THE BENEFIT TO THE DEAF OF REAL-TIME CAPTIONS IN A MAINSTREAM CLASSROOM ENVIRONMENT

by

Aaron Steinfeld

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Industrial & Operations Engineering
in The University of Michigan
1999

Doctoral Committee:

Associate Professor Yili Liu, Chair
Adjunct Associate Professor Paul Green
Professor Elliot Soloway
Professor Emeritus Daniel Weintraub

DEDICATION

To my grandparents for getting me into science,

To my parents for getting me into art,

And to my brother for getting me into trouble.

ACKNOWLEDGMENTS

I sincerely thank everyone who helped me complete this dissertation. Yili Liu provided funds for payments to study participants. The Center for Ergonomics and Paul Green of the University of Michigan Transportation Research Institute made critical equipment available for the experiments.

Leslie McHenry helped with the fabrication of the stimulus videotapes and Lisa Goldstein helped with the development of the testing procedure. Joan E. Smith of the Services for Students with Disabilities office provided insight on the use of captioning equipment in classroom settings. Finally, my committee members, my family, and friends provided continuous motivation and support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

1 INTRODUCTION

1.1 Overview of the problem

In the near future, speech recognition will progress to the point that portable and accurate processors will be feasible. Additionally, there is an increasing market for software that translates a foreign language to English. It is logical that, at some point, the two will combine and a personal translator will be born. As a forerunner to a mass-marketed product, personal captioning devices for hearing-impaired people are likely to be developed in the field of Assistive Technology. In the past, other technologies that were originally developed as aids to the disabled community became useful in the mainstream population. Such technologies include the telephone, e-mail, optical character recognition, and speech recognition. It is important that human factors professionals be prepared for the likely integration of captioning technology into the daily activities of the whole population.

Even today, real-time captioning (RTC) with human translators is often found at conferences and on television. At movies, personal closed-captioning devices are beginning to be used (Rifkin, 1993). It is also becoming common to find television closed-captioning turned on at noisy bars.

For RTC products to be successful, not only are advances in speech recognition technology needed, but greater understanding of the user requirements. To provide such, the goal of this proposed research is to examine, in depth, the cognitive and perceptual demands such RTC devices will generate and to develop appropriate design principles to aid human factors professionals. It is assumed that the shape and controls for such devices will be accommodated by existing ergonomics design practices. This research is solely focused on the presentation display and its information format.

This research is an attempt to develop a ubiquitous, yet "non-addicting" form of information presentation for RTC. The continuous presence of a readily accessible and easier to assimilate stream of visual information will tempt the user to ignore the speaker's presentation and focus solely on the display. As such, one goal of this research is to provide guidelines for a device that will be primarily used as a reference tool. As a captioner, the device is meant to improve the individual's reception of spoken information by supplying the words that may have been misperceived or not perceived through unaided communication. In translation facilities it is hoped that the device will be used to supplement the user's level of fluency. In both scenarios, a user who ignores the speaker in favor of the information display will not obtain important non-verbal communication.

An extension of the work described here is the imminent integration of speech recognition into hand-held personal computers. Such devices would be of great benefit to professionals (e.g., doctors, lawyers, etc.) and students. One can easily foresee a future where students tote "smart" notebooks to class and rarely write a word. Their notebooks (with speech recognition) will dutifully transcribe whole lectures while the student jots down important drawings or supplemental thoughts. The first students likely to use such devices would be hearing-impaired. Hearing loss exacerbates the previously mentioned addiction problem since the hearing-impaired users will probably attempt to follow the lecture in real time. A similar problem could occur in foreign language classes. It is widely known that the more students exercise their communication skills while learning a language, the more fluent they will become. Thus, it is important that interpreting on a smart notebook be readily accessible for quick reference, yet not be an intrusive element in the immediate task of communication.

The eventual availability of communication aids like the ones described above will require the use of solid human factors principles from the beginning. These devices will need to produce real-time language information in a visual format that is highly accessible yet does not result in most users tuning out the non-verbal communication flow. This constraint is combined with the need to provide the information in a manner that does not impair the primary task. Providing the language information must not interfere with the working memory processes devoted to the user's primary task. Additionally, the perception of the information provided must not increase the workload of the individual to the point of discomfort or a performance decline.

The idea of a real-time voice-to-text translation device for the deaf is not a new one (Gates, 1971; Houde, 1979; McCoy and Shumway, 1979; Stuckless, 1981; Block and Okrand, 1983; and Cutler, 1990). Stuckless (1981, p.292) refers to the concept as the "computerized near-instant conversion of spoken English into readable print." He also describes the possibility of using real-time captioning in the classroom. Furthermore, he points out that text displays of this type are not bound by the same temporal characteristics as speech since there can be a visual buffer of some sort (akin to "instant replay"). This feature should help the device fill the role of a reference tool as missed information is often needed shortly after it is originally presented.

When a deaf user is only paying attention to the text of the captions they are, in effect, participating in a one-way TDD (Telecommunication Device for the Deaf) conversation. This process is devoid of normal non-verbal communication cues such as the gestures, facial expressions, and pictures presented to the recipient during a communication process. These supplements often contain information that has an impact on what the person is saying. In addition, RTC is devoid of any emotion based icons (e.g., ":-)" happy face) that may normally be present in a TDD conversation.

As indicated above, RTC is already used for news broadcasts and certain public speaking events. This method is rather expensive since a specially trained stenographer must be hired. In some cases, special equipment may also be necessary. For speaking events, the captions are usually placed over a video image of the speaker and projected onto a large screen. Students who are hearing-impaired using such systems in college classrooms have indicated that real-time captions lead to greater understanding than signed interpretation (Stinson et al., 1988). In addition, these students rated hard-copy transcripts to be more helpful than notes taken by assistants. However, RTC is usually too costly for day-to-day classroom activities at schools with small populations of hearing-impaired students.

Simpler personal devices are already being utilized as a low cost option. In one application, a fast typist enters spoken material on an extra keyboard (or laptop computer) that is linked to the student's computer. The text on the student's screen appears as captions and is then saved as class notes (Everhart et al., 1996). Anecdotal comments indicate that while such arrangements are generally beneficial, counterproductive effects (e.g., drowsiness) may develop when students ignore the speaker, including non-verbal cues, in favor of the captions (Smith, 1996). However, without further research, it would be difficult to say that the negative effects were strictly due to the presence of captions and not some other factor (e.g., the lecture material). Computer-based notetaking systems attached to television monitors have also been used to provide real-time information ("computer-assisted notetaking," Youdelman and Messerly, 1996).

1.2 Studies on use of captions in educational environments

Speechreading is lipreading with whatever auditory assistance that is available. In a strictly oral environment, the person with a hearing impairment who can use speechreading to follow the conversation may be using a process similar to reading to perceive content. Instead of reading characters from a page, lipreaders read "visemes" (the building blocks of mouth movements during speech, Jackson, 1990). The important difference is the potential loss of information during speechreading. Only exceptional speechreaders are able to perceive spoken words near 90% accuracy.

Good speechreaders use educated guesses to fill in the words they are not able to perceive. By introducing a text-based supplementary information device, the user can refer to another source of information under difficult speechreading conditions. Thus, there should be less demand for working memory resources during the speechreading process (e.g., less guessing). As a result, a RTC device should free up additional capacity for higher level working memory processes. (Appendix A contains a summary of non-text based assistive devices for the deaf.)

The closest approximation of RTC is the use of captions in videotaped educational material. Several studies describe test procedures where captioned presentations were compared to non-captioned presentations.

An early study by Gates (1971) examined the recall of deaf students who watched video presentations (no audio) with all combinations of speaker, signed translation, and captioned formats. The seven different combinations grouped into a superior (combinations that used captioning) and an inferior set. The combination of all three at the same time produced the same performance as the speaker with captions combination. Interestingly the combination of all three inputs did not overload the visual channel.

Other studies have also found that deaf students demonstrate a significant improvement when captions are included. Boyd and Vader (1972) found that deaf students who viewed a captioned program did better on tests than a matched group that viewed a program that was not captioned. Murphy-Berman and Jorgensen (1980) also observed that comprehension increased when captions were included in video presentations to deaf students.

In another study, Nugent (1983) found that both hearing and deaf students performed better on presentations with visuals and captions than they did on presentations with either component alone. Additionally, Nugent states that neither the deaf nor the hearing groups showed unique abilities to learn from the presentations, implying that both groups have similar cognitive processes. This led to a conclusion that the differences in performance between the deaf and hearing groups were due to inferior entry level knowledge for the deaf students. For example, Van Biema (1994) reported that due to communication barriers, the deaf community had a substantially reduced awareness of AIDS and HIV. The improvement by the hearing students suggests that they may also benefit from redundant, captioned information.

The use of redundant, captioned information to improve learning capabilities in learning-disabled hearing students was also examined by Koskinen, et al (1986). The experiment revealed no statistical difference between captioning and no captioning. However, several other findings led the authors to speculate that the captioned presentations may help develop word recognition skills.

In a later study Neuman and Koskinen (1992) examined bilingual seventh and eighth grade students. They found that captioning presentations led to significant advances in word knowledge. They also found higher scores on sentence anomaly tests (detection of improper word usage) for the captioned group. Markham (1989) and Seriwong (1992) also found that English as a second language students exhibited better performance when captions were present.

Based on these studies, it is apparent that the use of captions in classroom environments provide significant, measurable assistance to a wide spectrum of students. These studies also support the basic hypothesis that a real-time captioned environment has the potential to increase the learning potential of all students.

1.3 The role of memory span in verbal comprehension

For learning to occur, information must be comprehended. Thus, it is imperative that some measure of comprehension is utilized to test the ability of different device formats to improve the user's potential to learn. It would also be useful to use a working-memory measure as it will eliminate concerns about bias due to differences in previous exposure to potential test material (as seen in Nugent, 1983). In addition, the ability to use working-memory resource theories to interpret the findings will help in examining performance differences in formats.

Daneman and Carpenter (1980) linked a measure of working memory span with verbal comprehension using a sentence processing task. They exposed their subjects to a series of sentences (either on sequential index cards or on audio tape). The memory span was defined as the number of sentences for which a person can remember the last word. This measure correlated well to verbal comprehension for both reading and listening tasks (from .42 to .86 at $p<.01$ or $p<.05$, depending on the test modality and the comprehension measures). Readers with high verbal comprehension had higher reading memory spans. Later studies showed that higher skilled readers recovered better from sentences with misleading context (Daneman and Carpenter, 1983) and had a better ability to develop inferences (Oakhill, 1984).

In addition, Pichora-Fuller, et al (1995) demonstrated that it is possible to lower this memory span score by making the stimuli more difficult to perceive (e.g., degradation of the signal). They theorized that under difficult conditions perception tasks have an increased demand on working memory resources.

Thus, the memory span score suffers due to lower working memory capacity available for the memory component of the test.

For the purposes of this study, the memory span defined by Daneman and Carpenter (1980) will be used as the baseline level for a memory span. The depressed memory span scores found by Pichora-Fuller, et al (1995) under adverse perceptual conditions can be defined as the effective memory span (EMS). Thus, EMS is the memory span that an individual exhibits under the perceptual conditions to which they are exposed. The EMS can equal the baseline memory span, but will vary as a function of the conditions imposed on the user.

By introducing a reference device to aid the perceptual processes during taxing conditions, the perceptual demand on the working memory capacity may be lowered. This should increase the available working memory capacity for the comprehension processes measured by the memory span. Conditions that reduce the perceptual load should raise the EMS. Thus, by examining the manner in which the EMS differs from the baseline memory span, it will be possible to measure the impact of a particular set of conditions on verbal comprehension.

Using the EMS as a measurement makes it possible to determine the degree to which a set of conditions will affect the subject's comprehension without using standard comprehension tests (e.g., paragraph comprehension). The less time consuming memory-span test should also shorten test sessions, thus reducing subject time in the study. In addition, using the EMS should eliminate problems due to unequal entry knowledge since working memory measures are less susceptible to knowledge bias.

1.4 Working memory, learning, and the deaf

To further analyze the working memory processes that may be occurring during RTC, this study utilizes terminology and definitions used by Baddeley (1992). He describes working memory as having three major components. One of the three is an attention controller called the central executive. The other two components are slaved to the central executive. They are the visuospatial sketch pad and the phonological loop. The visuospatial sketch pad is responsible for manipulating visual images while the phonological loop manages speech-based information. The phonological loop is further divided into the short-term phonological store (STPS) and an articulatory control process ("somewhat analogous to inner speech"). According to Baddeley (1992, p.558), the STPS is commonly perceived as a "backup system for comprehension of speech under taxing conditions" and is not as important when the material is simple and clearly presented.

It is known that prelingual, profoundly deaf students can utilize phonological processes during reading (Hanson and Fowler, 1987; Gibbs, 1989). However, some high school level deaf readers have not displayed use of phonological processes during reading (Locke, 1978). It is also known that the STPS is critical during the learning of new words (Baddeley, et al, 1988; Baddeley, 1992). It is possible that the well known vocabulary deficiency of deaf readers may be due to a poorly developed or utilized STPS. It should be noted that, for deaf subjects, Gibbs (1989) found no relationship between the level of phonological encoding and reading comprehension. Unfortunately, she did not test her subjects' vocabulary.

If deaf readers use none of the possible STPS they may have access to, their reading process loosely mimics the process exhibited by the case study in Baddeley, et al (1988). The authors conducted a series of working memory tests on a woman who exhibited a severe handicap in her short-term working memory. The authors also draw the conclusion that her memory deficiency was most likely isolated to her STPS. In one experiment she was shown to be unable to learn words from an unfamiliar foreign language when they were presented in an auditory manner. In a later experiment, she was able to learn similar words when they were presented visually. However, her performance was dramatically impaired when compared to a matched, unimpaired control group.

As discussed above, it is possible that some deaf people access the STPS to a lesser degree than the norm in the hearing population. The decreased use of the STPS may result in a reduced ability to gain new vocabulary in oral settings. This may help explain the traditionally low vocabulary scores among the deaf. However, in visual settings the ability to learn new words might be impaired, but still feasible (as seen in Baddeley, et al, 1988). Thus, the translation of oral material into text could provide some assistance in developing a better vocabulary.

1.5 The purpose of this work

As described above, there is a solid foundation in the belief that a text-based reference device can provide assistance in an educational setting. It can be hypothesized that for all users, through an increased EMS, the use of such a device will assist building vocabulary, comprehension, and drawing inferences. Furthermore, there is evidence that several factors can have a significant impact on the ability of a person to gain benefit from a RTC device. The primary purpose of this study is to determine the impact of specific RTC factors (e.g., rate of presentation, number of lines of captions) on comprehension. A secondary goal is to identify the key cognitive constructs (e.g., visual buffer) that affect the user's ability to benefit from the device.

To assist human factors professionals in developing design guidelines and a better understanding of working memory activities during RTC, the key questions to answer are:

- Will captioning devices increase the effective memory span?
- Which RTC design factors have an impact on performance? How much of an impact do these factors have?
- What are the key cognitive constructs that affect the user's ability to benefit from the device?

2 PAST RESEARCH ON SELECTED FACTORS

This chapter is an examination of previous research on a selected set of factors that could affect the manner in which a captioning device could be used. These factors can be categorized as display, subject, or information based. There are other possible factors, however, those discussed below have been identified as relevant for the purpose of this study. The factors with their own headings are of particular interest to this study. (Appendix B contains a summary of this section.)

2.1 Display factors

2.1.1 Rate of Presentation

The rate at which the captions should be presented to the users is a somewhat well-researched subject. Shroyer and Birch (1980) addressed this issue by examining the reading speed of students at a residential school for the deaf (reading levels of grades 2-6). They found that the mean for the slowest group of students was 124 words per minute (wpm). In addition, they found that 84% of the students tested had reading rates below that of 159 wpm (the rate of normal speech, Kelly and Steer, 1949 as cited in Shroyer and Birch). With the exception of this paper, the research seems to focus on rates of 120 wpm or less. This is unfortunate because typical speaking rates are near 160 wpm.

The mean of 124 wpm for the slower group is interesting since Braverman and Hertzog (1980) examined rates of captioning at 60, 90, and 120 wpm. They found that elementary and secondary hearing-impaired students showed no significant difference in comprehension across these rates. Braverman (1981) later stated that in a similar study, rates of 90 and 120 wpm were also found to produce equivalent comprehension scores.

In a study of deaf Japanese elementary and secondary students, Okada, et al (1985) found that caption presentation time (3, 5, or 7 seconds; Japanese text) did not have a main effect. In a follow-up study with the same students, caption rate was also found to have no main effects. However, caption rate did have an interaction with developmental level (elementary, secondary). Isihara, et al (1989) commented that the display time should be flexible to allow adjustment to the user's needs but did not state whether the caption rate was a significant factor (in the English abstract, the article has not been translated at this time). Note that the differences between the English and Japanese languages should be kept in mind when considering these results.

In a related paper where the rates were not reported, Kolers, et al (1981) found that people are more efficient readers at rates that scroll 20% faster than their preferred rate of scroll. In addition, a static page of text was read more efficiently than text scrolled at the subject's preferred speed.

2.1.2 Lines/Characters Shown

Braverman (1981) refers to the amount of captions seen as the density of the captions. In her study, she observed that comprehension was equal for low and high density captions (290 and 339 "captions" respectively, no further definition of units provided). It should be noted that she measured comprehension in a manner that is not commonly found in the working memory literature. She provided her subjects with multiple choice questions next to pictures corresponding to the relevant video portion. This technique seemed to be well suited to measuring the comprehension of captioned information that had large variations in the visual part of the recording (e.g., different scenes and backdrops). This technique will not be used in the current study since there is little change in the visual image of a teacher speaking.

While these findings suggest that there should be no effect due to the density of captions, there is a possible explanation for the lack of effects. Braverman's test involved a video presentation that was originally designed to be entertaining. Thus, the possibility that the program contained continuous, detailed, lecture-style monologues is low. A different information format may require caption densities higher than those of Braverman's study.

When a more scholastic form of material (a psychology textbook) was tested for readability on a cathode ray tube (Duchnicky and Kolers, 1983), density, line length, and number of lines shown were all found to have significant effects. This study only provided one source of information, but it is useful for identifying the important aspects of density. The density variable in this study was characters per line, a value that was varied between 40 and 80 characters per line. The higher line density scenario was read 30% faster. This supported the findings of an earlier study that found similar results (Kolers, et al, 1981). The analysis of line length showed that longer lengths (full screen width and 2/3 screen width) were read 25% faster than the shorter length (1/3 screen width).

The finding by Duchnicky and Kolers (1983) that is most useful to the current study was the analysis of the number of lines of scrolling text (1, 2, 4 or 20). No significant difference in reading performance was found between the one and two line formats. The 4 and 20 line formats were also found to have no significant differences. However, the larger group (4 and 20) was found to be 9% faster than the smaller group (1 and 2). Thus, there is a benefit if the number of lines is increased to 4, but not if it is increased further.

2.1.3 Location

Wickens and Carswell's (1995) proximity compatibility principle predicts that the location of the display will have an impact on the performance of the subject. Other research has also identified spatial proximity as one of the factors that affects performance in visual search tasks (Liu and Wickens, 1992; Liu, 1996a; and Liu, 1996b). Larger distances are predicted to increase the cost of a visual search. This relationship applies to the task being examined in the current study. When a particular word is missed in the perception of the primary source (the speaker), the user will probably attempt to search for the word in the display. Thus, there is a component of visual search in this activity.

Flannagan and Harrison (1994) examined the effects of Head-Up Display (HUD) location on performance in an automobile task. Performance was measured by the ability to spot a pedestrian in the road scene and the ability to answer a navigation question. HUD locations at 4, 9 and 15 degrees below the horizon were examined. They found that performance on the pedestrian task significantly decreased as the HUD was moved farther below the horizon. Navigation performance showed only a slight change in performance.

In another study on navigation display formats in automobiles (Steinfeld and Green, 1995), a full field-of-view HUD produced significantly faster subject response times when compared to an instrument panel mounted display. In addition, a small field-of-view HUD located between the road scene and the instrument panel produced subject response times between the other two locations (Green and Williams, 1992; Williams and Green, 1992).

2.1.4 Other Display Factors

Other display factors that are likely to have an effect are caption type and legibility. Video programs that are captioned in advance usually utilize static captions (pop up) that are presented like subtitles in a foreign language film. Due to pauses in the speaker's delivery, this format leads to variation in the information lag during real-time situations. Thus, real-time captions usually utilize a scrolling format (roll up). This allows the information immediately prior to a speaker's pause to be captioned without having to wait for the next words. There will likely be different results due to the semi-continuous movement generated by roll-up captions. This movement may distract the user when they are not attending to the captioned text. A related issue is that text with smoother scrolling movements is read more efficiently than text with instantaneous jumps from one location to another (Kolers, et al, 1981).

Legibility is similar to the signal-to-noise ratio. It is a measure of the increased difficulty for perceptual processors to acquire information. As seen in the study by Pichora-Fuller, et al (1995), a poor signal to noise ratio can result in reduced performance in a memory span test. The increased overhead needed to acquire the information reduces the working memory resources available to the higher level processes. This result can be generalized to situations where the distal stimulus requires more attention to be properly perceived. Also, when the distal stimulus is only partially perceptible, more working memory resources are necessary to fill in the missing information. Thus, it is reasonable to assume that reduced legibility of the text leads to poorer performance.

2.2 Subject factors

2.2.1 Baseline Reading Memory Span

Daneman and Carpenter (1980) reported the last-word memory span measures of 21 native English-speaking university undergraduates. They found that the reading aloud, silent reading, and listening memory spans were quite comparable (means of 2.76, 2.38, and 2.95 last words respectively) and had similar amounts of variability (standard deviations of .80, .70, and .72). They found that the lower boundaries of the memory spans ranged from 1.5 to 2 last words and the upper boundaries varied from 4 to 5 for the different test modes. Due to the structure of their test method, the lowest feasible span is a 1.5. This presents a potential problem for the current study. It will be necessary to modify the test sentences because there must be room for a score worse than baseline. Individuals who would normally have a baseline score of 1.5 under Daneman and Carpenter's method cannot drop to a lower performance level during modified conditions since they are already at the lowest score. Fortunately, the scores can be easily modified by shortening the test sentences. This will result in less memory decay by reducing the time between last words (i.e., reduced overall trial time). Another method is to simply look at the overall accuracy over a uniform number of sentences. (Daneman and Carpenter's experiment used a varying number of sentences.)

The individual's baseline memory span should have an impact on the benefit of the device being examined since the baseline memory span is correlated to the natural aptitude to comprehend verbal information. However, captioning devices may not assist people with different memory spans equally. This relationship will likely be an important factor in the overall performance of the device.

2.2.2 Reading Ability

One of the more commonly known characteristics of deaf students is their lower level of reading performance compared to hearing students of the same age (Conrad, 1977). Conrad reported that deaf high school students have a reading level about seven years younger than the reading level of matched hearing students. In later studies, reduced reading levels of reading were found to negatively affect ability to comprehend captioned programs (Braverman, 1981; Maxon and Welch, 1992).

However, it is important to point out that a greater degree of hearing loss has not been linked to a decreased level of language skills (Maxon and Welch, 1992). In addition, Caldwell (1973) found that students who were exposed to captions that were higher than their reading levels displayed a significant jump in reading level after a five-week period. Interestingly, there was no decrease in the interest levels of the students over this time period. Thus, continued exposure to more difficult reading levels of captions produced a rapid increase in performance.

This is particularly relevant to the previously mentioned work by Conrad. Conrad's paper was written prior to the heavy use of TDD's and captions. The extensive use of captions and heavy television usage by the deaf (Austin and Myers, 1984) should have improved reading performance since Conrad's work. In a more recent study, Murphy-Berman and Jorgensen (1980) reported that hearing-impaired students still have comprehension problems with increased levels of linguistic complexity. Although this work is more recent, it is still not recent enough to generalize to the reading performance of people who were raised on extensively captioned television. A more recent paper by Daneman, et al (1995) found that hearing ability was unrelated to reading achievement in mainstreamed, orally educated hearing-impaired children.

A study by Gibbs (1989) strengthens the argument that there may not really be that much difference in the potential for hearing-impaired students to achieve the same reading levels as their hearing counterparts. The key difference between deaf and hearing readers is probably the use of phonological processes. Gibbs found that the relationship between use of phonological information, as measured by a Cancel E's test, and reading performance was not significant for deaf high school students (average reading grade level of 6.7). A Cancel E's test consists of instructing subjects to cross off all silent E's present in a document. Thus, advanced deaf readers do not need to use the cognitive process that sets them apart from their hearing classmates. Gibbs also stated that good readers display a strong awareness of contradictions and errors regardless of hearing ability.

People with lower reading performance may utilize more of the non-textural sources of information. Maxon and Welch (1992) commented that deaf people with a decreased ability to use captions due to poor reading levels may need to rely on other sources of information. The authors targeted the auditory signal as the possible source of supplemental information when speechreading and other visual cues are not present. Of course the auditory signal would only be useful to those who have the ability to perceive it.

2.2.3 Hearing Ability

Individuals who are hearing-impaired usually have some combination of two different impairments. The first is the simple need for auditory stimuli to be louder in order to perceive them. The second is the degree of distortion between the real and perceived stimuli. One simplistic technique to mimic the latter is to introduce some degree of noise in the auditory signal, as occurs with environmental deafness (e.g., being on a noisy factory floor).

Pichora-Fuller, et al (1995) found that when the signal-to-noise ratio was decreased, subjects were less able to remember the last words in a memory span task. This finding supports those by Hyde and Power (1992) who examined the effect of combinations in audition, speechreading, fingerspelling, and sign (except for the four-way combination). For all the combinations of the conditions, the scores for the severely deaf group were significantly better than those for the profoundly deaf group with the exception of the four combinations that included sign. (The severely deaf have better hearing than the profoundly deaf. Thus, people who are profoundly deaf often have more difficulty differentiating sounds.) The importance of hearing loss is supported by Maxon and Welch (1992) who found that the degree of deafness had a significant effect on comprehension for one of the television programs tested in their study.

Another relevant study is the one conducted by Olsson and Furth (1966). They found that the visual memory span for nonsense forms of deaf and hearing subjects were the same, but that hearing subjects had superior scores for digit span. In addition, when the stimuli were presented simultaneously, the deaf group performed better than the hearing group. Inversely, when the stimuli were shown successively, the hearing group did better. They also suggested that the poor performance of the deaf on the digit span task might have been related to reduced exposure and experience with numbers. If this statement is true, the increased use of TDD's since 1966 (when this study was conducted) has probably improved digit span abilities in deaf users due to the increased use of telephone numbers. It is also possible that the deaf group's poor performance in the successive tasks could have been related to a lower level of reading ability when compared to the hearing subjects. As digit span test was suggested to be affected by experience, the linear, successive nature of reading may have led to higher successive tasks scores for the hearing subjects.

Hearing ability also may be related to the amount of television viewed. Austin and Myers (1984) reported that hearing-impaired college students indicated, on the average, that they watch television 221 minutes a day compared to 164 minutes a day for their hearing counterparts. Shows with captioning were watched more frequently and enjoyed more by the hearing-impaired students. This means that hearing-impaired students are probably extremely familiar with captions and could be described as having an "expert" status. Most hearing people would more likely be classified as novice users of captions.

2.2.4 Sex

The subject's sex is not expected to be a major factor in this research due to suspicion that other factors will have a larger impact. However, sex has been known to show significant differences in other cognitive studies (e.g., Steinfeld and Green, 1995). Thus, experiments will be designed to measure any impact that sex may have.

2.2.5 Other Subject Factors

Other subject factors that may have an effect are English fluency and age. There are several studies that examined the impact of fluency on caption comprehension. Beginning, intermediate, and advanced English-as-a-Second-Language (ESL) students' comprehension performance when using captions was tested by Markham (1989). All levels of students benefited from the presence of captions and seemed to improve at comparable rates. However, Neuman and Koskinen (1992) found that students who were more fluent displayed a greater improvement in vocabulary than their less fluent counterparts. It should be noted that Markham's subjects were college students and 92% of them had lived in the United States for less than a year. Neuman and Koskinen studied bilingual seventh and eighth graders. The difference in benefits gained from captions could be explained by the differences in subject populations.

Past work on spatially based displays (Green and Williams, 1992; Williams and Green, 1992; Flannagan and Harrison, 1994; and Steinfeld and Green, 1995) has shown significant differences between young (college) and older (65 and up) subjects. It is acknowledged that age will probably affect performance in this task independently of hearing ability. However, fluency and age will not be studied so that the scope of the present study can be limited.

2.3 Information factors

2.3.1 Difficulty of the Material

During the process of recording pop-up (non-scrolling) captions, there is the opportunity to edit the text to three specific levels. Braverman (1981) provides a good description of these levels. Level 1, the simplest and most basic, is traditionally used for programs where the viewers are expected to have low levels of reading ability. The sentences are very simple and short. This sometimes frustrates deaf people who have high reading levels since they can see the people on the screen talking and can tell there is a loss of information. Level 3 is the most advanced level and is close to (but usually not) a verbatim representation of the information. Level 2 is between Levels 1 and 3. Most prime-time television shown in the United States are captioned verbatim.

During real-time captioning, there is no opportunity to edit the information in a consistent manner. When the person captioning the presentation (usually a stenographer) is woefully lagging behind the speaker, they will often make judgment calls and edit on the fly. This usually only occurs when the speaker is talking quickly or the vocabulary contains difficult words. Thus, the difficulty of the material cannot be completely controlled for real-time captioning.

In a study of edited captions, Braverman and Hertzog (1980) found that the difficulty of the captions did have a significant effect on comprehension and inferential performance. Level 1 was found to produce higher scores than Level 3 for both performance measurements ($p<.001$ and $p<.01$, respectively). Braverman (1981) was able to show that caption Level was also significant in a later study.

However, in two experiments with deaf Japanese elementary and secondary students, Okada, et al (1985) did not find language level to significantly affect comprehension. They did find a significant interaction between language level and education level (elementary, secondary) in one experiment. The difference between the two studies is most likely due to the differences between the English and Japanese languages and educational systems.

An earlier study by Blatt and Sulzer (1981) reported data from 1,745 hearing-impaired people responding to a mailed survey. When asked why they did not watch the *Captioned ABC News* more often, the most common responses selected by subjects were "Inconvenient time", "Not on in viewing area", and "Prefer another show". (Prior to real-time captioning of the news, the WGBH Caption Center re-broadcasted ABC's *World News Tonight* with captions at 11 PM). The next most common responses were related to information content. "Captions leave out information" was selected 6.0% of the time and "Cannot understand captions" was selected 1.6%. The latter may be due to poor English skills as 31% of the survey population reported an education level between grades 1 through 11.

2.3.2 Time/Sentence Lag

This can be defined as the degree to which the captions are lagging behind the primary information source. This is often defined as the elapsed time (or number of sentences) between the audio track and the corresponding caption text. Braverman (1981, Reference Note 3) described an unpublished study by Braverman and Hertzog on this topic. Subjects showed no change in comprehension when the captions were not synchronous with the audio portion of the program. This could quite possibly be due to low usage of the audio track by the subjects or the lack of a face to speechread.

2.3.3 Other Information Factors

Other information factors that are likely to have an effect are the amount of spatial information and the delivery and pause rate of the speaker. Studies related to display formats (e.g., automobile navigation studies) have traditionally measured people's ability to work with spatial material. The examination of the mix of verbal and spatial information was briefly discussed by Braverman and Hertzog (1980). They found that the information in the pictures of the presentation did not seem to have an impact on the ability of their subjects to understand the captions. This finding is only marginally relevant to the present study in that it addresses the impact on the understanding of the captions. It would be more interesting to know how the mix affected the ability of the subjects to understand the whole presentation.

The rate at which the speaker talks (delivery rate) will likely affect the degree to which a redundant information device is used. This impact will probably be highly dependent on the rate of the display and the ability of the user to comprehend the speaker without the device. If the device is slower than the speaker, the user will be motivated to only attend to one source of information. The loss of information due to a switch in attention would be a large deterrent to switch often. The impact of a fast rate of delivery should be somewhat reduced by a speaker who pauses often (pause rate). These pauses, if they are long enough, will provide the user with time to quickly review what the speaker has just recently said. This should increase performance with the device.

Again, due to the need to limit the scope of this study, these factors will not be addressed in this experiment. Furthermore, all of the stimuli will be verbal material. This is based on the understanding that a vast majority of academic interactions are with verbal material.

3 PROPOSED MODEL OF INTERACTION

3.1 An overview of the model

The usability of a redundant, parallel information device is based on the user's ability to utilize working memory to process information perceived through the primary and secondary sources (e.g., speaker and captions). Working memory is a finite resource used by several perceptual and cognitive processes at once. When the scenario requires heavy utilization of working memory, higher level functions (e.g., the ability to develop inferences) can be starved of the capacity they require. If device usage reduces the demands on working memory, then the higher level functions can access the resources they need.

The use of an assisting device (e.g., real-time captioning) will result in an effective memory span (EMS). The EMS will likely be affected by direct effects (e.g., baseline memory span) as well as the level of two proposed aggregates (buffer and speed). Figure 3.1 describes the hypothesized interactions between the factors, buffer, speed, and the EMS. The relationship between the memory span and learning ability (as measured by verbal comprehension) was described earlier. This model is preliminary. It was developed as a conceptual framework to guide the formation of the experimental design. Considerable research, including the work here, will be necessary to develop a complete and empirically validated model.

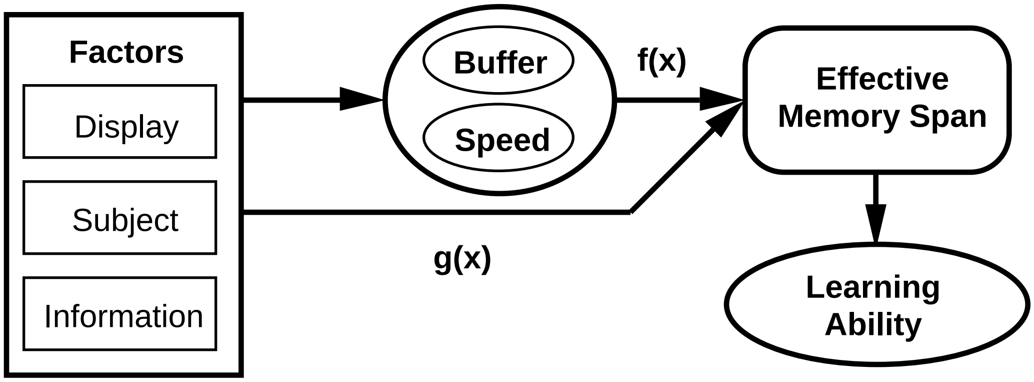


Figure 3.1. Chain of factors

The buffer and speed aggregates are representations of multiple device factors (e.g., number of lines and sentence lag). Aggregates are preferred as examining a single construct rather than several factors allows easier analysis of a potential design. This will permit developers and providers of captioning systems to determine the optimal format arrangements for particular users. In addition these constructs provide direction on the research value of untested format factors.

The buffer aggregate refers to those factors that would increase the size of a visual buffer external to the user (the captions). It is a measure of how much material in working memory can be refreshed at a later time. The speed aggregate describes how long information takes to be presented, processed, and recalled. This construct is akin to the inverse of traditional memory decay as influenced by the scenario. Figure 3.2 presents the two constructs and their hypothesized impact on memory span. The vertical position on the graphs of perception tasks and higher processes in relationship to each other is for the sake of presentation simplicity. Actual locations of the curves in respect to each other could be quite different.

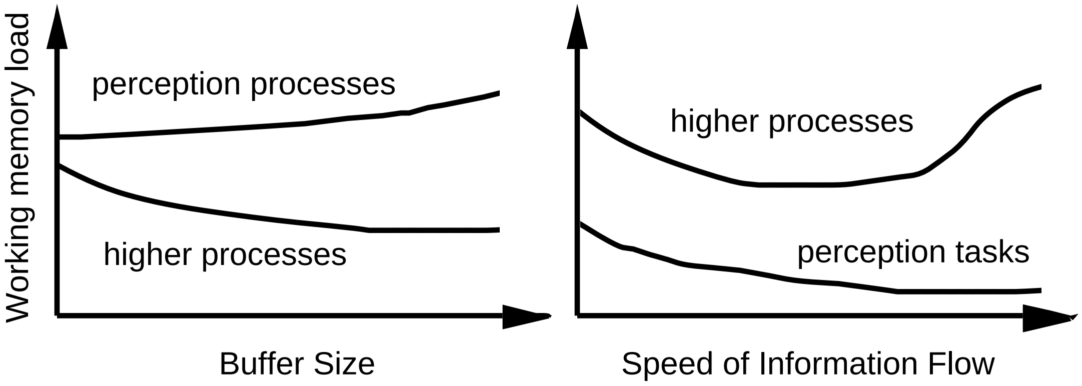


Figure 3.2. Hypothesized effects of buffer and separation
on different output variables

An increase in buffer size is expected to increase perception costs. This is due to the increased cost to search for the target information. The higher processes are expected to require less working memory resources as the size of the external buffer grows. The ability to refresh items into working memory should reduce the demands on memory processes.

At lower speeds, there will be a perception cost incurred due to semantic interference during attempts to recover information lost due to decay. At a certain point, the speed will be fast enough to prevent decay and there will be less information loss and thus less retrieval. After this occurs, the perception costs will level off as the user stops devoting attention to the task of refreshing the lost information. The higher processes will follow the perception tasks as increased decay requires additional load. The more information is lost, the more the user will be required to fill in the blanks. There will be an upward hitch at the higher end when the user reaches the limits of their ability to process information at high speeds.

Figure 3.3 displays the expected recall performance curves for the aggregates. This figure was based on the graphs in Figure 3.2 and incorporates the expected effects described above. The y-axis has been inverted since recall performance increases as working memory load decreases.

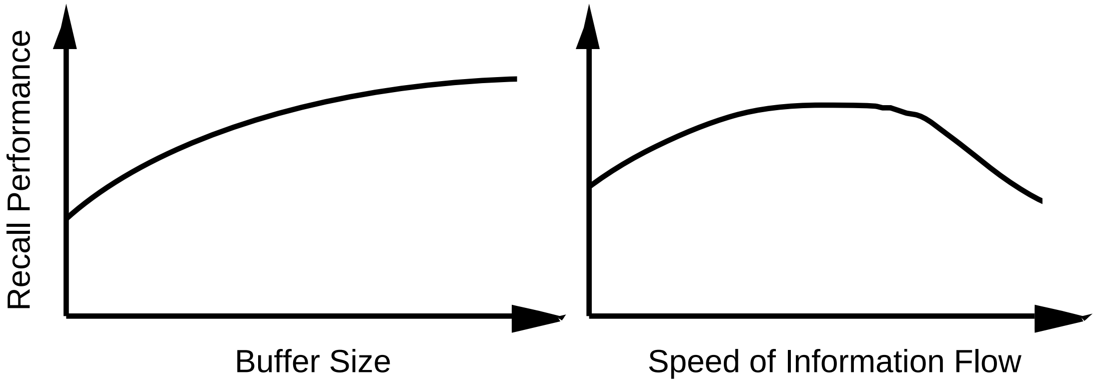


Figure 3.3. Expected performance for the aggregates

By examining the effects of the factors and the aggregates on EMS, it should be possible to develop some guidelines for appropriate design. If feasible, it may be possible to identify the impact of core units for the factors (e.g., presentation rate is a function of time and the number of characters displayed). Using knowledge of the impact of the core units, it may also be possible to predict the aggregates and the EMS for factors not tested in the experiments. Both the aggregates and potential core units will be directly examined in the data analysis.

A potential side benefit of larger buffers is the possibility of an increase in non-verbal cue (NVC, e.g., speaker facial expressions) reception. An increase in buffer size may provide the opportunity for the user to perceive more of the external stimuli (Figure 3.4). There will certainly be less NVC reception at lower buffer sizes due to the time pressure to acquire information before it disappears. The ideal device would maximize the EMS, thus maximizing higher level task performance, while maintaining a high level of NVC reception. Guidelines could be developed by determining the relationship between the buffer size and the reception of NVC's. However, such guidelines will probably require a data collection technique that is different than the one used to monitor comprehension.

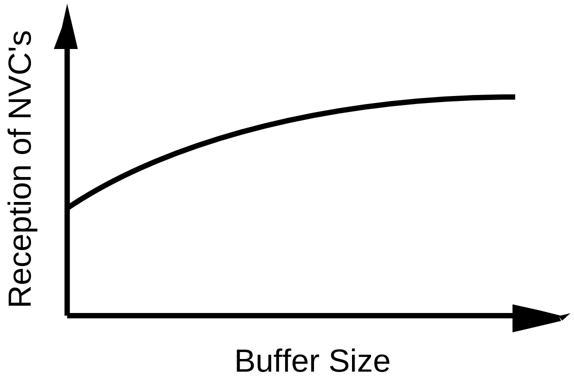


Figure 3.4. Hypothesized improvement in NVC reception due to the buffer aggregate

3.2 The hypotheses for each link of the model

3.2.1 The link between the factors and the aggregates

The impact the factors have on the aggregates can be summarized by the following hypothesis:

Hypothesis (1) The aggregates can accurately describe a set of factors.

This is the manner in which most of the factors are expected to affect the final performance of the device. This hypothesis can be described pictorially in Figure 3.5.

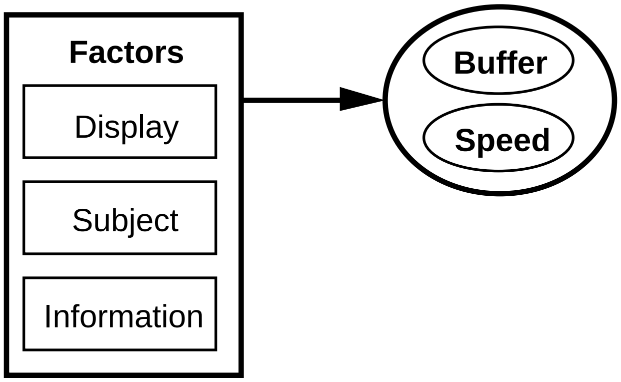


Figure 3.5. The interactions between the factors and the aggregates

The aggregates will be described as functions incorporating selected factors as variables. These functions will be based on knowledge of which factors have an influence on each aggregate. The buffer aggregate will most certainly include the number of lines displayed. The rate of presentation will likely be influential in the speed aggregate. Further discussion of the role of each factor in this model will be addressed after the model is described.

3.2.2 The link between the aggregates and effective memory span

If the factors that have a direct effect on EMS can be identified and their effects measured, then it should be possible to examine the importance of the aggregates. This can be accomplished by controlling the factors with a direct impact while initiating changes in the buffer and speed. With controls on direct factors, the model is simplified to Figure 3.6.

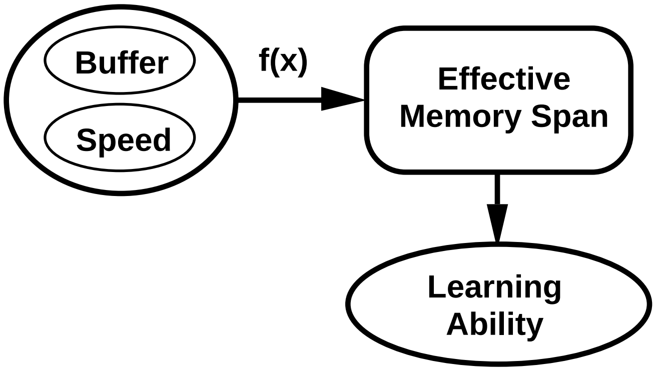


Figure 3.6. The model with direct effects stabilized

The relationship between the aggregates and the EMS will be referred to as f(x), with the aggregates as the independent variables and the EMS as the dependent variable. Due to varying initial baselines, it will be necessary to examine f(x) as it deviates from baseline memory span. (It may be possible to acquire the relative influence levels from a properly configured regression model.)

The possibility of the aggregates having an impact on the EMS, described as f(x), can be described as the following:

Hypothesis (2) Buffer size and the speed of the information flow will affect the EMS.

These aggregates should provide a concise, simple expression of how the device and scenario impact the eventual EMS levels. By their nature, certain subject factors will be difficult, if not impossible, to fold into these aggregates.

3.2.3 The link between the factors and effective memory span

Certain factors may affect the EMS directly. Under steady aggregate levels the model simplifies to Figure 3.7. The direct effect of the factors on the EMS will be described as g(x), a function where the factors that have a direct effect are the independent variables and the EMS is the dependent variable. As previously mentioned, the subject factors will likely fall into this category.

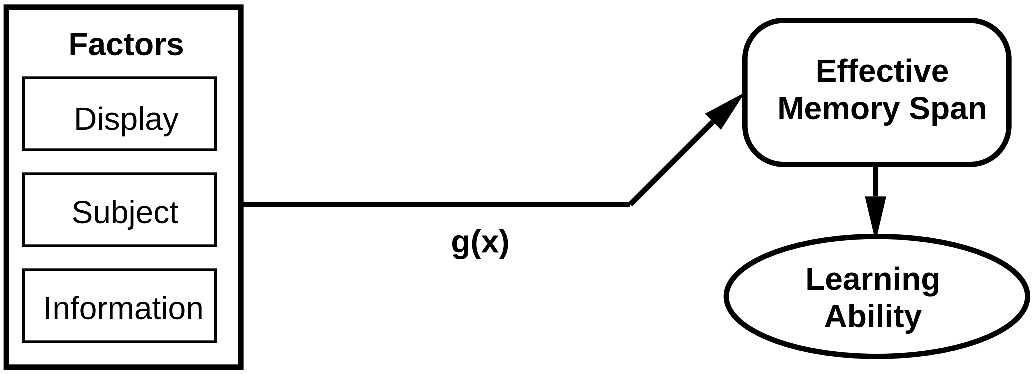


Figure 3.7. The model under steady usage levels

The relationship described as g(x), the direct link between the factors and EMS, can be summarized as Hypothesis (3).

Hypothesis (3) Some factors will directly affect the EMS regardless of the aggregates.

3.3 The impact of the factors on Effective Memory Span

An attempt to make initial predictions of the influence of each factor on EMS is feasible due to the literature described earlier. Some factors are expected to only influence EMS through the aggregates, while others will only apply influence through a direct route. It is possible that some factors will follow both routes.

3.3.1 Rate of Presentation

Past research has shown that elementary and secondary school hearing-impaired students are not affected by modifications in caption rates below 120 wpm (e.g., Braverman and Hertzog, 1980 and Braverman, 1981). These rates could be on the level portion of the speed graph in Figure 3.2. Considering that most speaking rates are probably above 120 wpm, it would be wise to examine the impact of faster rates. The faster rates should approach the limits of the user's ability, where there is a higher demand on capacity by the perceptual processes. At these rates the working memory capacity available to the higher level processors should be reduced. The rates closer to 120 wpm should produce the better memory span scores.

However, the rate will have an impact on how fast the task is completed. With faster rates less time will expire. Thus, less information will be forgotten. Slower rates will incur a higher decay cost due to their inherently longer trials.

3.3.2 Lines/Characters Shown

There are two possible outcomes for this factor. In the first case, the amount of text presented to the subject may be too much and finding target words may be too hard (the number of captions is too big). The high search requirement in this situation may dissuade subjects from using the device. The reverse would occur when a low amount of text leads to the target information scrolling off the screen (the number of captions is too small). Since captions have a fixed-pitch font, the width of the characters is not an issue. The only aspect of this factor that is readily adjustable is the number of lines shown. As described in the reviewed research (Duchnicky and Kolers, 1983), a larger number of lines is expected to produce slightly superior performance.

3.3.3 Location

The location of the information display is expected to have a direct effect on the performance of the device. Performance is expected to rise as the location of the display is brought closer to the primary information source. This is supported by research on Head-Up Displays in the automotive navigation literature (Green and Williams, 1992; Williams and Green, 1992; Flannagan and Harrison, 1994; and Steinfeld and Green, 1995). Another possibility is that when the device is close to the primary information source, the motion of the captioning (or subtitle change) will be distracting and may lead to higher perception costs.

This factor is not expected to impact the EMS through the aggregates. The theoretical linkage between location and buffer size or speed is somewhat tenuous. The most likely possibility would be that larger buffers would reduce performance in conjunction with display locations farther away due to the increased cost to look down and search for information.

3.3.4 Baseline Reading Memory Span

Information acquisition problems may prevent people with high baseline memory spans from utilizing the full potential of their capacity. The work by Pichora-Fuller, et al (1995) suggests that a device that reduces perception difficulty should enable the individual to dedicate more memory span capacity strictly to higher level processes. However, the introduction of the device may have no effect for people with low baselines. (The limited information from the primary source may be enough to completely utilize their memory span capacity.) Another possibility is that the device may be distracting and actually lower the person's EMS.

Since baseline memory span is a subject factor and it has little to do with buffer size or speed directly, it should not be incorporated into the aggregates. However, it is possible that there will be an interaction effect between this factor and the aggregates. If so, two versions of a descriptive model could be developed. The first would include this factor, providing an assessment tool for finding the right set of parameters for each potential user. A second version, without this factor, could provide a generic model that would describe the best set of parameters for the general population.

3.3.5 Reading Ability

The degree to which the user is able to process the text presented on the display will likely affect the EMS. When the reading level is set too high, the user may begin to focus working memory resources on the other sources of information. The level of speechreading and attention to non-verbal cues may increase to accommodate the inability to utilize the text. Inversely, use of the device may increase when the user's reading ability is higher than the level of the text. Since the text will be more easily perceived than speechreading, the user may ignore the speaker in favor of the captions. Due to the wide range in reading ability within the deaf population, this factor could drastically reduce the homogeneity of the deaf subject pool. Thus, it may be necessary to find subjects who are similar in reading ability.

3.3.6 Hearing Ability

A major challenge in studies with hearing-impaired subjects is assembling an adequate sample of participants. If the working memory performance of deaf and hearing subjects is deemed to be similar, hearing subjects may be used in place of deaf subjects. This drastically reduces recruiting difficulty by reducing the number of deaf subjects needed for the study. However, past research suggests that hearing ability has an affect on performance in related studies (Olsson and Furth, 1966; Hyde and Power, 1992; Maxon and Welch, 1992).

There are also concerns with using hearing subjects to predict the performance of their hearing-impaired counterparts. If the sound of the speaker's voice is included in the presentation, the hearing subjects may not use the redundant device. Thus, they will not mimic the deaf subjects in an appropriate manner. The study will also examine how hearing users with electronic notebooks that transcribe lecture material ("smart notebooks") would behave.

It is also important to include an audio track since deaf people who utilize hearing assistance devices (e.g., hearing aids, cochlear implants, etc.) will be able to add supplementary auditory information to the speechreading process. The inclusion of sound will more closely approximate a real classroom environment. However, this will also reduce the level of homogeneity in the deaf subject pool.

The hearing ability of the user is expected to have a direct effect on EMS. It is possible that the hearing and deaf subjects in previous caption studies utilized the captioning device at different levels. This difference could have led to different results. However, the work of Pichora-Fuller, et al (1995) suggests that there will be a main effect based on hearing ability. Difficult perception tasks have been shown to reduce memory span when compared to easier perception conditions. Hearing subjects will also have the benefit of redundancy through the combination of the audio and visual channels. Thus, hearing subjects will probably display higher EMS scores.

It is unlikely that hearing ability has an impact on the buffer aggregate although it is possible that the speed aggregate could be affected by this factor. This would be due to possible differences in short-term phonological store (STPS) usage. However, such effects are difficult to identify due to the confounding effect of redundant information streams for the hearing users. There will no doubt be interaction effects due to the presence or absence of redundant information. Thus, it would probably be best to isolate this factor as one that has a direct effect on EMS.

3.3.7 Sex

As previously mentioned, it is unknown whether sex will have an impact on this study. As a precautionary measure, the experiment should be designed to allow the impact of the user's sex to be examined. This practice is common within the human factors profession.

3.3.8 Difficulty of the Material

From the research on edited captions it is clear that the reading level of the captions could affect the performance of the subjects (Braverman and Hertzog, 1980 and Braverman, 1981). Therefore, the difficulty level of the information presented in a display could affect performance in a real-time captioning condition. This effect would likely be related to the user's reading ability and the rate of presentation. However, the effect may have some independent component as well.

Where the subject's reading ability is below that of the source (and therefore, the caption) information, the subject is expected to show improvement when the device is being used. The redundant supply of information should help reduce the perception costs, allowing the user to devote working memory capacity to the higher level processes that the more difficult information will require. If the reading difficulty is below the user's ability, the impact of the device should be reduced. The easier material will require less working memory capacity for the higher level processes.

As the difficulty of the material is tied to rate of presentation, it would be wise to consider incorporating it into the speed aggregate. The processing of difficult material will likely be strongly affected by information decay.

3.3.9 Time/Sentence Lag

When the captions are lagging behind the initial spoken information, the benefit of the device may be either great or none at all. The user have the benefit of the buffer if they rely heavily on the device. However, they will also likely suffer from semantic interference from the audio track. This interference and/or awareness of potential information loss during a switch between the speaker and the captions could lead the user to avoid using the device altogether.

Due to potential semantic interference from the audio track when reading the buffer material, this factor could be incorporated into the buffer aggregate. It may be wise to include it in the speed aggregate too. The amount of lag could introduce a variable in the decay time between the perception of the two information sources (speaker and captions). Placement of this factor into an aggregate will be deferred until data analysis is conducted since this factor could reside in either one.

4 EXPERIMENT 1 - EFFECTS OF LOCATION AND LINE NUMBER

4.1 Overview

The goal of this study is to examine the validity of using the last-word method to measure effective memory span (EMS) and to examine a subset of factors that may impact the EMS.

The basic procedure for both this experiment and the next involves subjects viewing videotaped presentations with real-time captioning (RTC). The format of the captions and the information were varied. The experiment mimics a classroom lecture where the students do not have the opportunity to ask questions. This situation is easily achieved by using a large television attached to a videocassette player (Figure 4.1). Different display formats can be simulated using monitors masked to show different text formats. The only concern is that a reasonable visual angle be attained for the speechreading portion of the task.

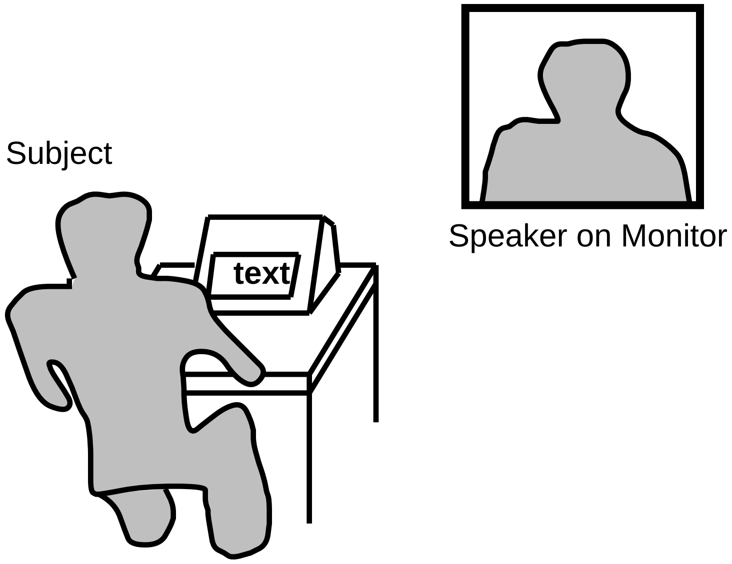


Figure 4.1. An example of a desk located display

This study will examine the impact of the number of lines and location of the RTC display. There has been little research on the location of the caption display. However, there is work in related fields. Studies of navigation displays for motor vehicles have shown that the proximity of the navigation display to the road scene is an important factor in effectiveness (Green and Williams, 1992; Flannagan and Harrison, 1994; and Steinfeld and Green, 1998). In these studies, performance declined as a graphic display was moved away from the driver's view of the road. Other research identified spatial proximity as one of the factors that affects the performance in visual search tasks (e.g., Liu and Wickens, 1992).

The number of lines shown on a captioning display is usually based on trial and error, aesthetics, or the limitations of existing equipment. A study by Duchnicky and Kolers (1983) indicated the number of lines presented on a cathode-ray tube affected reading speed (4 lines was faster to read than 1 or 2). These findings suggest similar results may be found in studies of real-time captioning.

There are also two methodology issues involved in this study. First, it would be beneficial to determine if there are factors that can be ignored in further experiments. The most important is if the performance of deaf subjects can be approximated by hearing subjects. It is not clear if such approximation will be possible, but the logistics of the second study will be much simpler if this is the case. Accordingly, the subject pool consisted of hearing and deaf students with subgroups of women and men. The subgroups allow analysis of the impact of both hearing ability and sex.

Secondly, it is necessary to validate that the variant of the memory span test used in this study produces results similar to the Daneman and Carpenter (1980) task when a parallel source of information is available (e.g., RTC). To test the validity of using the memory span test for RTC conditions, memory span tests will be obtained under (1) a simulated face-to-face condition (hearing and speechreading) similar to Daneman and Carpenter's listening memory span and (2) a captions-only condition that approximates a *scrolled* version of the reading memory span (SRMS). If both of these two measures are good predictors of the RTC conditions, then it may be acceptable to forgo one or both in subsequent studies. It will be important to repeat one of the RTC conditions in later work to allow comparisons of the findings. Duplicating the method should reduce the amount of testing time and allow for more conditions to be examined in subsequent work.

The actual tests will also utilize the Daneman and Carpenter (1980) memory span test. This has already been shown to correlate well with more traditional measures of verbal comprehension. As previously mentioned, the memory span scores were modified by shortening the test sentences and measuring accuracy as a percentage. By doing so, the method prevented "floor" effects for individuals with low baseline memory spans. This test provided a value for effective memory span (EMS).

To ensure compliance with the instructions of the memory span test (process the entire sentence, not just the last word) a simple decision task was added to the test method. The subjects were required to flag sentences where the last word had high context with the rest of the sentence by slightly raising one of their hands. The goal of this modification is to validate that the subject is processing the sentences, not to test comprehension.

The working-memory span test is preferable to a traditional paragraph reading comprehension test. A major concern in comparing performance of hearing and hearing-impaired subjects is the potential for differences in entry knowledge on traditional reading-comprehension tests (e.g., Nugent, 1983). Questions that involve references to pop culture, music, or uncaptioned movies would likely place the subjects who are hearing-impaired at a disadvantage. The working-memory span test increases objectivity, reduces subject testing time, and reduces stimulus fabrication costs.

While the working-memory span test has been used in a variety of single information channel studies (e.g., reading cards), its ability to measure performance in a multiple channel scenario (text, voice, and lip motions) is uncertain. Thus, it was decided to compare performance on two baseline components (the speaker alone and the captions alone) to performance on the real-time captioned scenarios (a speaker with real-time captioning). The baseline scores approximate the reading and listening scores developed by Daneman and Carpenter (1980). If the scores from the RTC conditions are correlated to the baseline scores then it is acceptable to use the last word method. Thus, the questions examined were:

- Is the last-word method appropriate for testing real-time captioned scenarios?
- How do the number of lines and the location of a RTC display affect last-word performance?

4.2 Experimental Method

4.2.1 Materials and Equipment

To maintain similarity to the experiment by Pichora-Fuller et al. (1995), sentences from the Speech Perception in Noise (SPIN) test (Kalikow et al., 1977 and Bilger et al., 1980) were chosen to be the stimulus material for this study. The SPIN test sentences are provided in eight forms of 50 sentences, where each form has a counterpart with the same collection of last words. See Table 4.1 for an example. However, the last words have either "high" or "low" context based on the whole sentence. Each form has an equal number of high and low context sentences, with the counterpart having the reverse context for each last word.

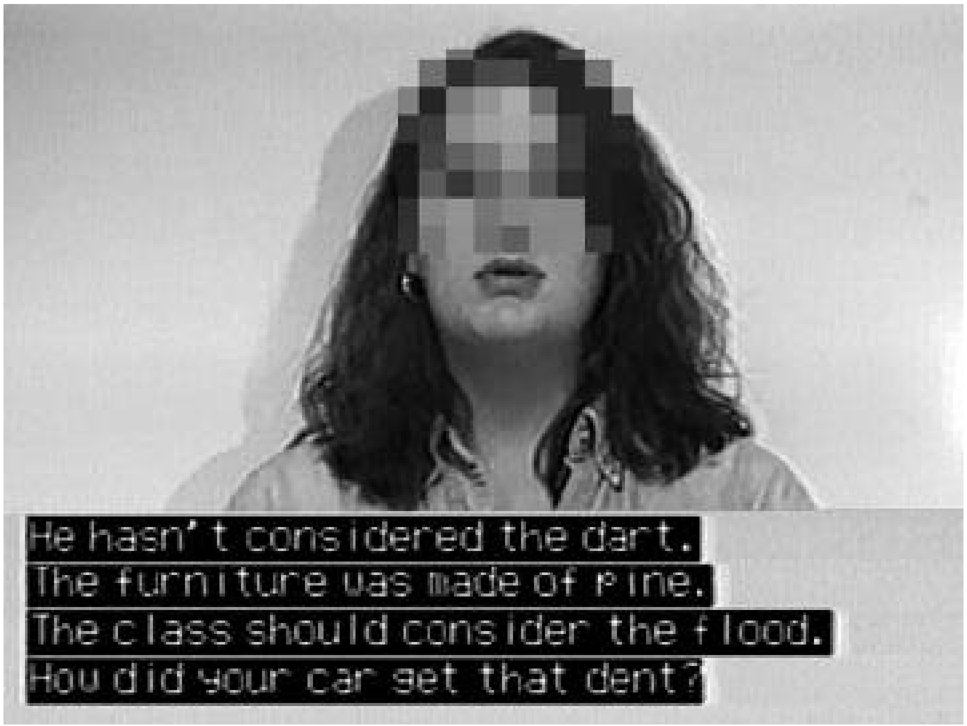
Table 4.1. Some sample sentences with context rating

Examples From Form 1	
Sentence	Context
The old train was powered by steam.	High
He caught the fish in his net.	High
Mr. Smith knew about the bay.	Low

Examples From Form 2 (counterparts to Form 1)	
Sentence	Context
We have not discussed the steam.	Low
Paul should know about the net.	Low
The boat sailed across the bay.	High

For the current experiments six forms of 60 sentences each (360 total) were made from the original eight SPIN forms of 50 sentences each (400 total). To provide the additional 10 sentences per form, sentence pairs were randomly selected from the last two SPIN forms and added to the first six. Semantic links between consecutive sentences were avoided. An equal number of high context and low context last words were randomly positioned within each form. The forms were then broken into sets of 2, 3, 4, 5, and 6 sentences. Three sentence sets were constructed for each set size. The resulting 15 sets for each form were ordered by increasing set size (2, 2, 2, 3, 3, 3, ... , 6, 6, 6).

A special program written in SuperCard 2.5 (Allegiant Technologies, Inc.) was developed to scroll these sentence sets at a rate of 160 wpm (a comfortable speaking speed). A videotape was recorded with the head and shoulders of a person (not the experimenter) reading the sentences at this rate. The speaker's head and shoulders occupied the top two thirds of the screen, while four lines of text scrolling at 160 wpm occupied the bottom third. Each sentence was synchronized with the speaker so that the sentence would scroll up the moment the speaker finished the last word. The end result was a simulated real-time captioning of a speaker with no sentence lag and no typing errors. Tapes were made of all six forms. Figure 4.2 shows a still from one of these tapes (the speaker's face was not masked when the subjects viewed the tape).



He hasn't considered the dart.
The furniture was made of pine.
The class should consider the flood.
How did your car get that dent?

Figure 4.2. A still from one of the tapes

The tapes containing both speaker and associated captions were shown to the subjects on a 33-inch color television and on a 9-inch monitor. The television and monitor were placed in front of the subject at 72 and 27 inches respectively (equivalent visual angle for the text). The monitor was placed on the desk in front of the subject while the television was positioned on a cart against the blackboard. These locations simulated a student sitting in the first row of a classroom. The speaker appeared on the television in all but one condition. The image was approximately life-size at a height similar to a speaker sitting on a stool. In the "Podium" location conditions, captions appeared on the television only as shown in Figure 4.2. For the "Desk" conditions, the captions on the television were masked off with paper and shown on the desk monitor instead. The image of the speaker was always masked off on the desk monitor. Figure 4.3 is a photograph of this arrangement (the television is black and the desk monitor is white). Figure 4.4 shows the plan view of the experiment set-up. The room used was a small carpeted college seminar room with typical classroom ambient lighting and acoustics.

Figure 4.3. The set-up of this study

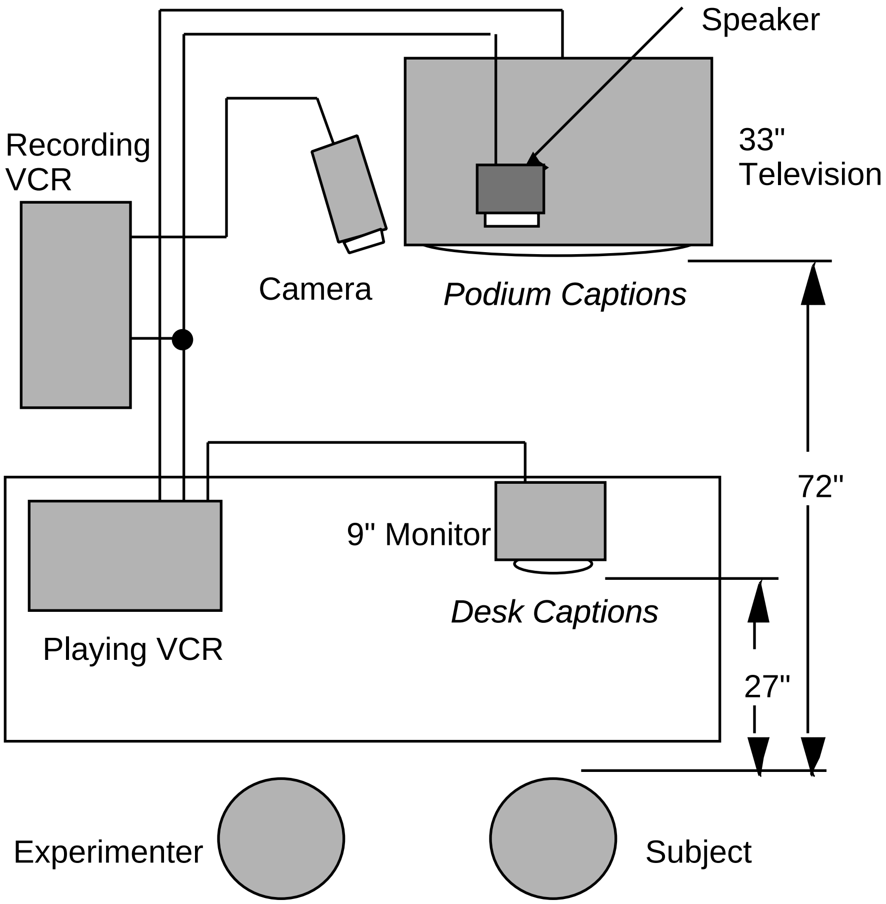


Figure 4.4. A plan view of the equipment layout

4.2.2 Experimental Design

The experiment was split into two parts. The conditions tested in Part I were the baseline conditions (one information source). Part II was for the RTC conditions (permutations of parallel information sources). Table 4.2 shows the six conditions used in this study. The "Captions Only" condition showed two lines of text at the podium location.

Table 4.2. Conditions examined

Part I - Baseline Conditions
Audio + Face
Captions Only

Part II - RTC Conditions
Desk, 2 Lines
Desk, 4 Lines
Podium, 2 Lines
Podium, 4 Lines

The baseline and RTC conditions were counterbalanced within each subject subgroup (Table 4.3). Part I was always run first. Note that the RTC blocks are the bottom three rows of a Latin Square (the study was reduced from 16 to 12 subjects for financial reasons).

Table 4.3. Block counterbalancing

Subject w/in subgroup	Part I - Baseline Block Order	
	1st	2nd
1	Captions Only	Audio + Face
2	Audio + Face	Captions Only
3	Captions Only	Audio + Face

Subject w/in subgroup	Part II - RTC Block Order			
	1st	2nd	3rd	4th
1	Desk, 4 Lines	Desk, 2 Lines	Podium, 4 Lines	Podium, 2 Lines
2	Podium, 2 Lines	Podium, 4 Lines	Desk, 2 Lines	Desk, 4 Lines
3	Podium, 4 Lines	Podium, 2 Lines	Desk, 4 Lines	Desk, 2 Lines

The six forms were counterbalanced over the six conditions (Table 4.4). Again, this was done within subject subgroup. Each form was kept adjacent to its counterpart, but the order was switched for one of the three rows.

Table 4.4. Form counterbalancing

Subject w/in subgroup	Form Order					
	Part I Baseline		Part II Format			
1	3	4	2	1	5	6
2	5	6	3	4	2	1
3	1	2	6	5	4	3

pattern:

b	a	c
c	b	a
a	c	b

4.2.3 Experimental Procedure

After a friendly greeting, subjects filled out consent and payment forms. They then provided biographical information (age and years in college) and rated their familiarity with captioning and interpreters.

The working-memory span test and the simple context decision task were then described to the subjects. First, they were asked to remember the last word for each sentence. Second, upon the completion of each sentence they were to raise their hand if the sentence had a high-context last word. The context decision was added to ensure that the subjects were reading the sentences and not just skimming the last words. Subjects were asked to write down the last words that they recalled upon the completion of each set. An example sequence of conditions is shown in Table 4.4.

Table 4.4. A possible sequence for a subject

Block	Type of block	Information location(s)	Location	Number of lines
0	Practice	On cards	Desk	1
1	Baseline	Speaker & Audio	Podium	0
2	Baseline	Captions	Podium	2
3	RTC	Captions, Speaker, & Audio	Desk	2
4	RTC	Captions, Speaker, & Audio	Desk	4
5	RTC	Captions, Speaker, & Audio	Podium	2
6	RTC	Captions, Speaker, & Audio	Podium	4

To provide practice, a four sentence set was presented to subjects using 3 x 5 inch cards, with one sentence per card. Cards were presented to the subject and then replaced with the next card at a slow, steady rate. After this practice the experimenter checked to make sure the subjects understood the tasks.

Subjects then watched the prepared videotapes. The subjects were not permitted to begin writing until two blank lines had scrolled up. (For the Audio + Face condition, the presence but not content of the captions was visible through the masking paper.) Thus, for the conditions with four lines, two lines were still visible when the subjects were instructed to begin writing. These last two lines were promptly blanked out by substituting a screen menu for the VCR that covered the whole screen. The pause was initially instituted so that subjects would have enough time to read the last line of captions. In retrospect, a single blank line may have been sufficient. The pause between the last spoken word and the scrolling up of two blank lines was consistent across all conditions so that there would always be the same audio and memory decay time.

Following the completion of testing, the subjects ranked the six conditions in order of their preference. Appendix C contains the experimenter instructions and the sequence for each subject.

4.2.4 Test Participants

During recruitment, the participants were informed they would be watching videotapes in a simulated classroom. All participants were paid volunteers and fluent in English. Twelve college students were tested (University of Michigan, 10; Michigan State University, 1; and Eastern Michigan University, 1). Of these 12, six were severely or profoundly deaf (mostly profoundly). The deaf students were all mainstreamed in university classes and were capable of interacting orally with the experimenter with only an occasional finger-spelled word. The pool included both undergraduate and graduate students. All groups were comparable across years of college experience except for the deaf females, who had less. The small deaf student population at these universities presented challenges to recruiting.

As seen in Table 4.5, the deaf students reported having much more experience with captioning than the hearing students. The experience with interpreters was greater for the deaf students than for the hearing students, but only two deaf students rated their use of interpreters as being a 4 or 5. Additionally, one deaf male commented that he thought he was the only student at the university who had real-time captioning by a stenographer on a regular basis.

Table 4.5. Participant information

Hearing Type	Sex	Number	Mean Age (years)	Years of College	Mean Rating of Experience w/ Captioning	Mean Rating of Experience Interpreters
Deaf	Female	3	23.3	1.3	4.3	3.7
	Male	3	35.0	9.3	4.7	1.7
	All		27.0	5.3	4.5	2.7
Hearing	Female	3	27.7	7.2	0.7	0.3
	Male	3	26.3	6.8	1.0	0.7
	All		29.8	7.0	0.8	0.5

(ratings: 0 being never, 5 all the time)

4.3 Results

4.3.1 Is the last-word method appropriate for testing real-time captioned scenarios? (Part I)

Words recalled were scored as accurate regardless of the tense the subjects wrote them in. Recall accuracy scores were calculated as the mean of correct responses for each set size within each condition. A mean of 1 corresponded to all last words for that set size trial being remembered, while 0 meant none were. As previously mentioned, the scores from the Daneman and Carpenter (1980) method were based on a different approach. The need to avoid a "floor" effect led to the measurement technique used in this study.

The Audio + Face condition score for the hearing subjects is similar to the scores developed by Daneman and Carpenter (1980) using spoken material (listening memory span). The major alterations to the task were the addition of the speaker's face, the simple context decision, and a pause at the end of the presentation. It should be noted that the Audio + Face scores for the deaf participants are actually memory spans that are confounded by the subject's ability to perceive oral information (speechreading with audio assistance). The Captions Only condition is a purer measure of memory span for the deaf participants. This is due to the higher probability of accurate perception (reading scrolled text as opposed to speechreading). The Captions Only condition can be described as scrolled reading memory span (SRMS) using the Daneman and Carpenter terminology.

A repeated measures ANOVA comparing recall accuracy of the deaf and hearing students over the Captions Only condition revealed no significant difference between the two groups (p>.40). This finding is important because it indicates that the two subject groups are similar when there are no perceptual differences.

Recall accuracy on the baseline conditions (Audio + Face and Captions Only) and the mean scores across the four RTC (real-time captioning) conditions were correlated. Recall accuracy under the Captions Only and the mean of the RTC conditions was found to have a strong, significant correlation (.83, p<.001). The relationship between Audio + Face and mean RTC condition was also significant (.76, p<.01). However, a plot by participant (Figure 4.5) revealed that performance on Audio + Face was cleanly split by hearing type, reflecting the fact that results for this condition are confounded by the ability to perceive the information. The two baseline conditions were not significantly correlated.

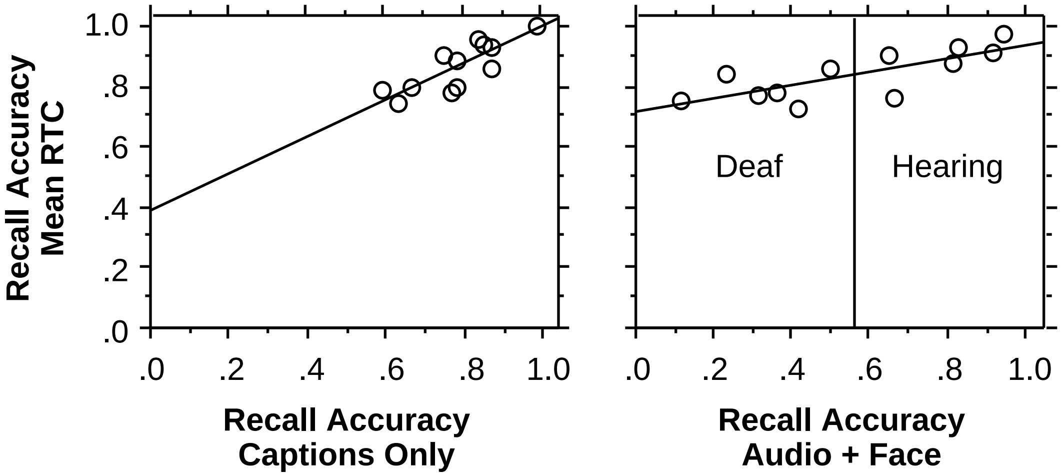


Figure 4.5. Comparisons of mean RTC condition to baseline conditions

Thus, performance on the RTC conditions is similar to performance under the more traditional working-memory span tests (baseline conditions). This is especially true when performance on the RTC conditions is compared to the Captions Only condition. Therefore, the last-word method is appropriate for testing real-time captioned scenarios.

4.3.2 How do the number of lines and the location of a RTC display affect last-word performance? (Part II)

As shown in Figure 4.6, real-time captioning increased recall accuracy. In terms of performance, the two conditions with lowest accuracy for the hearing students were the individual conditions (Face + Audio and Captions Only). The deaf participants displayed a remarkably low average score for the Audio + Face condition as they tried to speechread the person talking. Thus, speechreading clearly incurs high perception and cognitive costs. In addition, there were probably many cases where the deaf participants could not even discern the words.

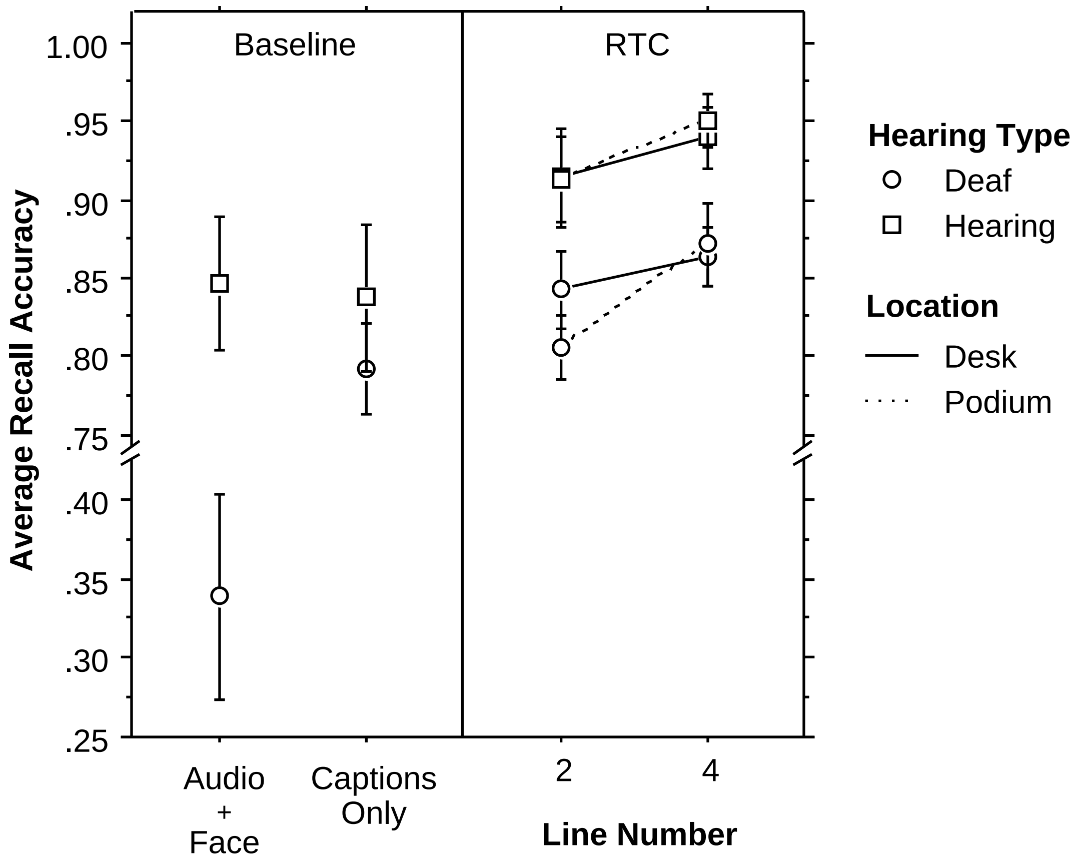


Figure 4.6. Performance of deaf and hearing subjects across conditions (Experiment 1, ±1 standard error)

The somewhat uniform shift between the hearing and deaf students for the RTC conditions seen in Figure 4.6 hints that there may not be interaction effects for Hearing Type by Line Number or Hearing Type by Location. This suspicion proved accurate. A repeated measures ANOVA (Table 4.6) demonstrated that neither interaction was significant (p>.6 and p>.3, respectively). Further examination of Figure 4.6 suggests that there was a Line Number effect but not a Location effect. The ANOVA analysis found that this was indeed true (p<.01 and p>.6, respectively).

Table 4.6. ANOVA for the format conditions

	Main Effect	Interactions			
		Sex	Line Number	Location	Set Size
Hearing Type Deaf, Hearing	+				+++
Sex Female, Male					
Line Number 2, 4	+			-	++
Location Desk, Podium					
Set Size 2, 3, 4, 5, 6	+++				

p<.1	p<.05	p<.01	p<.001
-	+	++	+++

The Set Size factor in Table 4.6 refers to the number of last words the subjects were asked to remember before writing them down. There were no significant 3-way interactions.

Figure 4.7 shows the significant effect for Set Size and the interaction between Set Size and Hearing Type. Note that once the subject was exposed to more than three sentences, the curves depart from near perfect recall. It is apparent that the deaf students displayed a more rapid departure from perfect recall as the set size increased.

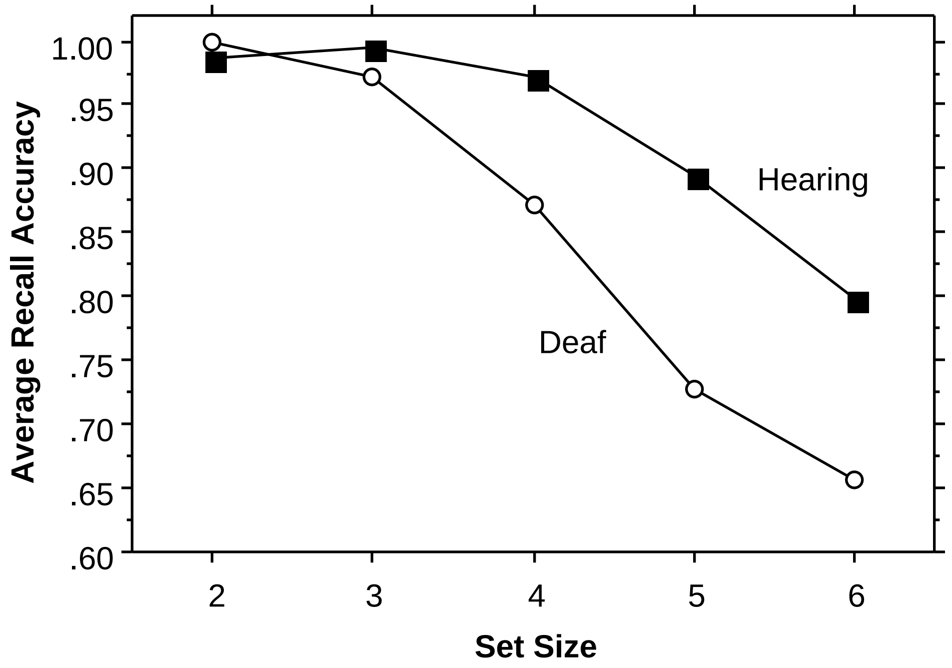


Figure 4.7. Performance differences due to Hearing Type and Set Size

Patterns similar to those seen in Figure 4.7 was also seen for the Set Size by Line Number interaction. For this case, sets sizes with two lines exhibited the more rapid departures from perfect recall. However, this interaction did not display differences as large as the one seen in Figure 4.7.

Due to the nature of the stimulus forms, combining Set Size and Context in the same analysis would lead to imbalanced means. Thus, a separate repeated measures ANOVA was conducted where Set Size was replaced by Context (see Appendix E for ANOVA table). Most of the common effects (those without Set Size or Context) showed similar levels of significance. The two exceptions were the Line Number*Sex (improved to p<.1) and the Line Number*Location (dropped to p>.1) interactions. The effects with Context mimicked their Set Size counterparts. The lone exception was the lack of significance for

Context*Hearing Type interaction which was somewhat surprising since deaf speechreaders are highly reliant on context.

The Line Number*Context interaction (Figure 4.8) showed that performance on low context words improved when more lines were shown; low context had a reduced impact when a larger external buffer was present. This finding is similar to one by Daneman and Carpenter (1983) where higher skilled readers (readers with higher memory spans) recovered better from sentences with misleading context.

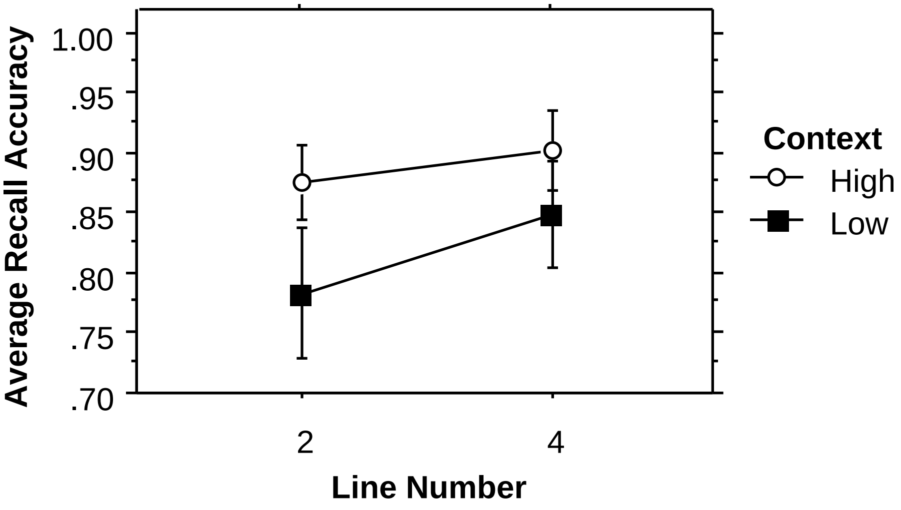


Figure 4.8. The impact of Line Number and Context (with 95% confidence intervals)

4.3.3 Additional Recall Findings (Parts I & II)

An interesting relationship was observed when all the conditions were pooled together. Figure 4.8 shows the interaction between sentence position and Set Size (no statistics were computed for this relationship). Note that "Percentage

Correct" is a raw score with no Set Size influence, thus it is not the same as "Average Recall Accuracy."

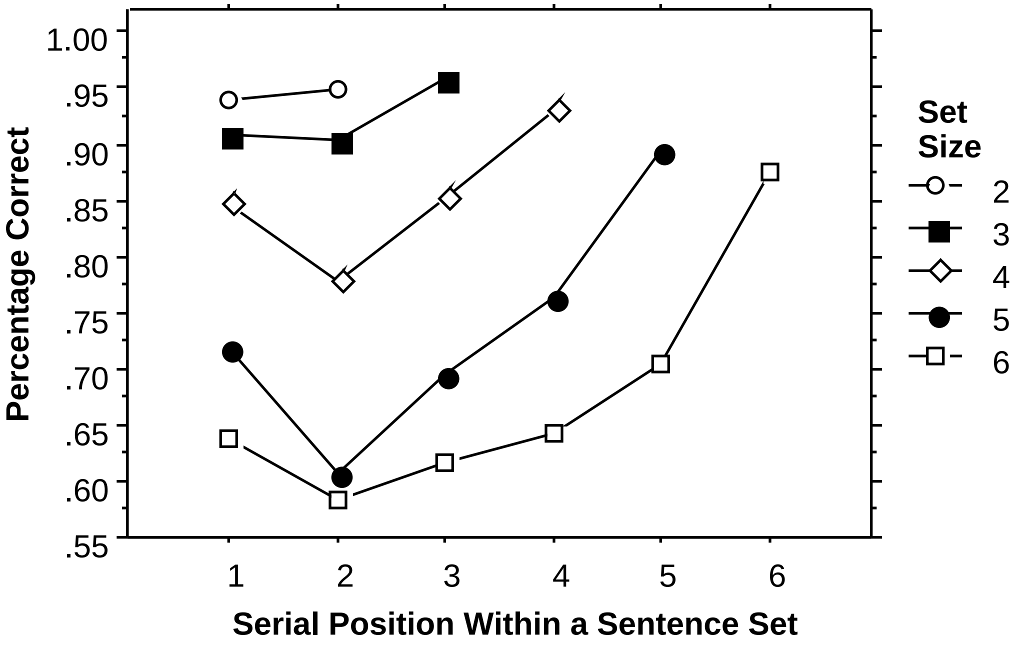


Figure 4.9. The impact of serial positioning within Set Size (all conditions)

Larger Set Sizes led to reduced performance along the serial curves. In addition, the first and last words recalled bracket the interior of the serial curves. These functions are reminiscent of standard serial position curves illustrating both primacy and recency effects (Baddeley, 1986, p. 9). It is quite apparent that the recency effect (the last word in any set) leads to the best recall.

4.3.4 Other Measures

Figure 4.9 displays the results of the subjects' mean preference rankings. The most interesting observation is that the trends are very similar to the recall performance results (Figure 4.6). The preference scores for each subject over each condition (12 x 6 = 72 cases) were significantly correlated with the corresponding memory recall scores (.53, p<.001). It is possible that the subjects were aware of conditions in which they did well and set their preferences accordingly at the end of the experiment. Note that the baseline conditions (Audio + Face and Captions Only) are considered somewhat equal by the hearing subjects. This relationship also seen in their memory span scores.

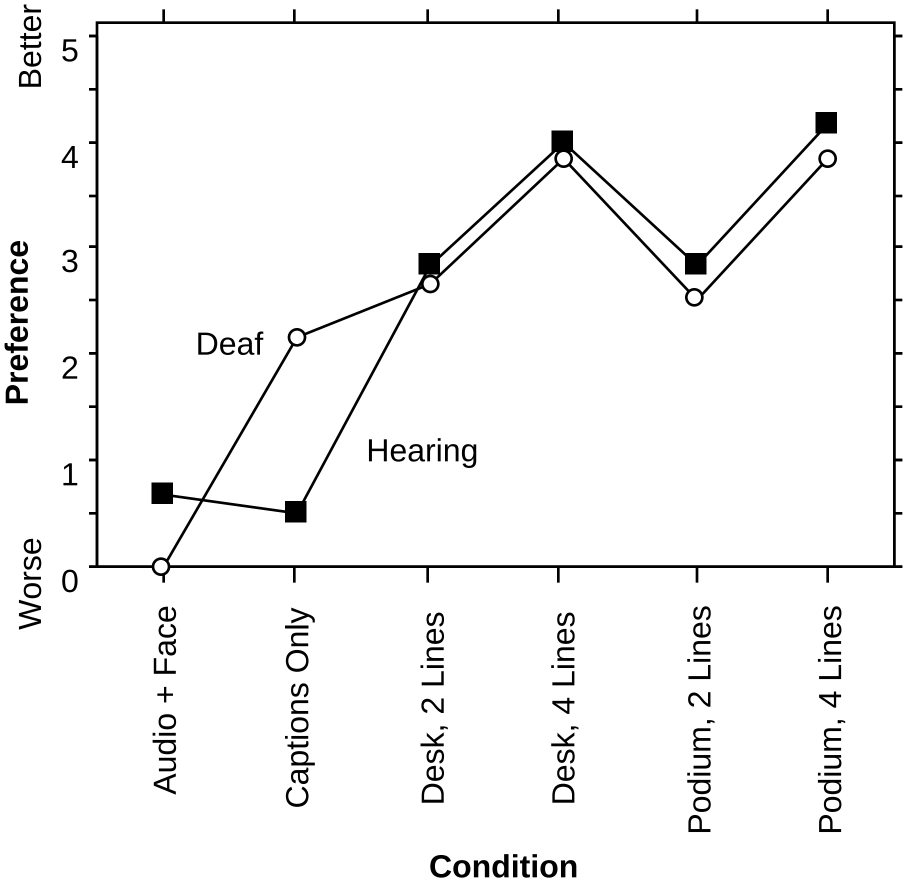


Figure 4.10. Preference ratings

Only a preliminary analysis of subject accuracy on high/low context decisions was done. These decisions were included only to insure that subjects were actually processing the whole sentence, rather than only attending to the last words. Additionally, there were cases where designations of high context by the sentence authors consistently did not match the perception of the subject pool (e.g., *The wedding banquet was a feast* and *Peter dropped in for a brief chat*).

An additional hand-raising measure was added to the context accuracy measure. The experimenter made a note when subjects exhibited a "checked swing" (a mid-stroke reversal of a hand-raising motion indicating they changed their mind). This was also recorded as a low context response. For the memory score, a miss was recorded if nothing was written for a particular word recall. A "hit" corresponded to a word that was recalled. In the event a word was not recalled accurately, it was examined for phonetic (or the speechreading equivalent) matching to the target word. A "close" score was granted if the answer was near the target. Table 4.7 shows the correlations between the measures for the RTC conditions and the traditional classroom condition (Audio + Face).

Table 4.7. Correlations between objective measures

	Context Response		Recall Response			
	Accurate	Checked Swing	Hit	Miss	Close	
Context Accurate		**-0.23**	0.01	-0.01	-0.01	RTC Only
Checked Swing	*-0.06*		0.00	0.00	-0.01	
Hit	***0.25***	*-0.04*		**-0.94**	**-0.20**	
Miss	***-0.18***	*0.02*	***-0.78***		-0.03	
Close	*-0.12*	*0.03*	***-0.33***	***-0.20***		
	Audio + Face only					

note: bold cells are significant at p<.05

If the subject only skimmed the last words, one would expect their Context Accurate score to be low. Since there would be a lower demand on working memory resources, the corresponding Hit score would be elevated. The lack of a significant correlation between these responses for the RTC conditions suggests that subjects did read through each sentence rather than just skim the last words. Thus, subjects performed the memory span task as instructed.

Note that the only strong correlations are those between hits and misses. This is not surprising as inaccurate memory responses take the form of either an erroneous response or a blank response (a miss). The majority of the inaccurate responses by subjects were blanks. It seemed that if a subject could not recall a word, they did not guess very often. The significant correlation between the context and memory scores for the Audio + Face condition was not surprising. A word that was not perceived by a deaf subject would neither be recalled nor flagged as high context.

The subjects' eye movements were recorded on videotape, but not analyzed. The experimenter was sitting immediately to the left of the subjects and was watching them for context decisions (hand motions). As a result, the experimenter had the opportunity to informally observe the subjects' eye fixations. This was especially easy to do for the desk conditions as the angle between the speaker and the text was rather large. Based on these informal observations, subjects tended to ignore the speaker in favor of the captions. This was even true for the hearing subjects. In fact, during a transition from a four line condition to a two line condition, one hearing subject remarked *"Oh no, you're taking away my words."*

Some subjects initially would watch the speaker until the sentence was completed, glance down at the text for redundancy or clarity, and then switch back to the speaker for the next sentence. However, this strategy was often phased out in favor of devoting their visual attention to the captions. Again, this was seen even in hearing subjects. Two subjects (one hearing, one deaf), who were particularly aware that they were failing to remember words, and thus not performing as well as they would like, consciously made this shift in strategy in an attempt to increase their performance. The observation of this strategy change was observed by the experimenter and later confirmed during post-testing discussions.

4.4 Discussion

4.4.1 General Conclusions

There was a clear benefit from real-time captions for both the hearing and deaf students. The hearing students probably benefited from the increase in stimulus redundancy as they were able to utilize both the speaker's voice and the captions. A 9.8% increase in recall accuracy was seen from a traditional presentation (Audio + Face) to the RTC conditions for the hearing subjects. The decrease in perception difficulty was clearly beneficial to the students who were deaf, with a 149.6% accuracy increase from the Audio + Face condition to the RTC conditions. The real world impact of these findings is that providing captions will clearly help deaf students. In addition, the captions will also assist their hearing classmates. This is especially true for rooms with poor acoustics where hearing students are similar to their hearing-impaired counterparts due to environment induced perception problems.

The two groups displayed similar scrolled reading memory spans as there was no significant difference between the two groups for the Captions Only condition. Daneman et al. (1995) reported similar findings where the degree of hearing loss in children did not affect reading memory span. Furthermore, the absence of significant interactions of display format factors with hearing ability for the RTC conditions lends merit to Nugent (1983) who showed that hearing ability does not affect one's ability to learn from captioned presentations.

The amount the subjects were asked to remember was the only factor that interacted with hearing ability. Deaf participants departed from perfect recall faster than hearing subjects for the RTC conditions. This difference is probably due to the greater degree of stimulus redundancy for the hearing students (text, lip movements, and voice as opposed to just text and lip movements).

The analysis also found a significant effect for the number of captioned lines but not for display location. These findings are important because they suggest what type of device should be used. The two locations were picked to simulate a personal system housed in a laptop computer (Desk) and a system installed into a classroom (Podium). Since there was no significant difference, the particular needs of the user and organization can dictate the selection of equipment. It should be mentioned that this study did not include the measurement of non-verbal cue (NVC) reception. A desk location might lead to a reduction in the student's awareness of the speaker's body or facial gestures.

The results clearly indicate that the equipment should display at least four lines of text. Scores on the four line conditions were 4.3% better than those for two lines. The reasons for this probably center on the larger buffer of information (stored on the screen). A larger buffer allows the user to re-read or slow down for difficult portions before the text scrolls off the screen. This result complements the findings of Duchnicky and Kolers (1983). Additionally, their work with a single source of information found no significant difference in reading speed between 4 and 20 lines; thus increasing beyond 4 may not be valuable. In fact, a greater number of lines may lead to longer search times when the student only needs a quick reference. However, the longer search times may be an acceptable sacrifice for a larger buffer. The question of whether to increase beyond 4 lines of text clearly needs further research.

The correlation between performance and preference is another finding that suggests potential for future study. A similar relationship was found by Steinfeld and Green (1998) in a study of navigational aids in automobiles. This suggests that, when given a set to evaluate, the user may be able to subjectively determine the best interface option. The second study should provide additional data on this relationship.

It should be noted that Daneman and Carpenter (1980) found that performance on the working-memory span test correlated well with reading comprehension for both reading and listening tasks. Later studies showed that readers with high scores were also better able to recover from sentences with misleading context (Daneman and Carpenter, 1983). This research, in conjunction with the lack of a significant difference between deaf and hearing subjects for the Captions Only condition, suggests that the two groups have similar capabilities to comprehend written material.

In general, this work supports earlier findings that providing real-time captions improves comprehension for deaf and hearing students. In addition, this work also demonstrates that presentation format cannot be ignored in studies of comprehension of RTC material. The format of information can have a measurable impact on comprehension. Since the number of lines shown was indeed important, real-time captioning displays should have at least four lines of text. However, the placement of the display (desk or podium) does not seem to affect performance. These findings should help determine appropriate display methods for real-time captioning, thus ensuring optimal application of current transcription and new speech recognition technologies.

4.4.2 Future Analysis

It would be valuable to determine if Line Number and Location factors had an impact on the amount of time users spent looking at the captions. The less time the user spends looking at the speaker, the less time they will be exposed to potential NVC's. This may lead to misinterpretations (e.g., facial expressions denoting sarcasm) and boredom. It is also possible that there is a "happy medium" where moderate use of the system leads to better performance. By analyzing looking behavior, it should be possible to examine this issue.

Attempting to measure a subject's usage of RTC through the percentage of time for each set of sentences spent looking at either the captions or the speaker may be misleading. For example, the subject may only switch twice; once from the speaker to the captions when they first appear and then back to the speaker after they have finished reading the last caption. (A preliminary observation of the videotapes suggest this strategy was not uncommon.) The same percentage may be found when a subject switches back and forth between the speaker and the captions. Similarly, counting the number of switches per set may be misleading for the cases where only three or four switches occur (especially for small set sizes). This will not provide information on which visual information source the subject is using. Thus, it is suggested that future analysis of eye movements use a strategy categorization approach. The experimenter would assign each set into a category that describes the way the subject utilized the captions. These categories could be relatively few in number and simplistic (e.g., Always Captions, Primarily Captions, Frequent Switching, Primarily Speaker, and Always Speaker). This would allow for easy analysis and more face validity.

Such an analysis would be useful to complete. However, it should be noted that the current method of videotaping a subject's face is inefficient and somewhat inaccurate. It was necessary to film a large area in order to accommodate subject head movements and posture shifts. As a result, slow motion analysis of small eye movements was difficult since the subject's eyes occupied a very small portion of the whole image. Thus, findings based on this method should be considered preliminary at best. The use of eye-tracking software and hardware in future studies would be more efficient and would provide a more accurate representation of the impact of looking behavior.

4.4.3 Impact on the model

Hearing ability was clearly identified as having an impact on effective memory span (EMS, as measured under the RTC conditions). This seems to be a manifestation of the increased amount of redundancy for the hearing subjects.

Additionally, the lack of an interaction between hearing ability and the number of lines shown suggests that there will not be an interaction between hearing ability and buffer size.

Another subject factor shown to have a distinct impact on EMS was the baseline memory span. As mentioned earlier, using the scores from Captions Only (SRMS) is more appropriate than Audio + Face as there is less information loss due to perception problems. SRMS was shown to have a significant correlation with EMS. In addition, a smooth upwards slope was observed for this relationship. Therefore, SRMS will likely have a positive impact on the EMS, thus satisfying one of the initial hypotheses for the model of interaction.

As previously mentioned, the number of lines shown was expected to influence the EMS through the buffer aggregate. There was clearly an impact with a 4.3% increase was seen from two lines to four. The findings of this study were rather similar to those by Kolers, et al (1981) who found a 9% increase when number of lines was increased (one and two lines vs. four and 20). Additionally, the interaction between Line Number and Set Size suggests that as subjects are asked to remember more, this difference is even more important to the EMS. This finding plus subject comments (e.g., as mentioned earlier, *"Oh no, you're taking away my words."*) suggests that the influence of the number of lines was through some form of an external buffer.

The remaining factors examined in this study were sex of the subject and location of the display. No significant effects were seen for either factor. The lack of a sex effect is not incredibly surprising even though a significant effect has been seen in other cognitive studies. However, the lack of effects related to location is more of a surprise. The angular difference between the two locations was large enough for a difference to be expected based on Flannagan and Harrison's (1994) work on HUD locations. This suggests that subjects did not switch that often. Less switching would reduce the cost of the increased angular distance under the desk condition. Informal observations of the subjects during the experiment led the experimenter to note that, during the RTC conditions, subjects tended to only attend to the captions, regardless of the location. Future studies of switching strategy will help determine if there was indeed a difference across the two locations.

5 EXPERIMENT 2 - EFFECTS OF SENTENCE LAG AND RATE

5.1 Overview

Although the first experiment showed the impact of hearing ability, display location, and number of lines on verbal comprehension in a captioned environment, two important factors were not examined. The rate of the presentation and the lag between the speaker's words and the captions are of key interest to people developing real-time captioning (RTC) systems. Presentation rates faster than 160 wpm (normal speech) need to be studied since the literature does not cover rates above 160 wpm. The literature suggests that speech rates slower then 160 wpm do not have a large effect on comprehension (Braverman and Hertzog, 1980 and Braverman, 1981). In addition, the sentence lag between speaker and caption that is introduced by a captioning system is a concern due to the limitations of available continuous speech recognition systems. Transcription by stenographers also has an inherent sentence lag due to the need for a human to convert speech into keystrokes.

Since the Captions Only condition (scrolled reading memory span) was found to be a good predictor of RTC condition performance it would be wise to repeat this condition in the second experiment. This will allow further examination of the validity of using the last-word method as a test of RTC conditions. In addition, it will facilitate linking this experiment with the first one.

If one of the blocks includes a repeated format condition, it would be possible to test up to five RTC conditions (only six stimulus forms are available). Repeating one of the RTC conditions would be useful as it would show any proportional shift in the benefit of the captions for this experiment's subject pool. This will consume two of the materials forms, leaving four for other conditions.

No interaction was seen between hearing ability and the format of the RTC conditions in the first experiment. It is possible that a similar result will occur in this study. A subset of deaf subjects will allow this topic to be examined. This subset will also permit the examination whether deaf users' performance can be predicted by hearing subjects.

As previously noted in the original presentation of the model, both rate of presentation and sentence lag are expected to produce significant effects. It is possible that the results will say otherwise (as they did for location of the display). Recall accuracy will probably decrease for faster rates due to the decreased amount of time the external visual buffer (the captioned sentence) is on the screen. However, the decrease in decay time may offset this effect. In addition, the faster rates may decrease performance due to time stress within the subject. Following similar logic, the slower rates could lead to the reverse effect. Overall, this will be an interesting result since working-memory theory can predict a performance shift in either direction.

The introduction of lag will probably decrease performance due to higher switching costs and the asynchronous reception of captions and speech. Semantic interference may also creep into the task. Due to memory decay between exposure to each source, performance will likely decrease as the amount of lag increases. An interaction between lag and rate is expected as the slower rates will exaggerate the time between the transient information of the speaker (audio and speechreading) and the buffered information in the captions.

Accordingly, four factor effects will be examined in this study. The subject factors are hearing ability and sex. The display and information factors of rate of presentation and sentence lag will be examined by altering conditions for each subject. Rate of presentation (160 and 200 wpm) and sentence lag (0, 1, and 2 lines of lag) will also be crossed to see if an interaction between the two is present.

Hypotheses to be examined are:

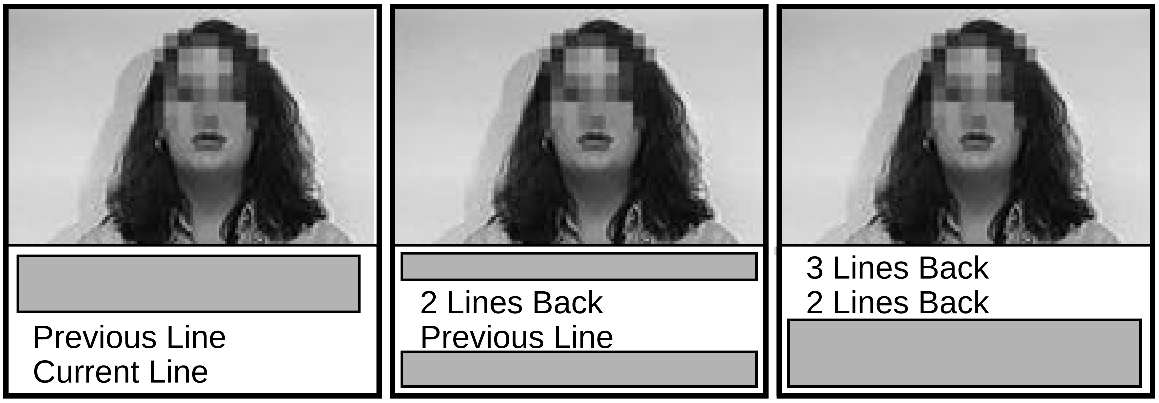
- The last-word method is still appropriate for testing RTC conditions.
- Captions will still improve verbal comprehension for both deaf and hearing students.
- Sentence lag and rate of presentation of the display will affect last-word performance.

5.2 Experimental Method

5.2.1 Materials and Equipment

The materials and equipment used in this study are identical to those used in the first experiment (subsection 5.2.1). Segments at 200 wpm were recorded when the original 160 wpm tapes were made. The smaller monitor used for the desk location was not used as all captions were shown at the podium location.

As previously mentioned, sentence lag was not addressed in the first experiment. Rather than rebuild the tapes for the six forms to include lag, sentence lag was simulated by masking of the captions (Figure 5.1). Unfortunately, this technique has one drawback. If the rate is crossed with the lag conditions, the time of the lag will change. This presents the need to be very careful when looking at the interaction between lag and rate.



No Lag

1 Line of Lag

2 Lines of Lag

Figure 5.1. Simulating sentence lag with masking

5.2.2 Experimental Design

As in the first experiment, the Captions Only condition was examined as a baseline. This allowed five RTC blocks to be run (rather than only 4). The Captions Only condition used in this study was identical to the one used in the first experiment.

Using the masking technique to simulate sentence lag required that the Podium, Two Line condition from the first experiment be the repeated RTC condition. Thus, the 160 wpm, no lag condition was the same as the Podium, Two Line condition from the first study (Table 5.1). The condition of 160 wpm and two lines of lag was run to provide additional information on lag times of over four seconds. This is useful to real-time captioning system designers as it provides an idea of just how much lag is acceptable.

Table 5.1. Conditions examined

Rate (wpm)	Sentence lag (sentences)	Time lag (sec)	Matches to Experiment 1	Part
160	0	0.0	Captions Only (SRMS)	I
160	0	0.0	Podium, 2 Lines	II
160	1	3.0	new	
160	2	6.0	new	
200	0	0.0	new	
200	1	2.4	new	

The baseline and RTC conditions were counterbalanced within each subject subgroup (Table 5.2). Captions Only (baseline) was always run first. The box placed around the middle four rows was the subset used for the deaf subjects. Only four deaf subjects were run due to financial concerns and difficulty finding subjects. Conditions with matching sentence lag were kept together in order to reduce the amount of set-up time needed for switching conditions.

Table 5.2. Block counterbalancing

Subject within subgroup	Format Order (wpm, lines of lag)					
	Baseline	RTC				
1	Cap Only	160, 0	200, 0	160, 1	200, 1	160, 2
2	Cap Only	200, 0	160, 0	200, 1	160, 1	160, 2
3	Cap Only	160, 1	200, 1	160, 2	160, 0	200, 0
4	Cap Only	200, 1	160, 1	160, 2	200, 0	160, 0
5	Cap Only	160, 2	160, 0	200, 0	160, 1	200, 1
6	Cap Only	160, 2	200, 0	160, 0	200, 1	160, 1

Common lines of lag kept together, 2x2 pattern within.

Pattern		
1	2	3
3	1	2
2	3	1

The boxed rows were used for the four deaf subjects.

Like the first experiment, the six forms were counterbalanced over the six conditions (Table 5.3). Again, this was done within subject subgroup. Each form was kept adjacent to its counterpart (1&2, 3&4, 5&6), but a 2x2 pattern was used within each pairing.

Table 5.3. Form counterbalancing

Subject within subgroup	Form Order					
1	1	2	**3**	4	**5**	6
2	2	**1**	4	**3**	6	5
3	5	6	**1**	2	3	**4**
4	6	**5**	2	1	**4**	3
5	3	4	5	**6**	1	**2**
6	4	3	**6**	5	**2**	1

Bold = 200 wpm

Counterpart forms kept together, 2x2 pattern within.

Pattern		
1	2	3
3	1	2
2	3	1

The boxed rows were used for the four deaf subjects.

5.2.3 Experimental Procedure

The procedure used in this study was identical to the one used in the first experiment (subsection 5.2.3) so that the experiments could be pooled for further analysis. Appendix D contains the experimenter instructions.

5.2.4 Test Participants

During recruitment, the subjects were informed they would be watching video tapes in a simulated classroom. Sixteen college students were tested (all paid volunteers and fluent in English). Four were severely or profoundly deaf (mostly profoundly). The deaf students were all in mainstream university classes and were capable of being independent in an oral environment. As seen in Table 5.4, the deaf students reported having more experience with captioning than the hearing students. The experience with interpreters was more for the deaf students than for the hearing students. However, all but one of the deaf students rated their use of interpreters as being below 4. The original plan was for deaf subjects of one sex to be run to reduce potential sex effects. A deaf male subject was substituted for a deaf female subject due to recruitment difficulty. This was deemed acceptable due to the lack of a sex effect in the first experiment.

Table 5.4. Participant information

Hearing Type	Sex	Number	Mean Age (years)	Years of College	Mean Rating of Experience w/ Captioning	Mean Rating of Experience Interpreters
Hearing	Female	6	21.5	4.2	0.5	0.5
	Male	6	22.2	4.4	1.7	1.0
	All		21.8	4.3	1.1	0.8
Deaf	Female	3	22.3	3.8	4.3	2.3
	Male	1	19.0	1.0	5.0	2.0
	All		21.5	3.1	4.5	2.3

(ratings: 0 being never, 5 all the time)

5.3 Results

5.3.1 Is the last-word method appropriate for testing real-time captioned scenarios? (Part I)

Words recalled were scored the same way as in the first study. They were marked as accurate regardless of the tense the subjects wrote them. Recall accuracy was calculated as a percentage of correct responses.

A repeated measures ANOVA comparing recall performance of the deaf and hearing students for the Captions Only condition revealed no significant difference between the two groups (p>.90). This finding is important because it indicates that there are no significant differences between the two subject groups when there are no perceptual differences.

Recall accuracy on the Captions Only and the mean scores across the RTC conditions were correlated. For the hearing subjects, the relationship between

Captions Only and mean RTC condition was found to have a significant correlation (.64, p<.05). The relationship was marginally significant (.48, p<.1) when the deaf subjects were included. Figure 5.2 displays the plots for this relationship without and with the deaf subjects.

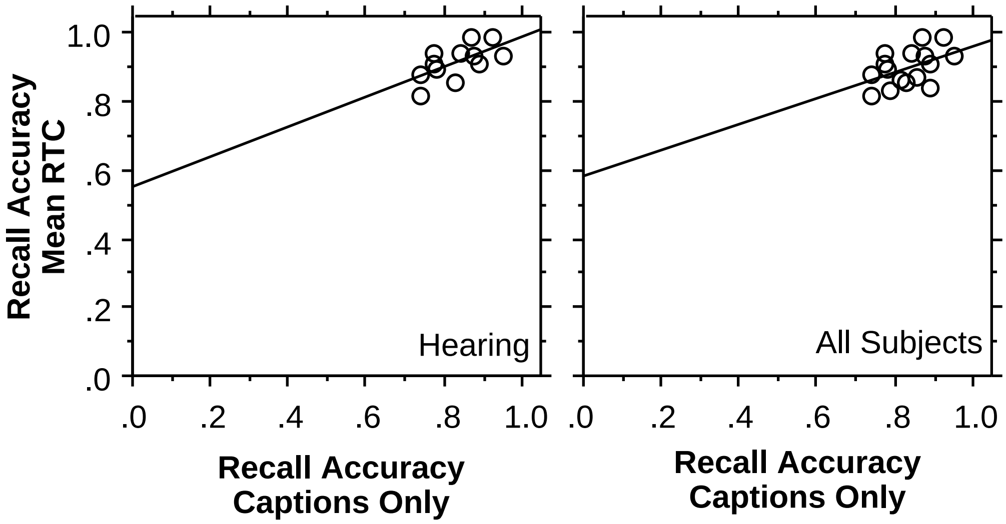


Figure 5.2. Comparisons of mean RTC condition to baseline condition for this study

While the correlation values were not as high as seen in the first experiment (.83, p<.001), they are still quite good for so few subjects. When the data sets for the two studies are merged this relationship shows a strong, significant correlation (.75, p<.001). Thus, performance on the RTC conditions is similar to performance under the more traditional working-memory span test (baseline condition). This confirms the first study's finding that the last-word method is appropriate for testing RTC scenarios.

5.3.2 How does sentence lag and rate of presentation of the display affect last-word performance? (Part II)

The first study showed that real-time captions improved performance for both the deaf and the hearing subjects. Since there was no Audio + Face condition in this study, it will not be possible to make this comment directly. In the first study, the hearing subjects displayed similar levels of performance for the Captions Only and Audio + Face conditions. In this study, the (different) hearing subjects showed improvement from the Captions Only condition (Figure 5.3). Thus, hearing students in this study were similar to their first study counterparts in that they benefited from captions. This statement is more valid for the deaf subjects. In the first study, the additional cognitive load of speechreading, the lower level of perception, and the partial lack of context led to recall of less than 35% for the deaf subjects under the Audio + Face condition. Presumably, the deaf subjects in this study would display a similar level. Figure 5.3 illustrates that their performance is much higher than 35% when captions are shown.

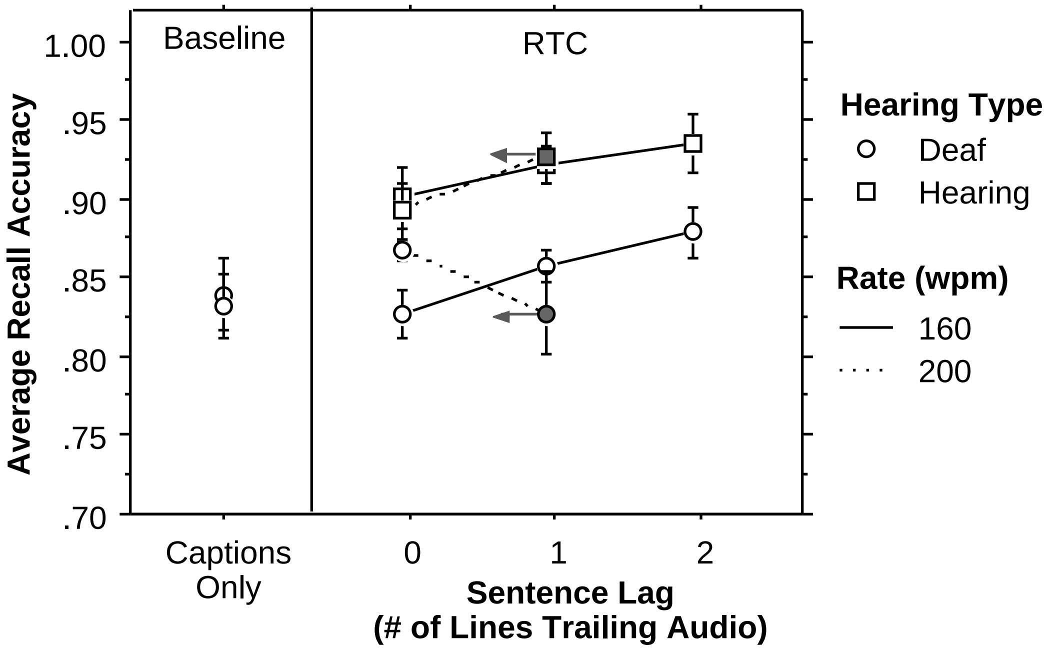


Figure 5.3. Performance of deaf and hearing subjects across conditions (Experiment 2, ±1 standard error)

Lag is represented in Figure 5.3 as Sentence Lag, rather than Time Lag. If Time Lag was used, the 200 wpm, 1 line of lag conditions would be shifted slightly to the right (2.4 seconds of lag as opposed to 3, see also Table 5.1). This is represented in Figure 5.3 as short gray arrows pointing in the direction of the shift. Note that the .6 second shift would be rather small compared to the overall range (0.0 to 6.0 sec).

In the first experiment there was no interaction between hearing type and format effect. This appears to be true for the 160 wpm conditions. However, the 200 wpm conditions do not display a uniform shift similar to the one seen in the first experiment (Figure 4.6). Note that the deaf subjects displayed behavior that is not consistent with their hearing counterparts for 200 wpm.

Due to the structure of the design, the data cannot be examined with a single repeated measures ANOVA over the five RTC conditions. Therefore, two subsets of the data conditions were analyzed separately with repeated measures ANOVA's. Table 5.5 shows the subgroups.

Table 5.5. Subgroup arrangements for ANOVA analysis

<table>
<tr><th>Analysis of Rate and
Rate x Lag interaction</th><th>Analysis of Lag</th></tr>
<tr><td>
<table>
<tr><td></td><td></td><td colspan="3">**Lines of Lag**</td></tr>
<tr><td></td><td></td><td>0</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">**Rate (wpm)**</td><td>160</td><td>■</td><td>■</td><td>□</td></tr>
<tr><td>200</td><td>■</td><td>■</td><td></td></tr>
</table>
</td><td>
<table>
<tr><td></td><td></td><td colspan="3">**Lines of Lag**</td></tr>
<tr><td></td><td></td><td>0</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">**Rate (wpm)**</td><td>160</td><td>■</td><td>■</td><td>■</td></tr>
<tr><td>200</td><td>□</td><td>□</td><td></td></tr>
</table>
</td></tr>
</table>

Table 5.6 displays the ANOVA results for the left cell of Table 5.5. There was one 3-way at $p<.05$ and one 4-way interaction at $p<.1$ (Rate*Lag*Hearing Type and Rate*Lag*Set Size*Hearing Type). As a reminder, there are only four deaf subjects in this study. Findings based on their data should be considered preliminary. Due to the fact that the time lag and sentence lag are not equal as rate changes, a Rate*Lag interaction was expected, but was marginal. The Set Size and Hearing Type*Set Size effects were similar to those seen in the first experiment. For the sake of brevity, discussions of Set Size will not be repeated here.

Table 5.6. ANOVA findings for Rate and Rate*Lag interaction analysis

	Main Effect	Interactions		
		Rate	Lag	Set Size
Hearing Type Deaf, Hearing	**+**			**++**
Rate 160, 200 wpm			**-**	
Lag 0, 1 line				
Set Size 2, 3, 4, 5, 6	**+++**			

p≥.1	p<.1	p<.05	p<.01	p<.001
empty cell	**-**	**+**	**++**	**+++**

The results over the 200 wpm conditions (Figure 5.3) suggest that the significant interaction of Rate*Lag*Hearing Type interaction is due to the performance of the deaf subjects. The inclusion of more deaf subjects may lead to a reduction of significance for this interaction since there are only four deaf subjects in the pool (Figure 5.3). Additionally, there is no Rate*Lag interaction when only the hearing subjects are in the model (Table 5.7). The effects changed slightly when the four deaf subjects were removed. There were borderline Lag and Lag*Set Size effects (p<.1), but no 3-way interaction.

Table 5.7. ANOVA findings for Rate and Rate*Lag interaction analysis (Hearing only)

	Main Effect	Interactions: Lag	Interactions: Set Size
Rate 160, 200 wpm			
Lag 0, 1 line	-		-
Set Size 2, 3, 4, 5, 6	+++		

p≥.1	p<.1	p<.05	p<.01	p<.001
empty cell	-	+	++	+++

As mentioned in Table 5.6, Hearing Type*Rate and Hearing Type*Lag were not significant. Figure 5.4 illustrates these relationships graphically. Note that in most cases the means are almost the same.

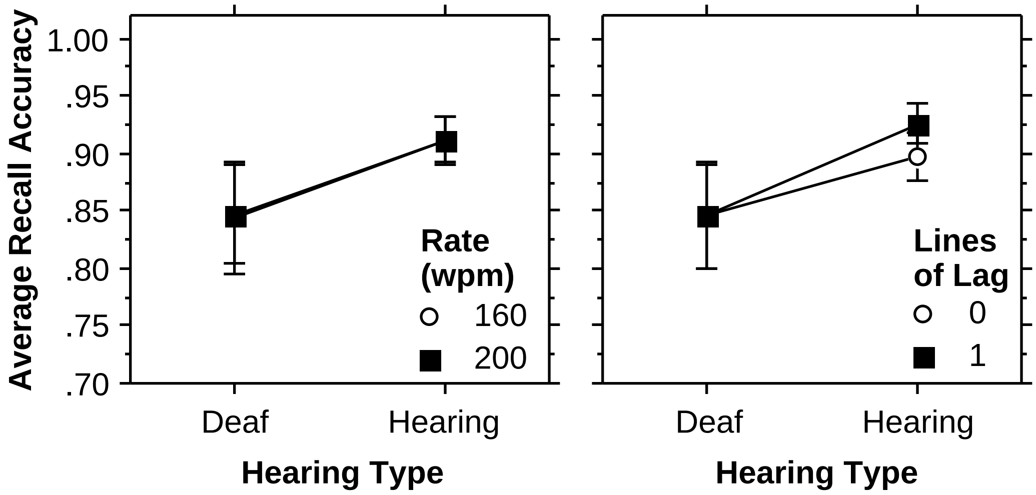


Figure 5.4. The lack of Hearing Type interaction effects (with 95% confidence intervals)

As previously mentioned, there was a borderline significant interaction between Rate and Lag. Figure 5.5 shows that the means were quite comparable, thus this interaction seems to have a smaller proportional impact when compared to other effects.

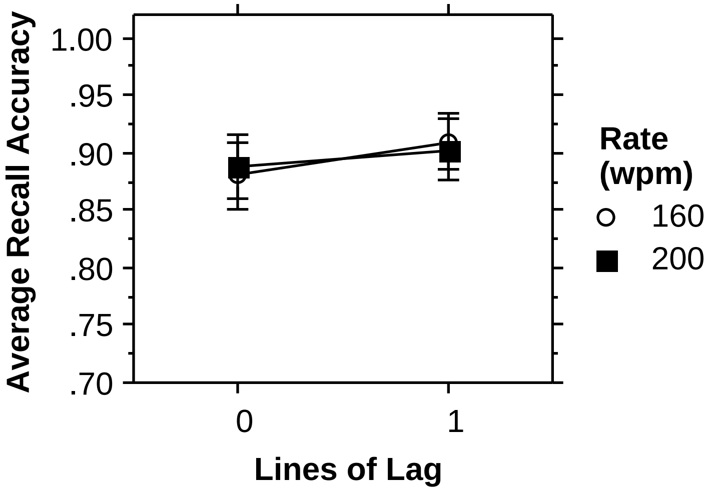


Figure 5.5. The interaction between Rate and Lag
(all subjects, with 95% confidence intervals)

One observation that is certainly evident in Figure 5.3 is that, with one exception, as sentence lag increases, so does recall performance. In addition, it appears that for the 160 wpm conditions, there was no interaction between hearing type and sentence lag. The repeated measures ANOVA findings for the analysis of sentence lag (right cell of Table 5.5) confirmed that there was no interaction between Lag and Hearing Type (Table 5.8). This means that the uniform shift seen in the first experiment seems to be present in the 160 wpm conditions in this experiment. There was no significant 3-way interaction.

Table 5.8. ANOVA findings for Sentence Lag analysis

	Main Effect	Interactions	
		Lag	Set Size
Hearing Type Deaf, Hearing	+		++
Lag 0, 1, 2 line	+		
Set Size 2, 3, 4, 5, 6	+++		

p≥.1	p<.05	p<.01	p<.001
empty cell	+	++	+++

Figure 5.6 shows the performance of all subjects for Lag over the 160 wpm conditions. Note that there is an upwards trend to the data (better recall with more lag).

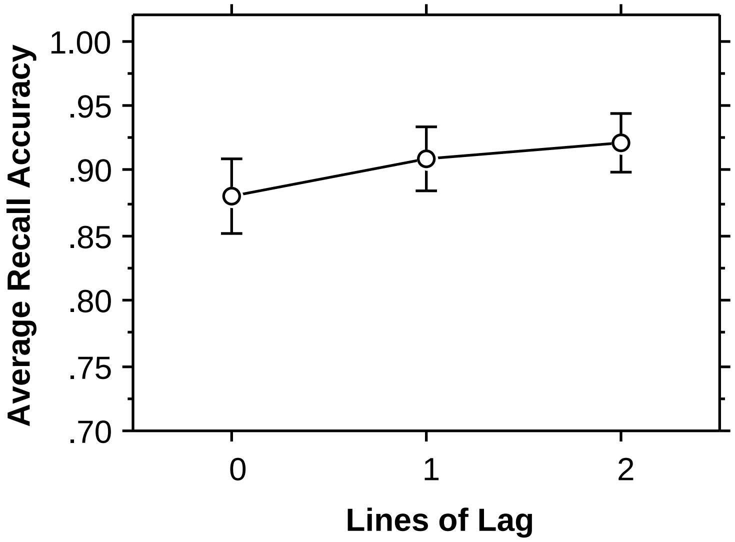


Figure 5.6. The main effect for Sentence Lag

(all subjects, with 95% confidence intervals)

However, when the deaf subjects were removed from the data set, the main effect for lag was no longer present (Table 5.9).

Table 5.9. ANOVA findings for Sentence Lag analysis (Hearing only)

	Main Effects	Interaction
		Set Size
Lag 0, 1, 2 line		
Set Size 2, 3, 4, 5, 6	**+++**	

p≥.1	p<.1	p<.05	p<.01	p<.001
empty cell	**-**	**+**	**++**	**+++**

As a check on the validity of using a male subject in place of a female one for the deaf subjects, all of the repeated measures ANOVA's shown above were run with a Sex effect. None showed Sex as significant.

As in the first experiment, repeated measures ANOVA's were run with Context in the place of Set Size (see Appendix E for ANOVA tables). Effect significance was very similar for the Lag analysis ANOVA (all subjects) with Context taking the place of Set Size for main and interaction effects. The Rate*Lag analysis also showed very similar effect significance with the exception of the Rate*Lag and Rate*Lag*Context*Hearing Type interactions (improved from p<.1 to p<.05).

Figure 5.7 shows the interaction between Context and Hearing Type for the Rate*Lag analysis (the impact was similar for the Lag analysis). While the first experiment did not show this interaction to be significant, it is not a surprise here. Context is important for speechreading and one would expect low context to have a negative affect on the deaf subjects.

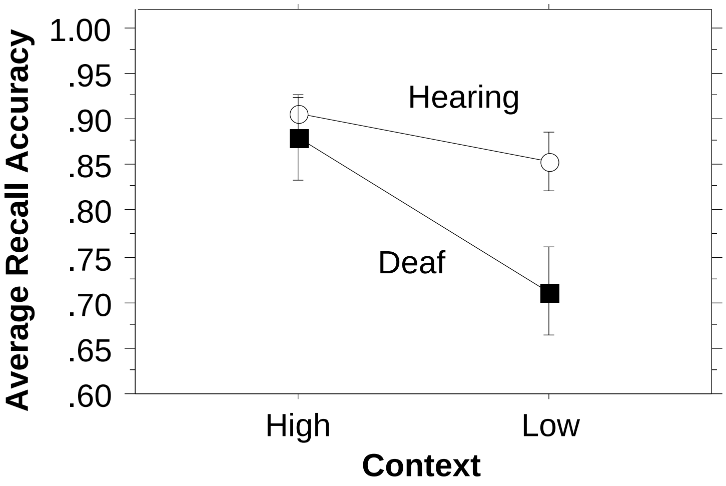


Figure 5.7. The impact of Hearing Type and Context
(Rate*Lag analysis, with 95% confidence intervals)

The significant four-way interaction (Rate*Lag*Context*Hearing Type) manifested as similar performance by the hearing subjects at 160 wpm across both levels of Context (gray arrow in Figure 5.8). Logic and past findings suggest that higher context should lead to higher performance, yet performance was similar to the low context level.

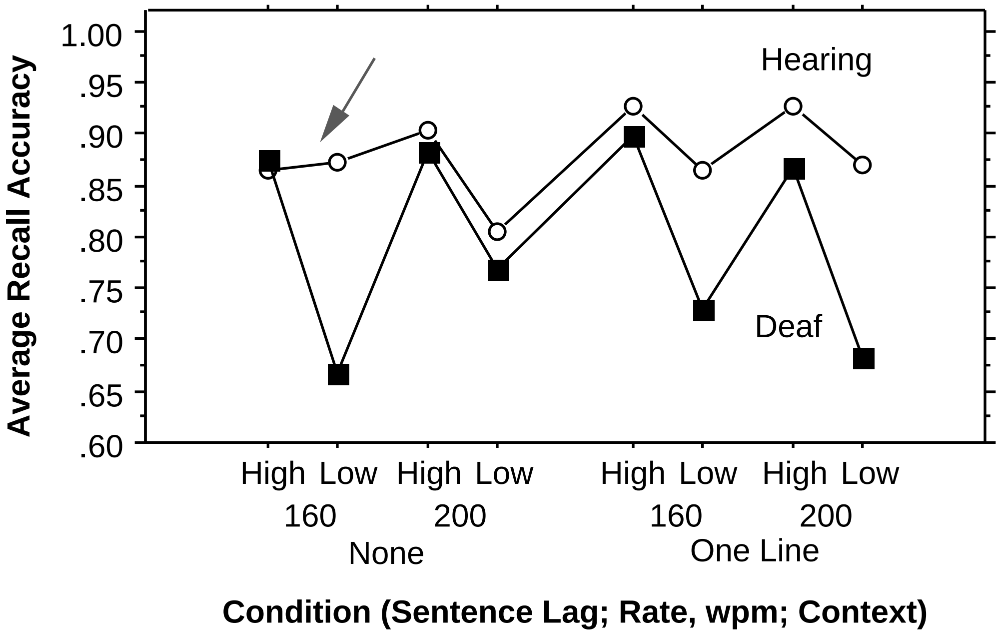


Figure 5.8. The four-way, significant interaction

(Rate*Lag analysis)

5.3.3 Prediction of Deaf Performance from Findings in Experiment 1

It was stated earlier that it may be possible to see if hearing subjects could predict deaf performance under conditions that there was no Hearing Type interaction. Two methods were developed using findings from the first study. These are shown in Equations 5.1 and 5.2. These are referred to as the percent and subtraction methods respectively.

Equation 5.1: $Recall_{Deaf} = Recall_{Hearing} - \left(.117 \times Recall_{Hearing}\right)$

Equation 5.2: $Recall_{Deaf} = Recall_{Hearing} - .11$

The numbers used were based on the separation between the hearing and deaf subjects in the first experiment. Figure 5.7 shows how the two methods compare to the data from the four deaf subjects in this experiment. Note that they seem to be somewhat acceptable for the 160 wpm and the 200 wpm, 1 line of lag conditions, but fail for the 200 wpm, no lag condition.

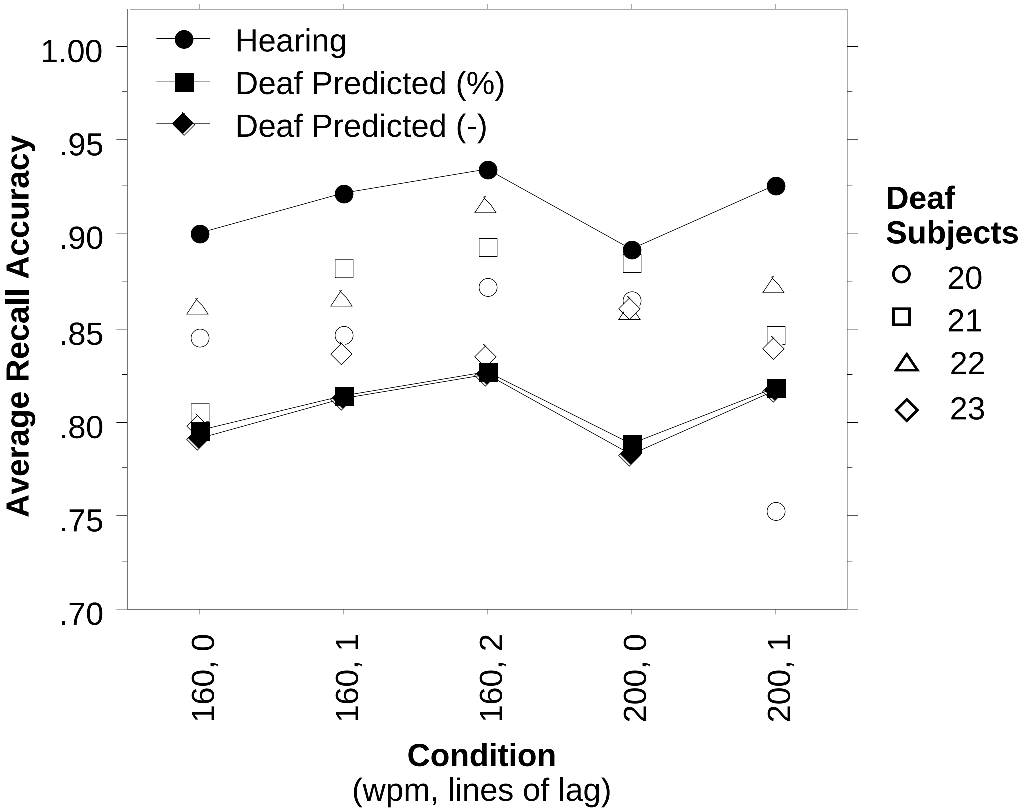


Figure 5.9. Prediction of deaf performance

Statistical comparisons of these predictions to the current deaf subjects are somewhat suspect due to there only being four subjects in this study. A correlation analysis found that there was no significant correlation between the predicted and actual results (.12, p>.6). However, there was a significant correlation (.65, p<.05) when only the 160 wpm conditions were examined.

(The two prediction methods provided identical results in both correlation analyses.) The significant correlation suggests that, for experiments with rates of 160 wpm (typical speaking speed), it is acceptable to predict performance of deaf subjects from the performance of hearing subjects. This is further reinforced by the finding that, under these conditions, there are no Hearing Type interactions with format factors.

5.3.4 Preference Findings

As seen in the first experiment, the preference scores of the hearing subjects (Figure 5.8) displayed an almost identical trend as that seen for recall accuracy (Figure 5.3). Preference and performance were significantly correlated for the hearing subjects (.32, p<.01).

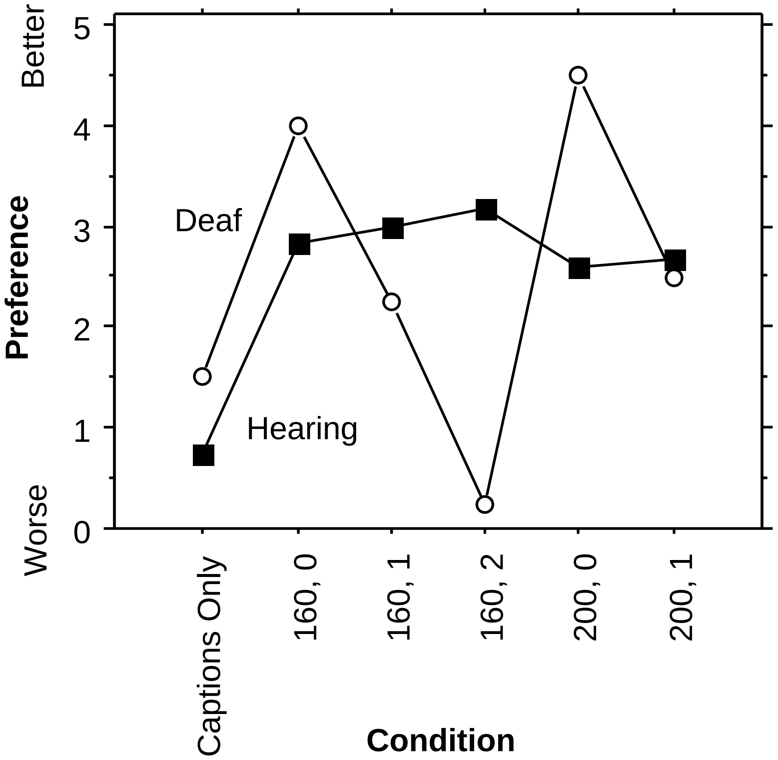


Figure 5.10. Preference findings for hearing and deaf subjects

However, the deaf subjects did not follow this pattern. In fact, the condition that showed the best performance (160 wpm, 2 lines of lag) was the least preferred. Interestingly, one deaf subject commented that this case reminded her of current real-time captioning systems due to the large amount of lag (6 seconds). In general, the deaf students preferred the no lag conditions over any condition with lag. Based on the author's experience with deaf individuals, this finding is in line with complaints heard over real-time captioned television shows with large amounts of lag.

5.4 Discussion

5.4.1 General Conclusions

The results from the second experiment support those from the first. The last-word test is acceptable for RTC conditions and captions improve performance for both hearing and deaf subjects.

The key format finding related to the design and implementation of real-time captioning systems is that for a typical speaking rate (160 wpm), some lag seems to be beneficial. Increasing the rate to 200 wpm appears to not be a major concern, but for deaf users it may erase the gains achieved through greater lag.

The increase in performance for trials with lag was not expected, but is not especially surprising. Analysis from the first study identified the visual buffer (the captions) as a means of refreshing working memory and reducing potential decay. The longer sentence lags would allow hearing subjects to skim the captions at the end of a set for items they forgot without semantic interference from the audio track. The deaf subjects would also gain from the chance to acquire the first word of each set twice; they could speechread the initial sentences that did not have captions and then switch to the captions when they began to scroll up. This would allow them to get redundant information from the initial sentences. As a side note, deaf users will probably have had a great deal of experience with lag from the use of current RTC systems.

There may be some other mechanism at work since the different methods of benefiting from lag might be reflected in a Lag*Hearing Type*Set Size interaction. However, this interaction was not significant in this study. It is possible that these different forms of benefiting from lag provide equivalent performance gains. Such a scenario may help explain the Rate*Lag*Hearing Type interaction. At 200 wpm, the deaf subjects may enter time stress or have trouble filling-in words that were not perceived when attempting to speechread the initial uncaptioned sentences. Due to this, they may not be able to process the last word in time for the next sentence. Thus, the performance gain that would occur under 160 wpm is lost. Unfortunately, this does not help explain the elevated level under the 200 wpm, no-lag condition. Speechreading may be avoided under this scenario due to the faster speaking rate and synchronized captions. Thus, this result may simply be due to the reduced amount of time between reading perception and recall.

As was suggested by the previous study, deaf subject performance can sometimes be predicted from the performance of hearing subjects. These findings should be considered preliminary due to the small sample size of deaf subjects in this study, but it does appear that such a technique can be used under conditions with 160 wpm. Predictions for conditions at 200 wpm should not occur without further research as the results do not suggest such predictions will be accurate.

As in the first experiment, users were able to select the format that is most appropriate for them from a set of options. A significant correlation between performance and preference was seen in the hearing subjects. However, deaf subjects tended to prefer conditions with no lag even though they may have had better performance in other conditions. This may reflect social and classroom interactions experienced by deaf students. In general, deaf people tend to be the last to receive information due to sentence lag from interpreting, captioning, and "phone-chain" interactions (e.g., "What did she say?"). These experiences would likely provide negative perceptions of lag in general. With no lag, a deaf user would be able to stay synchronized with everyone else.

5.4.2 Future Analysis

As a future project, the videotapes collected of the subject's faces could be examined for a rough estimate of looking behavior. Preliminary experience with videotapes from the first experiment suggest that this data will be imprecise. The visual angle between the speaker's face and the captions is rather small. Thus, detection of what the subject is looking at is rather difficult. However, such an exploration may be worthwhile at a later date.

5.4.3 Impact on the model

This study showed that presentation rate might not have an impact on EMS. In addition, it is apparent that any effect rate has is though interactions with Hearing Type and Lag. Thus, it would be wise to either pay close attention to any Lag*Speed interaction or consider folding lag into the speed aggregate. It was also observed that lag may be elevating EMS through a manipulation of the buffer. The inclusion of lag into the buffer aggregate would be valuable as well. This leads to a possible problem.

By including lag in both aggregates, there is the potential for problems to crop up in regressions or other statistical analyses. For example, including lag in both aggregates may result in lost degrees of freedom. It may be wiser to place lag in the buffer aggregate and limit this factor's impact on the speed aggregate to any potential Buffer*Speed interaction. This would probably be more appropriate due to the marginal Rate*Lag interaction observed in this study.

6 COMBINED ANALYSIS OF THE EXPERIMENTS

6.1 The link between the experiments

Both experiments used the same equipment, materials, and protocol. In addition, the subject pools were drawn from similar student populations. As a means of linking the two studies for analysis of a larger sample, two identical conditions were run in both studies. The baseline condition of Captions Only was the repeated baseline, while the common RTC condition was Podium, 2 Lines (160 wpm, no lag). A repeated measures ANOVA for recall accuracy found no significant difference due to the experiment for the Captions Only (p>.6) and the common RTC (p>.9) conditions. Thus, the two subject pools were essentially equivalent across the two conditions (Figure 6.1). Note the rather narrow range on the y-axis.

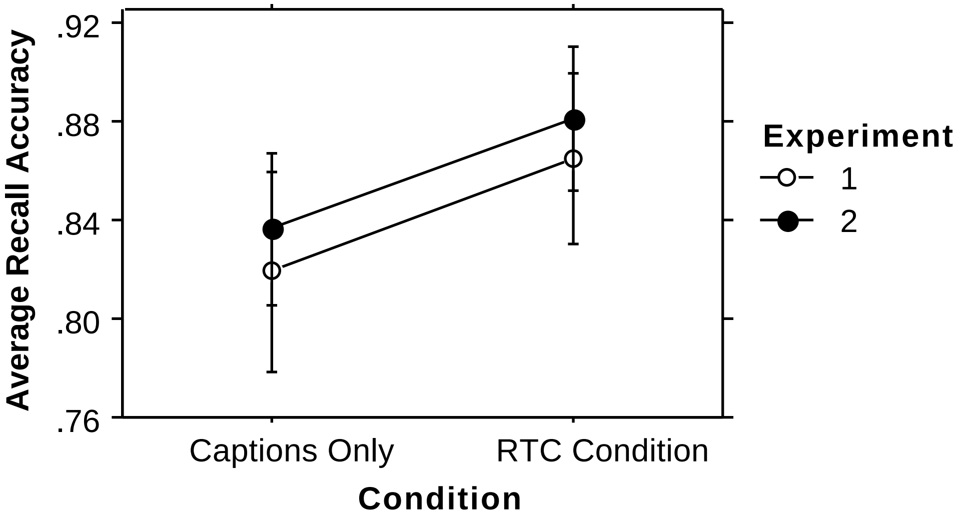


Figure 6.1. Performance the common conditions for each study (with 95% confidence intervals)

The RTC parts of the two data sets were combined. This new data set cannot be examined appropriately using repeated measures ANOVA's due to the incomplete cells (e.g., there were no Desk, 200 wpm conditions). Thus, regression and correlation analyses were used to examine the combined data set.

6.2 Findings with individual factors

An ideal outcome of analyses of the pooled data would be the discovery of useful performance prediction formulas. These would allow potential users and manufacturers to examine what format combinations are ideal. These formulas should have two forms. The first is an assessment tool that would be a user-dependent prediction that would allow optimization of the format arrangement for a particular person. The second would be a more generic, user-independent formula that would suit a broader population. A series of regressions were run on the pooled data set to determine the two sets of expressions. The baseline conditions were removed from the pool since they do not describe RTC scenarios but Captions Only performance (Scrolled Reading Memory Span) was applied as a subject factor for assessment analyses. In addition, location and sex were omitted due to lack of significance.

Table 6.1 shows the ability of some different models to predict the observed data. Each "Full" standard regression was used to develop a backwards stepwise regression counterpart. The mean number of words recalled was used as the dependent variable for the models with Set Size ("Size" in the table). This was computed using average recall accuracy and set size. Mean words recalled was chosen as it allows a more intuitive comparison to the observed data. Since the number of words recalled is dependent on Set Size, the average recall accuracy was used for the models without Set Size as a factor. The factors in italics in the "Significant Effects" column were at a significance level of $p<.1$ while all others were $p<.05$ or better. Hearing Type ("HT" in the table) was kept in the generic models due to the fact that RTC systems will likely be deployed primarily to benefit deaf users. Including Hearing Type in the model will also assist developers in determining the difference in performance between the deaf and hearing users. LagTime is the amount of lag expressed in seconds. The R^2's observed when lag was expressed in number of sentences were not as high as those in Table 6.1. A natural log was applied to Set Size due to the shape of the data when plotted by Set Size. This modification makes sense since memory performance decreased when subjects were asked to remember larger sets (i.e., decay). This was confirmed when regressions without a log applied to Set Size displayed lower R^2's.

Table 6.1. Regression with individual factors

		R^2	Effects in model	Significant Effects
Full w/ Size	Assessment (w/ SRMS)	.831	main and 2-way using: ln(Size), Lines, Rate, LagTime, HT, and SRMS with confounds removed	ln(Size), *HT*, ln(Size)*Lines, ln(Size)*LagTime, ln(Size)*HT, ln(Size)*SRMS
	Generic (w/o SRMS)	.788	main and 2-way using: ln(Size), Lines, Rate, LagTime, and HT with confounds removed	ln(Size), ln(Size)*Lines, ln(Size)*LagTime, ln(Size)*HT
Backwards Stepwise w/ Size	Assessment	.830	ln(Size), Lines, LagTime, HT, SRMS, ln(Size)*Lines, ln(Size)*LagTime, ln(Size)*HT, ln(Size)*SRMS	ln(Size), Lines, *LagTime*, HT, SRMS, ln(Size)*Lines, ln(Size)*LagTime, ln(Size)*HT, ln(Size)*SRMS
	Generic	.787	ln(Size), Lines, LagTime, HT, ln(Size)*Lines, ln(Size)*LagTime, ln(Size)*HT	ln(Size), Lines, *LagTime*, HT, ln(Size)*Lines, ln(Size)*LagTime, ln(Size)*HT
Full w/o Size	Assessment	.553	main and 2-way using: Lines, Rate, LagTime, HT, and SRMS with confounds removed	none
	Generic	.336	main and 2-way using: Lines, Rate, LagTime, and HT with confounds removed	Lines
Backwards Stepwise w/o Size	Assessment	.544	Lines, LagTime, HT, SRMS	Lines, LagTime, HT, SRMS
	Generic	.335	Lines, LagTime, HT	Lines, LagTime, HT

Note that the removal of Set Size dramatically reduces the ability of the models to describe the data. This reduction is probably not due to the use of recall accuracy as a dependent model for these models since the mean number of words recalled is simply recall accuracy multiplied by Set Size. It is more likely due to the loss of data points due to the compression of the data set when Set Size was eliminated as a factor.

The backwards stepwise models seem to be rather effective when compared to their full model counterparts. For the sake of simplicity, it makes sense to use the backwards stepwise prediction equations when developing prediction formulas. This will reduce the number of terms in the equation and should make future computations less error prone. The prediction equations for the backwards stepwise models can be found in Appendix E.

One interesting observation related to the prediction formulas is the tendency of models that use Set Size to provide unrealistic numbers at the lower set sizes (better than perfect). Figure 6.2 shows a graph of a prediction equation (backwards-stepwise with Set Size model) and the actual data collected. The diagonal line splits the graph into two regions. The top region is best described as "unrealistic" as points in this region indicate more words were recalled than were in a memory set. Note that the prediction formula crosses into the unrealistic area. Using a simple logic filter (Equation 6.1), it is possible to "clip" off the unrealistic region of the prediction formula. This is shown on the right in Figure 6.2.

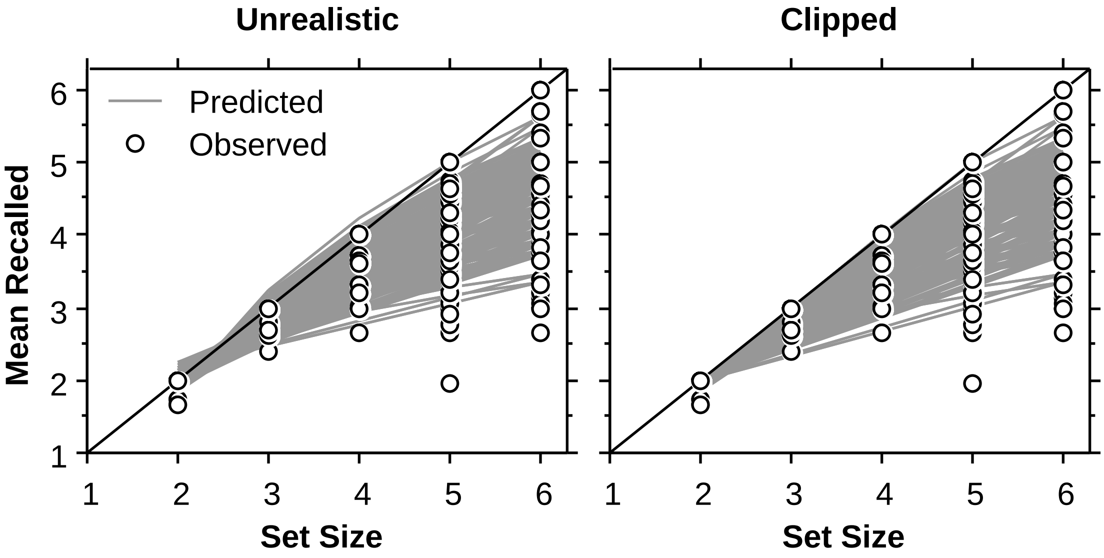


Figure 6.2. Unrealistic and realistic predictions of user performance (backwards-stepwise assessment with Set Size model)

Equation 6.1:

$$\begin{cases} \textit{Set Size} & \textit{Set Size} < \textit{prediction} \\ \textit{prediction} & \text{else} \end{cases}$$

It would be difficult to compute R^2's for the clipped formulas. Thus, an analogous measure would be the squared correlation between the predicted (clipped) and the observed values. Table 6.2 shows the comparison between the ability of the unrealistic and the clipped formulas to describe the observed data.

Table 6.2. R^2 and squared correlation for backwards stepwise models

		R^2	Clipped (squared correlation)
With Size	Assessment	.830	.834
	Generic	.787	.789

Another interesting observation is that the generic models tend to cover less of the observed region. For example, the backwards-stepwise generic with Set Size model is shown in Figure 6.3 (this model is similar to the one in Figure 6.2, but without SRMS). Note that the equation does not go above 5 words for the 6 word set size. This suggests that models without SRMS involved will not be valuable for depicting users who may be on the upper or lower end of the spectrum.

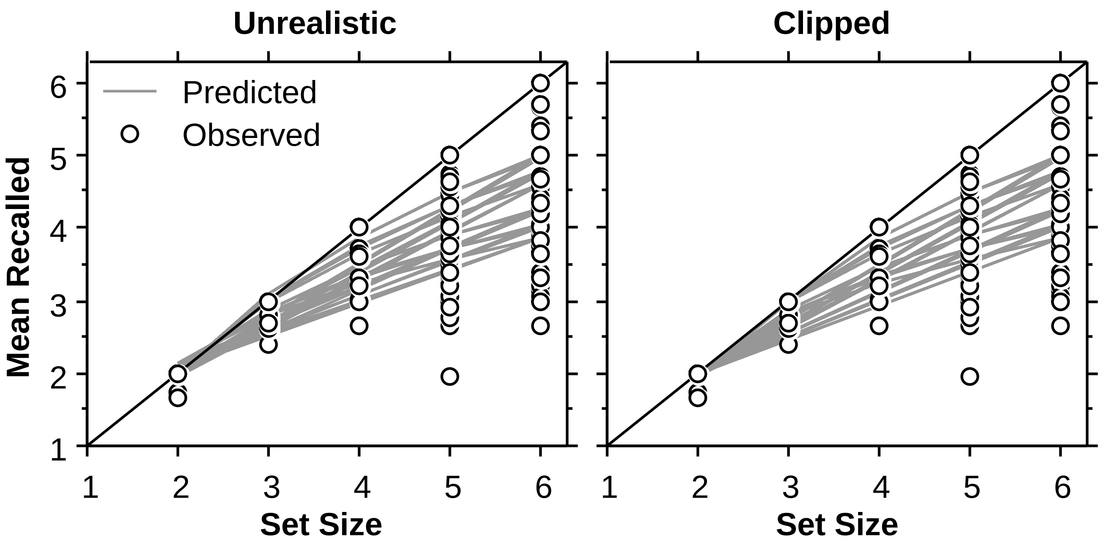


Figure 6.3. Unrealistic and realistic predictions of user performance (backwards-stepwise generic with Set Size model)

6.3 Findings with aggregate constructs

In the discussion of a proposed model of interaction (Chapter 3), the aggregate constructs of Buffer and Speed were introduced. A common finding from the two studies is support for the existence of these constructs. The manner in which the number of lines and the amount of lag affect performance suggests that they may be good candidates for inclusion in the buffer aggregate. Rate of presentation is certainly a prime component of speed.

Table 6.3 provides equations to describe Buffer and Speed. The use of a natural log function for the buffer aggregate was based on the findings of Duchnicky and Kolers (1983). They found that there was a point of diminishing returns past four lines of text during a scrolled reading task. The application of a natural log function in the Speed aggregate is based on similar reasons. At a certain point, faster speeds will also reach a point of diminishing returns. The log function is a logical choice since it is the inverse of a typical decay function. This has a high degree of face validity as the Speed aggregate is loosely based on the inverse of memory decay costs.

Table 6.3. The aggregate constructs

With Set Size	Without Set Size
$Buffer = \ln\left(\frac{Lines + \frac{Lag}{2}}{Set\ Size}\right)$	$Buffer = \ln\left(Lines + \frac{Lag}{2}\right)$
$Speed = \ln\left(\frac{Rate}{Set\ Size}\right)$	$Speed = \ln(Rate)$

Set Size was included in the aggregates during analyses that included Set Size. The decision to divide by Set Size, rather than multiply, was made due to higher R^2's compared to alternate models. The method of dividing Lag by 2 was similar (an integer was preferred to keep the equations simple). Note that

Lag in the buffer construct is lag expressed as number of sentences. This approach maintained a common unit for the buffer aggregate.

The aggregate equations without Set Size were plotted against performance to ensure that a logical trend was evident. Figure 6.4 shows that performance rises as the buffer increases. This pattern is similar to previous hypotheses on the relationship between buffer size and performance (see Figure 3.3). The somewhat flat slope for the speed graph is in line with previous analysis indicating that rate of presentation is not a significant main effect. It's impact was found in interactions. Since speed is only a function of rate at this point, a level slope is expected. It is also possible that these two levels of Speed are in a level part of the performance vs. Speed relationship (see Figure 3.3).

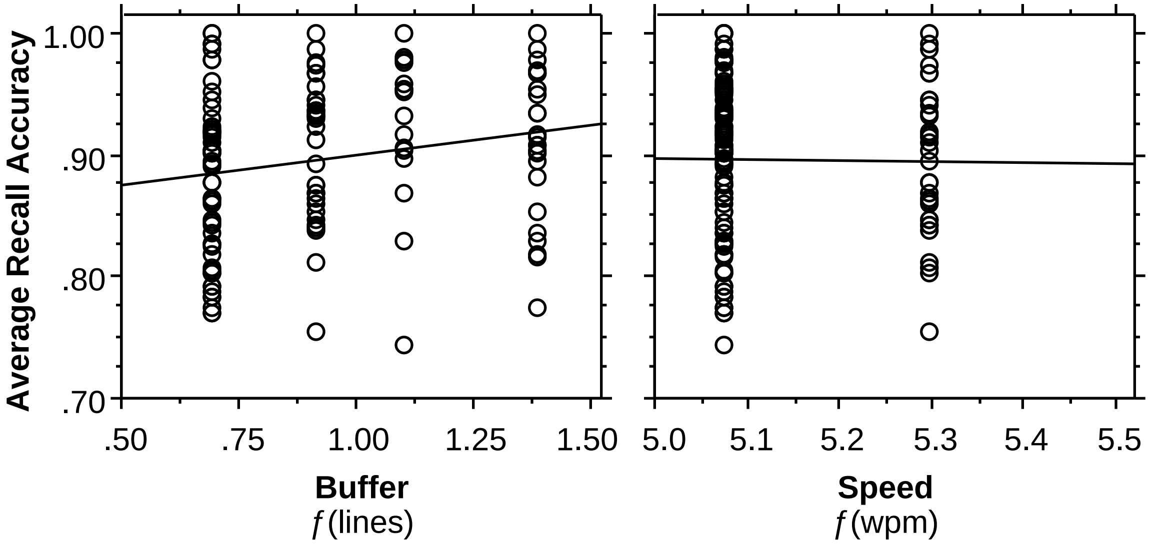


Figure 6.4. Performance as a function of the aggregate constructs

Regression models similar to those conducted for the individual factors were run with the aggregates (Table 6.4). In general, prediction performance was similar to that seen for the individual factors. A small drop in R^2's of about .03 was observed for the models with Set Size. This drop was essentially insignificant for the models without Set Size (.003). As in the previous section, the factors in italics in the "Significant Effects" column were at a significance level of p<.1 while all others were p<.05 or better. The lack of significance for all of the factors in the full models without Set Size was rather remarkable. The main effects, other than Speed, were significant when some of the effects were removed during the backwards stepwise regression. This highlights the value of using the backwards stepwise method.

Table 6.4. Regression with aggregates

		R^2	Effects in model	Significant Effects
Full w/ Size	Assessment (w/ SRMS)	.803	main and 2-way using: Buffer, Speed, HT, SRMS	Buffer, HT, SRMS, Buffer*Speed, Speed*HT, Speed*SRMS
	Generic (w/o SRMS)	.758	main and 2-way using: Buffer, Speed, HT	Buffer, Speed, HT Buffer*Speed, Speed*HT
Backwards Stepwise w/ Size	Assessment	.801	Buffer, Speed, HT, SRMS, Buffer*Speed, Buffer*SRMS, Speed*HT, Speed*SRMS	Buffer, HT, SRMS, Buffer*Speed, *Buffer*SRMS*, Speed*HT, Speed*SRMS
	Generic	.758	Buffer, Speed, HT, Buffer*Speed, Speed*HT	Buffer, Speed, HT, Buffer*Speed, Speed*HT
Full w/o Size	Assessment	.551	main and 2-way using: Buffer, Speed, HT, SRMS	none
	Generic	.332	main and 2-way using: Buffer, Speed, HT	none
Backwards Stepwise w/o Size	Assessment	.542	Buffer, HT, SRMS	Buffer, HT, SRMS
	Generic	.330	Buffer, HT	Buffer, HT

As in the individual factor analysis, the backwards stepwise models were comparable to the full models in their ability to predict performance. The prediction equations for the backwards stepwise models can be found in Appendix E.

The individual factor analysis demonstrated an improved ability to describe performance when the unrealistic region was clipped for the models with Set Size. This trend was also evident for the aggregate models (Table 6.5). Note that there was a greater improvement due to clipping for the aggregates when compared to the individual models (an average increase of .020 as opposed to .002).

Table 6.5. R^2 and squared correlation for backwards stepwise models

		R^2	Clipped (squared correlation)
With Size	Assessment	.801	.823
	Generic	.758	.775

Figure 6.5 gives a graphical example of clipping for an aggregate model. Note the log-shaped trend in the unclipped model. This is due to the log portion of the aggregate equations. It is useful to note that the aggregate pattern is very similar to the pattern displayed by the individual factors.

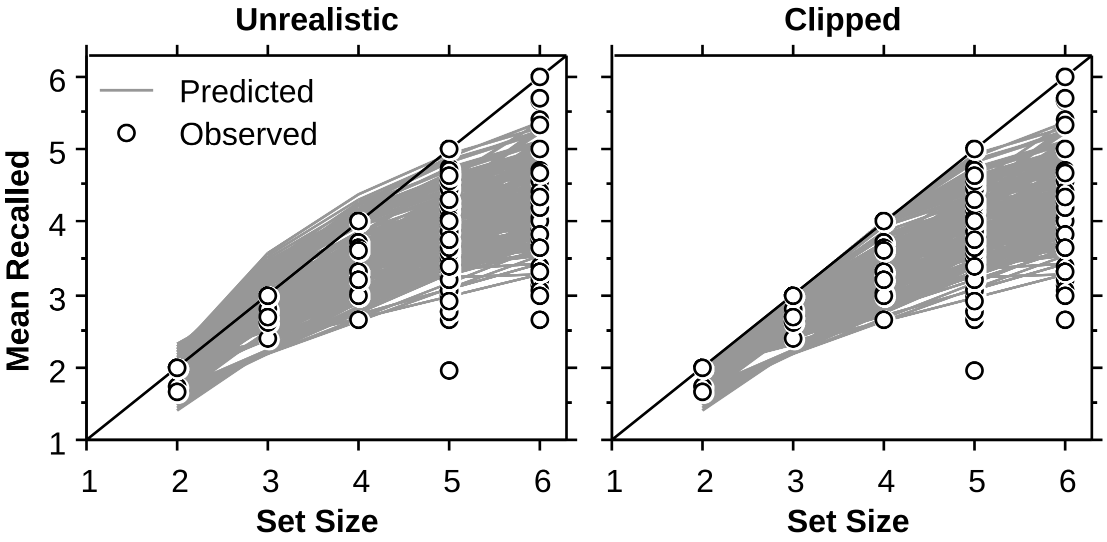


Figure 6.5. Unrealistic and realistic predictions of user performance (backwards-stepwise assessment with Set Size model)

In general, the aggregates were quite similar to the individual factors in their ability to predict performance. This is quite valuable in that the ability to examine a single construct rather than several factors allows easier analysis of a potential design.

At this point the only factor included in the speed aggregate is the rate of presentation. There are likely to be other factors strongly tied to rate that can be rolled into the speed aggregate. Since slow readers will be strongly affected by faster rates, the most likely candidate is the reading ability factor. This factor was not examined directly in this study. While SRMS has been shown to approximate reading comprehension, it does not approximate reading speed.

6.4 Examining potential core factors

It was hoped that these experiments might shed light on potential core factors. Core factors can be described as components of the factors examined in the experiments. For example, rate and number of lines are descriptors of how long an item is on the screen. A set of potential core factors (each with units) were developed (Table 6.6). This assisted the process of developing a data set from the two experiments.

Table 6.6. Potential constructs

Core Factor	Unit	Definition
Time on screen	seconds/word	$\frac{Rate\,(wpm)\times Lines}{60}$
Time of separation	seconds	$Seconds\ of\ Lag$
Visual density	words/screen	$8\times Lines$
Redundancy	channels	$\begin{cases} 2 & Speaker \text{ and } Captions \\ 1 & Speaker \text{ or } Captions \end{cases}$
Baseline Memory Span	words	$SRMS\ (CaptionsOnly)$

The "Time on screen" factor is a description of how long an item is displayed in the caption buffer. "Time of separation" describes the amount of time between the point the speaker speaks the word and the point at which the word is first displayed on the screen. To determine the cost of searching for a particular word, the construct of "Visual density" can be used. This factor is the amount of words shown on the screen (Each line has an average of 8 words since the screen was 40 characters wide.) "Redundancy" is a method of counting the number of sources to which a user has access. Hearing subjects can simultaneously follow two sources, while deaf subjects will only be able to follow one. Finally, a baseline level of performance (or what the user starts out with) is necessary. Thus, the value of SRMS for each subject is valuable.

These core factors were included in a regression analysis of performance (average recall accuracy). The model included 2-way interactions between the core factors. The R^2 for the model over average recall accuracy was very low (.14) and no effects were significant for the full model (the R^2 was lower for mean recalled). A backwards stepwise model did not show improvement in the R^2 but it did show that SRMS and Redundancy were significant. The low R^2's suggests that searching for core factors may not be a viable option. Although other constructs might better describe potential core factors, the ones tested are good candidates. They grasp some of the more essential issues involved in this type of memory task.

6.5 Further examination of preference and performance

Findings from the two studies suggested that there was a relationship between performance and preference. The data for how well subjects did on each condition and their corresponding preference ranking for that condition for the two studies were combined. Performance (average recall accuracy) and preference rankings were significantly correlated for the pooled data set (.36, $p<.0001$). The second study indicated that the deaf subjects did not show a correlation for this relationship. However, when the deaf subjects from both studies were pooled, this relationship did show a significant correlation (.44, $p<.001$). The results for hearing subjects were similar (.40, $p<.0001$). These findings suggest that the subject may accurately select the most appropriate format arrangement, given the opportunity to experience the different options under equivalent scenarios. If this were true, it would allow a more streamlined protocol to be used as a means of finding the best option for a particular user. Selecting the format on their own would also provide users with a greater sense of involvement in the process.

7 GENERAL CONCLUSIONS

7.1 Revisiting the proposed model of interaction

Earlier in this document a proposed model of interaction was presented (Chapter 3). Three hypotheses were identified during the discussion of the model. These hypotheses can now be evaluated and refined with respect to the findings of the two experiments. The picture describing the model is duplicated in Figure 7.1 to assist the reader.

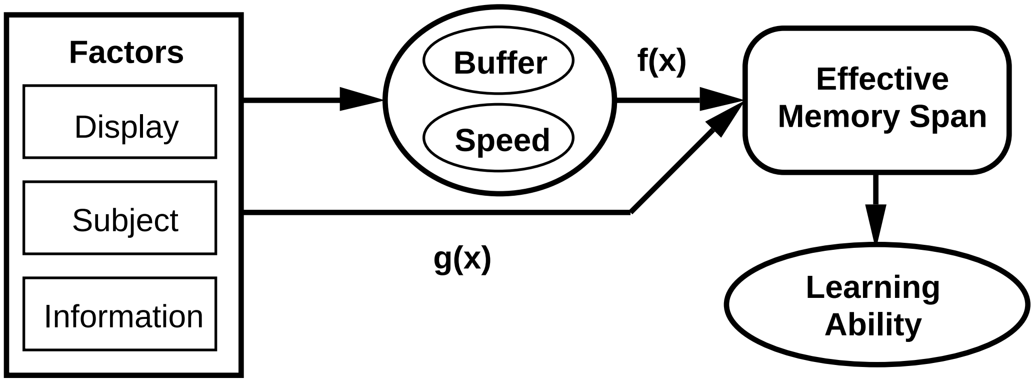


Figure 7.1. Chain of factors

7.1.1 Discussion of Hypothesis (1)

Hypothesis (1) The aggregates can accurately describe a set of factors.

This hypothesis represents the upper left arrow in Figure 7.1 that links the individual factors with the two aggregates. The main concern is that combining selected individual factors into aggregates will decrease the ability to describe changes in the effective memory span (EMS). The aggregates are preferred to individual factors since they more closely approximate standard theoretical cognitive constructs. Buffer size and the speed of the information flow are more general and powerful than individual factors as descriptors of how cognitive processes are affected.

This hypothesis does hold. The regression analyses of the combined data set (Chapter 6) show that the aggregate equations proposed are comparable to the individual factors in their ability to model the data. The better performing models were those with Set Size included in the analyses (R^2's in the vicinity of .80). There was only a drop of 0.03 in the R^2's when the aggregates were used instead of the individual factors. The drop for models without Set Size was even smaller (.003). These small drops represent a decrease in R^2 of only 3.6% (with Set Size) and 0.8% (without Set Size) from the individual models. The drop in squared correlation was even smaller (1.5%) when clipping was introduced to the models with Set Size. These findings indicate that the aggregates can successfully describe the individual factors. Much steeper drops were seen during development of the aggregate equations.

7.1.2 Discussion of Hypothesis (2)

Hypothesis (2) Buffer size and the speed of the information flow will affect the EMS.

The high R^2's seen in the regression models indicate that changes in the aggregates do indeed result in changes in the EMS. The prediction equations for these models (Appendix E) also show that the coefficients on the aggregate main effects are comparable in magnitude to the coefficients for Hearing Type. Hearing Type is known to have a significant, substantial impact on performance. In addition, buffer size was observed to have significant effects in all of the backwards stepwise regression models. Speed also demonstrated significant influence in the models with Set Size. Thus, the aggregates do affect the EMS.

7.1.3 Discussion of Hypothesis (3)

Hypothesis (3) Some factors will directly affect the EMS regardless of the aggregates.

Two subject factors were seen to have significant main effects during the aggregate regression analysis. Hearing Type and the subject's baseline SRMS were significant as main effects in all of the backwards stepwise models. In addition they exhibited large coefficients in the prediction equations. This was especially true for SRMS. This suggests that while the aggregates have a significant impact on EMS, it is essential to account for these two subject factors.

7.2 Revisiting the purpose of this work

Three key research questions were identified in the introduction of this work (subsection 1.5). Each question will be addressed individually to facilitate discussion.

7.2.1 Will captioning devices increase the effective memory span?

Improved performance with the introduction of captions for both deaf and hearing students has been observed in past research (Gates, 1971; Boyd and Vader, 1972; Murphy-Berman and Jorgensen, 1980; Nugent, 1983; Koskinen, et al, 1986; Markham, 1989; Neuman and Koskinen, 1992; and Seriwong, 1992). However, the direct impact of captioning on memory span scores has not been documented. This is important since higher memory spans have been linked to greater ability in comprehension, recovery from sentences with misleading context, and drawing inferences (Daneman and Carpenter, 1980; Daneman and Carpenter, 1983; and Oakhill, 1984). Subjects in both of the experiments presented here illustrated significant memory span improvements when captions were introduced.

As a potential positive side effect, the introduction of captions may also assist building vocabulary. Some research shows that deaf subjects utilize phonological processes during reading (Hanson and Fowler, 1987; Gibbs, 1989). Since the captions provide an opportunity to read the information (rather than speechread), there may be an increased usage of the short-term phonological store. This store has been shown to be important during the learning of new words (Baddeley, et al, 1988; Baddeley, 1992). Thus, captioning may indirectly help deaf users build on their vocabulary. A valuable follow-up study would be to test the vocabulary of control and experimental groups of deaf students over long-term exposure to captioning.

7.2.2 Which RTC design factors have an impact on performance? How much of an impact do these factors have?

The first experiment clearly indicated that the number of lines shown and hearing ability had a significant impact on EMS. Sex of the user and the location of the display were identified as not being significant factors. Sentence lag was identified in the second experiment as another factor with a borderline significant main effect. Rate was seen to have a limited impact through the significant Rate*Lag*Hearing Type interaction and the borderline Rate*Lag interaction. The finding that hearing ability was significant was reinforced by the second experiment.

An interesting finding in both studies was that two factors present due to the materials used, Set Size and Context, were also significant. Although this finding is useful to comprehension research, it has limited application in a captioning environment. A typical conversation contains multiple sentences with varying levels of context that are often longer than those used here. Set Size and Context are more useful as tools to examine cognitive processes than as guidelines to follow in an applied setting. It is unlikely that a teacher would be willing to deliver entire lectures in high context sound-bytes.

7.2.3 What are the key cognitive constructs that affect the user's ability to benefit from the device?

The regression analyses have identified the external buffer size and the speed of the information flow as viable cognitive constructs when describing redundant, parallel-source information processing. Under a captioning arrangement, an external text-based buffer is present. This buffer allows the user to reduce the amount of information that is retained in working memory during processing. This frees up capacity for overhead processes and allows the user to acquire the textual information in a manner similar to "Just-In-Time" manufacturing. This line of reasoning is supported by findings of increased recall performance as buffer size increased.

The speed construct provides a measure of the rate of information flow during these activities. In the current studies, this construct only includes the speed of the external information flow. However, in future studies it may be possible to incorporate measures of internal cognitive processes similar to those used in GOMS analysis (Card, Moran, and Newell, 1986). A simulation presented in Appendix F makes a preliminary attempt to model these processes.

The analysis of these cognitive constructs also has direct application value. The aforementioned regression analyses led to prediction equations that have a strong ability to depict subject recall performance. These equations allow developers and providers of captioning systems to determine the optimal format arrangements for particular users. In addition these constructs provide direction on the research value of untested format factors. Unfortunately, the examination of core constructs did not lead to a solid ability to predict user performance. However, the exercise did reinforce the importance of including baseline SRMS performance in any future research on untested format factors.

7.3 The impact of this work

As previously noted, the inclusion of captions in a face-to-face presentation dramatically increases a hearing-impaired person's ability to comprehend the speaker. In addition, providing captions to hearing people also seems to enhance verbal comprehension. The increased comprehension for both hearing and deaf students will likely lead to a better learning environment and improved information transfer between the teacher and the students. Similar work in the past with captioned video presentations (educational movies, etc.) has already demonstrated this potential (Gates, 1971; Boyd and Vader, 1972; Murphy-Berman and Jorgensen, 1980; Nugent, 1983; Koskinen, et al, 1986; Markham, 1989; Neuman and Koskinen, 1992; and Seriwong, 1992). It should be noted that the work described in this dissertation utilized face-to-face presentations rather than presentations with graphics and text (e.g., slides) to support the information.

This work shows that real-time captioning can be useful in spontaneous, unscripted scenarios where the teacher is talking in front of the class. In addition, the finding that the location of the captioning device (desk or podium) has no impact on performance suggests that providers of real-time captioning can use either a personal desktop device or a podium mounted display.

The major recommendation at this time is that captioning devices should display at least four lines of text so that there is a larger visual buffer for the user. In addition, higher amounts of sentence lag seem to improve performance. It should be noted that relatively small amounts of lag were examined (up to two lines or 6 seconds) in the present study. It is possible that larger levels of lag may result in decreased performance. The rate of presentation also appears to have a limited impact on performance except through the Rate*Lag*Hearing Type interaction. (Specifically, the 200 wpm, one sentence of lag condition showed a drop in performance for the deaf subjects.) This suggests that it may be necessary for captioning devices to recognize shifts in rate and modify the display or lag accordingly.

The participants' subjective rankings were correlated rather well with their performance. This suggests that users can be given a decision-making role in determining their ideal device parameters without danger of performance decrements. However, deaf users may perform better with lag even though their preference for it is low.

There seems to be a great deal of wiggle room for such developers. The findings suggest that people can benefit from a relatively wide range of parameters. Ideally, the work described here will reduce the likelihood that early systems do more harm than good. To this end, some guidelines for the developers of captioning devices can be made:

(1) Real-time captions are beneficial for both deaf and hearing students.
(2) Location of a RTC display (desk or podium) has no impact on recall performance.
(3) RTC devices should display at least four lines of text. However, there is only theoretical support for displays of more than four lines of text.
(4) Higher amounts of sentence lag seem to improve performance (up to two lines or 6 seconds). It is possible that larger levels of lag may result in decreased performance.

(5) A higher rate of presentation (200 vs. 160 wpm) appears to have a negative impact on performance for deaf subjects when lag is present.

(6) Users are reasonably good at determining their ideal device parameters given the opportunity to experience their choices. However, deaf users may perform better with lag even though their preference for it is low.

The theoretical implications of this work revolve around a better understanding of how people process redundant, parallel information. This study demonstrates that there is no interaction between the inclusion of an easy to process audio channel and the number of lines and the location of a captioning display. This suggests that altering the size of an external visual buffer and modifying the location of that buffer affect the visual processes within working memory and not their auditory counterparts. Preliminary results based on a small deaf subject pool also suggest that changing the rate and sentence lag of such displays has a performance impact isolated to the visual processes. This means that it may be possible to offset impairments within the phonological loop with visually based parallel information. Work by Baddeley, et al (1988) with a woman who could only learn new words when they were presented visually suggests that this may indeed be true.

An early attempt to model this flow of redundant, parallel information with simulation is presented in Appendix F. A more precise and robust simulation model can be developed in the future as additional studies provide more insight on the elements and assumptions of the model.

7.4 Future work

Besides the aforementioned simulation model, there are many possibilities for future research on this subject. Since a great portion of the hearing-impaired population is in the older generation, a study using subjects of varying ages (e.g., a group of 65 and up) would be useful to determine the impact of age. The study can also be transferred to other populations. Introducing the device as a translation aid in a foreign language environment may be of interest. Professional scenarios could also provide some valuable data (e.g., legal proceedings, business meetings, etc.). Of key interest is whether the cognitive constructs developed in this dissertation are also important under these other scenarios.

Besides expanding the subject population and the purpose of the device, there are also some modifications to the current study that may be of interest. Introducing varying degrees of spatial information (e.g., graphs and pictures) would be useful. This becomes more relevant for classroom topics that have a higher degree of such information (e.g., physics and art).

The findings presented here are limited due to the fact that only a small portion of teachers utilize stationary positions and avoid spatial information (e.g., chalkboards). As such, the application of this work to classrooms in general should be reinforced by additional studies incorporating spatial information and moving speakers.

Preliminary attempts to measure looking behavior in this study suggest that applying an accurate method of eye tracking to this experiment may be of interest. The eye tracking method used in this study (a camera fixed on the subject's face) proved too limited in its ability to provide anything more than a vague impression of device usage. While this vague measurement may lead to some preliminary conclusions in the future, a more precise depiction of when and how a subject uses the device would be more valuable. In a scenario where the user becomes dependent on the captions, the user may not observe the speaker at all. This reduction in non-verbal cue reception may lead to boredom. A good example of this is the effect of an emotionless delivery of a speech without body language. Boredom may be part of the reason some current users of captioning systems have been observed to fall asleep while using a device of this nature (as reported by Smith, 1995).

Another experimental variant would be to collect data over longer durations. One possible arrangement would be to wire a real-time captioning system (with a stenographer) to a personal display. The device could then be provided to a group of deaf students for an extended period of time in their normal classroom environment. By comparing their performances to a matched set of classmates it may be possible to test for increased vocabulary, reading ability, and ability to draw inferences. Even though the effect will probably be smaller, this study could also be completed with hearing students.

As can be seen by these suggestions, there are a number of additional experiments that may be conducted to build upon the findings from this particular study. In closing, there is great promise for valuable information about reference displays under a multitude of user scenarios to be found in future experiments.

APPENDICES

Appendix

APPENDIX A - ASSISTIVE COMMUNICATION DEVICES FOR THE DEAF

This study is concerned with text displays designed to assist understanding. This appendix is a brief description of non-textural assistance devices. Speech perception devices for the hearing-impaired usually can be categorized as being designed to help understand speech or designed to help improve a deaf person's ability to speak. The majority of the devices described are designed to improve understanding, but some visual tools for improving speech are also included. See Pickett (1986) for a review of some of the devices being tested. Please note that many of the accuracy figures quoted are not based on the same protocol for measuring understanding. Thus, comparing values reported in papers by different authors may be inappropriate.

Visual displays

Devices for understanding speech

An initial attempt to improve the ability to speechread using computer-generated visual aids was made by Upton (1968). He built a pair of glasses that had five miniature light bulbs in the periphery of the right eye's field of view. These lights were designed to indicate certain speech features. The hard-of-hearing author wore the glasses intermittently for several months. Upton conducted no formal quantitative tests, but made some qualitative observations. He felt that the device, while initially difficult to learn, was easy to use and provided a great deal of useful information.

Another peripheral device was examined by Ebrahimi and Kunov (1991). This device was also mounted to a pair of glasses. A small bank of LED's were placed on the right edge of the right eye's field of view. Certain consonant groups would trigger patterns in the bank of LED's. This led to an overall accuracy level in assisted speechreading of 76%.

Using signed cues in the area surrounding a speaker's lips to improve communication has been attained using a form of sign called Cued Speech. A speaker who knows the appropriate signs can position hand symbols near their lips to help perception of hard to discriminate phoneme groupings. The disadvantage of this technique is that the average person on the street does not know Cued Speech. Uchanski, et al (1994) examined the assistance a computer generated series of cues could provide to a hearing-impaired user. Their work supports the argument that assisted speechreading can achieve accuracy levels close to 75-90%. They stipulate that user skill, complexity of the task, and the specific computer system will have an impact on this number.

Devices for improving speech

A series of studies starting in the late 1960's led a number of authors to report that spectrographic displays provided good feedback for improving speech in deaf students during speech lessons (Stark, et al, 1968; House, et al, 1968; Pronovost, et al, 1968; and Stewart, et al, 1976). Watanabe, et al (1985) went further and added color to the display. They found the addition of color was very effective in providing sound information in a visual form. All of these studies utilized just the spectrographic display; there was only one source of visual information.

Tactile displays

Using a variety of channels, body locations, and vibrating mechanisms, a number of papers have identified tactile indications of sound structure as being highly beneficial (Saunders, et al, 1976; Goldstein, et al, 1976; De Filippo and Scott, 1978; Brooks and Frost, 1983; Brooks, et al, 1986a; Brooks, et al, 1986b; Boothroyd and Hnath, 1986; and Weisenberger, et al, 1989). Many of the later devices produced accuracy scores under an assisted speechreading condition in the vicinity of 80%. It is important to remember that while the person may be using touch rather than sight to perceive the pattern, these are still spatial displays.

References

Brooks, P. L. and Frost, B. J. (1983) Evaluation of a tactile vocoder for word recognition. Journal of the Acoustical Society of America, 74(1), 34-39.

Brooks, P. L.; Frost, B. J.; Mason, J. M.; and Gibson, D. M. (1986a) Continuing evaluation of the Queen's University Tactile Vocoder. I: Identification of open set words. Journal of Rehabilitation Research, 23(1), 119-128.

Brooks, P. L.; Frost, B. J.; Mason, J. M.; and Gibson, D. M. (1986b) Continuing evaluation of the Queen's University Tactile Vocoder. II: Identification of open set sentences and tracking narrative. Journal of Rehabilitation Research, 23(1), 129-138.

Boothroyd, A. and Hnath, T. (1986) Lipreading with tactile supplements. Journal of Rehabilitation Research, 23(1), 139-146.

De Filippo, C. L. and Scott, B. L. (1978) A method for training and evaluating the reception of ongoing speech. Journal of the Acoustical Society of America, 63(4), 1186-1192.

Ebrahimi, D. and Kunov, H. (1991) Peripheral vision lipreading aid. IEEE Transactions on Biomedical Engineering, 38(10), 944-952.

Goldstein, M. H.; Stark, R. E.; Yeni-Komshian, G.; and Grant, D. G. (1976) Tactile stimulation as an aid for the deaf in production and reception of speech: Preliminary studies. IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, 598-601.

House, A. S.; Goldstein, D. P.; and Hughes, G. W. (1968) Perception of visual transforms of speech stimuli: Learning simple syllables. American Annals of the Deaf, 113, 215-221.

Pickett, J. M. (1986) Speech communication for the deaf: Visual, tactile, and cochlear-implant. Journal of Rehabilitation Research, 23(1), 95-99.

Pronovost, W.; Yenkin, L.; Anderson, D. C.; and Lerner, R. (1968) The voice visualizer. American Annals of the Deaf, 113, 230-238.

Saunders, F. A.; Hill, W. A.; and Simpson, C. A. (1976) Speech perception via the tactile mode: Progress report. IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, 594-597.

Stark, R. E.; Cullen, J. K.; and Chase, R. A. (1968) Preliminary work with the new Bell Telephone Visible Speech Translator. American Annals of the Deaf, 113, 205-214.

Stewart, L. C.; Larkin, W. D.; and Houde, R. A. (1976) A real time sound spectrograph with implications for speech training for the deaf. IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, 590-593.

Uchanski, R.; Delhorne, L.; Dix, A.; Braida, L.; Reed, C.; and Durlach, N. (1994) Automatic speech recognition to aid the hearing impaired: Prospects for the automatic generation of cued speech. Journal of Rehabilitation Research, 31(1), 20-41.

Upton, H. W. (1968) Wearable eyeglass speechreading aid. American Annals of the Deaf, 113, 222-229.

Watanabe, A.; Ueda, Y.; and Shigenaga, A. (1985) Color display system for connected speech to be used for the hearing impaired. IEEE Transactions on Acoustics, Speech, and Signal Processing, 33(1), 164-173.

Weisenberger, J. M.; Broadstone, S. M.; and Saunders, F. A. (1989) Evaluation of two multichannel tactile aids for the hearing impaired. Journal of the Acoustical Society of America, 86(5), 1764-1775.

APPENDIX B - FACTOR SUMMARY

Display Factors

Factors	Literature	Findings	Predictions
Rate of Presentation	Shroyer and Birch (1980)	• lowest group of hearing-impaired students read at 124 wpm • 84% read slower than 159 wpm (rate of normal speech)	As speeds increase beyond 120 wpm the user may approach the extreme levels of usage. This is due to the increased load that the faster speeds should impose when the reading speed exceeds the users' ability.
	Braverman and Hertzog (1980) & Braverman (1981)	• no significant difference in comprehension between 60, 90, and 120 wpm caption rates for hearing-impaired secondary students	
	Okada, et al (1985)	• caption presentation time (3, 5, and 7 seconds) was not a main effect • caption presentation time interacted with developmental level	
	Kolers, et al (1981)	• scroll rates 20% faster than preferred speed are read more efficiently than the preferred speed • a static page is read more efficiently than the preferred scrolled speed	
Lines/ Characters Shown	Braverman (1981)	• comprehension was equal for low and high caption densities (290 and 339 captions per presentation)	In cases were a small portion is needed to be found, larger amounts of text should produce lower performance. When there is not much of text shown, the user may miss information that they needed. Strong interactions with other factors are expected due to the dependency on the user scenario.
	Duchnicky and Kolers (1983)	• density, line length, and number of lines shown were all found to have significant effects • 80 characters/line was found to be read 30% faster than 40 characters/line • no significant difference between 1 and 2 lines • no significant difference between 4 and 20 lines • larger group of line lengths were 9% faster than the smaller group	
	Kolers, et al, (1981)	• longer line formats were read 25% faster	

Location	Wickens and Carswell (1995)	• display location will have an effect due to the proximity compatibility principle	The closer the display is to the primary information source (the image of the speaker), the more the device is expected to be used. This is due to a reduced cost of the visual search. Higher performance on the closer display is expected even under steady usage levels. When the device is close to the primary source, the motion of the captions may be distracting and lead to unintended usage.
	Liu and Wickens (1992); Liu (1996a); & Liu (1996b)	• larger distances will increase the cost of a visual search	
	Flannagan and Harrison (1994)	• performance on a pedestrian detection task decreased as HUD location was moved from 4 to 9 to 15 degrees below the horizon • navigation performance showed only a slight change in performance	
	Steinfeld and Green (1995); Green and Williams (1992); & Williams and Green (1992)	• response times were faster as a navigation display was moved from the instrument panel to a small HUD to a full field-of-view HUD	

Subject Factors

Factors	Literature	Findings	Predictions
Baseline Reading Memory Span	Daneman and Carpenter (1980)	• oral reading, silent reading, and listening memory spans were quite comparable and had similar amounts of variability • the memory spans were found to be strong predictors of verbal comprehension	Users with a high baseline memory span are expected to be less likely to display high usage levels. Users with a low baseline are expected to display a lower benefit.
Reading Ability	Conrad (1977)	• deaf high school students are seven years behind in their reading level (ed.: prior to extensive captioning and TDD usage, may not be accurate for current generation)	This performance will be examined through a memory span score similar to those described above.
	Maxon and Welch (1992)	• reduced levels of reading ability were seen to affect caption comprehension • a greater degree of hearing loss was not seen to lead to a decreased reading ability	
	Caldwell (1973)	• after five weeks, students exposed to captions beyond their reading levels were shown to have a significant jump in ability with a consistent level of interest	
	Murphy-Berman and Jorgensen (1980)	• hearing-impaired students have comprehension problems at higher reading levels (ed.: same comment as seen above)	
	Gibbs (1989)	• in deaf readers, the relationship between use of phonological information and reading ability was not significant • advanced deaf readers have displayed that their hearing-impairment has no effect on their reading ability	

Hearing Ability	Pichora-Fuller, et al (1995)	• memory span performance decreased as the signal to noise ratio decreased	Under silent conditions, the hearing subjects will probably perform worse than their deaf counterparts due to their limited experience with speechreading. If sound is included, the hearing subjects may not use the reference device at all. The ability to perceive the primary information source is expected to have a significant effect. Difficult perception tasks have been shown to reduce memory span when compared to easier perception conditions.
	Hyde and Power (1992)	• a severely deaf group was found to be significantly better than a profoundly deaf group when sign was not present	
	Olsson and Furth (1966)	• the hearing have better digit spans then the deaf • the deaf have better visual memory spans for simultaneous nonsense forms • the hearing have better visual memory spans for successive nonsense forms (ed.: see comment in Reading Ability)	
	Austin and Myers (1984)	• hearing-impaired college students view 221 minutes a day of television (164 for hearing)	
Sex	e.g., Steinfeld and Green (1995)	• sex has been shown to be a significant effect in cognitive research studies	The effect is unknown.

Information Factors

Factors	Literature	Findings	Predictions
Difficulty of the Material	Blatt and Sulzer (1981)	• in a mailed survey, the information based reasons for not watching the *Captioned ABC News* were "Captions leave out information" (6.0%) and "Cannot understand captions" (1.6%) • 31% of responses were from people with education levels between 1 and 11	The redundant supply of information should help reduce the perception load, allowing the user to dedicate more capacity to the higher level processes that the more difficult information will require. In cases where the difficulty of the material is below the user's ability, the device is expected to provide less assistance.
	Braverman and Hertzog (1980) & Braverman (1981)	• level of captions did have an effect on comprehension and inferential measurements • the lower language level produced higher scores	
	Okada, et al (1985)	• language level was not found to be a significant main effect • there was an interaction between language level and education level (ed.: experiment conducted in Japanese)	
Time/ Sentence Lag	Braverman and Hertzog cited in Braverman (1981)	• no change in comprehension was seen when the captions were not synchronous with the audio portion of the program (ed.: may be due to a low usage of the audio track by the subjects)	In the presence of lag, there will likely be a move towards extreme levels of usage due to the penalty for switching between sources.

APPENDIX C - INSTRUCTIONS AND SEQUENCE FOR EXPERIMENT 1

Evaluation of Selected Design Factors for Real-Time Captions in a Mainstream Classroom

GREETING

Greet the subject and **thank** them for coming.

Explain that this study is to help determine the **manner in which real-time captions should be used in a classroom environment**.

Answer **VERY briefly any questions** the subject has about the ultimate goal of the study.

Inform the subject that you will be **happy to tell them more when the session is done**.

FORMS

Have subject fill out:
- **Consent**
- **Bio**

EXPOSURE TO RTC

Show **video tape** with real-time closed captions on and explain that **in this study**, there will be **no captioning errors**.

LAST WORD INSTRUCTIONS

Give the subject the **LW Instruction Sheet** and make sure they **do the example**.

The practice sentences are:	The judge is sitting on the bench. I did not know about the chunks. Bob could consider the pole. For dessert he had apple pie.

Make sure the subject **raises their hand for the HIGH** context/predictability sentences.

BASELINE

For each block, give the subject the **Response Sheet**.

Run the **captions only** and **audio + face** conditions. The counterbalance order is:

L	audio + face
R	captions only

	Sex	Num.	Condition		Form	
Deaf	~~**F**~~	~~**1**~~ *	~~**L**~~	~~**R**~~	~~**6**~~	~~**5**~~
	F	2	R	L	3	4
	F	3	L	R	5	6
	F	4	R	L	1	2
	~~**M**~~	~~**5**~~	~~**L**~~	~~**R**~~	~~**6**~~	~~**5**~~
	M	6	R	L	3	4
	M	7	L	R	5	6
	M	8	R	L	1	2
Hearing	~~**F**~~	~~**9**~~	~~**L**~~	~~**R**~~	~~**6**~~	~~**5**~~
	F	10	R	L	3	4
	F	11	L	R	5	6
	F	12	R	L	1	2
	~~**M**~~	~~**13**~~	~~**L**~~	~~**R**~~	~~**6**~~	~~**5**~~
	M	14	R	L	3	4
	M	15	L	R	5	6
	M	16	R	L	1	2

***** (subject slots 1, 5, 9, and 13 were not run in order to reduce the cost of this study)

FORMAT TESTING

Prior to the first block, inform the subject that they may take a **break** if/when necessary.

For each block, give the subject the **Response Sheet**.

Run the **format testing** blocks. The counterbalance order is:

A	desk	2	lines
B	desk	4	lines
C	podium	2	lines
D	podium	4	lines

	Sex	Num.	Condition Order				Form Order			
Deaf	~~F~~	~~1~~	~~A~~	~~B~~	~~C~~	~~D~~	~~4~~	~~3~~	~~1~~	~~2~~
	F	2	B	A	D	C	2	1	5	6
	F	3	C	D	A	B	3	4	2	1
	F	4	D	C	B	A	6	5	4	3
	~~M~~	~~5~~	~~A~~	~~B~~	~~C~~	~~D~~	~~4~~	~~3~~	~~1~~	~~2~~
	M	6	B	A	D	C	2	1	5	6
	M	7	C	D	A	B	3	4	2	1
	M	8	D	C	B	A	6	5	4	3
Hearing	~~F~~	~~9~~	~~A~~	~~B~~	~~C~~	~~D~~	~~4~~	~~3~~	~~1~~	~~2~~
	F	10	B	A	D	C	2	1	5	6
	F	11	C	D	A	B	3	4	2	1
	F	12	D	C	B	A	6	5	4	3
	~~M~~	~~13~~	~~A~~	~~B~~	~~C~~	~~D~~	~~4~~	~~3~~	~~1~~	~~2~~
	M	14	B	A	D	C	2	1	5	6
	M	15	C	D	A	B	3	4	2	1
	M	16	D	C	B	A	6	5	4	3

FORMAT PREFERENCE

Ask the subject to **rank** the six conditions by preference.

Ask if they have any **general comments about RTC**.

DEPARTURE

Thank them for their time.

Get their **mailing address** if they want a copy of a paper.

APPENDIX D - INSTRUCTIONS AND SEQUENCE FOR EXPERIMENT 2

Evaluation of Selected Design Dimensions for Real-Time Captions in a Mainstream Classroom

GREETING

Greet the subject and **thank** them for coming.

Explain that this study is to help determine the **manner in which real-time captions should be used in a classroom environment**.

Answer **VERY briefly any questions** the subject has about the ultimate goal of the study.

Inform the subject that you will be **happy to tell them more when the session is done**.

FORMS

Have subject fill out:
- **Consent**
- **Bio**

EXPOSURE TO RTC

Show **video tape** with real-time closed captions on and explain that **in this study**, there will be **no captioning errors**.

LAST WORD INSTRUCTIONS

Give the subject the **LW Instruction Sheet** and make sure they **do the example**.

The practice sentences are:	The judge is sitting on the bench. I did not know about the chunks. Bob could consider the pole. For dessert he had apple pie.

Make sure the subject **raises their hand for the HIGH** context/predictability sentences.

TESTING - HEARING

For each block, give the subject the **Response Sheet**.

Prior to the first block, inform the subject that they may take a **break** if/when necessary.

The counterbalance order is:

Condition		
R	captions only	
A	160	0
B	200	0
C	160	1
D	200	1
E	160	2

Hearing

Sex	Num	Formats					
Female	1	R	A	B	C	D	E
	2	R	B	A	D	C	E
	3	R	C	D	E	A	B
	4	R	D	C	E	B	A
	5	R	E	A	B	C	D
	6	R	E	B	A	D	C
Male	7	R	A	B	C	D	E
	8	R	B	A	D	C	E
	9	R	C	D	E	A	B
	10	R	D	C	E	B	A
	11	R	E	A	B	C	D
	12	R	E	B	A	D	C

Sex	Num	Forms					
Female	1	1	2	3	4	5	6
	2	2	1	4	3	6	5
	3	5	6	1	2	3	4
	4	6	5	2	1	4	3
	5	3	4	5	6	1	2
	6	4	3	6	5	2	1
Male	7	1	2	3	4	5	6
	8	2	1	4	3	6	5
	9	5	6	1	2	3	4
	10	6	5	2	1	4	3
	11	3	4	5	6	1	2
	12	4	3	6	5	2	1

TESTING - DEAF

For each block, give the subject the **Response Sheet**.

Prior to the first block, inform the subject that they may take a **break** if/when necessary.

The counterbalance order is:

Condition		
R	captions only	
A	160	0
B	200	0
C	160	1
D	200	1
E	160	2

Deaf

	Num	Formats					
Female	20	R	B	A	D	C	E
	21	R	C	D	E	A	B
	22	R	D	C	E	B	A
	23	R	E	A	B	C	D

	Num	Forms					
Female	20	2	1	4	3	6	5
	21	5	6	1	2	3	4
	22	6	5	2	1	4	3
	23	3	4	5	6	1	2

FORMAT PREFERENCE

Have subject **rank the blocks**.

Ask if they have any **general comments about RTC**.

DEPARTURE

Pay subject.

Thank them for their time.

Get their **mailing address** if they want a copy of a paper.

APPENDIX E - FULL ANOVA TABLES

Experiment 1

Repeated Measures ANOVA - Type III Sums of Squares **Set Size**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Hearing Type	1	0.419	0.419	7.581	0.0249		
Sex	1	0.001	0.001	0.019	0.8930		
Hearing Type * Sex	1	0.019	0.019	0.352	0.5693		
Subject(Group)	8	0.442	0.055				
Line Number	1	0.083	0.083	9.605	0.0147	0.0147	0.0147
Line Number * Hearing Type	1	0.002	0.002	0.240	0.6372	0.6372	0.6372
Line Number * Sex	1	0.030	0.030	3.447	0.1005	0.1005	0.1005
Line Number * Hearing Type * Sex	1	0.008	0.008	0.984	0.3503	0.3503	0.3503
Line Number * Subject(Group)	8	0.069	0.009				
Location	1	0.001	0.001	0.199	0.6671	0.6671	0.6671
Location * Hearing Type	1	0.006	0.006	0.906	0.3691	0.3691	0.3691
Location * Sex	1	0.006	0.006	0.906	0.3691	0.3691	0.3691
Location * Hearing Type * Sex	1	2.2E-04	2.2E-04	0.034	0.8579	0.8579	0.8579
Location * Subject(Group)	8	0.051	0.006				
Set Size	4	2.538	0.634	54.285	0.0001	0.0001	0.0001
Set Size * Hearing Type	4	0.284	0.071	6.076	0.0009	0.0265	0.0105
Set Size * Sex	4	0.016	0.004	0.340	0.8492	0.6292	0.7199
Set Size * Hearing Type * Sex	4	0.007	0.002	0.151	0.9611	0.7692	0.8635
Set Size * Subject(Group)	32	0.374	0.012				
Line Number * Location	1	0.012	0.012	3.738	0.0893	0.0893	0.0893
Line Number * Location * Hearing Type	1	0.004	0.004	1.259	0.2944	0.2944	0.2944
Line Number * Location * Sex	1	0.012	0.012	3.689	0.0910	0.0910	0.0910
Line Number * Location * Hearing Type * Sex	1	0.004	0.004	1.287	0.2894	0.2894	0.2894
Line Number * Location * Subject(Group)	8	0.026	0.003				
Line Number * Set Size	4	0.080	0.020	5.491	0.0018	0.0102	0.0018
Line Number * Set Size * Hearing Type	4	0.022	0.005	1.498	0.2262	0.2490	0.2262
Line Number * Set Size * Sex	4	0.027	0.007	1.826	0.1481	0.1841	0.1481
Line Number * Set Size * Hearing Type * Sex	4	0.018	0.005	1.269	0.3027	0.3087	0.3027
Line Number * Set Size * Subject(Group)	32	0.116	0.004				
Location * Set Size	4	0.015	0.004	0.505	0.7323	0.6086	0.7111
Location * Set Size * Hearing Type	4	0.008	0.002	0.263	0.8994	0.7667	0.8799
Location * Set Size * Sex	4	0.008	0.002	0.259	0.9021	0.7700	0.8829
Location * Set Size * Hearing Type * Sex	4	0.028	0.007	0.951	0.4473	0.4054	0.4407
Location * Set Size * Subject(Group)	32	0.235	0.007				
Line Number * Location * Set Size	4	0.004	0.001	0.142	0.9653	0.8164	0.9111
Lines * Location * Size * Hearing Type	4	0.016	0.004	0.566	0.6890	0.5387	0.6173
Lines * Location * Size * Sex	4	0.044	0.011	1.543	0.2135	0.2489	0.2360
Lines * Location * Size * Hearing Type * Sex	4	0.028	0.007	0.993	0.4254	0.3769	0.4050
Lines * Location * Size * Subject(Group)	32	0.229	0.007				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Line Number	1	1.429
Location	1	1.429
Set Size	0.325	0.507
Line Number * Location	1	1.429
Line Number * Set Size	0.592	1.174
Location * Set Size	0.488	0.885
Line Number * Location * Set Size	0.386	0.639

Dependent: Recall Accuracy

NOTE: Probabilities are not corrected for values of epsilon greater than 1.

Repeated Measures ANOVA - Type III Sums of Squares **Context**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Hearing Type	1	0.267	0.267	7.231	0.0275		
Sex	1	0.001	0.001	0.016	0.9029		
Hearing Type * Sex	1	0.010	0.010	0.269	0.6178		
Subject(Group)	8	0.295	0.037				
Lines	1	0.052	0.052	11.575	0.0093	0.0093	0.0093
Lines * Hearing Type	1	4.1E-05	4.1E-05	0.009	0.9268	0.9268	0.9268
Lines * Sex	1	0.016	0.016	3.549	0.0963	0.0963	0.0963
Lines * Hearing Type * Sex	1	0.007	0.007	1.605	0.2409	0.2409	0.2409
Lines * Subject(Group)	8	0.036	0.005				
Location	1	0.001	0.001	0.227	0.6465	0.6465	0.6465
Location * Hearing Type	1	0.002	0.002	0.360	0.5649	0.5649	0.5649
Location * Sex	1	0.003	0.003	0.650	0.4433	0.4433	0.4433
Location * Hearing Type * Sex	1	3.3E-04	3.3E-04	0.076	0.7893	0.7893	0.7893
Location * Subject(Group)	8	0.034	0.004				
Context	1	0.126	0.126	18.94	0.0024	0.0024	0.0024
Context * Hearing Type	1	0.008	0.008	1.196	0.3059	0.3059	0.3059
Context * Sex	1	3.2E-04	3.2E-04	0.048	0.8324	0.8324	0.8324
Context * Hearing Type * Sex	1	0.006	0.006	0.944	0.3598	0.3598	0.3598
Context * Subject(Group)	8	0.053	0.007				
Lines * Location	1	0.004	0.004	1.774	0.2195	0.2195	0.2195
Lines * Location * Hearing Type	1	0.002	0.002	0.868	0.3789	0.3789	0.3789
Lines * Location * Sex	1	0.011	0.011	4.387	0.0695	0.0695	0.0695
Lines * Location * Hearing Type * Sex	1	0.004	0.004	1.809	0.2156	0.2156	0.2156
Lines * Location * Subject(Group)	8	0.020	0.002				
Lines * Context	1	0.009	0.009	6.674	0.0324	0.0324	0.0324
Lines * Context * Hearing Type	1	0.002	0.002	1.397	0.2712	0.2712	0.2712
Lines * Context * Sex	1	8.9E-05	8.9E-05	0.065	0.8047	0.8047	0.8047
Lines * Context * Hearing Type * Sex	1	0.003	0.003	2.368	0.1624	0.1624	0.1624
Lines * Context * Subject(Group)	8	0.011	0.001				
Location * Context	1	0.004	0.004	1.714	0.2268	0.2268	0.2268
Location * Context * Hearing Type	1	0.001	0.001	0.392	0.5485	0.5485	0.5485
Location * Context * Sex	1	9.7E-05	9.7E-05	0.045	0.8376	0.8376	0.8376
Location * Context * Hearing Type * Sex	1	0.006	0.006	2.827	0.1312	0.1312	0.1312
Location * Context * Subject(Group)	8	0.017	0.002				
Lines * Location * Context	1	0.008	0.008	0.987	0.3495	0.3495	0.3495
Lines * Location * Context * Hearing Type	1	0.009	0.009	1.103	0.3243	0.3243	0.3243
Lines * Location * Context * Sex	1	0.001	0.001	0.106	0.7530	0.7530	0.7530
Lines * Loc * Context * Hearing Type * Sex	1	0.004	0.004	0.504	0.4980	0.4980	0.4980
Lines * Location * Context * Subject(Group)	8	0.062	0.008				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Lines	1	1.429
Location	1	1.429
Context	1	1.429
Lines * Location	1	1.429
Lines * Context	1	1.429
Location * Context	1	1.429
Lines * Location * Context	1	1.429

Dependent: Recall Accuracy

NOTE: Probabilities are not corrected for values of epsilon greater than 1.

Experiment 2 - Rate x Lag (all subjects)

Repeated Measures ANOVA - Type III Sums of Squares **R*L, All Subjects**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Hearing Type	1	0.257	0.257	6.410	0.0239		
Subject(Group)	14	0.560	0.040				
Rate	1	1.9E-04	1.9E-04	0.036	0.8519	0.8519	0.8519
Rate * Hearing Type	1	0.001	0.001	0.128	0.7255	0.7255	0.7255
Rate * Subject(Group)	14	0.074	0.005				
Lag	1	0.008	0.008	0.835	0.3763	0.3763	0.3763
Lag * Hearing Type	1	0.015	0.015	1.573	0.2303	0.2303	0.2303
Lag * Subject(Group)	14	0.134	0.010				
Set Size	4	3.383	0.846	74.962	0.0001	0.0001	0.0001
Set Size * Hearing Type	4	0.177	0.044	3.914	0.0071	0.0398	0.0305
Set Size * Subject(Group)	56	0.632	0.011				
Rate * Lag	1	0.012	0.012	4.081	0.0629	0.0629	0.0629
Rate * Lag * Hearing Type	1	0.026	0.026	8.422	0.0116	0.0116	0.0116
Rate * Lag * Subject(Group)	14	0.043	0.003				
Rate * Set Size	4	0.002	4.8E-04	0.110	0.9785	0.9283	0.9610
Rate * Set Size * Hearing Type	4	0.006	0.001	0.337	0.8517	0.7581	0.8119
Rate * Set Size * Subject(Group)	56	0.244	0.004				
Lag * Set Size	4	0.011	0.003	0.396	0.8109	0.7188	0.7723
Lag * Set Size * Hearing Type	4	0.036	0.009	1.260	0.2964	0.3007	0.2996
Lag * Set Size * Subject(Group)	56	0.399	0.007				
Rate * Lag * Set Size	4	0.011	0.003	0.684	0.6062	0.5384	0.5747
Rate * Lag * Set Size * Hearing Type	4	0.033	0.008	2.147	0.0870	0.1236	0.1044
Rate * Lag * Set Size * Subject(Group)	56	0.217	0.004				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Rate	1	1.077
Lag	1	1.077
Set Size	0.425	0.513
Rate * Lag	1	1.077
Rate * Set Size	0.612	0.805
Lag * Set Size	0.617	0.813
Rate * Lag * Set Size	0.606	0.795

Dependent: Recall Accuracy

NOTE: Probabilities are not corrected for values of epsilon greater than 1.

Repeated Measures ANOVA - Type III Sums of Squares **R*L, All Subjects, Context**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Hearing Type	1	0.164	0.164	6.074	0.0273		
Subject(Group)	14	0.379	0.027				
Rate	1	7.0E-05	7.0E-05	0.022	0.8849	0.8849	0.8849
Rate * Hearing Type	1	0.001	0.001	0.315	0.5836	0.5836	0.5836
Rate * Subject(Group)	14	0.045	0.003				
Lag	1	0.006	0.006	0.986	0.3375	0.3375	0.3375
Lag * Hearing Type	1	0.009	0.009	1.563	0.2317	0.2317	0.2317
Lag * Subject(Group)	14	0.083	0.006				
Context	1	0.295	0.295	19.46	0.0006	0.0006	0.0006
Context * Hearing Type	1	0.080	0.080	5.296	0.0373	0.0373	0.0373
Context * Subject(Group)	14	0.212	0.015				
Rate * Lag	1	0.009	0.009	4.933	0.0434	0.0434	0.0434
Rate * Lag * Hearing Type	1	0.018	0.018	9.950	0.0070	0.0070	0.0070
Rate * Lag * Subject(Group)	14	0.026	0.002				
Rate * Context	1	1.7E-04	1.7E-04	0.024	0.8785	0.8785	0.8785
Rate * Context * Hearing Type	1	0.012	0.012	1.767	0.2051	0.2051	0.2051
Rate * Context * Subject(Group)	14	0.099	0.007				
Lag * Context	1	0.001	0.001	0.391	0.5418	0.5418	0.5418
Lag * Context * Hearing Type	1	2.5E-07	2.5E-07	6.7E-05	0.9936	0.9936	0.9936
Lag * Context * Subject(Group)	14	0.052	0.004				
Rate * Lag * Context	1	4.2E-06	4.2E-06	0.001	0.9704	0.9704	0.9704
Rate * Lag * Context * Hearing Type	1	0.018	0.018	6.235	0.0256	0.0256	0.0256
Rate * Lag * Context * Subject(Group)	14	0.041	0.003				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Rate	1	1.077
Lag	1	1.077
Context	1	1.077
Rate * Lag	1	1.077
Rate * Context	1	1.077
Lag * Context	1	1.077
Rate * Lag * Context	1	1.077

Dependent: Recall Accuracy

NOTE: Probabilities are not corrected for values of epsilon greater than 1.

Experiment 2 - Rate x Lag (hearing subjects)

Repeated Measures ANOVA - Type III Sums of Squares **R*L, Hearing Subjects**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Subject(Group)	11	0.543	0.049				
Rate	1	1.5E-04	1.5E-04	0.027	0.8719	0.8719	0.8719
Rate * Subject(Group)	11	0.060	0.005				
Lag	1	0.045	0.045	4.246	0.0638	0.0638	0.0638
Lag * Subject(Group)	11	0.117	0.011				
Set Size	4	2.188	0.547	42.584	0.0001	0.0001	0.0001
Set Size * Subject(Group)	44	0.565	0.013				
Rate * Lag	1	0.002	0.002	0.872	0.3704	0.3704	0.3704
Rate * Lag * Subject(Group)	11	0.030	0.003				
Rate * Set Size	4	0.008	0.002	0.429	0.7867	0.6882	0.7367
Rate * Set Size * Subject(Group)	44	0.193	0.004				
Lag * Set Size	4	0.069	0.017	2.315	0.0722	0.1229	0.1095
Lag * Set Size * Subject(Group)	44	0.327	0.007				
Rate * Lag * Set Size	4	0.021	0.005	1.479	0.2249	0.2501	0.2459
Rate * Lag * Set Size * Subject(Group)	44	0.154	0.004				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Rate	1	1
Lag	1	1
Set Size	0.378	0.425
Rate * Lag	1	1
Rate * Set Size	0.591	0.763
Lag * Set Size	0.496	0.604
Rate * Lag * Set Size	0.490	0.596

Dependent: Recall Accuracy

Experiment 2 - Lag (all subjects)

Repeated Measures ANOVA - Type III Sums of Squares **Lag, All Subjects**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Hearing Type	1	0.188	0.188	6.484	0.0233		
Subject(Group)	14	0.406	0.029				
Lag	2	0.057	0.029	3.436	0.0463	0.0592	0.0492
Lag * Hearing Type	2	0.002	0.001	0.133	0.8760	0.8309	0.8663
Lag * Subject(Group)	28	0.233	0.008				
Set Size	4	2.208	0.552	66.78	0.0001	0.0001	0.0001
Set Size * Hearing Type	4	0.172	0.043	5.217	0.0012	0.0149	0.0094
Set Size * Subject(Group)	56	0.463	0.008				
Lag * Set Size	8	0.049	0.006	1.207	0.3016	0.3187	0.3126
Lag * Set Size * Hearing Type	8	0.030	0.004	0.737	0.6589	0.5626	0.6135
Lag * Set Size * Subject(Group)	112	0.571	0.005				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Lag	0.799	0.950
Set Size	0.451	0.550
Lag * Set Size	0.467	0.704

Dependent: Recall Accuracy

Repeated Measures ANOVA - Type III Sums of Squares **Lag, All Subjects, Context**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Hearing Type	1	0.128	0.128	6.666	0.0217		
Subject(Group)	14	0.270	0.019				
Lag	2	0.038	0.019	3.751	0.0360	0.0462	0.0366
Lag * Hearing Type	2	0.002	0.001	0.158	0.8546	0.8148	0.8523
Lag * Subject(Group)	28	0.141	0.005				
Context	1	0.207	0.207	23.801	0.0002	0.0002	0.0002
Context * Hearing Type	1	0.107	0.107	12.322	0.0035	0.0035	0.0035
Context * Subject(Group)	14	0.121	0.009				
Lag * Context	2	0.001	4.2E-04	0.095	0.9097	0.9013	0.9097
Lag * Context * Hearing Type	2	0.010	0.005	1.095	0.3484	0.3464	0.3484
Lag * Context * Subject(Group)	28	0.124	0.004				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Lag	0.825	0.988
Context	1	1.077
Lag * Context	0.951	1.176

Dependent: Recall Accuracy

NOTE: Probabilities are not corrected for values of epsilon greater than 1.

Experiment 2 -Lag (hearing subjects)

Repeated Measures ANOVA - Type III Sums of Squares **Lag, Hearing Subjects**

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Subject	11	0.381	0.035				
Lag	2	0.037	0.018	1.835	0.1832	0.1933	0.1887
Lag * Subject	22	0.221	0.010				
Set Size	4	1.275	0.319	35.802	0.0001	0.0001	0.0001
Set Size * Subject	44	0.392	0.009				
Lag * Set Size	8	0.076	0.009	1.732	0.1020	0.1785	0.1536
Lag * Set Size * Subject	88	0.482	0.005				

Dependent: Recall Accuracy

Table of Epsilon Factors for df Adjustment	G-G Epsilon	H-F Epsilon
Lag	0.778	0.882
Set Size	0.381	0.429
Lag * Set Size	0.381	0.544

Dependent: Recall Accuracy

APPENDIX F - PREDICTION EQUATIONS FOR MERGED REGRESSIONS

Only the backwards stepwise equations are shown. These are simpler and less likely to lead to computational error than the equations for the full models. All coefficients have been reduced to 3 significant figures for ease of use. Some variables have been factored out to simplify computation. For the dependent terms, MR refers to mean number of words accurately recalled and RA refers to average recall accuracy.

Individual Factor Models

Table F1. Definition of individual regression terms

Term	Definition
Lines	Number of lines shown on the screen
LagTime	Amount of lag expressed in seconds
HT	Hearing type (0 hearing, 1 deaf)
Size	Set size
SRMS	Scrolled reading memory span (Captions Only score)

Table F2. Individual regression equations

Assessment with Set Size: $MR = 4.52 - .200Lines - .051LagTime + .554HT - 4.42SRMS$ $+ \ln(Size)[.235Lines + .061LagTime - .668HT + 4.96SRMS - 2.29]$
Generic with Set Size: $MR = .791 - .183Lines - .056LagTime + .637HT$ $+ \ln(Size)[.216Lines + .067LagTime - .761HT + 1899]$
Assessment without Set Size: $RA = .527 + .021Lines + .006LagTime - .064HT + .401SRMS$
Generic without Set Size: $RA = .866 + .020Lines + .006LagTime - .072HT$

Aggregate Factor Models

Table F3. Aggregate regression equations

Assessment with Set Size: $MR = .812 - 2.642HT + 14.54SRMS + Buffer\left[-1453SRMS + 3.962\right]$ $+ Speed\left[-.751Buffer + .601HT - 3.402SRMS + .399\right]$
Generic with Set Size: $MR = 12.985 + 2.736Buffer - 3.101HT$ $+ Speed\left[-.748Buffer + .710HT - 2.45\right]$
Assessment without Set Size: $RA = .525 + .064Buffer - .065HT + .403SRMS$
Generic without Set Size: $RA = .865 + .061Buffer - .074HT$

APPENDIX G - A PRELIMINARY SIMULATION MODEL OF THE INFORMATION FLOW

Overview of the model

This section describes an attempt to develop a preliminary model of the flow of information in a real-time captioning scenario. It may also be possible to expand this model to other applications involving redundant, parallel information source scenarios (e.g., RTC assisted language translation) with future development and research findings.

There are three basic stages of the information flow:

1. Delivery of incoming information,
2. Perceiving and differentiating incoming information,
3. The flow of information through working memory.

Figure G.1 describes this preliminary model in a graphical manner. The aggregates developed in the proposed model of interaction (Chapter 3) can be incorporated into this model of information flow. The most obvious linkage is the buffer aggregate and the "Still on Screen?" decision point. The buffer size will determine the outcome of this question. By definition, the "Lag" step in the model will also be described by the buffer size. The speed of the information flow aggregate is analogous to the rate at which new words are delivered and are processed by the user. This will essentially be how fast information can flow through the model. The steps in Figure G.1 that are tied to the buffer aggregate are underlined, while those tied to the speed aggregate are in *italics*. Rectangles indicate modes of information, rounded rectangles are events, pointed items are logical routings, and ovals are inputs and outputs.

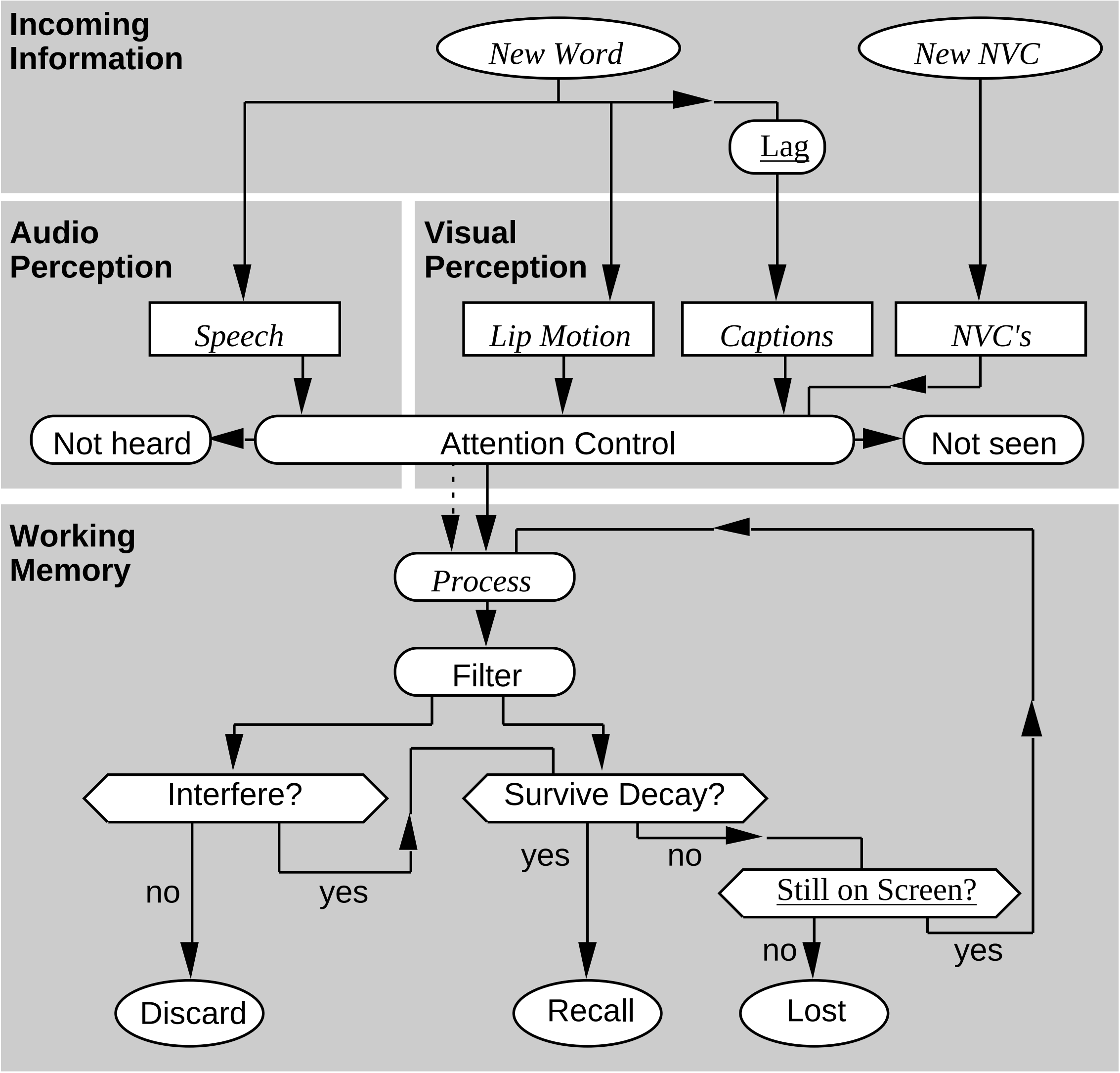


Figure G.1. A preliminary information flow model

(items related to the aggregates are marked: buffer and *speed*)

The first step is the delivery of incoming information. In a captioning scenario, the delivery of words will be in the form of audio, lip motions, and captions. In addition, there will be non-verbal cues (NVC's) delivered at somewhat random intervals. These NVC's will generally be visual in nature (e.g., facial expressions, hand waving, etc.).

The next stage of the model is the perception of the incoming information and the differentiation of what information is perceived from what information is not. This procedure is similar to existing descriptions of human capabilities (e.g., GOMS as described by Card, Moran, and Newell, 1986). Governing what information is perceived and what information is not is an attention control construct. This is not the same as Baddeley's central executive (1992). The central executive is described as being a part of working memory. The attention control construct described here refers to the process through which the user pays attention to certain information and ignores other information. It determines which source(s) enter working memory. Since all three of the word sources cannot be synchronized, it is possible that up to two can be passed into working memory at the same time (hence the dashed line in Figure G.1). Even if the text and the speaker's face could be in foveal vision at the same time, the work by Neisser and Becklen (1975) would suggest that only one of the visual sources can be attended to at a time.

The final stage of the model is the flow of the information through working memory. This model is somewhat different when compared to other models of working memory in that the major concern is not how fast the person processes information, but whether or not they are able to retain it. After processing the information into working memory, the user then strips out the information to be stored for recall, which in this case is the last word of each sentence. It is then necessary to determine if the retained information survives memory decay. If it cannot, the user can refresh the word by skimming the external buffer (the captions). However, if the word has scrolled off the screen, then the word is lost. There is also the possibility that words designated to be discarded cannot be suppressed and will remain in working memory.

Discussion of a simulation model

Overview

A preliminary simulation model was developed as an attempt to depict the flow of information in a quantitative manner. MicroSaint 1.2 by Micro Analysis and Design, Inc. (for the Macintosh) was selected as the software tool for constructing the model.

Some creative techniques were required to develop certain parts of the model due to software limitations. Thus, the preliminary simulation model does not appear the same as the logical diagram presented in the previous section. The most important change in appearance is the feeding of each new word into the working memory branch at the same time the word is sent to each perception branch. There are also branches for the attention control construct and for controlling caption scrolling. Each perception branch determines if the word has successfully been perceived. If so, the word is released into working memory. Once a word enters working memory, it is processed and filtered. The word then follows a path similar to the one described in the logical diagram.

Figure G.2 shows a screen shot of the simulation model. (For the sake of brevity, the computer coding of each object will not be shown later in this

Appendix.) The triangle next to New Word indicates the start of the model. MicroSaint requires each branch to be seeded individually due to syntax requirements. Variables are used to maintain synchronization and to transfer information across branches. The rectangles represent subnetworks. The subnetwork for Line Scroll keeps track of what line the captions are currently displaying. The Attention subnetwork determines to which visual source the person is currently attending. The Audio (aka Speech), NVC, Lip Motion, and Caption branches all determine if their respective sources were perceived. These branches, along with the Attention subnetwork correspond to the Perception layers in Figure G.1. The Release tasks at the end of each perception branch trigger entry into the Working Memory (WM) subnetwork (this is controlled within the Perceived? subnetwork). The Recalled subnetwork simply filters the key words from the extra, interfering words for scoring purposes. Individual cell descriptions can be found at the end of this Appendix.

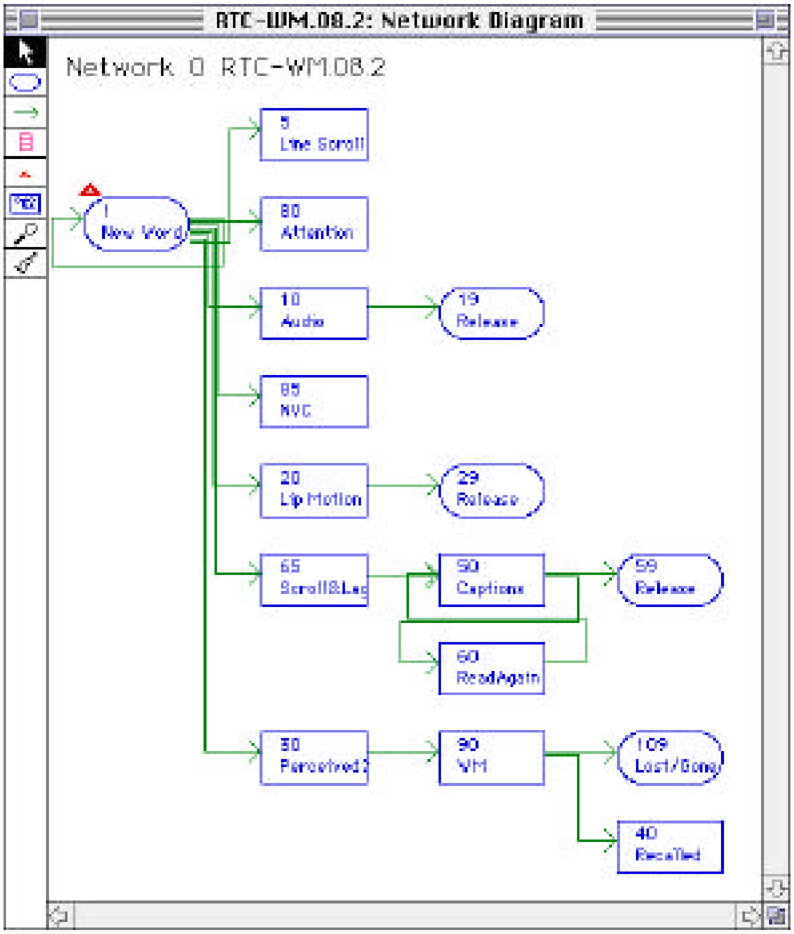


Figure G.2. The whole simulation model

Within the Working Memory subnetwork (Figure G.3, a detail of box 90 in Figure G.2) is a pattern similar to the one described in the working memory layer of the conceptual model (Figure G.1). The Into WM task filters out the non-key words and sends them to the Discard task. The Discard task uses a simple probability test to determine if a particular word will interfere with the retention activity. The Exhausted task determines if a discarded word has exhausted all chances to be processed (e.g., captions roll off screen). The last words are sent to Store instead of Discard. An exponential decay test is run to see if the word can survive until the trial is over. A source redundancy factor is applied to increase the likelihood a word will be recalled (Task 40). If a last word does not survive the decay test (Forget), it may cycle back through the decay test after being re-read off the captions (Refreshed).

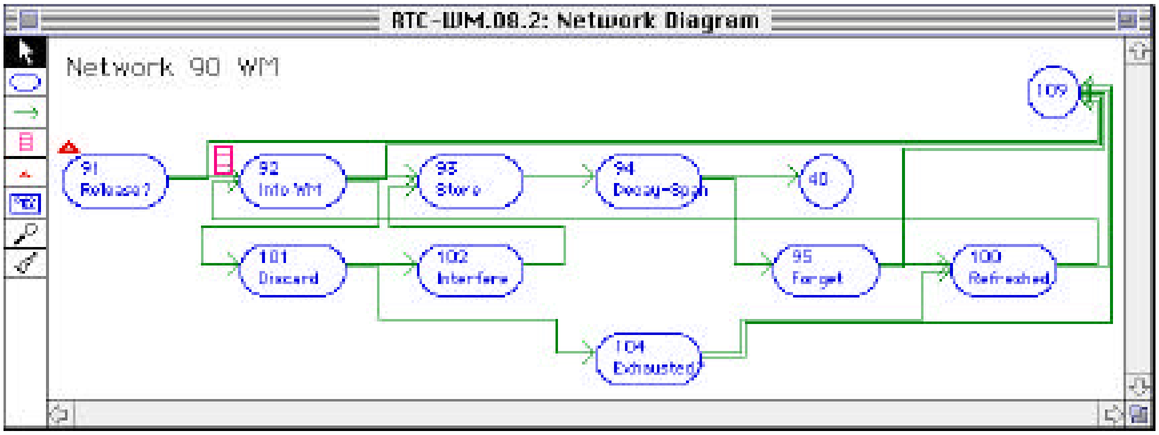


Figure G.3. The working memory subnetwork

Many of the processing times and decay functions for the model shown above were acquired through GOMS analysis (Card, Moran, and Newell, 1986). However, the attention control component requires additional research so that this simulation can generate accurate results. Decision rules are needed to define when and for how long attention is devoted to each source. For the purpose of testing, a simple probability test was used to see which visual source the person was looking at (the sources were not weighted equally).

Additional issues that need further investigation and development are related to syntax and queuing mechanisms (Liu, 1996c, 1997). This is especially true for the first portion of the working memory branch where words either enter into working memory or are set aside to wait for perception through an information source. This process has been particularly difficult to model adequately with the software. It may be necessary to restructure this entire portion or even switch software packages as a solution.

Simulation assumptions

The most important assumption in this simulation is the modeling of decay. The time a word needs to be retained for is tested against an exponential decay half-life of 7 seconds. This was based on the GOMS value of 7 seconds for 3 chunks present in memory (Card, Moran, and Newell, 1986). Preliminary simulations used set size functions that were extrapolations of GOMS values (73 seconds for 1 chunk and 7 seconds for 3 chunks). Since the simulations were returning very exaggerated performance results this simpler approach was deemed to be more appropriate.

Some of the basic assumptions used during the development of the current simulation are listed in Table G.1. Many of these assumptions can be modified or simplified after more research is completed. Some of the assumptions that appear to be easy to modify are, in fact, rather difficult. For example, the stipulation that most of the queues in the model do not need to be "first-in-first-out" (FIFO) is necessary at the present time due to difficulty with MicroSaint syntax and queuing mechanisms.

Table G.1. General simulation assumptions

Perception	• Short term auditory & visual store decays can be modeled by a simple probability. • Differences in hearing type can be emulated using P(hear the word) in the audio branch. • Attention control can be modeled by a simple probability, rather than scenario based. • Time for eye movements between visual sources can be ignored due to audio/visual overlap. • Queues do not need to be FIFO right now.
Working Memory	• Refresh of non-key words will occur once, at most. Key words will be refreshed as long as they are still visible in the captions. • An exponential decay can be used to see if a word survived decay. • After it is determined if a word has passed a basic decay test: P((85+Redundancy/3)/100) is reasonable. (The equation was developed during the tuning of the simulation model.) This says that you have an 85.67% chance to retain it on the first time if you've perceived 2 sources. It rises to 86% after the third source (e.g., saw lips, hear voice... then saw captions). Since key words cycle back through if they fail, it will be a 98% chance the second time it is processed (refreshed by captions) and a 99.7% the third time (re-read). So, the critical question is: Can it survive decay? • No rehearsal. • Chance of interference from non-key words can be modeled by a simple probability using set size to determine if it will go through the "retain" branch. Higher set sizes increase likelihood that this will occur. • Queues do not need to be FIFO right now.

Table G.2 supplies some of the assumptions for specific variables. Some of these are very simplistic (e.g., each word is one chunk) as this model is very preliminary. The variable assumptions below were utilized to acquire some preliminary findings. The working memory capacity and the baseline reading memory span were kept fixed during data generation for the sake of simplicity.

Times for perception and cognition were taken from GOMS middleman values (Card, Moran, and Newell, 1986).

Table G.2. Simulation variable assumptions

Variable	**Value**	**Rationale**
Chunks/Word	1	The words are short. Set for simplicity.
Words/Line	8	40 characters screen width, 5 characters a word.
P(hear) for deaf	.10	Assuming limited aural capabilities.
P(hear) for hearing	.95	Assuming slightly poor acoustics (classroom).
Short Term Auditory & Visual Stores	.9	Simple probability on surviving short term stores. Set for simplicity.
P(attend to captions)	.8	Set for simplicity. Educated guess.
P(attend to face)	.2	Set for simplicity. Educated guess.
P(attend to NVC)	0	Set for simplicity. Educated guess.
Working memory decay	7	Half-life in seconds. Card, Moran, and Newell (1986).
P(interference)	.005	Multiplied by Set Size since larger sets would have more opportunity for interference. Educated guess. Accurate value to be determined.
Baseline reading memory span	4	Set and fixed for simplicity.
Working memory capacity	7	Card, Moran, and Newell (1986). Fixed for simplicity.

Some preliminary findings

Since the flow model and the simulation were developed with the number of words recalled as the primary measure, the output of this simulation can be directly compared to the data collected in the experiments. It is possible to test the model's ability to accurately describe the flow of information by adjusting the caption presentation parameters. Preliminary confirmation of the model can be achieved if the simulation provides findings comparable to the data collected.

The simulation was run under real-time captioning (RTC) conditions from the first study with enough iterations to create a data set for the same number of subjects (2 conditions, 12 subjects/iterations). Experiment 1 was chosen as a good study to compare to as the findings provided a precise description of working memory effects. Please note that the simulation does not examine display location. The lack of a significant Location effect in the first experiment indicated that other factors are more important. As shown in Figure G.4, the simulation over-estimates for the higher set sizes. Refinement of the assumptions would likely reduce this effect. However, it does accurately depict better performance with four lines of captions.

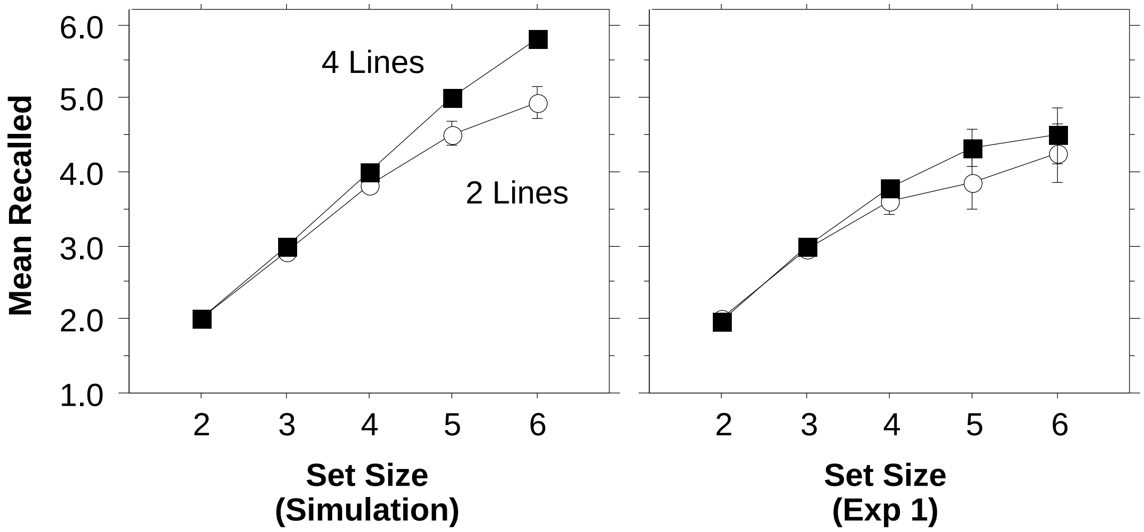


Figure G.4. Comparison of findings over the number of lines

(with 95% confidence intervals)

As seen in Table G.3, the simulation results are highly correlated to the experimental findings (all were significant at p<.001). It seems the simulation is better when the comparison is over four lines of captions or hearing subjects. This is most likely due to the fact that the simulation over-estimates performance. Experimental conditions with higher performance will be closer to the simulation, and thus will exhibit higher correlations.

Table G.3. Correlation between simulation and first study for RTC conditions

Comparison	**Correlation**
All	.838
2 Lines	.786
4 Lines	.856
Deaf	.807
Hearing	.912

Figure G.5 illustrates the over-estimation. The diagonal line represents perfect alignment between the simulation and the experimental findings. The data clearly resides below this line, especially at higher set sizes. Each circle represents a subject-iteration with matched hearing type and number of lines.

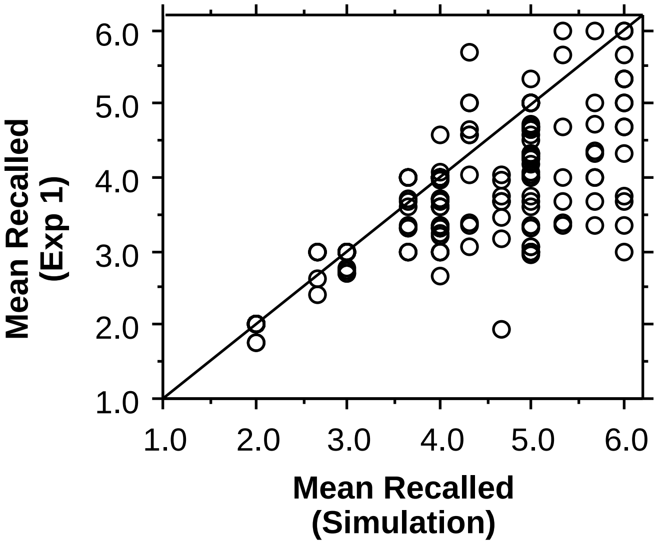
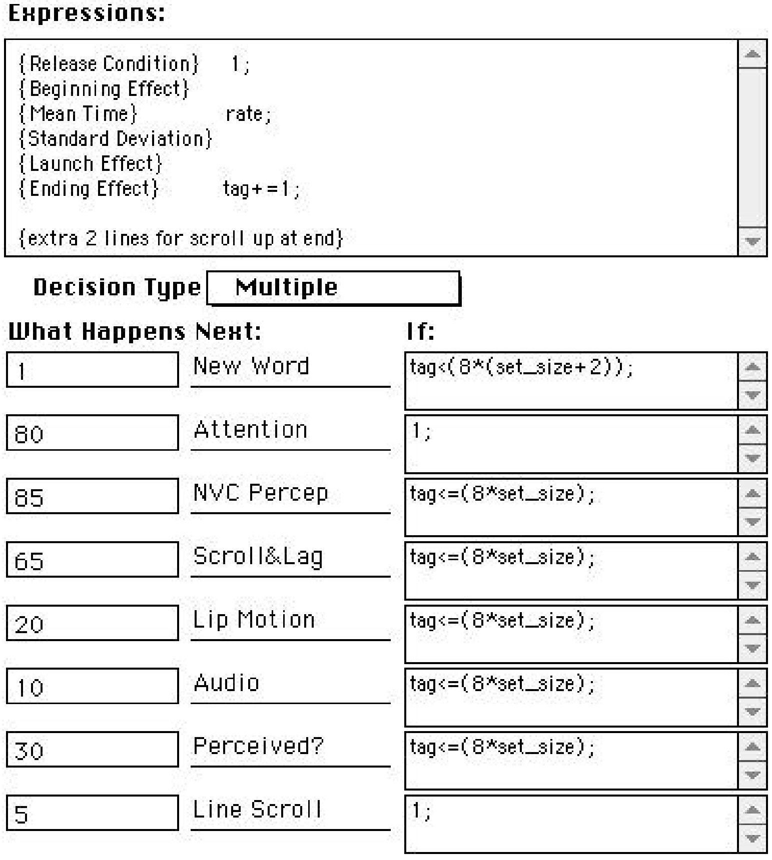


Figure G.5. Mean recalled of simulation vs. first study

The high, significant correlations suggest that it may be possible to refine the assumptions and make the simulation more accurate (e.g., reducing the over-estimation of performance). It is possible that useful data can be acquired when the attention control mechanism is better understood and the software limitations are resolved. This model clearly illustrates the potential to be a rather powerful descriptor of how parallel information is processed. Thus, a continuation of this work would be an obvious choice for future study.

Cell descriptions

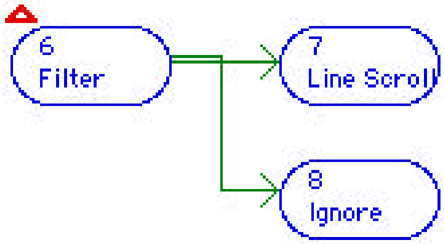
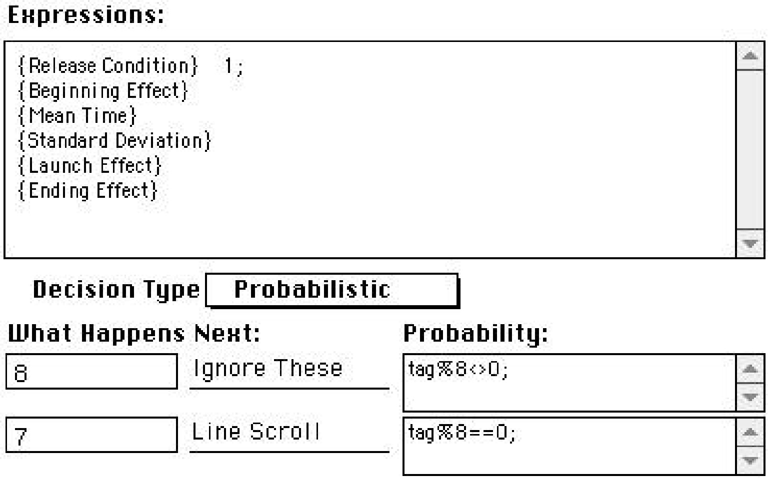
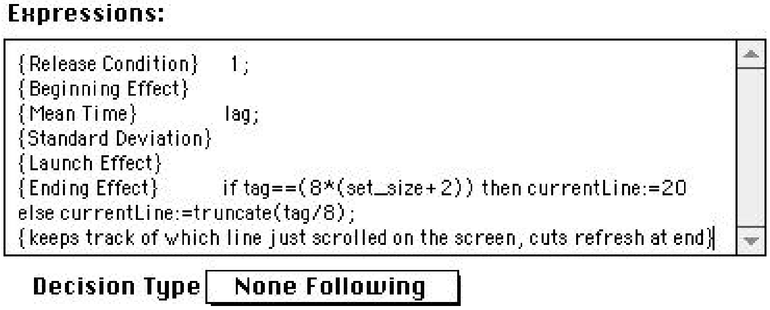
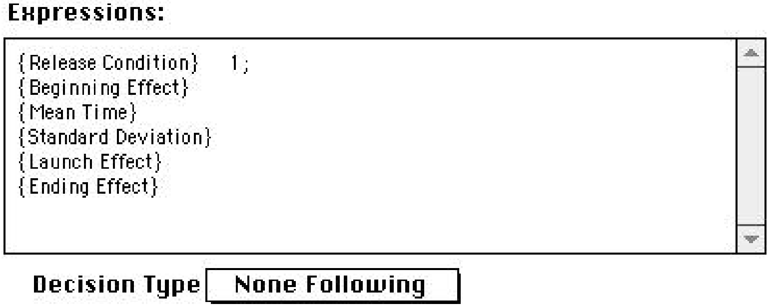
#	Name	Time Dist.	Logic & Routing
1	New Word	Normal	

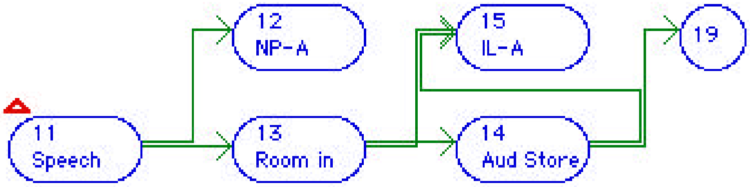
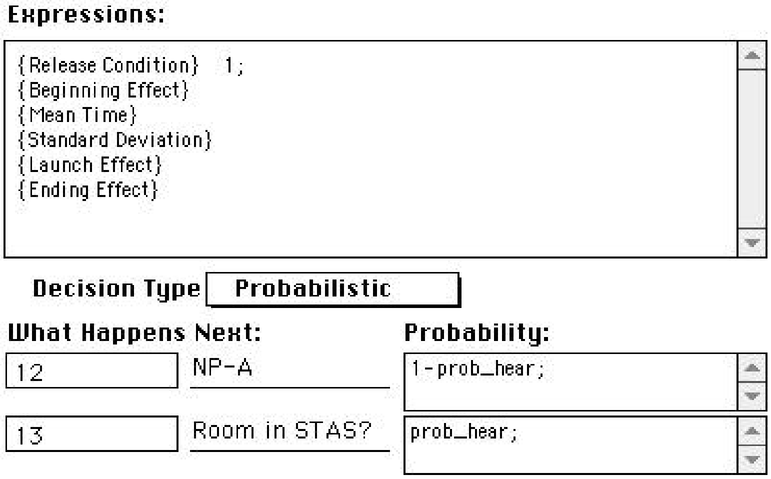
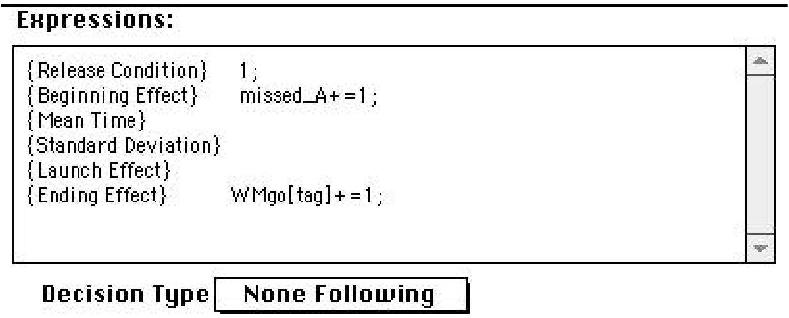
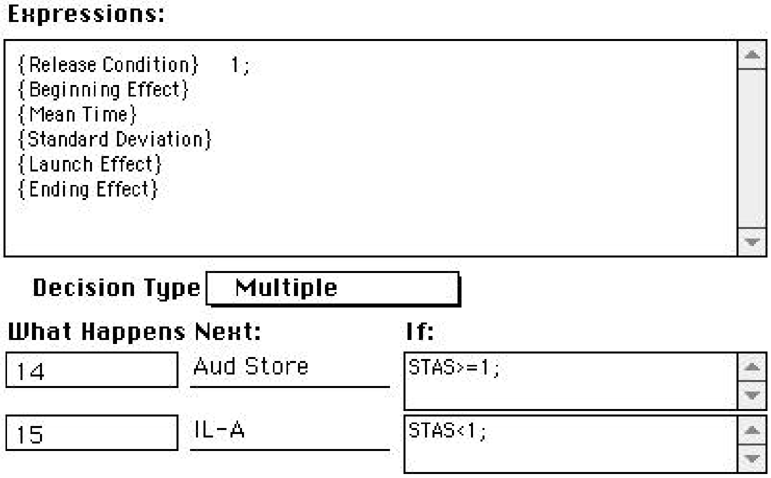
Expressions:

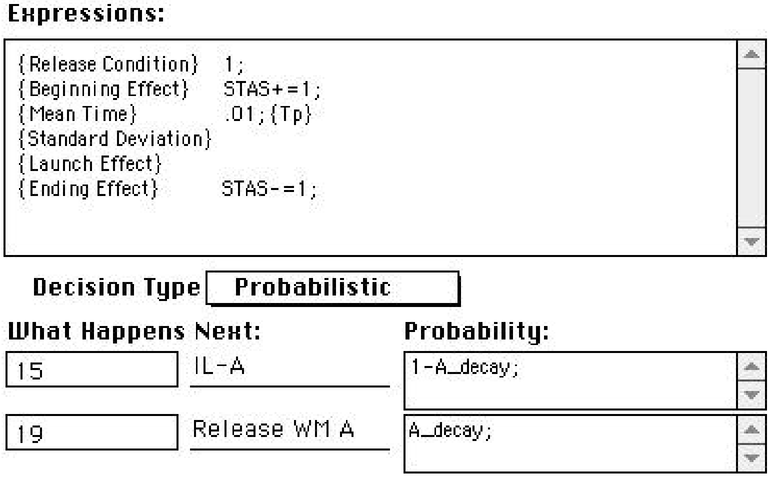
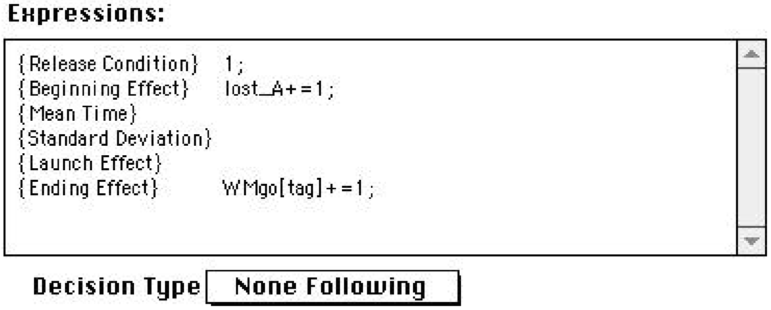
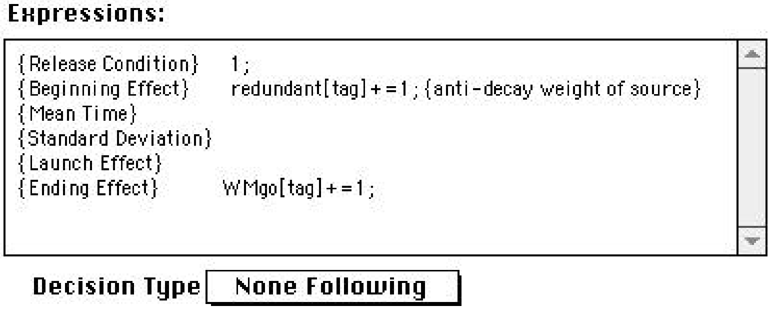
```
{Release Condition}    1;
{Beginning Effect}
{Mean Time}            rate;
{Standard Deviation}
{Launch Effect}
{Ending Effect}        tag+=1;

{extra 2 lines for scroll up at end}
```

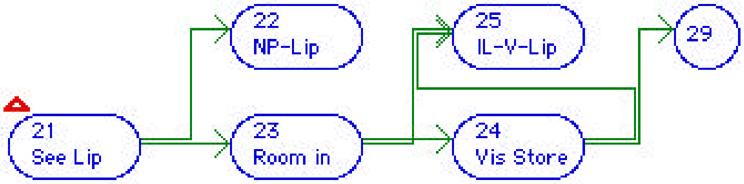
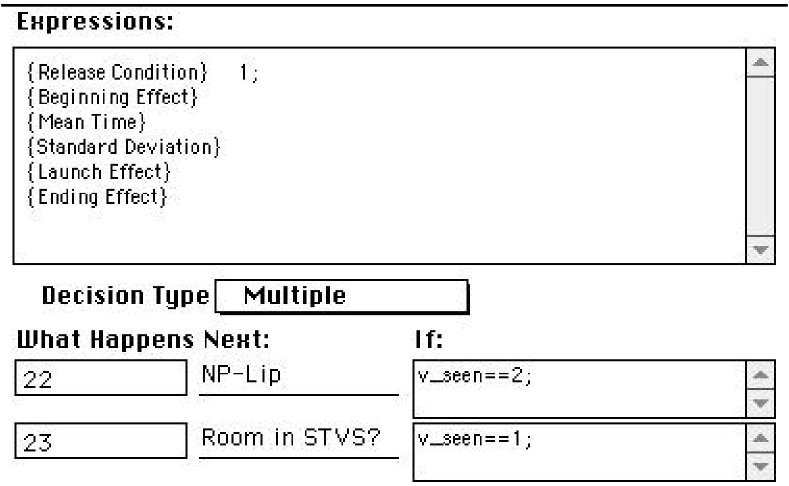
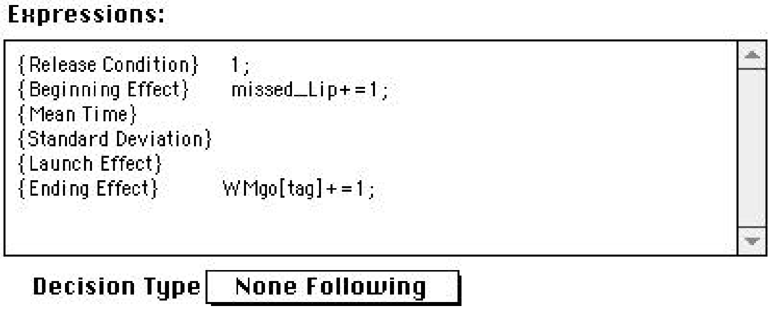
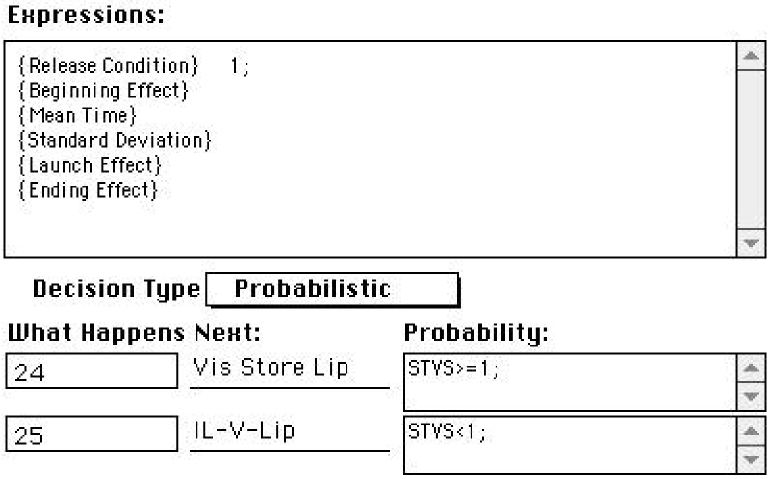
Decision Type Multiple

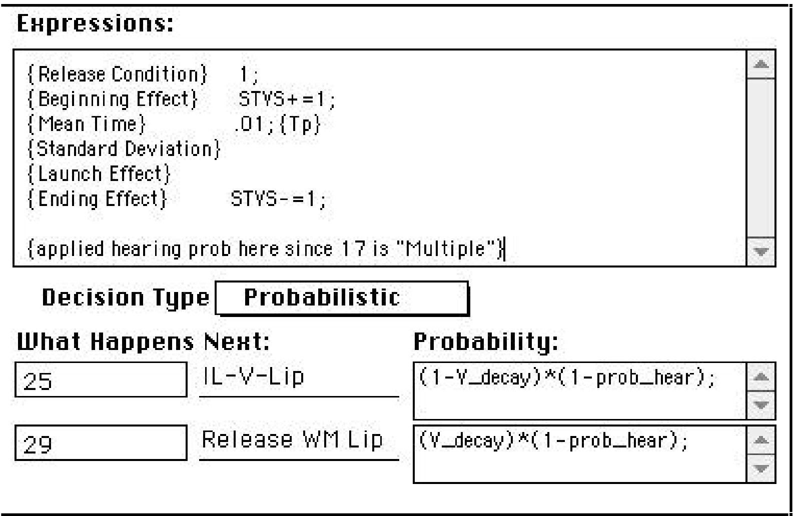
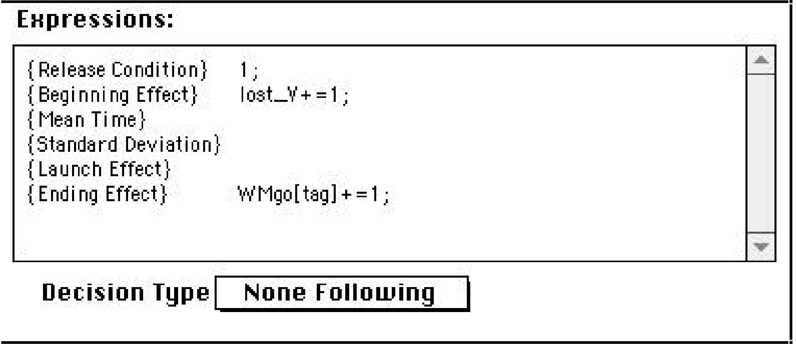
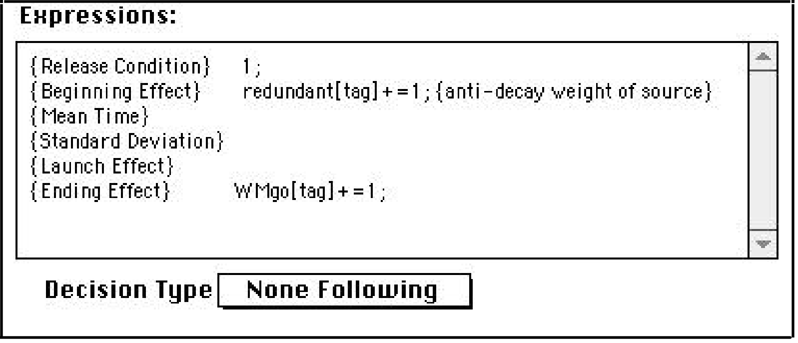
What Happens Next:		If:
1	New Word	tag<(8*(set_size+2));
80	Attention	1;
85	NVC Percep	tag<=(8*set_size);
65	Scroll&Lag	tag<=(8*set_size);
20	Lip Motion	tag<=(8*set_size);
10	Audio	tag<=(8*set_size);
30	Perceived?	tag<=(8*set_size);
5	Line Scroll	1;

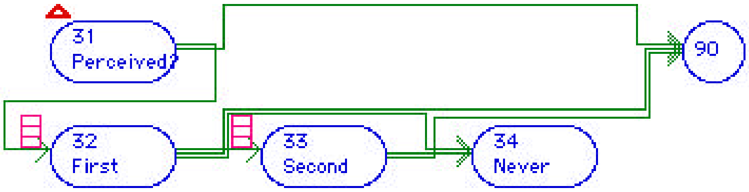
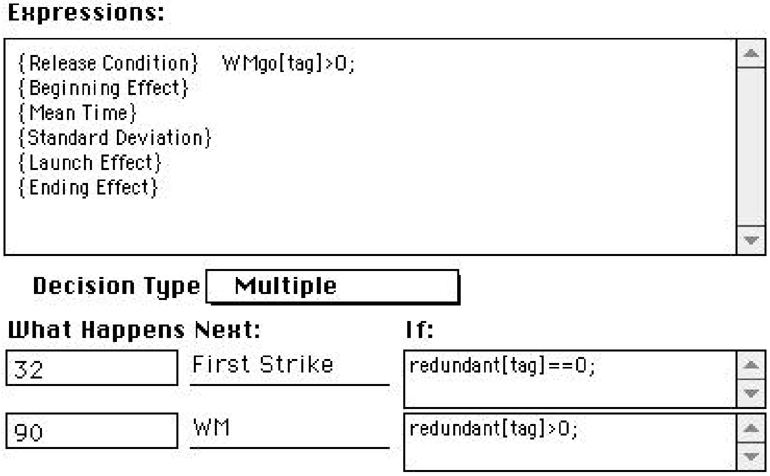
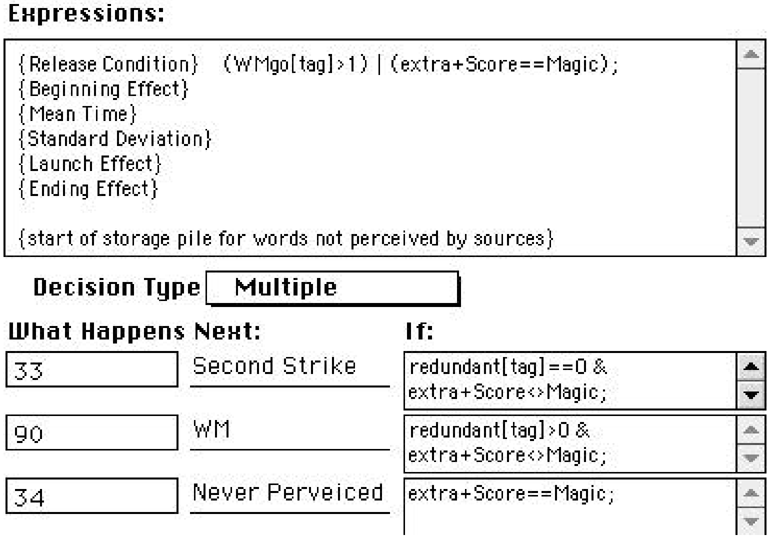
#	Name	Time Dist.	Logic & Routing
5	Line Scroll	Sub-network	
6	Filter	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Probabilistic **What Happens Next:** **Probability:** 8 Ignore These tag%8<>0; 7 Line Scroll tag%8==0;
7	Line Scroll	Normal	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} lag; {Standard Deviation} {Launch Effect} {Ending Effect} if tag==(8*(set_size+2)) then currentLine:=20 else currentLine:=truncate(tag/8); {keeps track of which line just scrolled on the screen, cuts refresh at end} **Decision Type** None Following
8	Ignore These	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** None Following

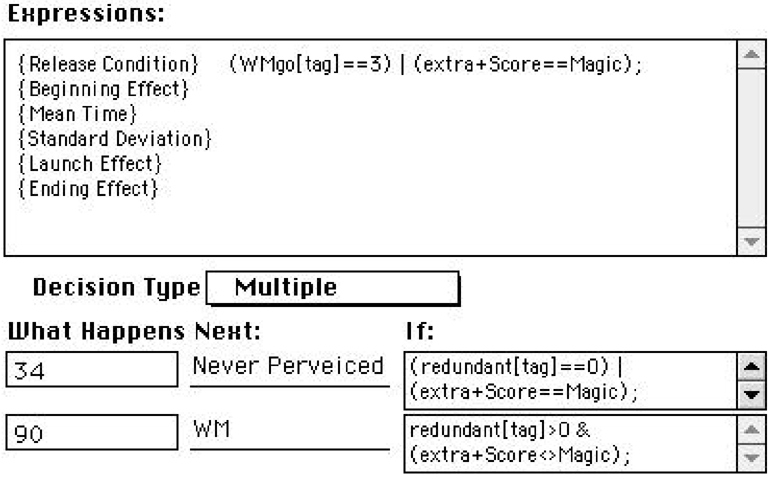
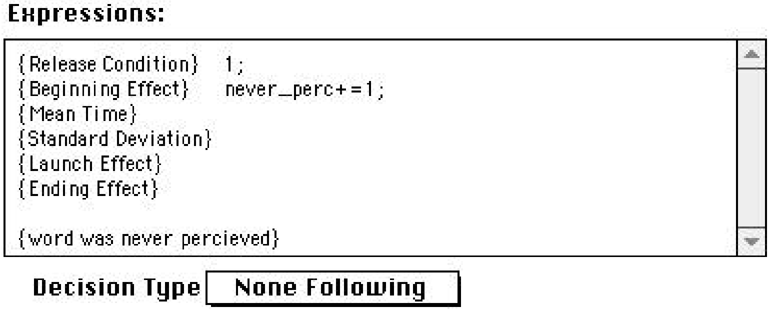
#	Name	Time Dist.	Logic & Routing
10	Audio	Sub-network	
11	Speech	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Probabilistic **What Happens Next:** **Probability:** 12 NP-A 1-prob_hear; 13 Room in STAS? prob_hear;
12	NP-A	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} missed_A+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; **Decision Type** None Following
13	Room in STAS?	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Multiple **What Happens Next:** **If:** 14 Aud Store STAS>=1; 15 IL-A STAS<1;

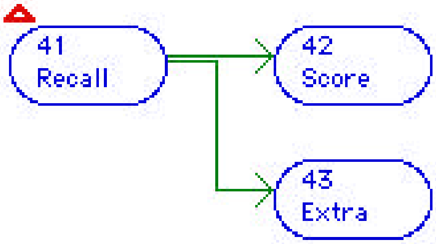
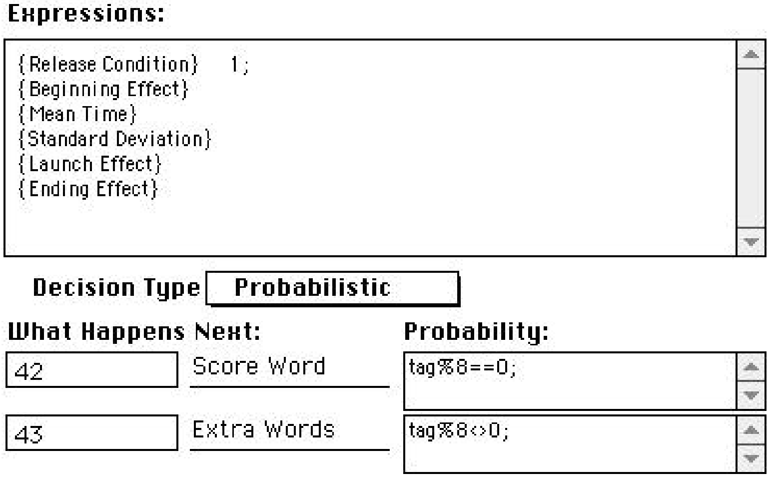
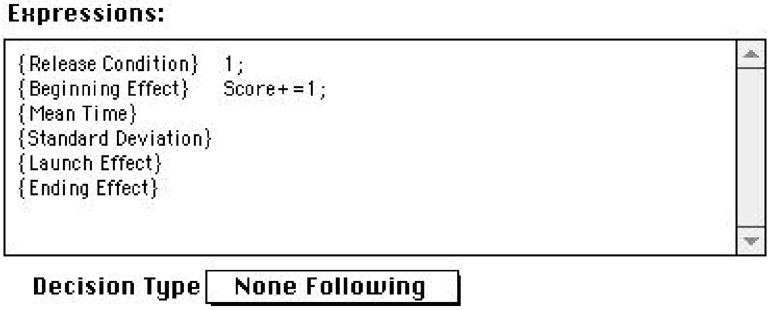
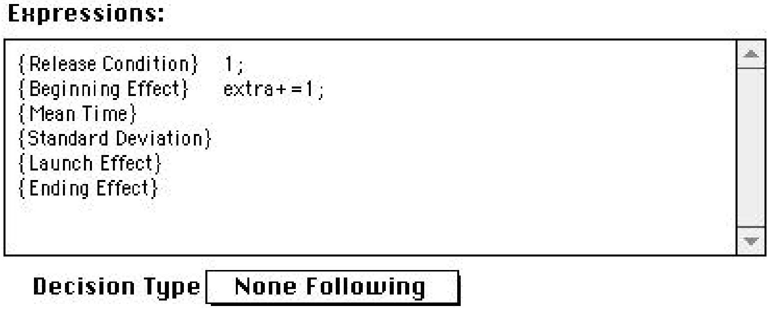
14	Aud Store	Normal	**Expressions:** {Release Condition} 1; {Beginning Effect} STAS+=1; {Mean Time} .01;{Tp} {Standard Deviation} {Launch Effect} {Ending Effect} STAS-=1; **Decision Type** Probabilistic **What Happens Next:** **Probability:** 15 IL-A 1-A_decay; 19 Release WM A A_decay;
15	IL-A	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} lost_A+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; **Decision Type** None Following
19	Release WM A	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} redundant[tag]+=1;{anti-decay weight of source} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; **Decision Type** None Following

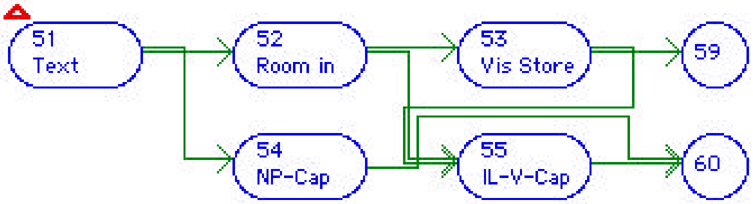
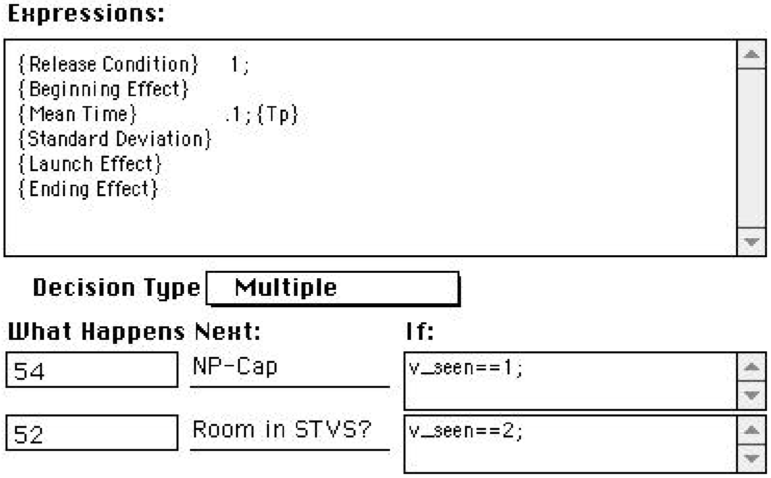
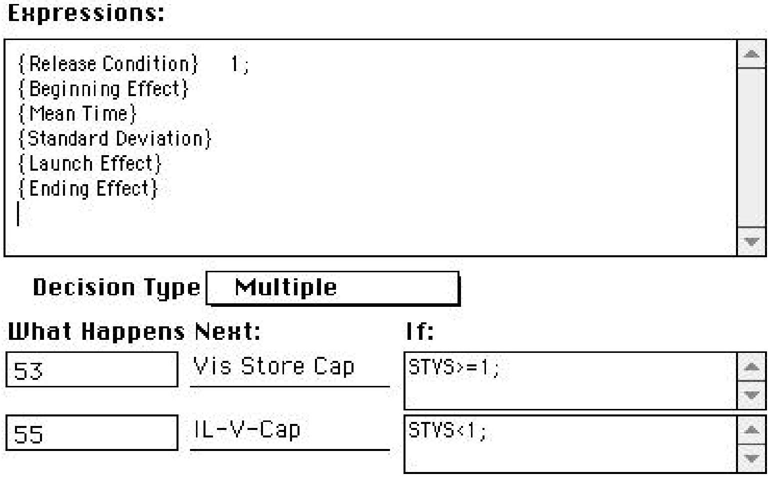
#	Name	Time Dist.	Logic & Routing
20	Lip Motion	Sub-network	
21	See Lip Motion	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Multiple **What Happens Next:** **If:** 22 NP-Lip v_seen==2; 23 Room in STVS? v_seen==1;
22	NP-Lip	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} missed_Lip+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; **Decision Type** None Following
23	Room in STVS? Lip	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Probabilistic **What Happens Next:** **Probability:** 24 Vis Store Lip STVS>=1; 25 IL-V-Lip STVS<1;

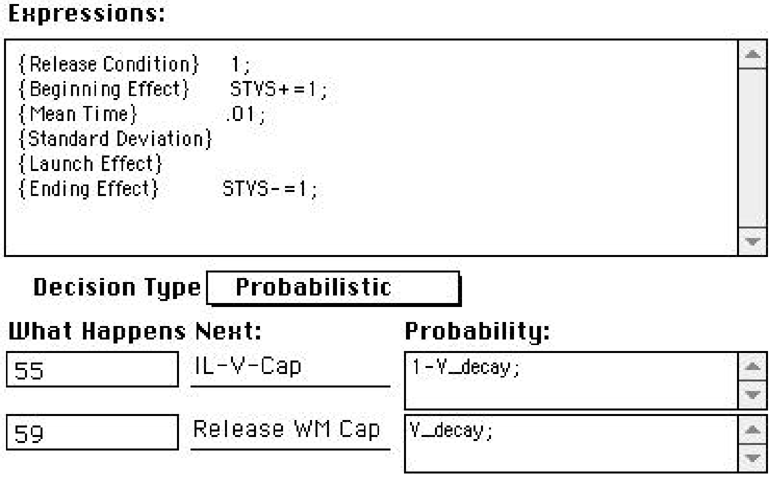
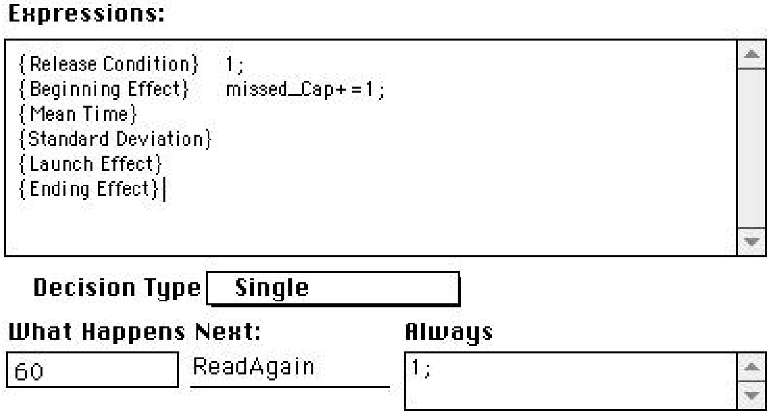
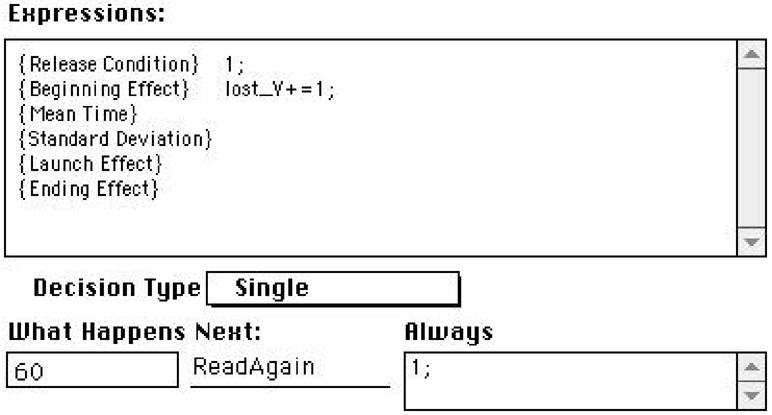
24	Vis Store Lip	Normal	 **Expressions:** {Release Condition} 1; {Beginning Effect} STVS+=1; {Mean Time} .01;{Tp} {Standard Deviation} {Launch Effect} {Ending Effect} STVS-=1; {applied hearing prob here since 17 is "Multiple"} **Decision Type** Probabilistic **What Happens Next:** **Probability:** 25 IL-V-Lip (1-V_decay)*(1-prob_hear); 29 Release WM Lip (V_decay)*(1-prob_hear);
25	IL-V-Lip	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} lost_V+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; **Decision Type** None Following
29	Release WM Lip	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} redundant[tag]+=1;{anti-decay weight of source} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; **Decision Type** None Following

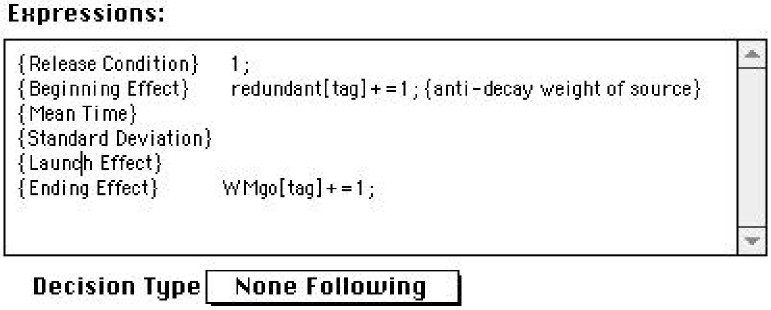
#	Name	Time Dist.	Logic & Routing
30	Perceived?	Sub-network	
31	Perceived?	n/a	**Expressions:** {Release Condition} WMgo[tag]>0; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: Multiple What Happens Next: / If: 32 First Strike — redundant[tag]==0; 90 WM — redundant[tag]>0;
32	First Strike	n/a	**Expressions:** {Release Condition} (WMgo[tag]>1) | (extra+Score==Magic); {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {start of storage pile for words not perceived by sources} Decision Type: Multiple What Happens Next: / If: 33 Second Strike — redundant[tag]==0 & extra+Score<>Magic; 90 WM — redundant[tag]>0 & extra+Score<>Magic; 34 Never Perveiced — extra+Score==Magic;

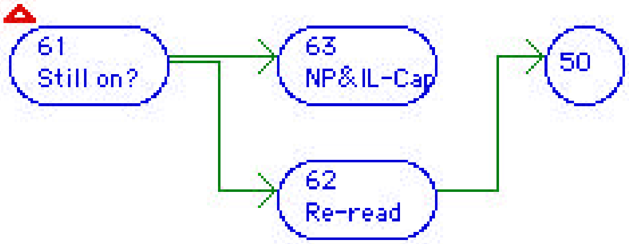
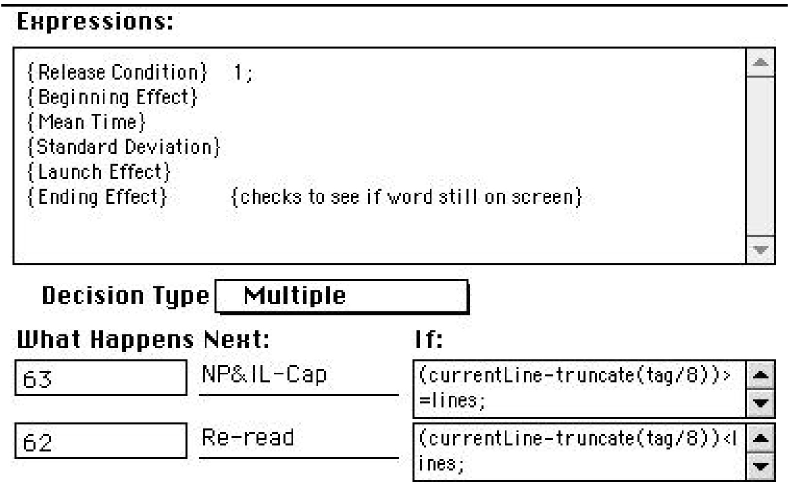
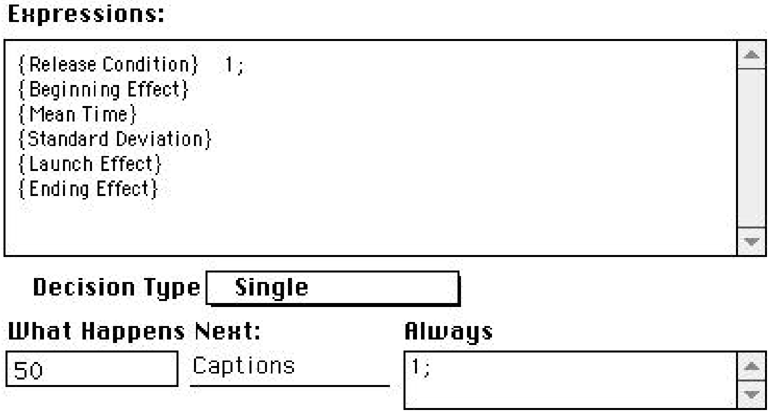
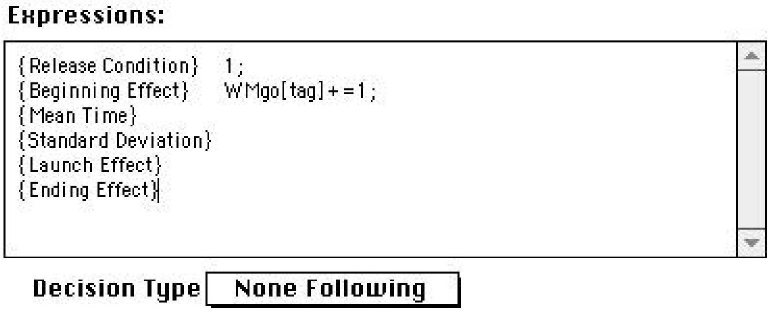
33	Second Strike	n/a	 **Expressions:** {Release Condition} (WMgo[tag]==3) | (extra+Score==Magic); {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Multiple **What Happens Next:** **If:** 34 Never Perveiced (redundant[tag]==0) | (extra+Score==Magic); 90 WM redundant[tag]>0 & (extra+Score<>Magic);
34	Never Perceived	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} never_perc+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {word was never percieved} **Decision Type** None Following

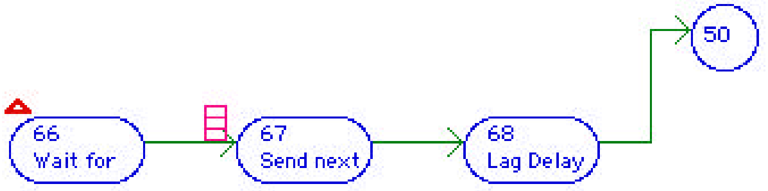
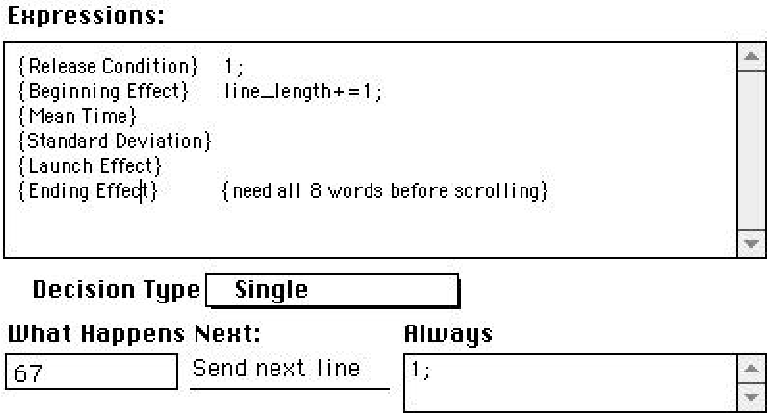
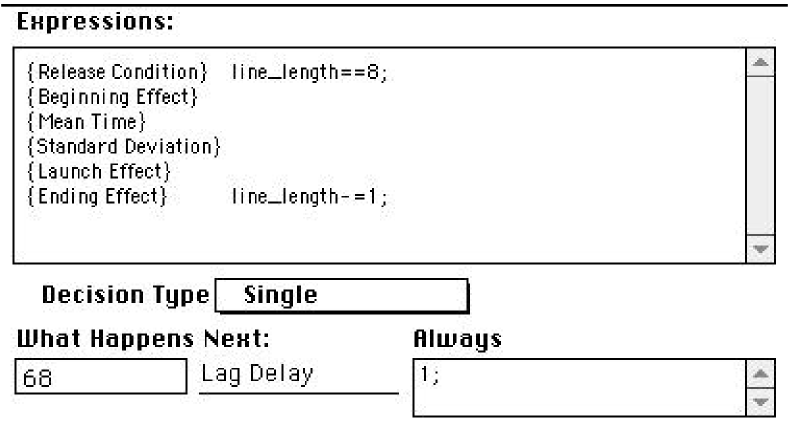
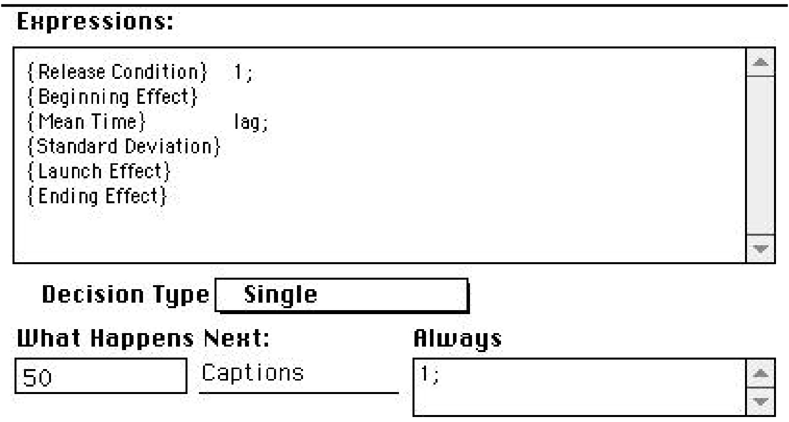
#	Name	Time Dist.	Logic & Routing
40	Recalled	Sub-network	
41	Recall Word	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: Probabilistic What Happens Next: / Probability: 42 Score Word: tag%8==0; 43 Extra Words: tag%8<>0;
42	Score Word	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} Score+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: None Following
43	Extra Words	Normal	**Expressions:** {Release Condition} 1; {Beginning Effect} extra+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: None Following

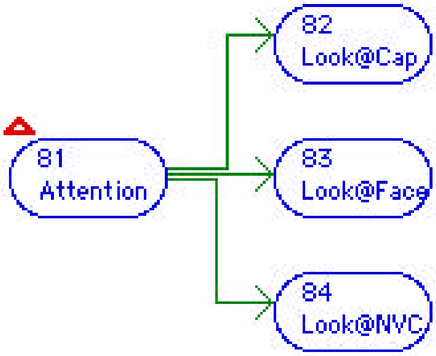
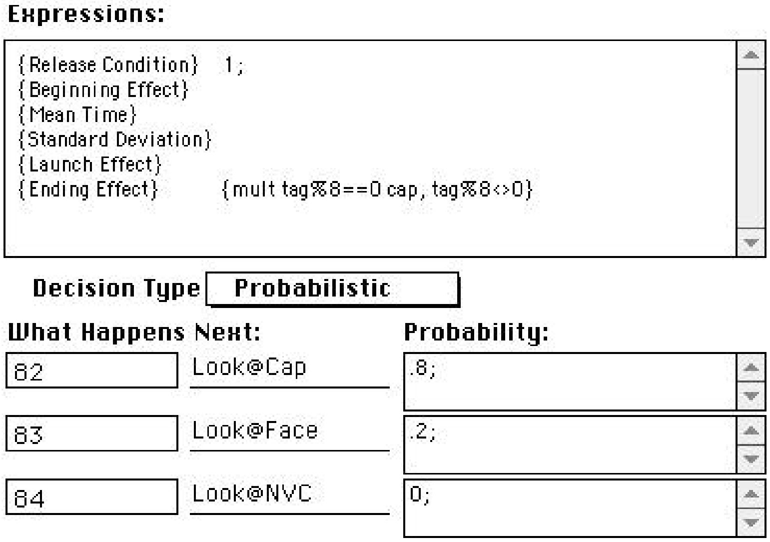
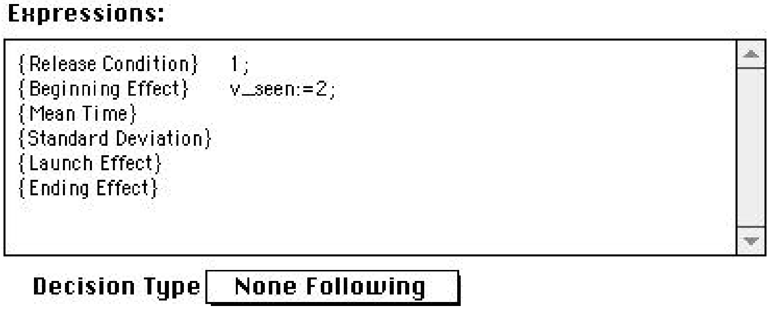
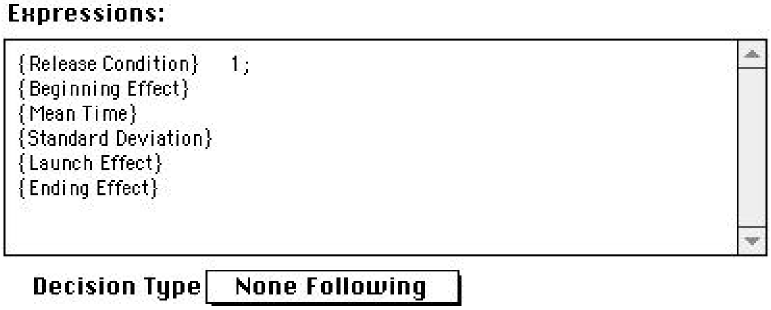
#	Name	Time Dist.	Logic & Routing
50	Captions	Sub-network	
51	Text	Normal	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} .1;{Tp} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Multiple **What Happens Next:** **If:** 54 NP-Cap v_seen==1; 52 Room in STVS? v_seen==2;
52	Room in STVS? Cap	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Multiple **What Happens Next:** **If:** 53 Vis Store Cap STVS>=1; 55 IL-V-Cap STVS<1;

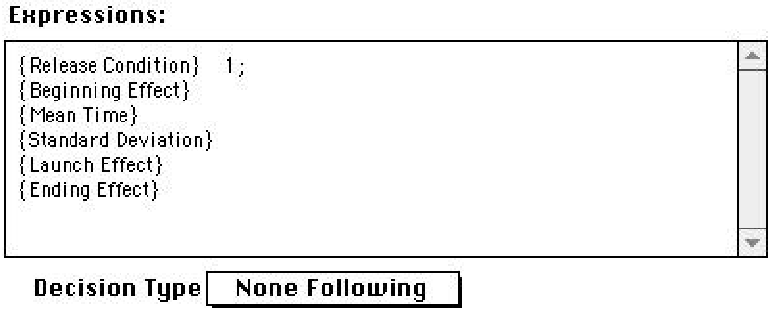
53	Vis Store Cap	Normal	 Expressions: {Release Condition} 1; {Beginning Effect} STVS+=1; {Mean Time} .01; {Standard Deviation} {Launch Effect} {Ending Effect} STVS-=1; Decision Type: Probabilistic What Happens Next: / Probability: 55 IL-V-Cap 1-V_decay; 59 Release WM Cap V_decay;
54	NP-Cap	n/a	 Expressions: {Release Condition} 1; {Beginning Effect} missed_Cap+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: Single What Happens Next: / Always 60 ReadAgain 1;
55	IL-V-Cap	n/a	 Expressions: {Release Condition} 1; {Beginning Effect} lost_V+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: Single What Happens Next: / Always 60 ReadAgain 1;

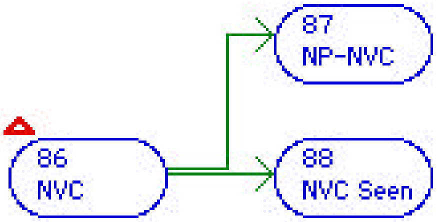
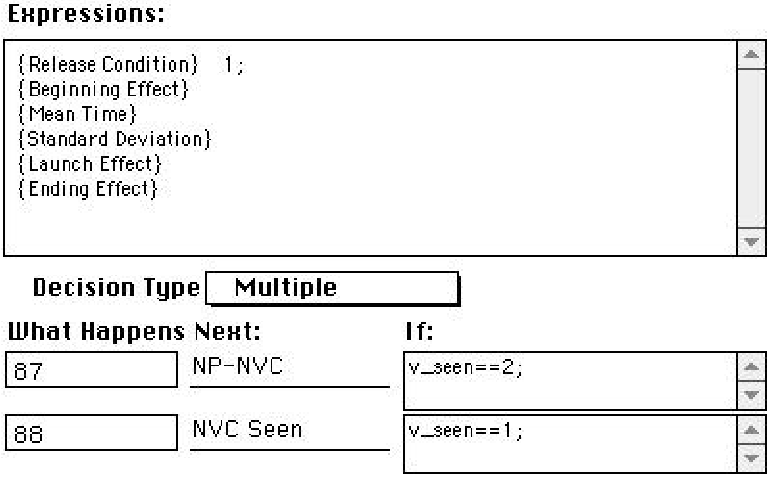
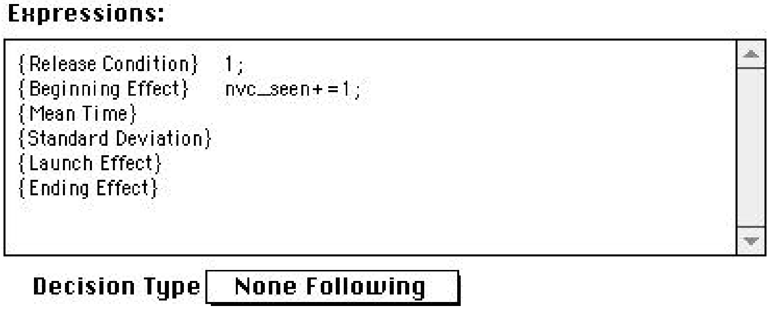
59	Release WM Cap	n/a	 Expressions: {Release Condition} 1; {Beginning Effect} redundant[tag]+=1;{anti-decay weight of source} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} WMgo[tag]+=1; Decision Type None Following

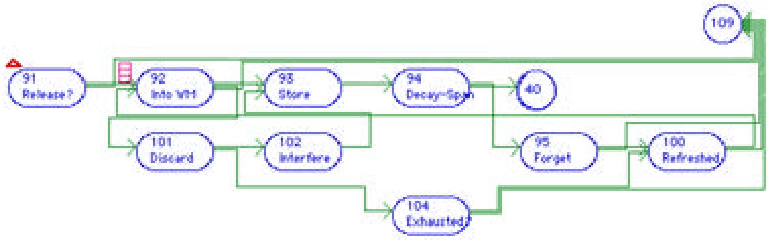
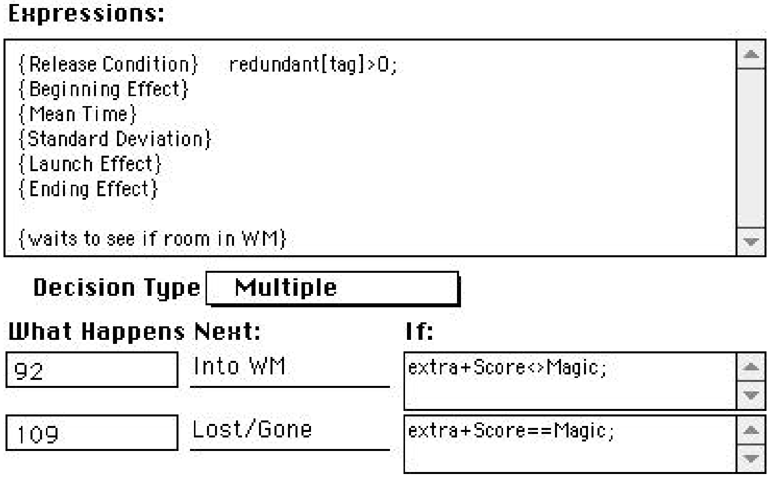
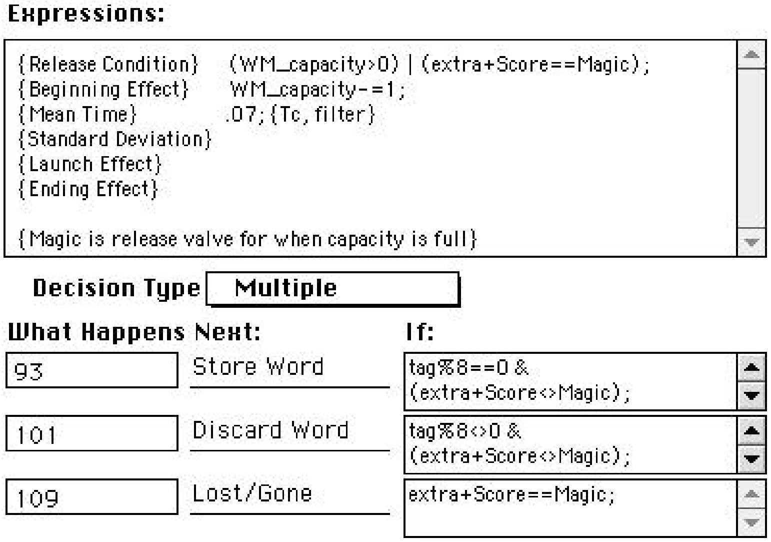
#	Name	Time Dist.	Logic & Routing
60	Read Again	Sub-network	
61	Still On?	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {checks to see if word still on screen} **Decision Type** Multiple **What Happens Next:** **If:** 63 NP&IL-Cap (currentLine-truncate(tag/8))>=lines; 62 Re-read (currentLine-truncate(tag/8))<lines;
62	Re-Read	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Single **What Happens Next:** **Always** 50 Captions 1;
63	NP&IL-Cap	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} WMgo[tag]+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** None Following

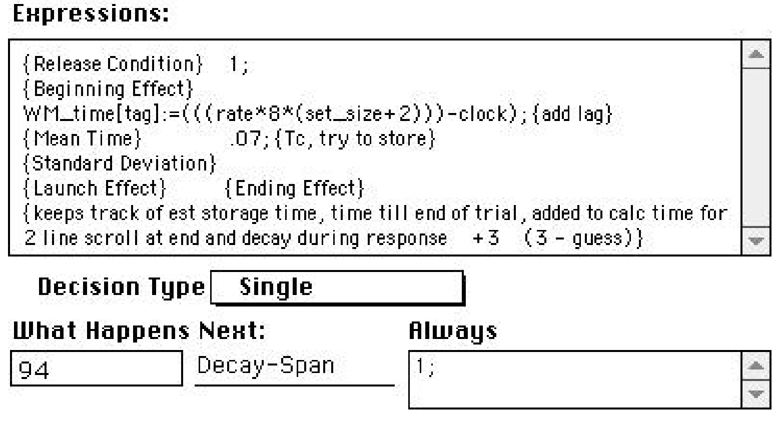
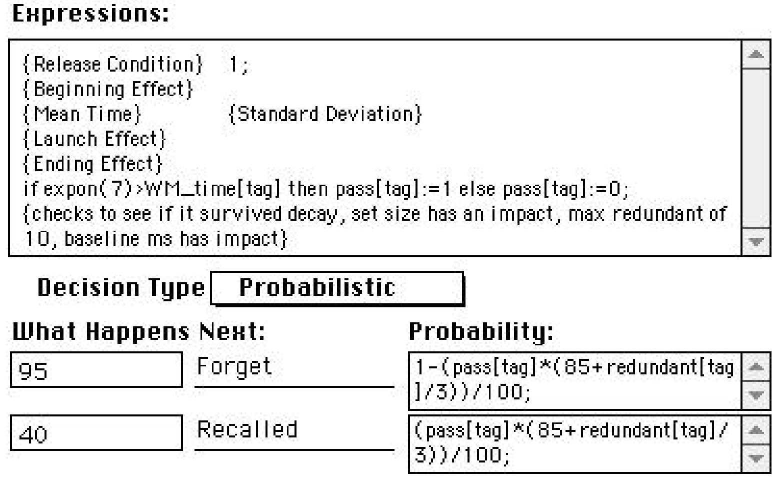
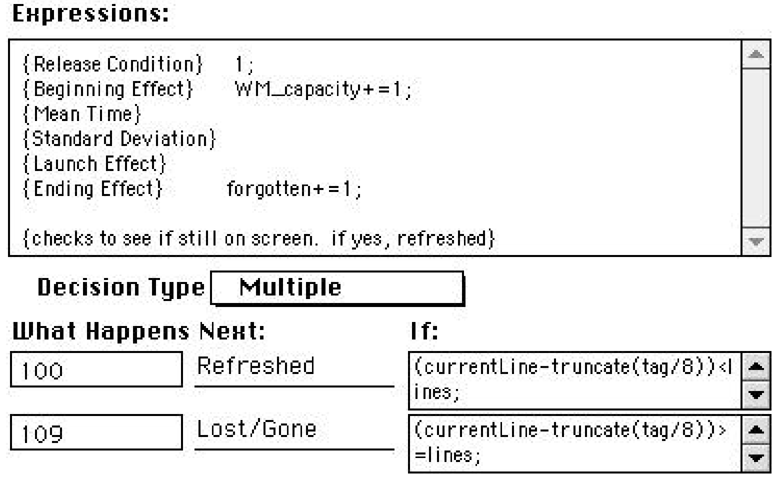
#	Name	Time Dist.	Logic & Routing
65	Scroll & Lag	Sub-network	
66	Wait for Line	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} line_length+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {need all 8 words before scrolling} **Decision Type** Single **What Happens Next:** 67 Send next line **Always** 1;
67	Send Next Line	n/a	**Expressions:** {Release Condition} line_length==8; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} line_length-=1; **Decision Type** Single **What Happens Next:** 68 Lag Delay **Always** 1;
68	Lag Delay	Normal	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} lag; {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Single **What Happens Next:** 50 Captions **Always** 1;

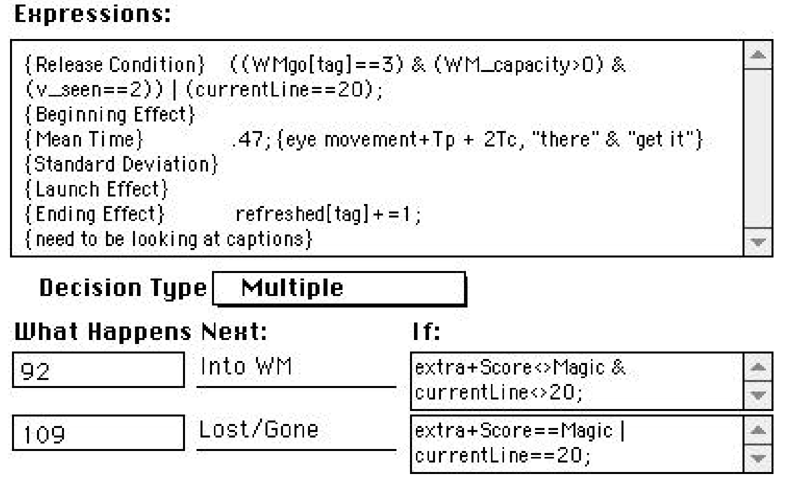
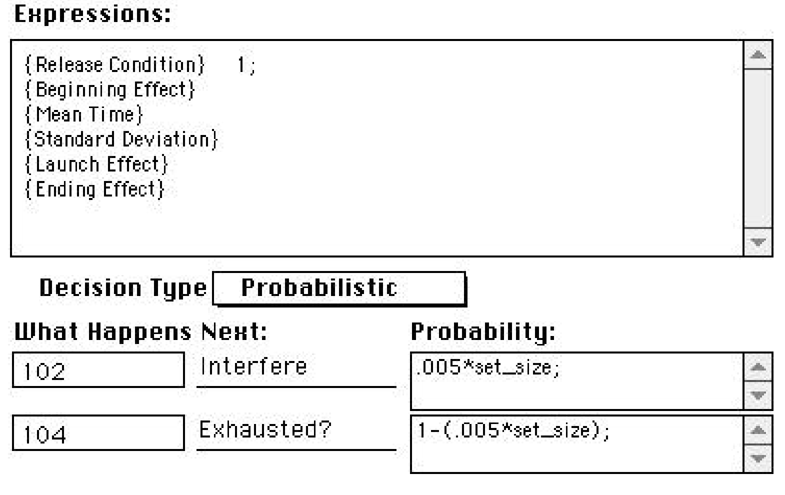
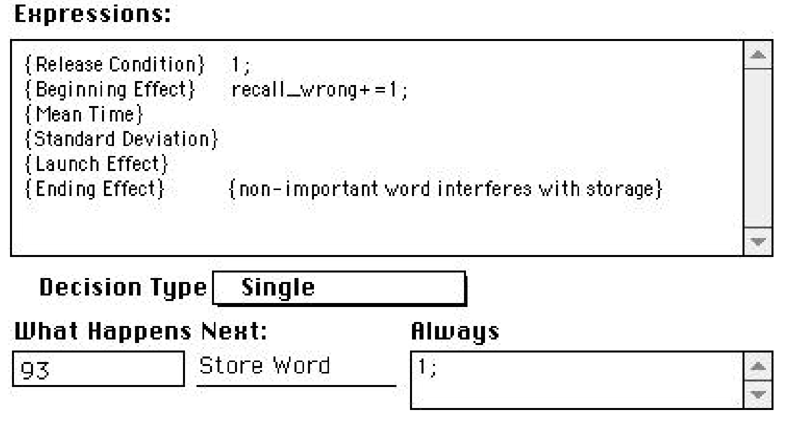
#	Name	Time Dist.	Logic & Routing
80	Attention Control	Sub-network	
81	Attention	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {mult tag%8==0 cap, tag%8<>0} Decision Type: Probabilistic What Happens Next: / Probability: 82 Look@Cap .8; 83 Look@Face .2; 84 Look@NVC 0;
82	Look at Captions	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} v_seen:=2; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: None Following
83	Look at Face	n/a	**Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type: None Following

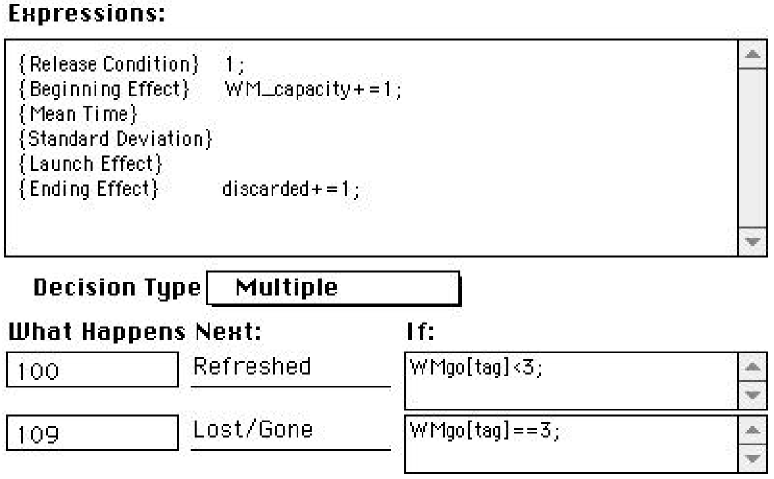
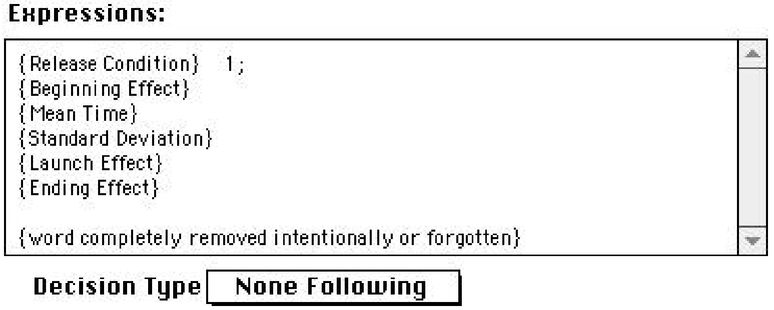
84	Look at NVC	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} Decision Type None Following

#	Name	Time Dist.	Logic & Routing
85	NVC Perception	Sub-network	
86	NVC	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Multiple **What Happens Next:** **If:** 87 NP-NVC v_seen==2; 88 NVC Seen v_seen==1;
87	NP-NVC	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} missed_NVC+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** None Following
88	NVC Seen	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} nvc_seen+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** None Following

#	Name	Time Dist.	Logic & Routing
90	WM	Sub-network	
91	Release?	n/a	 **Expressions:** {Release Condition} redundant[tag]>0; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {waits to see if room in WM} **Decision Type** Multiple **What Happens Next:** **If:** 92 Into WM extra+Score<>Magic; 109 Lost/Gone extra+Score==Magic;
92	Into WM	Normal	 **Expressions:** {Release Condition} (WM_capacity>0) | (extra+Score==Magic); {Beginning Effect} WM_capacity-=1; {Mean Time} .07;{Tc, filter} {Standard Deviation} {Launch Effect} {Ending Effect} {Magic is release valve for when capacity is full} **Decision Type** Multiple **What Happens Next:** **If:** 93 Store Word tag%8==0 & (extra+Score<>Magic); 101 Discard Word tag%8<>0 & (extra+Score<>Magic); 109 Lost/Gone extra+Score==Magic;

93	Store Word	Normal	 **Expressions:** {Release Condition} 1; {Beginning Effect} WM_time[tag]:=(((rate*8*(set_size+2)))-clock);{add lag} {Mean Time} .07;{Tc, try to store} {Standard Deviation} {Launch Effect} {Ending Effect} {keeps track of est storage time, time till end of trial, added to calc time for 2 line scroll at end and decay during response +3 (3 - guess)} **Decision Type** Single **What Happens Next:** 94 Decay-Span **Always** 1;
94	Decay-Span	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} if expon(7)>WM_time[tag] then pass[tag]:=1 else pass[tag]:=0; {checks to see if it survived decay, set size has an impact, max redundant of 10, baseline ms has impact} **Decision Type** Probabilistic **What Happens Next:** / **Probability:** 95 Forget — 1-(pass[tag]*(85+redundant[tag]/3))/100; 40 Recalled — (pass[tag]*(85+redundant[tag]/3))/100;
95	Forget	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} WM_capacity+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} forgotten+=1; {checks to see if still on screen. if yes, refreshed} **Decision Type** Multiple **What Happens Next:** / **If:** 100 Refreshed — (currentLine-truncate(tag/8))<lines; 109 Lost/Gone — (currentLine-truncate(tag/8))>=lines;

100	Refreshed	Normal	 **Expressions:** {Release Condition} ((WMgo[tag]==3) & (WM_capacity>0) & (v_seen==2)) | (currentLine==20); {Beginning Effect} {Mean Time} .47; {eye movement+Tp + 2Tc, "there" & "get it"} {Standard Deviation} {Launch Effect} {Ending Effect} refreshed[tag]+=1; {need to be looking at captions} **Decision Type** Multiple **What Happens Next:** / **If:** 92 Into WM — extra+Score<>Magic & currentLine<>20; 109 Lost/Gone — extra+Score==Magic | currentLine==20;
101	Discard Word	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} **Decision Type** Probabilistic **What Happens Next:** / **Probability:** 102 Interfere — .005*set_size; 104 Exhausted? — 1-(.005*set_size);
102	Interfere (Wrong Word)	n/a	 **Expressions:** {Release Condition} 1; {Beginning Effect} recall_wrong+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {non-important word interferes with storage} **Decision Type** Single **What Happens Next:** / **Always** 93 Store Word — 1;

104	Exhausted	n/a	Expressions: {Release Condition} 1; {Beginning Effect} WM_capacity+=1; {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} discarded+=1; Decision Type: Multiple What Happens Next: / If: 100 Refreshed — WMgo[tag]<3; 109 Lost/Gone — WMgo[tag]==3;
109	Lost/Gone	n/a	Expressions: {Release Condition} 1; {Beginning Effect} {Mean Time} {Standard Deviation} {Launch Effect} {Ending Effect} {word completely removed intentionally or forgotten} Decision Type: None Following

Variable definitions

Name	**Purpose**
A_decay	STAS decay
clock	system variable
currentLine	current line delivered
discarded	# of words intentionally not remembered
duration	system variable
extra	# of words remembered that are not last words
forgotten	# of words forgotten
lag	Lag (in seconds)
lines	Number of Lines
line_length	logic gate, fill line before scrolling
lost_A	# of words decayed out of STAS
lost_V	# of words decayed out of STVS
Magic	release valve for when capacity is full
mem_span	SRMS
missed_A	# of words not heard
missed_Cap	# of words not seen in captions
missed_Lip	# of words not lipread
missed_NVC	# of NVC's not seen
never_perc	# of words never perceived
nvc_seen	# of NVC's seen
pass[i]	logic gate for working memory decay
prob_hear	Hearing Ability, p(word was heard)
rate	Rate (seconds per word)
recall_wrong	# of words interfering with storage
redundant[i]	impact of redundancy
refreshed[i]	# of times a word is re-read from captions
run	system variable
Score	# of last words remembered
seed	system variable
set_size	Set Size
STAS	short term auditory store capacity
STVS	short term visual store capacity
tag	system variable
V_decay	STVS decay
v_seen	code for what user is looking at
WMgo[i]	logic gate, word has been perceived
WM_capacity	working memory capacity
WM_time[i]	time in working memory

Name[i] indicates the variable is an array (1x50).

An example of variable starting values:

```
prob_hear:=.95;  lines:=2;  A_decay:=.9;  V_decay:=.9;   mem_span:=4;  rate:=.375;
set_size:=6;  STAS:=1;  STVS:=3;  WM_capacity:=7; Magic:=7; lag:=0;
```

BIBLIOGRAPHY

Austin, B. A. and Myers, J. W. (1984) Hearing-impaired viewers of prime-time television. Journal of Communication, 34(4), 60-71.

Baddeley, A. (1992) Working memory. Science, 255, 556-559.

Baddeley, A.; Papagno, C.; and Vallar, G. (1988) When long-term learning depends on short-term storage. Journal of Memory and Language, 27, 586-595.

Baddeley, A. (1986) Working Memory, Oxford Psychology Series No. 11. Clarendon Press, Oxford.

Bilger, R.; Nuetzel, M.; Rabinowitz, W.; and Rzeczkowski, C. (1980) Standardization of a test of speech perception in noise. Journal of Speech Hearing Research, 27, 32-48.

Block, M. H. and Okrand, M. (1983) Real-time closed-captioned television as an educational tool. American Annals of the Deaf, 128(5), 636-641.

Braverman, B. B. and Hertzog, M. (1980) The effects of caption rate and language level on comprehension of a captioned video presentation. American Annals of the Deaf, 125, 943-948.

Braverman, B. B. (1981) Television captioning strategies: A systematic research and development approach. American Annals of the Deaf, 126, 1031-1036.

Blatt, J. and Sulzer, J. S. (1981) Captioned television and hearing-impaired viewers: The report of a national survey. American Annals of the Deaf, 126, 1017-1023.

Boyd, J. and Vader, E. A. (1972) Captioned television for the deaf. American Annals of the Deaf, 117, 34-37.

Caldwell, D. C. (1973) Use of graded captions with instructional television for deaf learners. American Annals of the Deaf, 118(4), 500-507.

Card, S. K.; Moran, T. P.; and Newell, A. (1986) The model human processor: An engineering model of human performance (Chapter 45) in Boff, K.; Kaufman, L.; and Thomas, J. (eds.) Handbook of Perception and Human Performance (Volume II - Cognitive processes and performance), New York: Wiley-Interscience.

Conrad, R. (1977) The reading ability of deaf school-leavers. British Journal of Education Psychology, 47, 138-148.

Cutler, W. B. (1990) Beyond the hearing aid: Assistive listening devices and systems. International Journal of Technology and Aging, 3(2), 101-109.

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

Daneman, M. and Carpenter, P. A. (1983) Individual differences in integrating information between and within sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9(4), 561-584.

Daneman, M.; Nemeth, S.; Stainton, M.; Huelsmann, K. (1995) Working memory as a predictor of reading achievement in orally educated hearing-impaired children. Volta Review, 97(4), 225-241.

Duchnicky, R. and Kolers, P. (1983) Readability of text scrolled on visual display terminal as a function of window size. Human Factors, 25(6), 683-692.

Everhart, V.; Stinson, M.; McKee, B.; and Giles, P. (1996). Evaluation of a speech-to-print transcription system as a resource for mainstreamed deaf students, Annual Meeting of the American Educational Research Association (AREA '96), New York.

Flannagan, M. and Harrison, A. (1994) The effects of automobile Head-Up Display location for younger and older drives. (Technical Report UMTRI 94-22), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Gates, R. R. (1971) The reception of verbal information by deaf students through a television medium -- a comparison of speechreading, manual communication, and reading. Proceedings of the Convention of American Instructors of the Deaf, Little Rock, 513-512.

Gibbs, K. W. (1989) Individual differences in cognitive skills related to reading ability in the deaf. American Annals of the Deaf, 134, 214-218.

Green, P. and Williams, M. (1992) Perspective in orientation/navigational displays: A human factors test. Conference on Vehicle Navigation and Information Systems (VNIS'92), (IEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, 221-226.

Hanson, V. and Fowler, C. (1987) Phonological coding in word reading: evidence from hearing and deaf readers. Memory and Cognition, 15(3), 199-207.

Houde, R. (1979) Prospects for automatic recognition of speech. American Annals of the Deaf, 124(5), 568-572.

Hyde, M. B. and Power, D. J. (1992) The receptive communication abilities of deaf students under oral, manual, and combined methods. American Annals of the Deaf, 137(5), 389-397.

Isihara, Y.; Tsukakoshi, H.; Nishikawa, S; and Obata, S. (1989) A study of superimposed captions on television for students of schools for the deaf. Japanese Journal of Special Education, 27(2), 25-37.

Jackson, P. L. (1988) The theoretical minimal unit for visual speech perception: visemes and coarticulation. Volta Review, 90(5), 99-115.

Kalikow, D.; Stevens, K.; and Elliot, L. (1977) Development of a test of speech intelligibility in noise using sentence material with controlled word predictability. Journal of the Acoustical Society of America, 61, 1337-1351.

Kelly, J. and Steer, M. (1949) Revised concept of rate. Journal of Speech and Hearing Disorders, 14, 222-226, pages as cited in Shroyer, E. H. and Birch, J. (1980) Captions and reading rates of hearing-impaired students. American Annals of the Deaf, 125, 916-922.

Kolers, P.; Duchnicky, R.; and Ferguson, D. (1981) Eye movement measurement of readability of CRT displays. Human Factors, 23(5), 517-527.

Koskinen, P.; Wilson, R. M.; Gambrell, L. B.; and Jensema, C. (1986) Using closed captioned television to enhance reading skills of learning disabled students. National Reading Conference Yearbook, 35, 61-65.

Liu, Y. and Wickens, C. D. (1992) Visual scanning with or without spatial uncertainty and divided and selective attention. Acta Psychologica, 79(2), 131-153.

Liu, Y. (1996a) Interactions between memory scanning and visual scanning in display monitoring. Ergonomics, 39,1038-1053.

Liu, Y. (1996b) Quantitative assessment of effects of visual scanning on concurrent task performance. Ergonomics, 39, 382-399.

Liu, Y. (1996c) Queueing network modeling of elementary mental processes. Psychological Review, 103(1), 116-136.

Liu, Y. (1997) Queueing network modeling of human performance of concurrent spatial and verbal tasks. IEEE Transactions on Systems, Man, and Cybernetics, 27(2), 318-328.

Locke, J. L. (1978) Phonemic effects in the silent reading of hearing and deaf children. Cognition, 6, 175-187.

Markham, P. L. (1989) The effects of captioned television on the listening comprehension of beginning, intermediate, and advanced ESL students. Educational Technology, 29(10), 38-41.

Maxon, A. B. and Welch, A. J. (1992) The role of language competence on comprehension of television messages by children with hearing impairment. Volta Review, 95, 315-326.

McCoy, E. and Shumway, R. (1979) Real-time captioning -- promise for the future. American Annals of the Deaf, 124(5), 681-690.

Murphy-Berman, V. and Jorgensen, J. (1980) Evaluation of a multilevel linguistic approach to captioning television for hearing-impaired children. American Annals of the Deaf, 125, 1072-1081.

Neisser, U. and Becklen, R. (1975) Selective looking: Attending to visually specified events. Cognitive Psychology, 7, 480-494.

Neuman, S. B. and Koskinen, P. (1992) Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. Reading Research Quarterly, 27(1), 95-106.

Nugent, G. C. (1983) Deaf students' learning from captioned instruction: The relationship between the visual and caption display. The Journal of Special Education, 17(2), 227-234.

Oakhill, J. (1984) Inferential and memory skills in children's comprehension of stories. British Journal of Educational Psychology, 54(1), 31-39

Olsson, J. E. and Furth, H. G. (1966) Visual memory-span in the deaf. American Journal of Psychology, 79, 480-484.

Okada, A.; Tsuzuki, S.; Sato, Y.; and Hamakado, N. (1985) An experimental study on the optimum of producing captioned television for the deaf. Japanese Journal of Psychology, 33(1), 22-32.

Pichora-Fuller, M. K.; Schneider, B. A.; and Daneman, M. (1995) How young and old adults listen to and remember speech. Journal of the Acoustical Society of America, 97(1), 953-608.

Rifkin, G. (1993) Courting a deaf movie audience with caption devices. New York Times, v. 143, sec. 3 (November 21), F11.

Seriwong, S. (1992) A pilot study on the effects of closed-captioned television on English as a second language students' listening comprehension (television). Doctoral dissertation, Southern Illinois University at Carbondale. University Microfilms, Inc.

Shroyer, E. H. and Birch, J. (1980) Captions and reading rates of hearing-impaired students. American Annals of the Deaf, 125, 916-922.

Smith, J. E. (1995) Conversations with, October 18.

Steinfeld, A. and Green, P. (1992) Driver response times to full-windshield, head-up displays for navigation and vision enhancement. (Technical Report UMTRI 95-29), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Stinson, M.; Stuckless, E. R.; Henderson, J.; and Miller, L. (1988). Perceptions of hearing-impaired college students toward real-time speech to print: RTGD and other educational services. Volta Review, 90(4), 336-348.

Stuckless, R. E. (1981) Real-time graphic display and language development for the hearing-impaired. Volta Review, 83(5), 291-300.

Van Biema, D. (1994) AIDS. Time, 143(14), 76-77.

Wickens, C. D. and Carswell, C. M. (1995) The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37(3), 473-494.

Williams, M. and Green, P. (1992) Development and testing of driver interfaces for navigational displays. (Technical Report UMTRI 92-21), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Youdelman, K. and Messerly, C. (1996). Computer-assisted notetaking for mainstreamed hearing-impaired students. Volta Review, 98(4), 191-199.