Title

Working Memory Failure in Phone-Based Interaction

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

Authors

Huguenard, Brian R.
Lerch, F. Javier

Publication Date

1993

Working Memory Failure in Phone-Based Interaction

Brian R. Huguenard **F. Javier Lerch**
. Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA 15213-3890
fl0c@andrew.cmu.edu

Abstract[1]

This paper investigates working memory failure in menu driven Phone-Based Interaction (PBI). We have used a computational model of Phone-Based Interaction (PBI USER) to generate predictions about the impact of three factors on WM failure: PBI features (i.e., menu structure), individual differences (i.e., WM capacity) and task characteristics (i.e., task format and number of tasks). Our computational model is based on the theory of WM proposed by Just and Carpenter (1992). This theory stipulates that the storage *and* the processing of information generate demands for WM resources. Our empirical results provide strong evidence for the importance of storage demands, and moderate evidence for the importance of processing demands as predictors of WM failure in PBI. In addition, our results provide evidence for the importance of individual differences in WM capacity as a predictor of WM failure in PBI. Finally, our results indicate that, contrary to general guidelines for the design of PBI, deep menu hierarchies (no more than three options per menu) do not reduce WM error rates in PBI.

Introduction

This paper investigates working memory (WM) failure in Phone-Based Interaction (PBI). Phone-based interfaces allow interaction between a human and a computer by means of a touch-tone telephone. In these interfaces users must compare their current goal against the menu options presented by the interface. The interaction with menu driven phone-based interfaces places high demands on WM because the comparisons of goals and menu options are performed in WM.

We have used a computational model of Phone-Based Interaction (PBI USER) to generate predictions about the impact of three factors on WM failure: PBI features (i.e., menu structure), individual differences (i.e., WM capacity) and task characteristics (i.e., task format and number of tasks). Our computational model is based on the theory of WM proposed by Just and Carpenter (1992). This theory stipulates that the storage *and* the processing of information generate demands for WM resources. Our empirical results provide strong evidence for the importance of storage demands, and moderate evidence for the importance of processing demands as predictors of WM failure in PBI. In addition, our results provide evidence for the importance of individual differences in WM capacity as a predictor of WM failure in PBI.

Just and Carpenter (1992) have proposed a theory of WM in which *both* storage and processing demands determine WM load and the probability of WM failure. Just and Carpenter (1992) describe their theory as "... a computational theory in which both storage and processing are fueled by the same commodity, namely activation. In this framework, capacity can be expressed as the maximum amount of activation available in working memory to support either of the two functions". PBI USER is built on top of the instantiation of this theory (CAPS). PBI USER keeps track of all the information encoded, maintained and processed in WM during Phone-Based Interaction. In this paper we only present the hypotheses generated by PBI USER and omit most of its architectural details.

As previously mentioned, activation is the general WM resource consumed by the processing and the storage of information in WM. Activation is used in CAPS: 1) for maintaining WM elements so they are not forgotten, and 2) for processing WM elements. A key feature of the architecture is that the total amount of activation available can be constrained to a preset amount. Given this constraint, if the total processing and storage demands for activation exceed the total amount of activation available, then all activation is scaled back so that the activation constraint is not exceeded. This "scaling back" has the effect of slowing down processing (since less activation is spread per cycle, more cycles will now be required to complete the processing), and the effect of causing a form of forgetting through displacement (since less activation is now available for storage, some WM elements may lose so much activation that they are effectively forgotten). It is important to note that it is the combined demand of storage and processing for a given task that determines whether or not the activation constraint will be exceeded. Therefore this combined demand determines the probability of WM failure.

---

[1] This work was supported in part by grants from Bellcore #1-41117 and #1-41239.

Phone-Based Interaction

Phone-Based Interaction (PBI) has considerable advantages as a task domain for building detailed cognitive models of WM failure and for empirically investigating error behavior due to WM limitations. First, PBI is conceptually simple so knowledge of the task is easy to acquire for the subjects and simple to represent in the cognitive model. This allows us to concentrate our effort on how information is encoded, maintained, and processed by the user. Second, PBI allows direct experimental manipulation of WM load. For example, PBI users can be given tasks with or without numeric modifiers (e.g., in a student registration system: add class 123 vs. retrieve schedule) or users can be asked to retain information about two tasks and to perform them consecutively. Third, PBI allows only two types of execution errors: a) *information loss errors*, that is, users forget information about the task (e.g., the code of the class to be added, the name of the object for which information is to be retrieved), and b) *choice errors*, that is, users select the wrong alternative when presented with a set of choice options.

In terms of the CAPS architecture the first type of PBI execution error, information loss, involves the *storage* of information in WM and is directly caused by WM capacity being insufficient to satisfy the activation requirements of the task. When storage and processing requirements of the task exceed the available WM capacity, CAPS will "scale back" the amount of activation used for all storage and processing, resulting in an increased probability of information being lost from WM due to an insufficient amount of storage activation. The second type of execution error, choice error, involves the *processing* required to select an alternative from a set of choices. A choice error can be caused by: 1) WM capacity being insufficient to satisfy the activation requirements of the task (again resulting in a "scaling back" of all activation, leading to an increased probability of choice error due to an insufficient amount of processing activation), and/or 2) lack of knowledge about how the goal (the task to be accomplished) relates to the current set of choice alternatives.

Computational Model

We have developed a computational model (PBI USER) based on the CAPS architecture that simulates the storage and processing of information in WM in Phone-Based Interaction. PBI USER has two main components: the Device Model and the User Model. The Device Model, written in Lisp, simulates the behavior of a specific telephone interface. The User Model is a production system written in CAPS. Once the User Model has encoded and stored (in WM) a goal task to be achieved (e.g., "add class 123"), the Device Model presents to the User Model the menu options of the simulated interface. As the options are presented, the User Model encodes and stores the option information, processes the available information to determine which (if any) of the options are appropriate for achieving the desired goal, and then chooses one option. After the User Model makes a menu option choice, the Device Model presents the new menu options, and the process continues until the goal is achieved, or an information loss or choice error occurs.

We have used PBI USER to generate a set of hypotheses about the impact of three factors on WM failure: menu structure, WM capacity and task characteristics. Menu structure refers to the topology of the menu hierarchy. We have tested PBI USER with two menu structures (PBI-DEEP and PBI-BROAD) that are identical in functionality but differ in topology. PBI-DEEP has a 3x3x3x3 structure, while PBI-BROAD has a 9x9 structure. The use of menu structure allows us to manipulate the *processing* requirements of the PBI task, since the three options at the top level of PBI-DEEP would be expected to be more ambiguous and require more processing than the nine options at the top level of PBI-BROAD.

WM capacity refers to the total activation available for storage and processing. Through the activation capacity constraint of the CAPS architecture, we can directly model the impact of individual differences in WM capacity on WM failure. We have also investigated two task characteristics: task format and number of tasks. These two characteristics address the amount of information that must be stored during the performance of a task. All tasks performed by PBI USER were given in an action/object/modifier format; for the task "add class 123", the action is "add", the object is "class", and the modifier is "123". We manipulated task format by changing the type of modifier found in the task. We have investigated three types of modifiers: 1) absent (e.g., "retrieve transcript"), 2) natural language (e.g., "retrieve events <u>community-service</u>"), and 3) numeric (e.g., "add class <u>123</u>"). Number of tasks was manipulated by having PBI USER perform either one single task or a pair of tasks. Through these two task characteristics, we are able to manipulate the *storage* requirements of PBI tasks.

Hypotheses

Information Loss Errors

The PBI USER architecture suggests that it is not the number of options per menu that determines the magnitude of WM load, but rather the amount of processing and storage required to evaluate the "goodness" of each individual option. Thus WM load should rise, peak, and fall during each option evaluation, and it is the height of each peak relative to total WM capacity that determines the probability of information loss. We began with the conservative assumption that the processing and storage requirements for each specific option evaluation were the same for PBI-DEEP and PBI-BROAD. Thus, **H1: We expected menu structure would not have an impact on information loss error rates.**

Even if we make the further restriction that the options in PBI-DEEP would require more processing than those of PBI-BROAD (due to the increased ambiguity of the options from the 3x3x3x3 structure), H1 would still hold, since we expected WM capacity to be allocated first to storage requirements (since the user knows exactly what task information – primarily the task description – must be remembered), and the remainder allocated to processing requirements (which are not known ahead of time). Note that H1 is in conflict with the widespread guideline that phone-based interfaces should limit the maximum number of options per menu to three (Gould & Boies, 1984).

Our model of phone-based interaction predicts WM errors when the demands for activation (due to processing and storage) exceed the activation constraint. Since we know humans have different WM capacities (Daneman & Carpenter, 1980), we should expect that users with lower WM capacity should have a greater probability of having their capacity exceeded by the activation demands of phone-based interaction, therefore, **H2: We expected subjects with higher WM capacity would have lower information loss error rates than subjects with lower WM capacity.**

In terms of WM storage requirements for the three levels of task format, we expected the absence of a modifier to create the least demand for storage, followed by natural language and numeric. We expected numeric modifiers to require more storage capacity than natural language modifiers because multiple symbols would be required to represent a random series of digits, while the natural language modifiers would be likely to be represented by a single symbol. Thus, **H3a: We expected increased complexity of tasks (operationalized by task format moving from "none" to "natural language" to "numeric") would increase information loss error rates.**

The two levels of number of tasks provide us with a straightforward means of manipulating the amount of information that must be held in WM during task performance: single tasks should require less storage activation than pairs of tasks. Thus, **H3b: We expected increased complexity of tasks (operationalized by number of tasks moving from "single" to "pair") would increase information loss error rates.**

Choice Errors

Our model of phone-based interaction (PBI USER) depicts menu option choice as the process of comparing each menu option to a description of the current task, and then selecting the most appropriate option. This comparison process should be more error prone when the menu options are ambiguous than when the options are specific (due to insufficient knowledge, not due to WM constraints), and so we expected more choice errors when menu options are ambiguous. In our study, the menu structure PBI-DEEP must categorize all of the functionality of the final 81 terminal nodes into 3 options at the top level menu, in contrast to the PBI-BROAD structure which has 9 options at the same level. Therefore, **H4: We expected choice error rates would be higher for PBI-DEEP than for PBI-BROAD.**

The PBI USER architecture tells us that WM errors do not occur unless task demands for activation exceed the activation constraint. We expected that the processing requirements for choosing options in the menu structure PBI-DEEP would be greater than for PBI-BROAD. This increased processing requirement would be the result of the additional semantic processing required to disambiguate the option labels in PBI-DEEP. This disambiguation process requires (in PBI USER) the retrieval of information from long-term memory that relates the ambiguous option label to the current task. This additional processing load should result in PBI-DEEP users being more likely than PBI-BROAD users to exceed their WM capacities during the selection of options. On the other hand, the more specific options of PBI-BROAD should require less processing for disambiguation than PBI-DEEP, and therefore we expected a weaker effect between WM capacity and choice error rates for the broad menu structure. **H5: We expected an interaction effect between WM capacity and menu structure for choice error rates. For PBI-DEEP, subjects with higher WM capacity would have lower choice error rates than subjects with lower WM capacity. We expected this effect to be weaker for PBI-BROAD.**

Experiment And Results

The Experiment

The purpose of the experiment is to provide empirical evidence for the impact of three factors on WM error rates in a typical PBI task: 1) structure of the menu hierarchy, 2) individual differences in WM capacity, and 3) WM demands induced by task characteristics.

Subjects. Eighty-seven students were recruited and paid $10.00 to participate in the experiment. Subjects were categorized by their dynamic WM capacity into three groups, based on their scores on the reading span test (Daneman & Carpenter, 1980): low span (20 subjects with scores of 2.0 2.5), medium span (45 subjects with scores of 3.0 - 3.5), and high span (22 subjects with scores of 4.0 - 6.0).

Materials and Apparatus. We have implemented two simulated telephone-based student registration systems (PBI-DEEP and PBI-BROAD) in the NeXT workstation. The two systems are functionally equivalent but differ in the topology of their menu hierarchies. The systems allow the performance of a variety of student registration tasks, such as adding/dropping classes, retrieving grades, paying fees, etc.

Experimental Design. In addition to the two menu structures (PBI-BROAD and PBI-DEEP), the other experimental factors were WM capacity (low, medium, and high) and task type (nine levels). The nine levels of task type result from the crossing of task format (no modifier, natural language modifier, numeric modifier) with the three levels of task order (task order is derived from number of tasks, and has the levels single task, first task in a pair of tasks, and second task in a pair). Thus, we have a 2x3x9 design, with menu structure and WM capacity being between-subjects factors, and task type being a within-subjects factor. A total of twenty-seven tasks were presented to each subject, allowing for three replications of each of the nine task types.

Procedure. Each subject was run in two separate sessions. In the first session, the reading span test was administered to each subject on an individual basis, using the guidelines given in Daneman and Carpenter (1980). Each span test took approximately ten minutes to complete. The second session was the main session of the experiment, in which the subjects interacted with either PBI-DEEP or PBI-BROAD. For this session, subjects received all options and feedback pertaining to the PBI systems through headphones. All subject input was made by clicking on a representation of a 12-key touch-tone telephone keypad on the screen. Task descriptions were briefly presented on the computer screen and then erased. After completing a series of training tasks, each subject performed twenty-seven tasks. Subjects were allowed to retrieve a single parameter (a parameter is an action, an object or a modifier) for each task if needed. All subject actions (mouse clicks), along with timing data were captured by the software. The time required to complete this session was approximately fifty minutes.

Data Analysis. Our theoretical classification of error types (information loss errors and choice errors) was operationalized in the following manner. Information loss was operationalized in two different ways: 1) rehearsal errors, in which the subject forgot a parameter of a task such as a modifier (e.g. add class **123**) or the object of the task (add **class** 123), but after making a request to review this parameter, the subject completed the task successfully (subjects were only allowed one such request per task), and 2) task failure, in which the subject failed to complete the task correctly (by failing to complete the task, or by completing the wrong task). Task failure may occur even after a request for a forgotten parameter has been made. In our experimental setting, rehearsal errors were a recoverable type of information loss, since the subject was able to complete the task correctly; on the other hand, we consider task failure to be a more severe type of information loss since we assume subjects have forgotten more than one task parameter. However, since both are forms of information loss, we expected hypotheses H1-H3b to hold for both rehearsal and task failure errors. Choice errors were operationalized as navigational errors, in which the subject made at least one incorrect option choice while traversing the menu structure but completed the task correctly by backtracking. We expected WM limitations to be an important factor on the frequency of navigational errors, and we expected hypotheses H4 and H5 to hold for navigational errors.

We have focused on developing statistical models and analyses from first principles rather than relying on traditional off-the-shelf statistical methods (Junker et al., 1993). These statistical models have been developed to address many of the features of error behavior research that make statistical analysis challenging in our experiment, including categorical dependent variables and the dependence of categorical responses within each subject (due to the repeated-measures design often used in error behavior research). Overall effects of the experimental factors were assessed by comparing nested models in GLIM using the deviance statistic; a model deviance difference is identical to both the log-likelihood ratio for the two models and the statistic ΔG^2 (Bishop, Fienberg and Holland, 1975) often used to compare loglinear models. This statistic is approximately chi-squared distributed with degrees of freedom equal to the difference in the number of parameters of the two models when the smaller model is correct (See (Junker et al., 1993) for more details). Some experimental effects were more easily assessed by performing a one-sided test for the hypothesis that the rate-predicted-to-be-high minus the rate-predicted-to-be-low was indeed positive (for these effects there was insufficient power, given the sample size, to do more detailed comparisons). The relevant test statistic can be calculated from the statistical model using the delta method (Bishop, Fienberg, & Holland, 1975). In the results section, this test will be referred to as a "high-low contrast".

Results

Overall Results. Table 1 shows the observed rates for the five response categories. Menu structure had a significant effect on overall response rates (GLIM deviance of 52, on 4 degrees of freedom, $p<.001$). These results suggest that the impact of menu structure is mainly on navigational errors (17.23% for PBI-DEEP, 7.74% for PBI-BROAD).

WM capacity was highly significant (368 on 8 d.f., $p<.001$). As an example of the impact of WM capacity on response rates, the observed rates for the response "no errors" were 71.21% for the high WM capacity subjects, 65.93% for the medium, and 59.81% for the low (combined for both menu structures).

Finally, task characteristics (task format and number of tasks) were also significant (167 on 8 d.f. for task format, p<.001; 200 on 8 d.f. for number of tasks, p<.001). The impact of task format on response rates is exemplified by the observed rates for the "no errors" response (combined for both menu structures): 73.18% for no modifier, 67.94% for natural language modifier, and 56.45% for numeric modifier. Number of tasks resulted in the observed "no errors" rates of 79.44% for single tasks, and 59.07% for pairs of tasks (again, combined for both menu structures).

Table 1: Overall Response Rates by Menu

Response	PBI-DEEP	PBI-BROAD
no errors	61.41%	70.20%
task failure	10.08%	10.19%
correct w/rehearsal	7.32%	8.25%
correct w/navigational	**17.23%**	**7.74%**
correct w/rehearsal & nav.	3.96%	3.62%

Task Failure. No significant differences were found between the two menu structures in terms of overall task failure error rates (0.3 on 2 d.f., ns), with PBI-DEEP and PBI-BROAD having observed task failure rates of 10.08% and 10.19%, respectively. Since task failure is a form of information loss, this result supports H1. The factors of WM level, task format, and number of tasks all had significant effects on the frequency of task failure rates for both menu structures (providing support for hypotheses H2, H3a, and H3b). For instance, the three levels of Working Memory capacity (high, medium, and low) resulted in observed task failure rates (combined for the two menu structures) of 5.73%, 11.19%, and 12.59% (19 on 2 d.f., p<.001). Task format (no modifier, natural language modifier, numeric modifier) resulted in observed task failure rates (combined for menu structure) of 5.49%, 10.73%, and 14.18% (32 on 2 d.f., p<.001). Finally, number of tasks (single task, pair of tasks) resulted in observed task failure rates (combined for menu structure) of 3.83% and 13.28%(56 on 2 d.f., p<.001).

In summary, all hypotheses concerning information loss were supported by the results of task failure error rates, that is, there is no difference between menu structures, a significant difference among WM capacity groups, and a significant impact of task characteristics.

Rehearsal Errors. No significant differences were found between the two menu structures in terms of rehearsal error rates (H1) , with PBI-DEEP and PBI-BROAD having observed rehearsal error rates of 12.55% and 13.21%, respectively (high-low contrast z=0.10, ns). The three levels of WM capacity (high, medium, and low) resulted in observed rehearsal error rates (combined for both menu structures) of 11.61%, 12.23%, and 15.89%, but this effect did not reach statistical significance (high-low contrast z=2.0, ns using the conservative Bonferroni correction for multiple comparisons). The task complexity features of task format and number of tasks both had significant effects on the frequency of rehearsal error rates (H3a and H3b). Combining rehearsal error rates for the two menu structures, task format (no modifier, natural language modifier, numeric modifier) resulted in observed rehearsal error rates of 4.19%, 12.16%, and 23.21% (high-low contrast z=10.0, p<.001). Number of tasks (single task vs. pair of tasks) resulted in observed error rates (combined for both menu structures) of 3.05% and 18.34% (high-low contrast z=12.5, p<.001).

In summary, with the exception of H2 all hypotheses concerning information loss were supported by the results of rehearsal error rates, that is, there is no difference between menu structures, and a significant impact of task characteristics. While the observed rehearsal error rates did increase as WM capacity decreased (as hypothesized in H2), this effect was not statistically significant.

Navigational Errors. Menu structure was significant in terms of navigational error rates (H4), with observed error rates of 23.56% for PBI-DEEP, and 12.65% for PBI-BROAD (high-low contrast z=6.6, p<.001). WM level did not have a significant effect on the frequency of observed navigational error rates (see Table 2) for PBI-DEEP (high-low contrast z=2.47, ns using the Bonferroni correction), or for PBI-BROAD (high-low contrast z=0.61, ns using the Bonferroni correction).

In summary, hypothesis H4 was supported by the results of navigational error rates, but hypothesis H5 was not supported. That is, there is a significant difference between menu structures, and no significant interaction between WM capacity and menu structure.

Table 2: Observed Navigational Error Rates by WM and Menu Structure

	Navigational Error Rates	
WM	PBI-DEEP	PBI-BROAD
high	19.93%	12.68%
medium	22.21%	11.81%
low	29.44%	14.52%

Discussion

Our hypotheses for information loss errors in PBI were supported by the results for task failure and rehearsal error rates, with the one exception of H2 (which involves the impact of WM capacity on information loss) for rehearsal errors. These results indicate that information loss errors are more likely to occur as the demand for WM storage increases, and that individual differences in WM capacity have a substantial impact on the ability of PBI users to maintain information in WM. Our results strongly indicate that two radically different menu structures (PBI-DEEP and PBI-BROAD) had no impact on the probability of information loss. Menu structure did not have an impact on severe (task failure) or moderate (rehearsal error) information loss. This supports our claim that the primary determinant of information loss in PBI is not the number of options per menu, but rather the complexity of the evaluation of each individual option.

Hypothesis H2, which predicted an increase in information loss error rates as WM level decreased, was not supported by the results for rehearsal errors. While observed rehearsal error rates did increase as WM level decreased, this effect was not strong enough to reach statistical significance. One explanation of this lack of significance could be our relatively small sample size, coupled with the conservative multiple comparison correction we used, but we offer an additional explanation. Consider the distribution of WM load over all the tasks given in the experiment. Figure 1 presents such a distribution, in which most of the tasks have low to moderate WM load, and the proportion of tasks decreases as WM load increases.

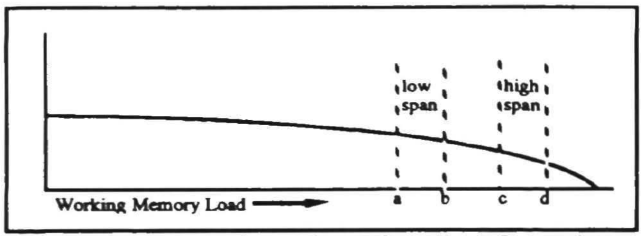

Figure 1: Relationship of WM Capacity Groups and WM Load to Task Frequency.

In Figure 1 two "bands" of WM capacity are represented, one extending from "a" to "b" for the low span subjects, and one extending from "c" to "d" for the high span. These bands indicate the relationships between WM capacity, WM load of tasks, and information loss errors. For instance, the low span subjects would generate 1) no information loss errors for tasks with WM load less than "a", 2) rehearsal errors for tasks with WM load between "a" and "b", and 3) task failure errors for tasks with load greater than "b". Similarly, high span subjects would generate 1) no information loss errors for tasks with WM load less than "c", 2) rehearsal errors for tasks with WM load between "c" and "d", and 3) task failure errors for tasks with WM load greater than "d". Notice that although the area representing rehearsal errors for the high WM capacity group is less than that for the low WM capacity group, the difference is not great. However, if we consider the area representing task failure for the low capacity group (all tasks beyond "b"), we note that it is much greater than the corresponding area for the high capacity group (all tasks beyond "d"). Thus, although the high capacity subjects do have lower error rates than the low capacity subjects for both rehearsal and task failure errors, it is easier to distinguish between the task failure error rates. Therefore, while there was not a significant effect of WM on rehearsal error rates, we believe that the true test of the importance of WM as a predictor of information loss error rates is with the impact of WM capacity on task failure rates. This belief is supported by our empirical results, which show a strong effect of WM on task failure error rates.

While menu structure had little impact on information loss, it was highly significant on choice (navigational) error rates, confirming our claim that menu structure should have an impact on *processing* demands in WM, but not on *storage* demands. Although we did not observe a significant interaction between WM capacity and menu structure as predicted by H5, choice (navigational) error rates for PBI-DEEP did increase by nearly ten percentage points as WM decreased from high to low, while choice error rates for PBI-BROAD increased by less than three percentage points in the same situation. Out inability to find a significant effect for the interaction may have been due to a lack of statistical power, or to an actual lack of an effect. If the interaction effect does exist, then it appears to be a weak effect, and may be due to strategic reallocation of WM resources between storage and processing, based on WM load incurred "on the fly" during menu navigation. We are in the process of refining the modeling of WM failure by accounting for alternative resource allocation strategies in WM.

Acknowledgements

We would like to thank Brian Junker and Richard Patz for their help in the statistical analysis, Marcel Just, Patricia Carpenter, Marvin Sirbu, and Rob Kass for their comments and suggestions, and Chris Maeda for developing the software for the experiment.

References

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. 1975. *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge, MA.

Daneman, M., and Carpenter, P.A. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Gould, J.D., and Boies, S.L. 1984. Human factors challenges in creating a principal support office system - The speech filing system approach. *ACM Transactions on Office Information Systems, 1*(4), 273-298.

Junker, B. J., Patz, R.J., Huguenard, B.R., Lerch, F.J. and, Kass, R.E. 1993. The Dutch identity and working memory failure: loglinear analysis of a mixture model in cognitive psychology. Working paper, Department of Statistics, CMU

Just, M. A. and Carpenter, P. A. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review. 99*(1), 122-149.