Eye Movement Patterns of Captioned Television Viewers

Carl J. Jensema, Sameh El Sharkawy, Ramalinga Sarma Danturthi, Robert Burch, David Hsu

American Annals of the Deaf, Volume 145, Number 3, July 2000, pp. 275-285 (Article)

Published by Gallaudet University Press
DOI: *https://doi.org/10.1353/aad.2012.0093*

EYE MOVEMENT PATTERNS OF CAPTIONED TELEVISION VIEWERS

Eye movement of six subjects was recorded as they watched video segments with and without captions. It was found that the addition of captions to a video resulted in major changes in eye movement patterns, with the viewing process becoming primarily a reading process. Further, although people viewing a specific video segment are likely to have similar eye movement patterns, there are also distinct individual differences present in these patterns. For example, someone accustomed to speechreading may spend more time looking at an actor's lips, while someone with poor English skills may spend more time reading the captions. Finally, there is some preliminary evidence to suggest that higher captioning speed results in more time spent reading captions on a video segment.

CARL J. JENSEMA, SAMEH EL SHARKAWY, RAMALINGA SARMA DANTURTHI, ROBERT BURCH, AND DAVID HSU

Jensema is vice president of the Institute for Disabilities Research and Training (IDRT), Silver Spring, MD. At the time the article was in preparation, El Sharkawy was a computer program developer at IDRT. Danturthi and Burch are research associates at IDRT, and Hsu is a graphic artist there.

The information presented in the present study represents the initial data from a government-funded research project (U.S. Department of Education Grant H026R70003), "A Study of the Eye Movement Strategies Used in Viewing Captioned Television," undertaken at the Institute for Disabilities Research and Training (IDRT). This project represents the first attempt in more than 20 years to examine how people move their eyes when they watch captioned television programs.

Captions are subtitles that display spoken dialogue in printed form, usually on the bottom of the screen. There is a considerable body of literature related to captioned television stretching back to the 1970s. Recent studies include Jensema (1998), Jensema, McCann, and Ramsey (1996), Kirkland (1999), and Loeterman and Kelly (1997). Unfortunately, almost none of this literature relates to how people actually see captions. In the mid-1970s, some attempts were made to measure the eye movement of caption viewers when the Public Broadcasting Service (PBS) was developing the technology for a national closed-captioned television system. This research was done about 1974 by Kenneth G. O'Bryan of the Ontario Educational Communication Authority for the WGBH Caption Center in Boston and resulted in two apparently unpublished manuscripts, *Eye Movement Research Report on Captioning Television Programs for the Deaf* and *Report on Basic Findings on Captioned News*. (Mardi Loeterman found these manuscripts in July 1999 in the WGBH archives.) A literature search has turned up no other subsequent research articles on eye movement and captioning.

Over the last 20 years, closed captioning has evolved from a PBS engineering development project to an established part of the overall television system in the United States. Millions of dollars are spent each year providing captioning services, but it is still not known exactly where people are looking when they watch captioned television (e.g., at the captions, at the picture) and how their eyes scan the screen to absorb the information. The research project described in the present article addresses that issue. We describe the eye movement measuring system that has been developed by IDRT and show some of the initial results from our exploration of some basic questions:

1. How does captioning change the way a television program is viewed?

2. Do all people view captioning the same way, or are there individual viewing strategies?

3. Does prior knowledge of a video's content influence how it is viewed?

4. How does the rate of captioning influence viewing?

Method

Equipment

To track eye movement accurately, we used a variation of the DOS-based Eyegaze Development System designed by LC Technologies, of Fairfax, VA. This system uses the Pupil-Center/Corneal-Reflection method to determine eye movement and fixation. The subject's eye is continually observed by a video camera located below a monitor screen. Our particular Eyegaze system uses a Sanyo CCD camera with a lens having f values from 1.3 to 22. To allow good exposure of the eye and its pupil with a wide aperture, the f value is set at 2.8.

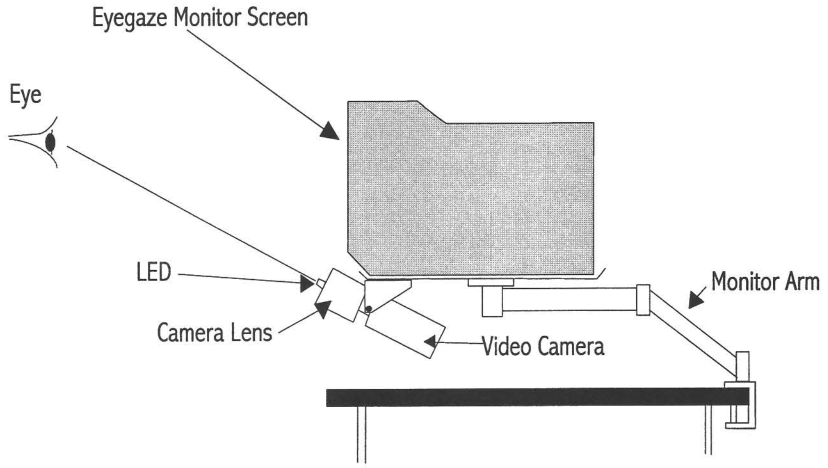


Figure 1
Eye Movement Tracking Arrangement

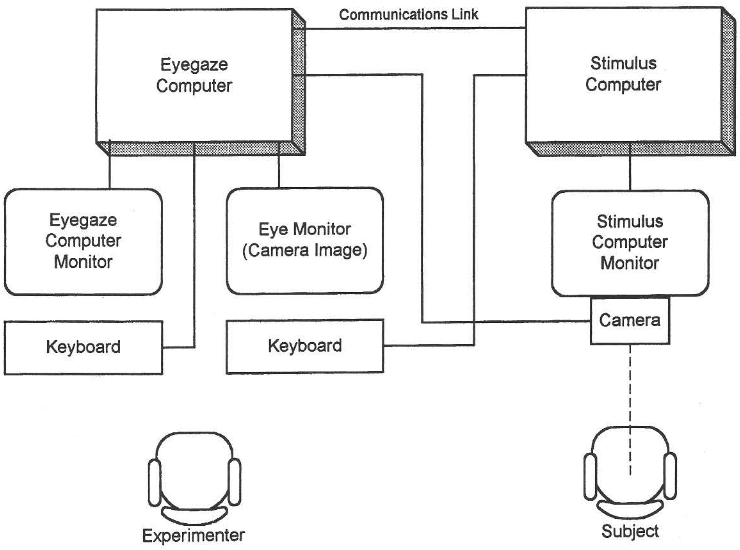


Figure 2
Schematic Layout of Visual Monitoring Equipment

The eye of the subject is illuminated by a small, low-power, infrared light-emitting diode (LED) located on the center of the camera lens. This LED generates corneal reflection and causes a bright-pupil effect that enhances the camera's image of the pupil. Using the center of the pupil and the corneal reflection within the video image, trigonometric calculations determine the subject's gaze point. Since a low-power infrared LED is used, the system requires a low level of ambient infrared light entering the subject's eye. Stray sources of infrared light obscure the LED and degrade the image of the eye. The Eyegaze system works best with people who do not wear corrective lenses, but performs acceptably well with people who wear soft contact lenses or eyeglasses without bifocals. Hard contact lenses and eyeglasses with bifocals can cause measurement problems. Figure 1 depicts the relationship between a subject's eye and the camera and monitor screen of the Eyegaze system.

To enhance the power of the DOS-based Eyegaze system for the project, the IDRT staff developed a Windows-based multimedia presentation program that runs as a client application. The new system consists of two connected computers, one for the DOS-based Eyegaze system and one for the Windows-based presentation system. The two computers are linked through RS-232 ports and programmed to work at a data transfer rate of 9,600 bits/sec. Figure 2 shows a schematic layout of the eye movement measuring system developed by IDRT for the present study.

The procedure for analyzing the eye movement of a subject involves three different steps:

1. calibration of the eye to the Eyegaze system;
2. projection of an image (either stationary or moving) on the screen and collection of the gaze points of the subject while the subject looks steadily at the projected image;
3. collection and plotting of the gaze point data on the screen in an appropriate and readable manner.

These three steps have been incorporated into a single Windows-based Visual C++ program module to ensure a smooth transition from one step to the next.

Calibration

Whenever a subject is tested on the Eyegaze system, calibration of the eye with the camera is necessary. Calibration is a process whereby the Eyegaze software learns a number of physiological parameters of the subject, such as radius of curvature of the eye's cornea and angular offset between the eye's optic and focal axes. The system requires these parameters in order to project the gaze point accurately. Calibration is done by projecting a number of points around the computer screen and collecting data while the subject looks steadily at these points. Once collected, the calibration data are shared among the Eyegaze software functions to project future gaze points. The calibration procedure takes 15 to 20 seconds.

Image Presentation and Gaze Point Data Acquisition

The process of image presentation and data acquisition begins with initialization of communications between the presentation program and the Eyegaze program to ensure that they are properly coordinated. The Eyegaze program detects the location of the subject's eye gaze and records this information at a rate of 30 eye locations per second. The presentation machine displays the stimulus material on the monitor screen and, at the same time, runs a background process that collects the data sent by the Eyegaze system. The eye movement data are stored on the presentation machine along with some configuration information for later use.

The data that have been collected can be played back for closer examination and analysis. The eye movement data and the stimulus are synchronized so each eye gaze data point is displayed on the exact video frame from which it was originally recorded. The program can also give a separate graphic trace of the data points without the stimulus material to provide an overall concept of how the subject's eye moves over the stimulus.

Initial Applications of the System to the Research Questions

The system was used to examine the eye movement of six subjects while they watched a total of eight video clips. The subjects were IDRT employees and had the following characteristics:

Subject 1: male, age 54 years, deaf since age 9 years, signs
Subject 2: male, age 37 years, deaf since age 8 years, speechreads
Subject 3: male, age 30 years, deaf since birth, signs
Subject 4: male, age 28 years, hearing
Subject 5: female, age 52 years, hearing
Subject 6: female, age 28 years, hearing

All subjects were familiar with captioned television, but only the deaf subjects watched it regularly.

In order to standardize technical accuracy during the project, subjects were seated in a firm but comfortable chair in front of the monitor, and the eye-to-camera distance was restricted to a range of 18 to 22 inches. To keep the subject's head stationary and comfortable, his or her head was positioned against a sponge pillow. Measurement of eye gaze location was

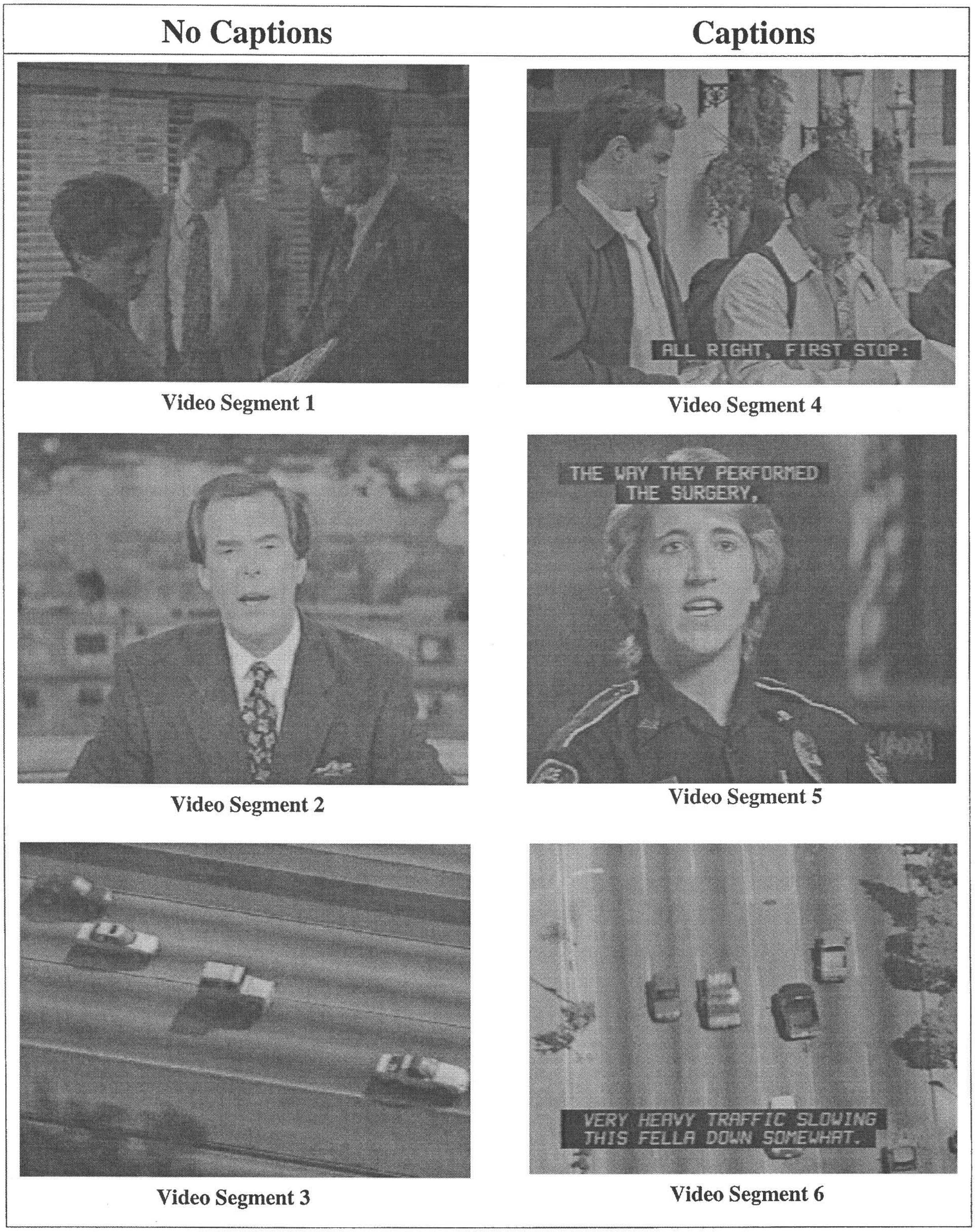


Figure 3a
Still Pictures From Video Segments, Captioned and Uncaptioned

done at the frame rate of the camera, that is, every 1/30th of a second. There is a possibility of error due to inaccuracies of measurement of head range, asymmetries of the pupil opening about the eye's optic axis, or astigmatism. The typical average bias error of measurement is 0.15 inch and the maximum average bias error is 0.25 inch, over the monitor screen range. While the measurements are made, the only condition the Eyegaze system requires is that the eye remain in the field of view of the camera. Because the stimulus picture could be set to any size up to a full monitor screen, it was possible to set the video image to 4 inches high and 6 inches wide. This size roughly approximates the visual scanning movement required of an eye watching a 27-inch television set at a distance of 8 to 10 feet.

For the first part of the experiment, six short segments of broadcast television programming were recorded with open captions. From this, 8-to-18-second samples of material were captured in digital files for use in eye movement experimentation. The digital samples were selected as being representative of some of the different program types and captioning styles frequently encountered by those viewing captioned television. The segments were taken from four different television programs, *Law and Order*, *ABC Evening News*, *Friends*, and *World's Wildest Police Videos*, and one movie, *Speed*. Because the study focused on captioning, none of the segments had audio. Figure 3a provides a still image from each of these six segments, referred to as Video Segment 1, Video Segment 2, etc.

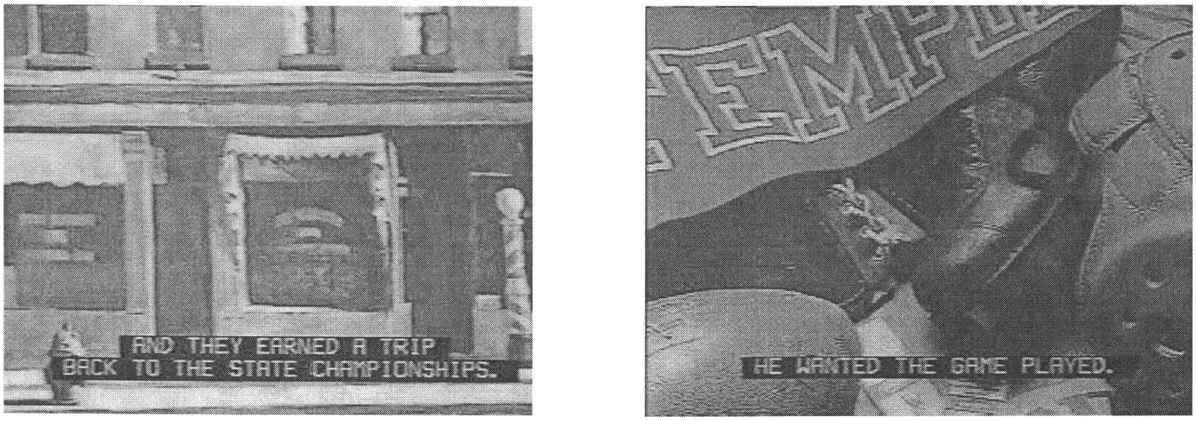
Two other segments used in our research were from another caption research study being done by the authors. These two 30-second segments had been custom created, and contained the exact number of words needed to yield captioning at 80 and 220 words per minute. Figure 3b provides still images from these two segments, referred to as Video Segment 7 and Video Segment 8.

Segments 1, 2, and 3 had no captions. Segments 4, 5, and 6 were captioned. The segments were paired, 1-4, 2-5, and 3-6, with each pair representing a very similar video scene, with one image being captioned and the other noncaptioned.

Segments 1 and 4 had actors talking with relatively limited movement. Segment 1 consisted of 11 seconds from an episode of Law and Order and had no captions. Segment 4 was from World's Wildest Police Videos and was captioned at 106 words per minute (15 words in 8.5 seconds), representing a typical low-difficulty caption reading situation.

Segment 2 consisted of 11 seconds of Peter Jennings presenting the *ABC Evening News*. Segment 5 was from *World's Wildest Police Videos* and featured a policewoman talking about injuries experienced in the line of duty. Both video segments represented a typical "talking head" situation. What makes Segment 5 unusual is that during the first half of the segment the captions were above the speaker, while in the second half they were below the speaker. The captioning rate was 122 words per minute (23 words in 11.3 seconds).

80 WPM | **220 WPM**



Video Segment 7 | Video Segment 8

Figure 3b
Still Pictures From Video Segments, 80 and 220 Words per Minute

Segments 3 and 6 both showed helicopter views of a police chase. Segment 3 consisted of 11 seconds from the movie *Speed*; Segment 6 was from *World's Wildest Police Videos*. The scenes were dotted with moving cars and presented a complex, continually changing image. The captioning on Segment 6 was very fast, with a rate of 197 words per minute (56 words in 17.1 seconds). This was a very difficult segment for most people trying to read the captions and watch the video.

Segments 7 and 8 were part of a series of eight 30-second video segments that told a story about a football coach. These segments were custom made for another caption research study and had carefully controlled caption speed and vocabulary. Segments 7 and 8 were captioned at 80 and 220 words per minute, respectively. They were used to determine how eye movement differed for fast and slow captions.

Each of the six subjects had his or her eye movement recorded while watching all eight video segments.

Results

Figure 4 depicts the eye movement patterns for each subject on the segments that were not captioned: Segments 1, 2, and 3. For Video Segment 1 (see Figure 3a), which shows a woman and two men talking, Subject 1 kept moving his eyes from face to face in the video and also glanced a few times at the paper the woman was holding.

In viewing Figures 4 through 7, the reader may find it helpful to remember that each small graphic plot indicates exactly where a subject was looking at the video screen every 1/30th of a second during the showing of a particular video segment. Each position measured is marked with a "+" and is connected by a line to the next "+" to indicate movement.

If one glances down the columns in Figure 4, it is apparent that the eye movement patterns tend to be consistent among subjects. In Segments 1 and 2, the subject's eye gaze focuses on the faces of the actors. In Segment 3, the gaze roams among the moving vehicles, as the subject apparently tries to extract as much information as possible from the complex scene. There is no captioning to draw the subject's attention in Segments 1, 2, or 3, so subjects focus their efforts on extracting information from the picture.

Figure 5 gives the eye movement patterns for each of the six subjects on the captioned segments: Segments 4, 5, and 6. Segments 4, 5, and 6 are similar to Segments 1, 2, and 3, respectively, except that they are captioned. Comparison of the patterns in Figure 4 and Figure 5 clearly indicates that the addition of captions to a video has a major impact on the eye movement of both hearing and hearing impaired viewers. The addition of captions apparently turns television viewing into a reading task, since the subjects spend most of their time looking at the captions and much less time examining the picture.

Figures 4 and 5 both indicate that the subjects had similar general viewing strategies, since the eye movement patterns are similar for all subjects who view a particular video segment. However, Figure 5 suggests that there were variations in caption- reading strategies on the basis of personal characteristics. Subjects 2 and 3, especially, exhibited unique caption-viewing strategies that appear directly related to specific individual traits. Subject 2 depended heavily on speech reading for his personal communication, and for Video Segment 5 (the "talking head" segment), he focused primarily on the speaker's lips and spent relatively little time reading the captions. In fact, he totally ignored the captions placed at the top of the screen. Subject 3 came to live in the United States about 12 years ago. English is not his native language, and he must concentrate when he reads English words. On Segments 4 and 6, Subject 3 spent all his time looking at the captions, and depended on peripheral vision to follow the screen action.

Each of the six subjects was tested again with the same captioned video segments (Segments 4, 5, and 6) a few days after the first testing. The results are given in Figure 6. Figure 6 has exactly the same information as Figure 5, except that testing was done a few days later. For the second testing, the subjects had a vague memory of the videos and knew what to expect. As a result, Figure 6 shows that the subject's eye movement patterns for particular videos became more similar. The unique characteristics of Subjects 2 and 3 that were found in Figure 5 are no longer apparent in the eye movement shown in Figure 6. It is particularly interesting to note that the hearing subjects had eye movement patterns similar to those of the deaf subjects, especially in Figure 6. As we have already noted, the hearing subjects were familiar with captioning but were not regular caption viewers.

The six subjects were shown Video Segments 7 and 8. Segment 7 was captioned at 80 words per minute and Segment 8 at 220 words per minute. Figure 7 shows the eye movement patterns for each subject for these two video segments. At the higher caption speed, the subjects spent more time reading and less time examining the picture.

Discussion

The present study is the first of several research studies that will be done on the eye movement of captioned television viewers. The data we have presented are limited in both the scope of the work done and our inability to clearly present the flow of eye movement over time on a printed page. However, the data do suggest

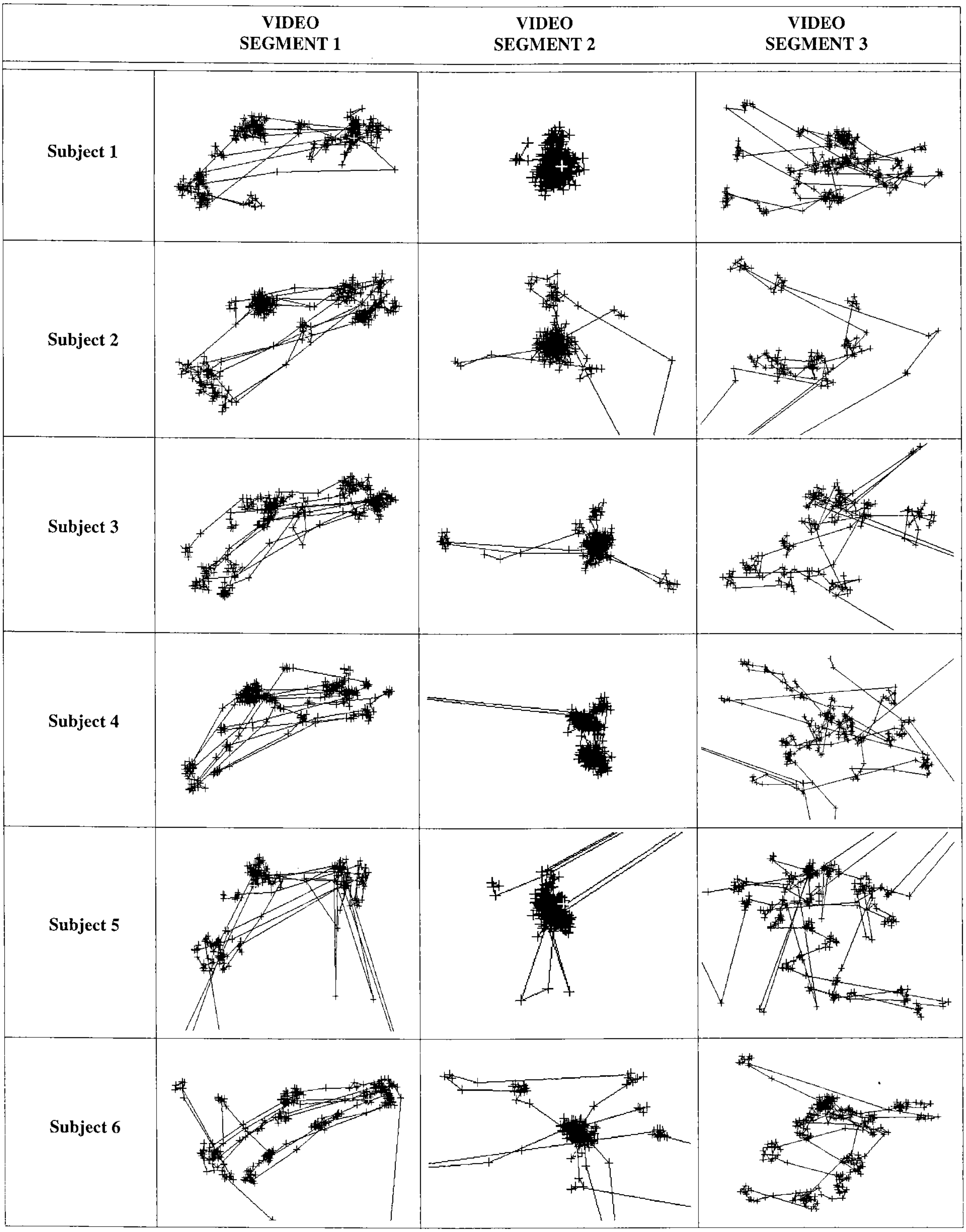


Figure 4
Eye Movement Patterns, No Captions

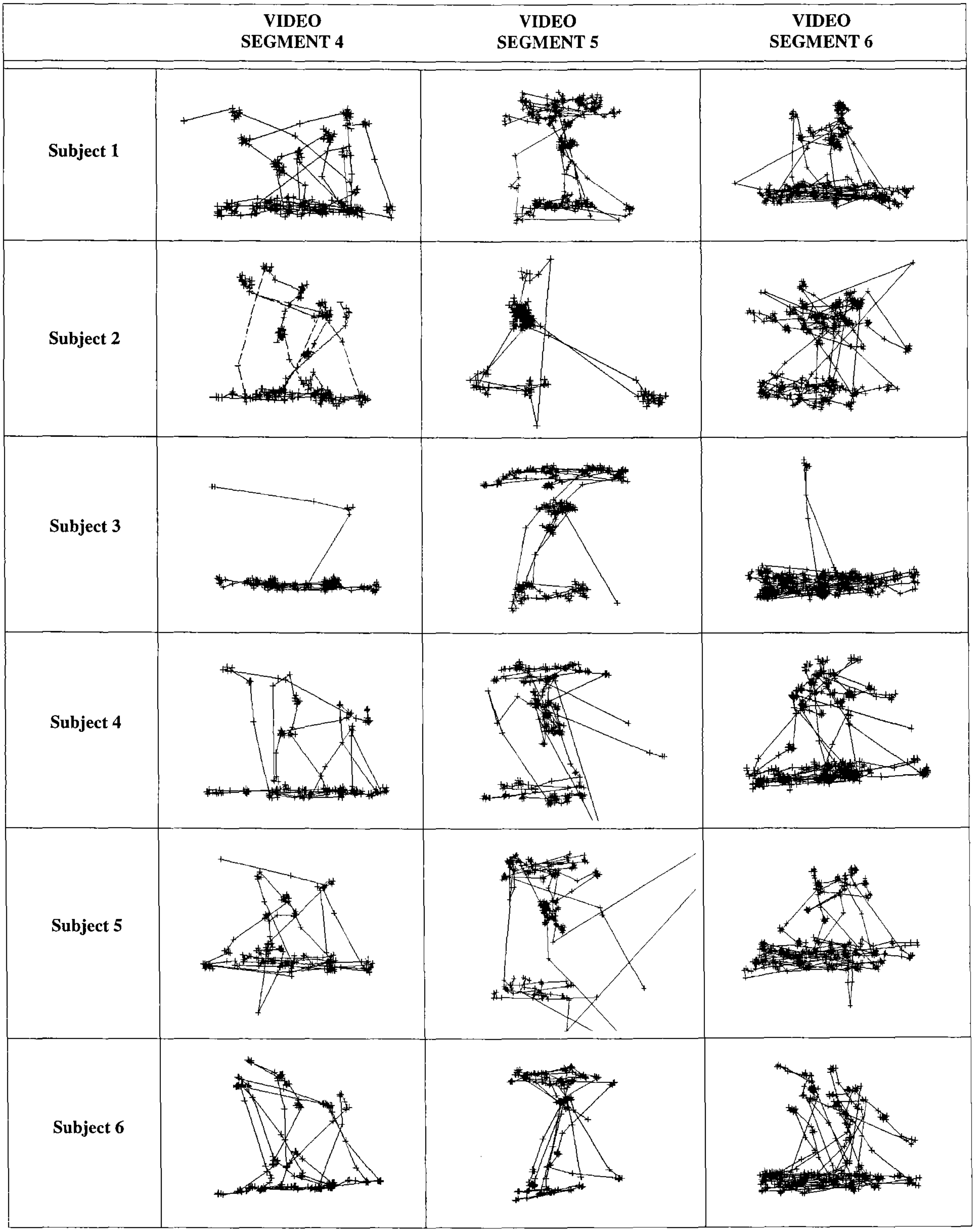


Figure 5
Eye Movement Patterns, Captions

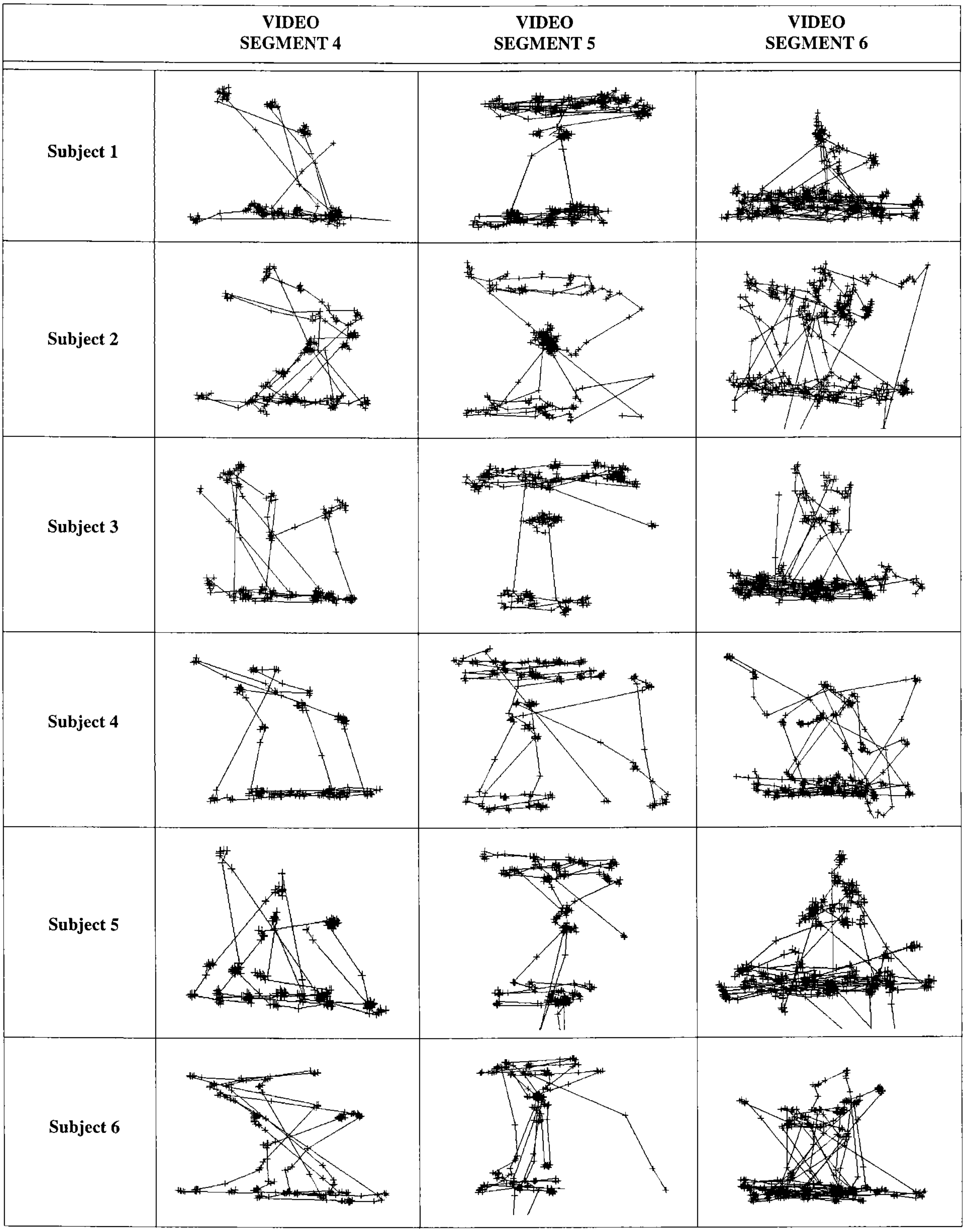


Figure 6
Eye Movement Patterns, Second Viewing, Captions

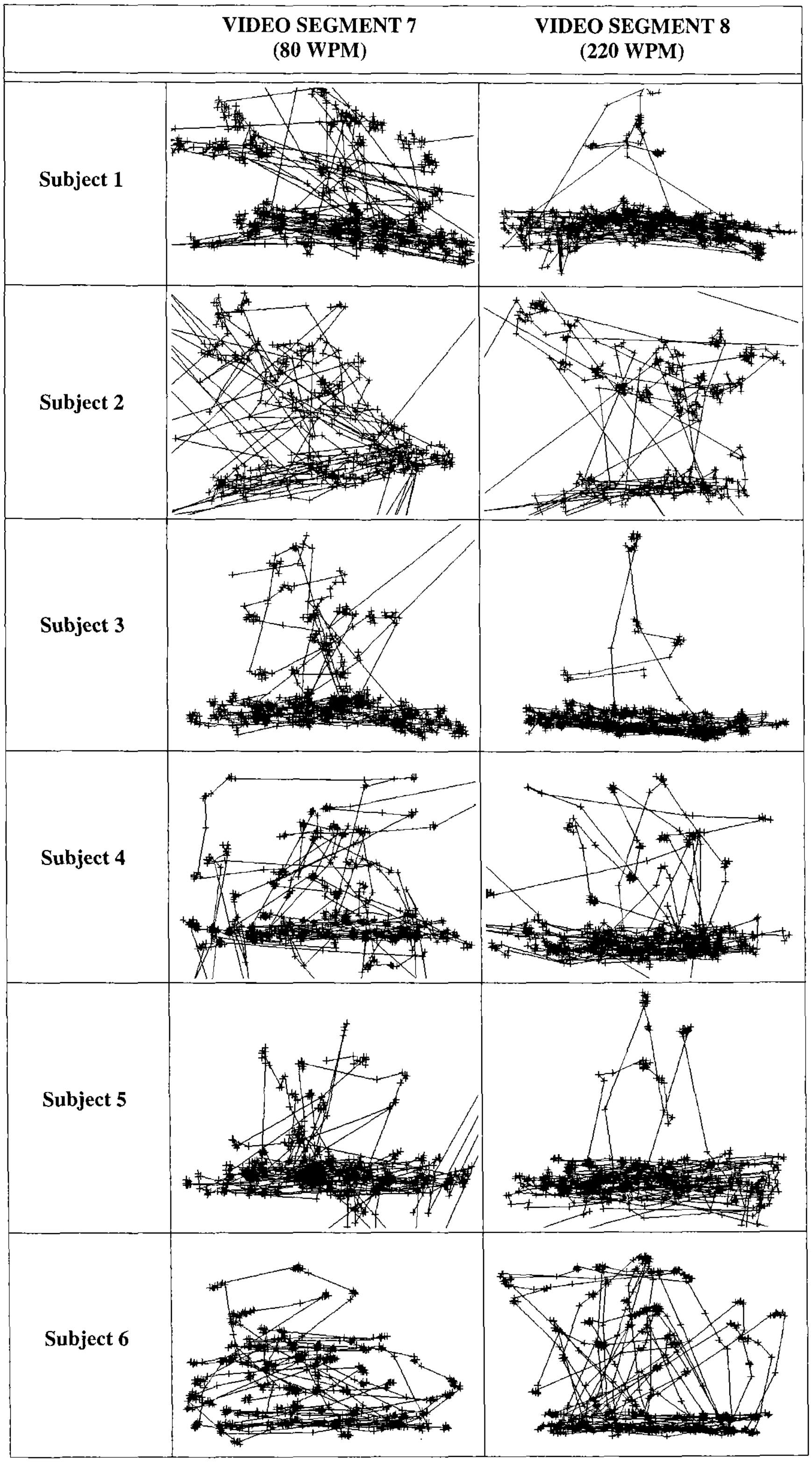


Figure 7
Eye Movement Patterns at Different Caption Speeds

some broad concepts concerning captioned television viewing.

In general, people who view a particular video segment have similar eye movement patterns. The addition of captions to a video results in a major change in eye movement patterns, and the viewing process becomes much more of a reading process. When captions are present, there appears to be a general tendency to start by looking at the middle of the screen and then moving the gaze to the beginning of a caption within a fraction of a second. Viewers read the caption and then glance at the video action after they finish reading. When a new caption comes on the screen, they move their gaze to the new caption and begin reading again. It appears that viewing captioned television is basically a reading task. Caption reading dominates eye movement; viewing the screen action tends to be secondary. It is not known yet if this finding also holds true for young children. If it does, it could have very important implications for the teaching of reading to both hearing and deaf children through the use of captioned video materials.

Of course, exceptions to our general findings can occur. For instance, the personal characteristics of the viewer can influence the approach that is used. The data indicated that a person who depends greatly on speechreading for communication might focus more on an actor's lips than on the captions. A person who does not have strong English skills may spend almost all of his or her time reading captions. Under conditions in which a viewer has some idea of what to expect, such as when the video segment has been seen some time in the past, the eye movement patterns of different subjects (as shown in Figure 6) seem to be more alike, as if these different patterns were converging on a single general pattern for viewing of that video segment.

The time spent reading captions

appears to increase as the speed of the captions increases (as shown in Figure 7). Reading captions is given first priority, and, at the highest caption speeds, relatively little time is spent looking at the video action. This balance between reading captions and examining the video action warrants further research. The best balance between reading and action viewing may vary with the material being presented, the skills of the viewer, and other factors.

Overall, much more work on the relationship between eye movement and captioned television needs to be done. There are still many issues that remain unresolved. For example, we have not yet examined how a caption's position on the screen influences eye movement, how eye gaze changes with scene changes, or how screen complexity and action influence caption reading. We have used a very limited number of subjects and do not know what degree of variation may exist among the general population. Many other issues also remain to be examined.

References

Jensema, C. (1998). Viewer reaction to different television captioning speeds. *American Annals of the Deaf, 143*(4), 318-324.

Jensema, C., McCann, R., & Ramsey, S. (1996). Closed-captioned television presentation speed and vocabulary. *American Annals of the Deaf, 141*(4), 284-292.

Kirkland, C. E. (1999). Evaluation of captioning features to inform development of digital television captioning capabilities. *American Annals of the Deaf, 144*(3), 242-260.

Loeterman, M., & Kelly, R. R. (1997). Using video and captioning to engage students. *Media & Methods, 34*(2), 10-12.

O'Bryan, K. G. (1974?). *Eye movement research report on captioning television programs for the Deaf.* Unpublished manuscript, WGBH archives, Boston.

O'Bryan, K. G. (1974?). *Report on basic findings on captioned news. Unpublished manuscript,* WGBH archives, Boston.