et al., 1980). Compared to the young
group who tended to have a higher ability on average to encode early contents of lists with higher accuracy (Sanders, 1980). Sanders et al. (1980) also discussed the impact of education based on the assumption that the level of education might have an impact on the types of strategies that participants use. Sanders et al. (1980) also stresses the point that strategy is best measured using specific tests aimed at testing the strategies at hand as opposed to inferring which strategy might have been utilized based on a more general performance metric. This point is relevant for discussing if transcription testing by itself is enough, from the perspective of retention strategy. But, of course, one has to keep in mind that the STTs are not free recall tests, and pose a different dynamic altogether. Other research delving into age related effects in memory tasks, found that younger participants tend to more readily utilize a more flexible set of association strategies which can include exploiting the association between words to infer or aid in recall (Bryan et al., 1999). Since this study showed that performance did not drop for the older participants, the fact that executive function, strategy and WM capacity is known to decline with age (Bryan et al., 1999), is perplexing. But an important reminder, that cannot be stressed enough, is that in order to adequately analyze any differences in strategy due to age or other demographics, a much greater sample size is required, and probably a much more comprehensive and refined test. Anderson (2003), for example, posits that the very act of remembering is problematic when it comes to retrieval because of the stress posed in inhibiting the influence of interference. Therefore it is of relevance to add a qualitative dimension to the test, or test specific strategic processes in unison to this type of testing. This additional inquiry could help us to understand how to interpret the transcription data. And cross comparisons could be done after the participants has been instructed to follow some established compensatory strategies for memory retention (West, 1995).
Is STT performance related to RIT performance?

This will unfortunately go unanswered, since the test of response inhibition used in this case, could not adequately measure response inhibition. This was touched upon in the methods section but will be explain in a more straightforward way here. The first factor is that response time is not the same as stop signal reaction time. The second is that the time variations involved in measuring response inhibition are on the magnitude of milliseconds and does therefore require a stable low latency testing apparatus. And finally the third is that the test was too short. Given these factors it would be fallacious and misleading to engage in any discussions based on the results of the RIT.

Are the differences in distractor stimuli significant compared to each other? And Does the STT test results reveal patterns that pertains to established principles of memory retention?

This part of the discussion is devoted to interpreting the transcription data, without any relation to sample variance based on age or anything else. And will delve into discussing the overarching patterns observed in the data.

The transcription data does seem to fit the notion that information is affected on two temporal levels. In close proximity (L1), to the spoken word, and in distant proximity (L2) to the spoken word. This in ways, does reinforce the notion of duality of temporal effects as proposed by Waugh and Norman (1965) and Atkinson and Shiffrin (1968). Based on the data in this case, bias based on syntactic positioning is more prevalent on the L1 level. Whereas, on the L2 level, syntactic-positioning is not imposing as much bias as the repetition of information (availability). Perhaps if these sentences were semantically coherent, semantic processing might have aided positioning of syntactically interchangeable n-vectors (which means that a word can fit in two places and still be syntactically correct).
Examples of the effects of syntactic positioning is seen in the interaction between the N2 and N7 vectors across all of the tests. Since N1 and N2 vectors (which could be subjects of both primacy and recency) are consistently correct for the most part. Indicating that the first two items in the sentence usually remains intact. This might impose additional bias on the N7 vector, due to the information in N2 being retained in close proximity, in addition to being syntactically interchangeable. N4 and N9 seems to remain intact, if they are remembered at all. The retention of the N9 vector might be interpreted as following the U-shaped curve, where the last items in a list usually are subject to recency. This however might not apply in this case due to this task not being a test of free recall. This notion has however been disputed by the works of Bjork and Whitten (1974), which contend that recency effects are not limited to free recall but rather can be exhibited as a result of repetition of temporal spacing. Another contributor could be the one proposed in the model of interference, in where both the N4 and N9 (particularly N9) vectors have comparatively low weight due to the vectors representing a multitude of unique items, both semantically and lexically. Therefore perhaps limiting the response competition, making hypothetical comparisons easier. Another effect that might apply to vector N4 could be a manifestation of the von Restorff effect (remembering novel stimuli in the middle of a sequence of homogenous stimuli). Studies by Baddeley et. al. (1984), has indicated that shorter words and words that are phonologically dissimilar then other words in sequence tend to be recalled more often. This would explain the persistence of the N4-vector recall seen across all the tests. And in the few cases where N9 had not been correctly transcribed, it was usually the case of duplicating it with the N4 position.

Based on the hypothesis that word sequences consisting of short words are recalled better then sequences of long words called the word-length effect (WLE) (Jacquemot & Scott, 2006). We should see increased performance for
the sentences with shorter words, given that the effects of distraction are static. The sentences were 59, 51, 53, 55, letters long, excluding spaces. STT2 was the shortest, and even though the participants performed better on STT2 on average. It was not a particularly large difference. However, when added together with the fact that the distractor was unintelligible speech sounds, this might constitute to the slight performance increase seen for STT2.

The N6 vector is the one that is most often forgotten. Perhaps this is due to it not containing much variety. Or being processed arbitrarily in some way. Or it could be due to the accumulated effect of non-linear information retention. As seen from the data, information across the tests follow a pattern of retaining N1-N4 and N9. If the process of transcribing creates a second account of information, i.e. that information comes from two sources, the input information and output information. Then this would skew the performance in the direction of the words that were written, regardless if they were correctly transcribed or not. This could further potentiate the interference effect the further into the testing one goes, leading to more replaced words in particular.

Figure 14. Illustration of how the errors and lost words increase through time.

The data reflects varying limitations set on perceptual short term storage. The tendency to lose or replace words increases for the second half of the sentence. When this occurs, three patterns are pertinent to this discussion; a: forgetting and b: replacing or c: guessing. The method of transcription is not as linear as a digit span for example, that mostly rely on rehearsal (see figure 15). In order to
understand the difference of this, as it pertains to control processes, the proposed mechanism of transcription in this case is: [1]. Hearing the sentence. [2]. At the onset of writing, the information is rehearsed, relying on a perceptual buffer (phonological store). [3]. When or if this perceptual buffer diminish (which seem to occur quicker if distractions are present) [4]. Three behaviors ensue to complete the task: A: Guessing (knowingly) or B: Replacing (unknowingly) or C: Stopping. Drawing back to Waugh and Norman (1965), if the information is not lost in the C-group, but rather, is replaced by interference, the fact that the participants did not answer might be due to them knowing that any answer would be false. If this is the case however, cannot be determined in the absence of asking the participants, if this indeed was the case. When it comes to the initiation or maintenance of the rehearsal loop, the notion of strategy becomes relevant. One strategy could be comparison, that is, to compare a couple of words in a specific blank-space, for fit. As well as doing a visual comparison of the written material (since the participants could reason that the same word does not occur twice, and if it does, they would have remembered that). Another avenue, perhaps utilized by individuals with low WMC, is to automatically compensate for their low capacity by fragmenting the sentence into two parts. Which in the latter case could make rhythmicity of vocal input a possible factor.

Figure 15. *Phonological P-STM, Based on* (Baddeley et. al., 1984)
In a discussion regarding this type of strategic control process retrieval by Reder (1988), it is mentioned that individual differences are found in both extrinsic and intrinsic factors. In this case, one intrinsic factor, as proposed by Reder (1988), might be at play: the feeling of knowing. Or the feeling of not knowing and guessing. To help us understand what role individuality might have had in this aspect. An examination of the results on an individual level (Figures 8-13.) reveals some variation. Some individuals who don’t know, guess, and continue to guess across all the tests, making more errors as they progress. While other individuals who don’t know, don’t guess, and do not try to guess. While a third category of individuals might believe a false word is the correct one. This tendency might reveal some individual difference in the processing of information. Perhaps the level of modulation of activity that occurs during information processing is related to the degree of motivation to perform on the task itself. This merits further research.

On the aspect of replacing words. As reflected in the data and proposed throughout this work, *retroactive interference*, is likely at play. One notion is that STM encodes the verbal information, and that interference can occur when codes are stored that are similar (Lewis, 1996). Also, the high degree of memory dropout for vector N9, in the STT4 (the test with distracting novel words) for both groups, could be due to the *suffix effect*. Described by Dallet (1965) and later by Crowder (1982), the suffix effect is the observation that recall of the last items in a word list is impaired by subsequent nominally irrelevant but not modally irrelevant
stimuli (def. by Bloom, 2006). It usually is most vivid for the last item (Spoehr & Corin, 1978), but can extend further back into the sequence. And in the context of the STT4, the novel influx of words might suppress the ability to recall the sentence, to be transcribed. Crowders (1982) study shows that this effect, is more vivid in the cases when the recall is likely dependent on auditory perceptual memory or echoic memory. The reason for the suffix effect is not clearly understood, but is thought to be related to the depth of processing and lateral inhibition engaged by an immediate stimulus item (Crowder, 1982). In this case, the existence of the suffix effect might bring insight into if these sentences are primarily processed perceptually by some and post-perceptually by others. The data in this case however, is small, and in places dubious, and is therefore of limited value when it comes to reveal anything on this issue however.

Overall, the concept of lexical competition can be applied to understand this pattern. In essence, shared and diverging lexical properties of words have been indicated to carry its own share of associative power (Gaskell & Dumay, 2003) and could therefor contribute to the interference effects that is posed throughout this work and Andersons (2003) view of response competition. Also, as Gordon et al., (2002) study shows, syntactic similarity effects contributes to a higher level of interference, and pose a greater toll on memory-load. To compensate for this, it is posited that we use schemata of meaning to aid us in handling this greater processing demand (Gordon et al., 2002). However, per design, this testing eliminated such processing by the use of nonsensical sentences which most likely potentiated the effect of interference.

As previously stated, due to its size and procedure (made at a distance with no researcher monitoring the procedure) this data is of limited value when it comes to judging its reliability and validity in general. But the notion of repeating and using common words in combination with a rigid syntactic arrangement does seem to produce large amounts of errors which might bring insight into how
information is processed. Making it appealing for future modelling of memory interference. But! to further refine this procedure, I would recommend against using interchangeable n-vectors, since that brings an additional layer of ambiguity over the data.

Based on these results, an avenue for future studies could be to use a refined version of this method in combination with electroencephalography (EEG), since language processing is very much a cortical activity. To further pinpoint the interaction of auditory buffering and higher order processing. Another approach would be to track eye movements. This could be done rather, inexpensively, using commercially available eye tracking hardware. The reason to which, might elucidate if the visuospatial sketchpad might be integrated into the task, by some participants, visualizing the verbal information as a way of creating a multimodal account of the information. Some tentative evidence of such a mechanism has been proposed by Pazzaglia (1999) and Beni et. al. (2005).

Another useful thing could be to ask the participant after testing, how they think they did, how they solved particular transcriptions, and what task felt more difficult than others. All these questions would have been explored in greater detail if not the testing was performed at a distance.

And perhaps most importantly the testing would have benefited from a bigger data set. Both in terms of a larger sample but also more iterations of the test. Perhaps, using both typing and verbal recall. This would in particular be important in understanding how stimuli might interact with information processing, and what role semantics might have had.

Conclusion

The testing in this case was inconclusive, most likely due to limitations of the testing procedure, the test itself, and the size of the data sample. Although patterns did emerge, these patterns could not be linked to any particular factor as measured
in this case. The transcription test does however show appeal because it clearly shows how memory is affected by previous information. And creates analytic space for understanding differences between individuals. It could also be used to reveal semantic/associative processing. Since the sentence are not processed semantically, any semantic “residue” encoded in the words by experience, might attribute to the misplacing of some words. Overall however, more controllability is needed over the testing situation. And more measures should be taken to account for variables that might interfere with the testing procedure. Therefore, further studies are merited to further explore the viability of this type of testing.

References


Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological review, 63(2), 81.


Digital links
(2020-10-23)

Working memory definition by American Psychologist Association:
https://dictionary.apa.org/working-memory

Executive functions definition by American Psychologist Association:
https://dictionary.apa.org/executive-functions

Selective attention definition by American Psychologist Association:
https://dictionary.apa.org/selective-attention

WAIS verbal test:

Appendix

Appendix A

DATA COLLECTION

The test data collected (based on example):

IP (Identification measure) = 82.196.111.201

Date of admission = DAY:MONTH:HH:MM:SS:MS: Coordinated Universal Time (UTC)*:YEAR

*Swedish time is UTC+1

VIEWS = SESSION (the amount of times the data has been sent to the server. It is a way of determining if a participant cheat during the test)

TRY = SESSION (the amount of times the page has been shown. It is the second way of determining if a participant cheat during the test)

----------------------
DATA FORM COLLECTED FROM EACH PARTICIPATION----------------------

Template

AGE = 18

82.196.111.201 - Sun Dec 6 17:24:23:1223 UTC 2020-VIEWS = 1

GENDER = FEMALE

82.196.111.201 - Sun Dec 6 17:24:48:1248 UTC 2020-VIEWS = 1

START RIT

PRESSED X1-82.196.111.201 - Sun Dec 6 17:24:56:1256 UTC 2020

PRESSED X2-82.196.111.201 - Sun Dec 6 17:24:58:1258 UTC 2020

PRESSED X3-82.196.111.201 - Sun Dec 6 17:25:00:1200 UTC 2020

PRESSED X4-82.196.111.201 - Sun Dec 6 17:25:04:1204 UTC 2020

PRESSED X6-82.196.111.201 - Sun Dec 6 17:25:13:1213 UTC 2020

PRESSED X7-82.196.111.201 - Sun Dec 6 17:25:16:1216 UTC 2020
ÖVNING1 SVAR = framför stadens långa maträtt.

ÖVNING2 SVAR = inuti deras minsta tåg hämtades långsamt skolans korta balkong.

ÖVNING3 SVAR = långa framför deras snabb bil långa deras öppnad.

STTPrelim SVAR = inuti husets minsta katt tvättades långsamt husets äldsta hund.

STT1B SVAR = framför hennes yngsta tåg tvättades långsamt hennes yngsta maträtt.

STT2B SVAR = inuti husets sista bil tvättades långsamt husets sista pojke.

STT4B SVAR = utanför hennes sista pojke tvättades långsamt hennes minsta katt.

STT3B SVAR = utanför husets äldsta hund tvättades långsamt husets äldsta hund.

ALTERNATIV8 = Den med två meningar
COMMENT (Optional) = The test was too easy. I aced it! I’d like to know what others have answered.

Appendix B

Randomly generated sentences

Inuti husets största katt arbetades långsamt deras största balkong
Inuti husets största tåg sågs långsamt husets första katt
Framför stadens första fågel kördes snabbt hennes minsta adress
Inuti skolans sista hund tvättades snabbt skolans första adress
Framför stadens långa maträtt lånades snabbt stadens långa hund
Inuti stadens sista bil träffades långsamt stadens korta fågel
Inuti dennes yngsta adress lagades snabbt hans långa bil
Framför dennes minsta bil arbetades snabbt stadens yngsta balkong
Framför hennes minsta tåg tvättades långsamt hennes yngsta maträtt
Inuti stadens äldsta balkong sågs långsamt stadens minsta adress
Framför hennes korta adress hämtades snabbt landets långa flicka
Framför hennes långa bil öppnades snabbt deras största bil
Framför hans yngsta balkong kördes långsamt landets första bil
Framför landets yngsta adress lagades snabbt skolans största katt
Inuti stadens sista bil tvättades snabbt husets sista pojke
Utanför hennes sista pojke träffades snabbt stadens minsta katt
Utanför landets äldsta hund kördes snabbt deras första flicka
Utanför skolans största hund arbetades snabbt dennes sista tåg
Inuti husets minsta katt tvättades långsamt hennes äldsta hund
Utanför hans äldsta balkong lånades snabbt landets största pojke
Appendix C

HEJ OCH VÅLKOMMEN!

Tack för att du tänka dig att göra detta test och vara med i denna undersökan. Tennat för denna undersökan är att testa vårt arbetsminne med hjälp av ett transkriptionsontes för att se hur vi påverkas av olika typer av stimuli. Och hur vi dynamiskt anpassar oss när vi utstås för förändring.

BAKGRUND


OBS! De behöver använda ett tangentbord för att hinna transkribera meningarna du hör.

PERSONLIG INFO

Detta test tar mellan fem och tio minuter att fullfölja.

Genom att trycka på länken och fortsätta till testet geskänner du att du testdata får användas för att besvara studiens följeställning.

Utöver det tanteras all testatta anonymitet och du kommer ej att kunna identifieras (av någon) stift från den information du lämnar här.
Appendix D

Figure 6.  This is the answers of group T1, in the order they were transcribed.

<table>
<thead>
<tr>
<th>Framför hennes minsta tåg tvättades långsamt hennes yngsta maträtt</th>
<th>Utanför hennes sista pojke träffades snabbt stadens minsta katt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framför hennes minsta tåg tvättades långsamt hennes yngsta maträtt</td>
<td>Utanför hennes sista pojke träffades snabbt hennes sista pojke</td>
</tr>
<tr>
<td>Framför hennes längsta tåg tvättades hennes långsta maträtt</td>
<td>Utanför hennes minsta stad träffades snabbt den minsta pojken</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades hennes yngsta maträtt</td>
<td>Inta</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades</td>
<td>Utanför hennes sista pojke träffades snabbt stadens största katt</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades hennes yngsta maträtt</td>
<td>Utanför hennes minsta pojke tvättades hennes minsta katt</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades långsamt hennes äldsta maträtt</td>
<td>Utanför hennes äldsta pojke tvättades snabbt stadens katt</td>
</tr>
<tr>
<td>Utanför landsåts äldsta hund kördes snabbt deras första flicka</td>
<td>Inuti stadens sista bil tvättades snabbt husets sista pojke</td>
</tr>
<tr>
<td>Utanför landets näst något tvättades snabbt stadens längsta flicka</td>
<td>Inuti stadens minsta bil tvättades snabbt hennes yngsta pojke</td>
</tr>
<tr>
<td>Utanför landsåts första hund hämtade stadens längsta bilen</td>
<td>Innanför sista tåg tvättades den längsta biken</td>
</tr>
<tr>
<td>Utanför landsåts äldsta hund kördes</td>
<td>Inuti hennes sista</td>
</tr>
<tr>
<td>Utanför landsåts längsta hund</td>
<td>Inuti stadens sista bil tvättades snabbt stadens sista pojke</td>
</tr>
<tr>
<td>Utanför stadens största hund hämtades snabbt landets länga flicka</td>
<td>Inuti stadens sista bil hämtades</td>
</tr>
<tr>
<td>Utanför landsåts minsta hund kördes långsamt landsåts minsta flicka</td>
<td>Inuti hennes näst omtvättades näst omtadens pojke</td>
</tr>
<tr>
<td>Utanför landsåts</td>
<td>Inuti stadens sista bil tvättades snabbt stadens äldsta pojke</td>
</tr>
</tbody>
</table>

The first test (STT1) is on the top left, second (STT2) on the top right, third (STT4) on the bottom left, and fourth (STT3) on bottom right.

Figure 7.  This is the answers of group T2, in the order they were transcribed.

<table>
<thead>
<tr>
<th>Framför hennes minsta tåg tvättades långsamt hennes yngsta maträtt</th>
<th>Utanför landsåts äldsta hund kördes snabbt deras första flicka</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framför</td>
<td>Inuti stadens sista bil hämtades stadens minsta pojke</td>
</tr>
<tr>
<td>framför hennes yngsta tåg tvättades hennes yngsta maträtt</td>
<td>inuti stadens sista bil</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades hennes yngsta maträtt</td>
<td>Inuti stadens sista bil tvättades stadens sista pojke</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades långsamt hennes minsta maträtt</td>
<td>Inuti stadens sista bil tvättades stadens första pojke</td>
</tr>
<tr>
<td>Framför hennes längsta tåg, tvättades långsamt hennes minsta maträtt</td>
<td>Inuti stadens minsta bil, tvättades långsamt stadens äldsta pojke</td>
</tr>
<tr>
<td>Framför hennes minsta tåg tvättades långsamt hennes minsta klapphåst</td>
<td>Inuti stadens minsta bil, tvättades långsamt stadens äldsta pojke</td>
</tr>
<tr>
<td>framför hennes minsta tåg tvättades långsamt hennes äldsta maträtt</td>
<td>Inuti stadens minsta klapphåst</td>
</tr>
<tr>
<td>Inuti deras längsta tåg tvättades deras minsta maträtt</td>
<td>Inuti deras minsta hus tvättades stadens pojke</td>
</tr>
<tr>
<td>framför hennes minsta tåg tvättades hennes minsta maträtt</td>
<td>Inuti deras minsta bil tvättades stadens sista pojke</td>
</tr>
<tr>
<td>framför hennes längsta tåg tvättades hennes yngsta maträtt</td>
<td>Inuti deras minsta bil tvättades stadens sista pojke</td>
</tr>
</tbody>
</table>
Utanför hennes sista pojke träffades snabbt stadens minsta katt
Utanför landets äldsta
framför hennes yngsta hund
Utanför deras första hund låg landets minsta flicka
Utanför landets största hund
främst landets högsta, hämtades den minsta flickan
Utanför landets första hund
utanför landets största hund hämtades
Inuti deras minsta hund...
främst landets första hund hämtades
utanför landets äldsta hund
Inuti stadens sista bil tvättades snabbt husets sista pojke
Utanför minsta näning näning, det här var svårt Sinal!
Utanför stadens minsta mening låg pojkens
Utanför hennes minsta stad
Utanför stadens minsta katt, hämtades långsamt hennes längsta pojkvän
Utanför hennes minsta stad, hämtades långsamt hennes längsta pojkvän
...
### Appendix E

<table>
<thead>
<tr>
<th></th>
<th>STT1BSCORE</th>
<th>STT2BSCORE</th>
<th>STT1BSCORE</th>
<th>RITSCOREMETHOD2</th>
<th>RITSCORE</th>
<th>INHIBCORRECT</th>
<th>RESPONSETIME</th>
<th>RW2</th>
<th>MWLW</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missed and/or</td>
<td>.160</td>
<td>.007</td>
<td>.125</td>
<td>.231</td>
<td>.475</td>
<td>.222</td>
<td>.028</td>
<td>.322</td>
<td>.269</td>
<td>.330</td>
</tr>
<tr>
<td>Replaced Words</td>
<td>-.781</td>
<td>-.321</td>
<td>-.403</td>
<td>-.263</td>
<td>-.040</td>
<td>-.170</td>
<td>.103</td>
<td>.357</td>
<td>-.904</td>
<td></td>
</tr>
<tr>
<td>Response time</td>
<td>.636</td>
<td>.099</td>
<td>.099</td>
<td>.356</td>
<td>.141</td>
<td>.073</td>
<td>.069</td>
<td>-.186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct inhibitions</td>
<td>-.255</td>
<td>-.579</td>
<td>-.174</td>
<td>-.409</td>
<td>-.445</td>
<td>-.699</td>
<td>-.046</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIT Score 1</td>
<td>-.083</td>
<td>-.253</td>
<td>-.298</td>
<td>-.146</td>
<td>.466</td>
<td>.726</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIT Score 2</td>
<td>.062</td>
<td>-.192</td>
<td>.350</td>
<td>.141</td>
<td>.674</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIT Score Method 2</td>
<td>.105</td>
<td>.211</td>
<td>.169</td>
<td>.146</td>
<td>.726</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.169</td>
<td>.333</td>
<td>.218</td>
<td>.216</td>
<td>.338</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.216</td>
<td>-.198</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Pearson correlation coefficients.

Figure 2. *Pearson correlation with STT-performance added.*
Appendix F

Figure 1. *Heat-map showing where bias is projected to occur (STT3, T1) according to model.*

Figure 2. *Heat-map showing for STT3, T1, according to model.*