A New Method for Speechreading Research: Tracking Observers' Eye Movements

Charissa R. Lansing
Department of Speech and Hearing Science
University of Illinois

George W. McConkie
Department of Educational Psychology
Beckman Institute for Advanced Science
University of Illinois

The feasibility of recording speechreaders' eye movements as a way of studying visual perception of spoken language is illustrated. New techniques for relating speech events in full-motion video sequences to information about the location, duration, and sequences of eye gazes are described. These methods are illustrated by using data obtained from a subject with profound hearing loss and excellent speechreading proficiency. The usefulness of this approach to developing and testing hypotheses about attentional and visual processes in speechreading is also discussed.

Individuals with normal hearing, as well as those with hearing loss, frequently rely on visual cues in adverse listening conditions. Additionally, many individuals with a profound hearing loss depend primarily on vision for speech perception. Speechreading (lipreading) proficiency, however, is not dependent on an individual's hearing status (Erber, 1972; Summerfield, 1991). Individual performance scores broadly differ but are typically estimated to range between 11-57% accuracy for English (Berger, 1972). Remarkably, data reported by Bernstein, Demorest, Coulter, and O'Connell (1991) demonstrate that some individuals

Charissa R. Lansing, PhD is a Fellow in the Center for Advanced Study and an Assistant Professor, Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, 901 S. Sixth Street, Champaign, IL 61820. George W. McConkie, PhD is a Professor in the Department of Educational Psychology, Beckman Institute for Advanced Science, and Center for the Study of Reading, University of Illinois at Urbana-Champaign.

achieve vision-only speechreading scores that exceed 80% correct word accuracy for connected discourse. It is not clear, however, what accounts for individual differences in speechreading proficiency.

Numerous investigators have attempted to systematically relate speechreading proficiency to other sensory, perceptual, and cognitive abilities (for reviews see Jeffers & Barley, 1971; O'Neill & Oyer, 1981). Summerfield (1991) concluded that correlations reported in the published literature between intelligence and verbal reasoning measures and speechreading performance were unexpectedly low and not significant for individuals with normal intellectual and language ability. Gailey (1987) reviewed data suggesting that performance on speechreading identification for syllables and isolated words correlated significantly with that of word recognition in connected speech. In the 1970s, initial results (Shepherd, DeLavergne, Frueh, & Clobridge, 1977) suggested that neurophysiologic responsiveness, measured by visual evoked response, accounts for speechreading proficiency. Conflicting results, however, have challenged this conclusion (Rönnberg, Arlinger, Lyxell, & Kinnefors, 1989; Samar & Sims, 1984). The clinical view is that speechreading is a complex multi-factor process (De Filippo, 1990).

Early published accounts of speechreading training recommended the use of visual materials to improve visual perception, increase attention, and develop concentration. For example, Bartlett (1949) outlined activities employing a tachistoscope, a lantern projector with a time-controlled shutter, to vary the exposure to written materials and pictures. O'Neill and Oyer (1981) cited visual training activities outlined more than 40 years ago by Renshaw and by Barnette and continue to recommend the use of timed exposure to complex pictures, printed words and letter strings, and abstract shapes to increase visual attention. Neither clinical nor basic studies, however, have *directly* examined perceptual and attentional processes in speechreading. Lesner and Hardick (1982) investigated the frequency of spontaneous eye blinks in a vision-only speechreading task. Electro-oculographic recording techniques were used to identify blinking, periods of time during which visual input was disrupted. Participants achieving higher passage comprehension scores for lipread material were observed to blink less frequently than those achieving lower scores. This supported their hypothesis that lipreading performance and blinking are inversely related as blinking may reduce visual attention. Another hypothesis related to visual attention is that proficient speechreaders may make better use of the talker's observable speech gestures (Summerfield, 1987, 1991). However, this hypothesis has not been tested. In fact, little is known about proficiency in visual attention to the complex moving face and the selection and use of phonetic visual cues for speech perception.

Several investigators have employed eye tracking to study on-line processing of visual information in a variety of cognitive/visual tasks. Some of these tasks

include the reading of printed text, human face and picture perception, and visual search (e.g., see O'Regan & Lévy-Schoen, 1987; Rayner, 1984; Walker-Smith, Gale, & Findlay, 1977). Our preliminary work has focused on determining whether information about how speechreaders direct attention and extract and use relevant speech cues from the moving face can be studied in real time by monitoring their eye movements.

Purpose

The purpose of the present paper is to report on the feasibility of eye tracking in lipreading from full-motion video images of a talker's face. Instrumentation, procedural modifications to increase data yield, calibration techniques, and data reduction procedures are described. Additionally, evidence of the sensitivity of eye tracking for the quantification of visual behavior in speechreading is provided, using data obtained from a subject with profound hearing loss.

METHOD

Eye-Gaze Behaviors

Typically, a speechreader makes several glances to inspect the talker's face or track facial motion by moving (rotating) the eyes. Physiological evidence demonstrates the importance of eye movement to achieve high acuity for fine visual detail (for a review see Hallett, 1986). Rotating the eyes takes advantage of the highly specialized, densely packed, cone receptor cells that lie within a small indentation in the retinal surface area known as the fovea. When inspecting visual detail, observers typically point their eyes so that the region of interest is projected onto the fovea for periods of time during which there are only miniature eye movements. These periods during which the eyes are relatively stable have been termed eye fixations or gazes. The part of the visual stimulus pattern being directly attended is called the fixation location or point-of-regard. An observer shifts eye gaze from one location to another by making quick, high-velocity eye movements called saccades during which visual input is largely suppressed. There is a strong tendency for people to direct their eyes toward spatial regions or objects being attended, though it is possible to attend to a different area during an eye fixation (Posner, 1980).

Instrumentation

System for eye tracking. Systems for eye tracking have taken many forms (for a review see Hall & Cusack, 1972; Young & Sheena, 1975). The classical vision research by Yarbus (1967) employed a customized cap-like contact lens held on the eye with suction. Methods are now available that are non-invasive and use natural landmarks of the eye (e.g., pupil or reflections from cornea or sclera) to accurately record regions of eye gaze.

In initial experiments, we have used equipment that tracks the center of the pupil, with the observer's head in a stable position, to determine eye-gaze location in reference to sequences of full-face, front-view, video images of a talker. The eyetracker (Stoelting wide-angle Eyetracker, Model 12861) employs a low-level (9000 A) infrared light source (ANSI Z136.1-1973) positioned to illuminate the speechreader's face. A half-silvered mirror reflects the illuminated image of the eye, which is recorded by a small video camera, fitted with a wide-angle lens, and positioned at about 90° to the direction of gaze. Only the left eye is tracked and one pupil monitored. Hardware identifies the x, y position of the pupil, together with its size, from each video frame of the recording video camera. The experimenter qualitatively monitors image quality by observing the reflection of the pupil on a small monochrome monitor. The system is interfaced to a personal computer to digitally process the data acquired by a Qua-Tech PXB-721 Parallel Expansion Board, at a sampling rate of 60 samples per second. Each sample yields three digital values, 7 bits for horizontal and vertical eye-gaze position and 8 bits for pupil size. Algorithms used for data reduction software were developed by McConkie, Scouten, Bryant, and Wilson (1988) and were checked for accuracy.

A second video camera (field-of-view camera), positioned directly above and behind the speechreader, monitors the experimental stimuli shown and positions a box-like cursor over each video frame to indicate the speechreader's point-of-regard. This qualitative information is displayed on-line for the experimenter on a second small monochrome monitor during eye tracking and can be observed to determine whether or not the eye-tracking system is operating properly.

The eyetracker includes an adjustable head rest and chin-support plate to stabilize the observer's head. For even greater stabilization, a bite-bar is used. This consists of a metal mouth plate covered with dental compound (Compound Impression Cake, Red, from Patterson Dental Company). A dental impression is taken, and the compound hardens. Biting back into the impression brings the head to its original position and holds it still. The stable-head (fixed-head) position is necessary because in the pupil-center method of eye tracking, the equipment can not distinguish between eye movements and head or body movements. Consequently, to insure that eye-rotations rather than head movements are recorded the head must remain in a fixed position.

Stimulus delivery. Sequences of full-motion video (30 frames/second) are played without sound on a video recorder/player (Panasonic S-VHS MTS AG1960) and displayed on a 20 in. (diagonal) color television monitor. The monitor is positioned 27.5 in. directly in front of the subject, at a height of 41 in. above the floor. This placement simulates a conversational viewing distance for life-like proportions of the talker's head when the speechreader is seated.

Procedural Modifications

Coding face motion events. A primary goal in the present research has been to understand the temporal relation between face motion as a talker produces speech

and the speechreader's eye-gaze behaviors. Therefore, it was necessary to develop software that could accurately link records of eye-gaze data with specific video frames for speech events (e.g., lips parting, cheeks puffing). For this application, a longitudinal linear time code (Society of Motion Picture and Television Engineers, SMPTE) was dubbed onto one audio channel of the video tape. The tape was played on an S-VHS, Hi-Fi editing recorder/player (Panasonic AG-7750), fitted with a time-code chip. Two observers viewed frame-by-frame sequences displaying the time code to independently identify specific video frames associated with particular speech events. Time-codes (cue-points) corresponding to these events were noted and stored in a file, together with event codes that indicated their nature, to be accessed on-line by the eye-gaze data collection software (McConkie et al., 1988). Supporting software was developed for data reduction.

Linking face motion and eye-movement data. To digitally process the time-code information in real time, a Musical Instrument Digital Interface (MIDI), the Music Quest MQX-32M board, read the SMPTE code recorded on the stimulus video-tape as eye movements were being recorded, making that information available to the computer. The data collection program then watched for the previously-stored time codes and, when one of these appeared, placed a mark containing its event code in the eye-movement data stream to indicate its occurrence.

Calibration Techniques

The data produced by an eyetracker, consisting of two numbers indicating horizontal, x, and vertical, y, position of the eyes at the time each sample is taken, does not, by itself, indicate the point-of-regard in the observer's stimulus field. In order to determine the location in the display toward which the observer's eyes are directed, it is necessary first to perform a calibration task. The purpose of this task is to determine what eyetracker data values are obtained when the observer directs his or her gaze toward certain specified locations in the stimulus area. If the eyetracker data values for these stimulus locations are known, then it is possible to estimate the stimulus locations corresponding to other eyetracker values, through mathematical interpolation (McConkie, 1981).

Stimulus materials. The calibration task we used required the subject to look at each of nine benchmark locations, recording the eyetracker data values corresponding to each. These locations were arranged as a 3×3 array, with each location indicated by an uppercase X, which was 14 pixels high and 8 pixels wide. The center of each X was taken as the location of the benchmark in the stimulus space. The grid was 384 by 384 pixels in size, the region within which the male talker's face appeared on the display screen. A 60 s video clip of the 3×3 array of Xs was included in the videotape prepared for the research at each point where the calibration task was to be carried out.

Calibration procedure. To carry out a calibration, the speechreader was

instructed to look directly at the intersection of each "X" displayed in the array on the 20 in. television screen. This process was repeated a second time. The computer then compared the two sets of eyetracker data values obtained for each of the benchmark locations to see if they were consistent. If the difference between the two samples for either the horizontal or vertical eye position values was greater than 2, then the speechreader was asked to look at that position again and an additional sample was taken. This procedure continued until two successive samples were obtained for each benchmark location that had both horizontal and vertical values that met the criterion for consistency. Because the eyetracker data values tended to be on a scale in which 1 unit of difference indicated a difference in eye position of about ±.25° of visual angle, this insured that the eyes were within .5° of visual angle on the two samples taken at each benchmark location. The average of the two values for both horizontal and vertical location was taken as the eyetracker data value that corresponded to that location in the display.

In order to carry out the calibration procedure with non-hearing subjects, it was necessary to develop a procedure to communicate to them at which X they were to look. An assistant indicated the target benchmark by touching pre-determined regions on the speechreader's hand that were assigned to each of the benchmark positions in the array. This procedure was adopted for all participants. The calibration task took less than 30 s.

The benchmark values obtained in the calibration task were then used to determine where in the display the eyes were being directed during the speechreading task. With each eye fixation, linear interpolation was used to determine toward which pixel of the display the eyes were being directed. In order to insure that the head had not moved significantly during a data collection period, the calibration task was performed before and after presenting three sentences to the speechreader. If the difference between the horizontal and vertical eyetracker values obtained for each of the benchmark locations was greater than 3, the data were considered to be not sufficiently accurate and were not analyzed further. This insured that eye position information in the data analyzed was within .75° of visual angle of the actual location of the eyes. This distance, .75° of visual angle, corresponded approximately to the width of a front tooth in the talker's mouth. Subjects vary in their ability to remain still during trials; data from about 25% of the trials are typically discarded. It should be noted that the spatial resolution of the type of video-camera-based eyetracker used in this study does not permit the reliable detection of very small eye movements (McConkie et al., 1988).

As a way of motivating the subjects to remain still while they were speechreading, they were shown the data values obtained in the calibration task when they did not meet the criteria stated above. This method was used to impress upon them the necessity of refraining from head and body movement during the data collection periods.

Considerations for Subject Selection

Initial experience with eye tracking of speechreading performance has helped us develop a set of guidelines for subject selection, some of which are well known to investigators involved in eye-movement research.

Physical characteristics. One primary consideration is that speechreaders can actively and comfortably participate in the calibration and experimental tasks. For example, one prospective participant was excluded due to habitual squinting and excessive blinking that interfered with the quality of the eye-tracking data. Additionally, due to the necessity of holding the head still, it is critical that every participant demonstrate full control of head movement and comfortably maintain a stable head and body position during a set of trials.

It is also important to establish that speechreaders have sufficient visual acuity and no history of visual pathology. Furthermore, because only one eye is monitored, accurate binocular coordination (the ability to move both eyes together) is required. Normal binocular visual acuity or visual acuity corrected to 20/30 was screened with a Bausch and Lomb Modified Orthorater. In general, speechreaders who wear eyeglasses (particularly bifocals) and some types of hard contact lenses are more difficult to track due to reflections from the corrective lenses. Finally, the use of a dental impression precludes speechreaders who wear dentures or other orthodontic dental appliances.

Experimental Task

Considerations. Because participants are generally restricted in movement and required to remain alert and cooperative, the eye-monitoring portions of the experiments are designed to sample only a subset of speech perception behaviors. The experimental tasks, however, are designed to actively engage participants in the speechreading process, and record baseline behaviors regarding general patterns of looking behaviors.

Due to the necessity of holding the head still, it is important to construct experimental protocols that do not require responses from the speechreader that might introduce head or body movement. Consequently, response tasks are limited to button presses or simple hand gestures and activities such as speaking, typing, signing, and writing are avoided. This restriction could be lifted by using other types of eye-tracking equipment that allow greater freedom of movement.

Speechreading proficiency. To determine subjects' speechreading proficiency and familiarize them with the talker prior to eye tracking, repeated measures of sentence identification and syllable recognition are obtained in initial sessions. We used several hundred stimuli sampled from the Bernstein and Eberhardt (1986) laser video-disc lipreading corpora for this purpose. These stimuli show a male and female talker using natural facial expression in speaking and are presented without auditory cues.

Recognition task for eye tracking. Following the familiarization and profi-

ciency identification sessions, additional sessions are scheduled in which subjects' eye movements are recorded as they carry out sentence recognition tasks. The subjects speechread a set of 18 Central Institute for the Deaf (CID) Everyday Sentences (Davis & Silverman, 1970), stored on laser video disc (Bernstein & Eberhardt, 1986) or other similar materials. Sentences are presented without auditory cues. Each spoken sentence is followed by a written sentence. The subject's task is to decide whether the written sentence does or does not match the talker's utterance and to signal the response with a finger movement. This procedure maintains the subjects' participation in the speechreading task during eye tracking. To insure that the eye-gaze positions are calibrated properly, the subjects are asked to look at the target (benchmark) locations used on calibration at the beginning of the task and following every third stimulus presented.

Data Reduction Procedures

Data reduction employs algorithms (McConkie et al., 1988) to identify locations and durations of eye-gazes, operationally defined as the interval of time from the beginning of one stable period in the data until the beginning of the next stable period. The data stream is inspected to identify irregularities due to blinks, eye-squinting, and head-movement. Such data are excluded from analysis; however, they are documented to determine the frequency with which they occur because it is plausible that they might affect performance.

RESULTS AND DISCUSSION

In order to illustrate more concretely the processes used in the study of speechreading through the use of eyetracking techniques, a small amount of data from a single person will be presented. This subject, age 43, has a bilateral sensori-neural profound hearing loss of congenital onset. She has excellent speechreading skills, scoring 69% words correct on several hundred unrelated sentences sampled from the Bernstein and Eberhardt (1986) lipreading corpora, and relies on speechreading in daily communication situations. She was enrolled in oral education classes during elementary and high-school years and is a graduate of the National Technical Institute for the Deaf. She reports using speech several hours each day with her family and friends, who have normal hearing.

Illustration: Consistency of Calibration Values

The first step in examining eye movement data is to determine whether the data values obtained during calibration tasks meet the consistency criteria stated above. This indicates whether there was sufficient head and body movement to invalidate the data for a given trial. As an example, Table 1 presents a set of calibration data for one trial in which three sentences were speechread and recognition was tested. It shows, first, the nine locations which were used as benchmark locations, at which the Xs were placed. There is one row in the table for each

Table 1

Eye Gaze Calibration Values Obtained From a Subject With Profound Hearing Loss at the Start and End of Data Collection

Array position	X, Y locations					
			Eyetracker data values			
	Location of benchmarks (in pixels)		*Start* before speechreading		*End* after speechreading	
	x	y	x	y	x	y
Row 1						
1	128	452	42	80	42	78
2	320	452	58	80	59	80
3	512	452	76	80	77	80
Row 2						
4	128	260	41	67	41	66
5	320	260	58	68	59	67
6	512	260	76	68	77	66
Row 3						
7	128	68	42	54	41	53
8	320	68	60	54	60	54
9	512	68	76	54	76	54

location. This specifies the 3 × 3 grid used in calibration. Next, the eyetracker data values for horizontal (x) and vertical (y) eye positions, obtained at the beginning of the trial, are listed. Finally, similar values obtained at the end of the trial are listed.

As Table 1 indicates, the first benchmark was at pixel location 128, 452, in the upper left region of the display, and the eyetracker data values obtained prior to the trial were 42, 80. At the end of the trial, when the speechreader looked at the same location, the eyetracker values were 42, 78, thus showing no change in horizontal position, but a small change in vertical position. Neither the x nor the y value showed a difference greater than 3 between the values at the beginning and end of the trial, so the data for this benchmark met the consistency criterion that was set. As can be seen, the values at all other benchmark locations also met this criterion, so the data from this trial were acceptable and were included in further analyses. Eyetracker data values obtained before and after each trial were averaged to obtain the values used in data reduction described below.

Illustration: Data Reduction

Number of samples. The second analysis examined whether the eyetracker data samples were accurately linked to each speech event in the video sequence.

We knew how much time passed between the events of interest on the video tape (e.g., the amount of time that elapsed from face onset to the onset of face motion). Calculations of the elapsed time between these events made it possible to anticipate the number of times eye position would be sampled during these intervals, because samples are taken every 16.7 ms, and with each sample the system checks for critical SMPTE time codes.

To illustrate, Table 2 shows the SMPTE time codes recorded on the video tape corresponding to the onset or offset of different events for the sequence produced by the male talker: "You can catch the bus across the street." In this example, the face appeared on the screen (see "face on" in column 1) at 3 min 16 and ⅞₀ s into the tape, which was ⅘₀ s after the previously-displayed text was removed, as shown in columns 3 and 4. Because eye position was being sampled 60 times per second, twice the video frame rate, it is anticipated that 8 samples would be taken during the ⅘₀ s prior to face onset, as shown in column 5. Column 6 indicates that sample number 1307 was taken at the time the text was removed. This was indicated in the eyetracker data by having a mark value of 6 in the data, as shown in column 2. On sample number 1315, a mark value of 1 was present in the data, indicating that the face had come on. Thus, 8 samples were taken during the interval between the disappearance of the text and the appearance of the face, as shown in column 7. This was the number anticipated.

As Table 2 indicates, each of the obtained values in column 7 is the same as the anticipated value in column 5, with the exception of the number of samples taken between onset of the face (face on) and the onset of facial motion associated with speaking (motion on). Here the obtained number of samples was only 1 less than anticipated, indicating good agreement, being off by no more than 17 ms in a 2.5 s period.

Locations and durations of gazes. The next analysis focuses on identifying the locations and durations of eye gaze. To illustrate, Table 3 lists some results from the data reduction algorithm (McConkie et al., 1988) that were linked to the events in the video sequence for the same sentence. This is only a partial listing of the matrix and does not include information about the lengths or directions of saccades.

A total of 21 eye gazes (GZNUM 15 to 35) were recorded for this stimulus from the time the talker's face appeared until its offset and Table 3 presents a row of data for each gaze. In each row, GZLX and GZLY (gaze locations X and Y) indicate the pixel location in the display to which the eyes were directed. GZDUR (gaze duration) indicates the number of samples taken during the time the eyes were directed at this location or were progressing to the next location. The number of samples taken during the part of this time that the eyes were still is given in GZSTA (stable gaze period). Values in GZDUR and GZSTA can be converted to seconds by dividing by 60, the number of samples per second. Thus, gaze 19 (GZNUM = 19) included 20 samples (GZDUR = 20), and, hence, lasted

Table 2

Temporal Events in the Video Display and Samples Collected for One Spoken Sentence

Column	1	2	3	4	5	6	7
	Video display event	**File mark**	**SMPTE time code**	**Elapsed time**	**Anticipated number of samples**	**Sample number at mark**	**Actual number of samples**
	text off	6	03:16:03			1307	
	face on	1	03:16:07	00:04	8	1315	8
	motion on	2	03:18:21	02:14	148	1462	147
	motion off	3	03:20:29	02:08	136	1598	136
	face off	4	03:22:10	01:11	82	1680	82

Note. SMPTE = Society of Motion Picture and Television Engineers.

Table 3

Eye-Movement Data Matrix Obtained From a Subject With Profound Hearing Loss Speechreading One Sentence

GZNUM	GZLX	GZLY	NAX	NAY	GZDUR	GZSTA	PFLAG	SAMPL	ABER	MCE1	SCE1	ONSCE1	MCE2	SCE2	ONSCE2
15	151.8	237.6	1.00	1.00	19	17	0	1298	0	2	2	17	0	0	0
16	262.1	261.2	1.00	1.00	13	11	0	1317	0	1	13	0	0	0	0
17	332.6	150.6	0.83	1.00	18	18	0	1330	0	1	18	0	0	0	0
18	344.0	187.5	1.00	0.56	18	18	0	1348	0	1	18	0	0	0	0
19	353.6	189.0	0.95	1.00	20	20	0	1366	1	1	20	0	0	0	0
20	344.9	206.2	1.00	1.00	14	14	0	1386	0	1	14	0	0	0	0
21	352.4	220.5	1.00	0.96	115	111	0	1400	1	1	115	0	2	53	62
22	369.1	188.2	1.00	1.00	10	10	0	1515	0	1	10	0	1	10	0
23	365.2	18.5	0.20	0.80	6	5	0	1525	4	1	6	0	1	6	0
24	0.0	500.0	0.00	0.75	7	4	100003	1531	4	1	7	0	1	7	0
25	361.3	192.0	1.00	1.00	4	4	3	1538	0	1	4	0	1	4	0
26	357.3	221.4	0.96	0.72	50	50	0	1542	0	1	50	0	1	50	0
27	349.8	256.7	1.00	1.00	7	5	3	1592	0	1	7	0	4	6	6
28	341.4	313.1	0.62	0.54	13	13	0	1599	6	1	13	0	0	0	0
29	650.7	240.0	1.00	1.00	9	7	0	1612	1	1	9	0	0	0	0
30	320.0	299.4	0.60	0.60	5	5	3	1621	2	1	5	0	0	0	0
31	321.5	316.5	1.00	1.00	8	7	0	1626	0	1	8	0	0	0	0
32	351.5	227.2	1.00	1.00	28	28	0	1634	0	1	28	0	0	0	0
33	330.3	226.7	1.00	1.00	10	10	0	1662	0	1	10	0	0	0	0
34	352.1	248.6	1.00	1.00	8	8	0	1672	0	1	8	0	0	0	0
35	339.1	286.9	1.00	1.00	5	5	3	1680	0	4	0	0	0	0	0

Note. GZNUM = gaze number, GZLX = gaze location on the x dimension, GZLY = gaze location on the y dimension, NAX = percent of x values in the gaze that are non-aberrant, NAY = percent of y values in the gaze that are non-aberrant, GZDUR = gaze duration, GZSTA = gaze duration of stable samples, PFLAG = problem flag, SAMPL = sample number of first sample in the gaze, ABER = number of samples in gaze with aberrant values, MCE1 = mark for continuous event 1, SCE1 = number of samples in gaze for continuous event 1, ONSCE1 = number of samples in gaze prior to which the mark occurred in continuous event 1, MCE2 = mark for continuous event 2, SCE2 = number of samples in gaze for continuous event 2, ONSCE2 = number of samples in gaze prior to which the mark occurred in continuous event 2.

for .333 s. SAMPL (sample number) indicates the number of the eyetracker data sample at which each gaze period began, and thus provides a continuous time record from the beginning of the data collection period.

Indicating the time of talker events. Software has also been developed that places in the data matrix an indication of when the specified events on the video tape occurred. In the data presented in our illustration, there were only five types of events to be marked in the data, as indicated in Table 2. These are conceptualized as three types of continuous events, each with an onset and an offset: text on, face on, and face motion on. Two of these are shown in Table 3, face on and face motion on. For each continuous event, three columns are included in the data matrix. One column, MCE (mark for the continuous event), indicates whether or not the event was on during this gaze. A value of 2 indicates that the event began during this gaze, 1 indicates that it was on during the entire gaze period, 4 indicates that it went off during this gaze period, and 0 indicates that it was never on during this gaze period. A second column, SCE (samples for a continuous event), indicates the number of samples during this gaze that the event was on. As in previous columns, the number of samples becomes a basis for timing, and can be converted to seconds by dividing by 60. The third column, ONSCE (onset of continuous event), indicates the time at which the event changed state if it either came on or went off during that gaze period.

In Table 1, event 1 was the presence of the face on the screen, and data for this event are given in columns labelled MCE1, SCE1, and ONSCE1. On gaze number 15 (GZNUM = 15), the value in column MCE1 is 2, which indicates that the face appeared during that gaze period. The value of SCE1 for that gaze is also 2, indicating that the face was present on only two samples. The value of ONSCE1 is 17, indicating that the face came on after the 17th sample in that gaze period. On gaze number 16 through 34 the value of MCE1 is 1, indicating that the face was on the screen throughout that period. In each of these gazes, the value of SCE1 is equal to the gaze duration, GZDUR, because the face was on for the entire period of each gaze, and the value of ONSCE1 is 0, indicating that no state change occurred during those gaze periods. Finally, on gaze 35, the value of MCE1 is 4, indicating that the face went off during that gaze period. SCE1 for this gaze is equal to 0, indicating that the face went off just as that gaze began, so was not on for any of the samples; ONSCE1 is also zero for the same reason.

Event 2 in the data presented in Table 3 was the presence of face motion, beginning at the first point at which motion associated with speaking the sentence could be observed, and ending with the end of the facial motion associated with speaking that sentence. Values in columns MCE2, SCE2, and ONSCE2 can be interpreted in the manner just described for face onset time.

The software has an additional capability, not illustrated here, of recording non-continuous events such as the moment in time at which the lips parted in

speaking a certain phoneme. Mark values would be assigned to them, similar to those shown in Table 2, and they would be recorded in additional columns, indicating on which sample in which gaze periods these events occurred.

Adding stimulus region information. It is also possible to specify certain areas of interest within the stimulus region and identify the gazes directing the eyes toward each of these regions. By giving each region a unique code, a column can be added to the data matrix that indicates, for each gaze, which region the eyes were directed toward during that gaze period.

Identifying blinks and other aberrations. Certain problems can occur in the process of preparing a data matrix of the type shown in Table 3, in which there is a row for each gaze period. Some of these are the result of non-optimal adjustment of the eye-tracking equipment, leading to unreliability in identifying the pupil boundaries as required in order to identify the center of the pupil. Others result from eye behavior of the speechreader, in the form of blinking and squinting, which distorts or occludes the pupil and can lead to false eye position readings. In the data matrix, the PFLAG (problem flag) column codes a number of potential difficulties that are occasionally detected during gaze periods that would suggest inaccuracy in the data. This includes periods when the pupil size becomes too small or too large, indicating occlusion by the eyelids, improper illumination of the face, or other problems. Any time the value in this column is 10 or greater, the data are not likely to be accurate. Gaze number 24 shows an example of such a case. Thus, the extreme GZLX and GZLY values for this gaze do not truly indicate eye position, but probably result from the speechreader blinking her eyes.

Other columns in the matrix, including NAX, NAY, and ABER, provide further diagnostic information to help detect times when the eyetracker has not been well adjusted and when droopy eyelids or squinting may have caused inaccurate data. As a case in point, the low value in the NAX column in gaze number 23 indicates that this gaze is also somewhat questionable. Because this is just before a blink, it is likely that the pupil was partially occluded by the eyelids during part of the gaze period, leading to an unreliable reading. The low value in the GZLY column, suggesting a saccade to the bottom of the screen, is thus probably spurious.

Summary. A data matrix is produced that can be readily analyzed with common statistical software or specially-developed programs. It consists of a row giving information about each separate gaze that the speechreader makes, indicating where the eyes were directed in pixel coordinates, how long the eyes remained at that point, which of the specified stimulus regions contained the point-of-regard, and the state change and time period of any events of interest occurring on the video display during that gaze. It also provides indicators that can be used to identify inaccurate data. With this matrix, it becomes easy to determine the mean duration of speechreaders' gazes when watching a talker, the

amount of time spent looking at different facial regions, and the time that elapses from the onset of facial motion until the eyes move to the speaker's mouth, for example. It is also easy to determine where the eyes tend to be directed at the time of, or immediately following, selected speech events. Assuming a close relation between where the eyes are directed and the part of the face being attended, this record can provide information about the dynamics of visual attention taking place as a person attempts to comprehend speech visually.

Illustration: Scan Path

Another way to describe the temporal and spatial patterns of eye gazes is in the form of a scan path diagram. To illustrate, Figure 1 displays the data presented in Table 3 for the sentence "You can catch the bus across the street." In this figure, VGA pixel coordinates that locate eight regions of the talker's face are outlined. The areas within the rectangles define the talker's (from top-to-bottom, left-to-right): forehead, left eye, right eye, left cheek, nose, right cheek, mouth, and chin regions. The point-of-regard for each of the gazes in Table 3, given by GZLX and GZLY, were plotted to illustrate eye-gaze location for known regions on the talker's face. The size of the triangles has been scaled to reflect the duration of each gaze location during face motion, and the circles, times prior to and following face motion. The chronological order in which the gaze durations occurred during the video sequence is shown by the connected lines. The solid lines are used to show the scan path during face motion, and the dotted lines, times prior to and following face motion.

Description of Subject's Performance

The information contained in Figure 1 and Table 3 make it possible to examine eye behavior in relation to the temporal sequence of speech events of interest.

Face onset to motion onset. At the onset of the talker's face (shown by MCE1 = 2), the speechreader's eye-gaze location (GZLX = 151.8, GZLY = 237.6) was just outside of the talker's left cheek region. This gaze (GZNUM = 15) was 19 samples, or .30 s, in duration (GZDUR = 19) and began on sample number 1298 (SAMPL = 1298). The actual sample number at which the face onset occurred, however, was 1315. Therefore, 17 samples were collected at this location prior to the face onset and once the face onset occurred, eye gaze was maintained at this same location for only an additional 2 samples (.03 s). From this time until the onset of face motion (MCE2 = 1), the speechreader made a total of six separate eye movements (outside of left cheek-nose-chin-mouth-mouth-mouth-mouth). Gazes were typically of short durations ranging in time from .22 s to .33 s.

Face motion. Within the seventh eye gaze (GZNUM = 21) the onset of face motion occurred (MCE2 = 2). Eye gaze was directed at the talker's mouth (GZLX = 352.4, GZLY = 220.5) for 1.03 s (ONSCE1 = 62) prior to the onset of

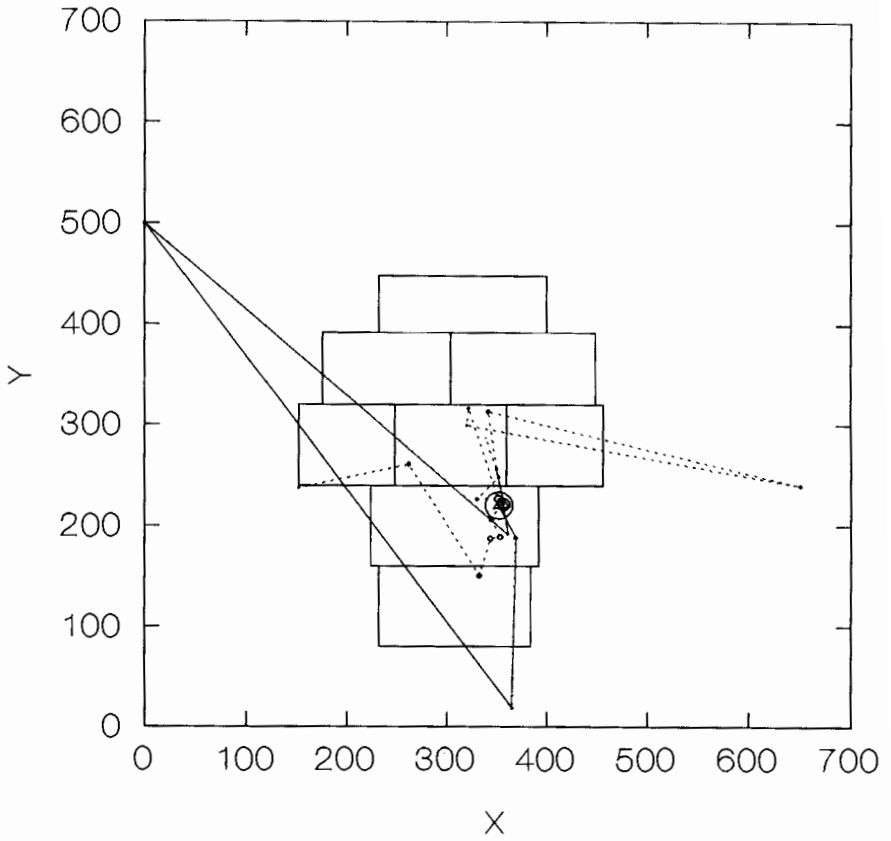

Figure 1. The chronological order and X, Y locations for sequences of eye gaze by a subject with profound hearing loss are displayed in relation to eight regions of a talker's face for the sentence "You can catch the bus across the street." Durations (seconds) of eye gaze during face motion are represented by the scaled size of the triangles connected by solid lines, and eye-gaze durations prior to and following face motion are represented by the scaled circles connected by dotted lines.

motion and then remained at that location for another .88 s (SCE2 = 53). Obviously, the subject was expecting, and waiting for, the speech to begin. This was followed by six additional gazes during face motion that ranged in duration from .067 to .833 s. The X, Y eye gaze locations (GZLX, GZLY), as plotted in Figure 1, suggest that the speechreader directed her gaze outside of the regions of the talker's face on two occasions during face motion (GZNUM = 23, 24). However, as previously noted, the information associated with these two gazes indicates that a blink actually occurred at this time, so the point-of-regard infor-

mation should be disregarded. However, Table 3 indicates that the blink lasted for a total of 13 samples, or .22 s.

The data pattern indicates that as the speechreader watched this sentence spoken, all her gazes were toward the region of the mouth. After an initial long gaze, she made a short saccade followed by a 2.17 s gaze, made another saccade followed by a very short gaze, then gazed at a new location for .83 s. A final gaze of .12 s ended near the time the face motion ended, with a saccade that took the eyes away from the mouth. Thus, the speechreader appears to have been aware that this was the end of the sentence before the face motion stopped.

Face motion offset to face offset. Following face motion (MCE2 = 4), the speechreader proceeded to make eight additional gazes returning to the mouth region on two occasions for a total time at the mouth of .63 s. Most of the gazes, however, clustered at the talker's nose region which is a common point-of-regard for people looking at a face as a whole. Gaze number 29 was directed out of the face region for .15 s and everything indicates that this was a real event rather than a blink. At face offset (MCE2 = 4) the speechreader returned to the talker's nose region, probably anticipating the appearance of the written test sentence which typically appeared at this location.

Summary. The present results illustrate the fact that shifts in eye gaze occur during speechreading. By interpreting the eye gaze patterns in relation to specific speech events or facial movement in the stimuli, various hypotheses may be developed. For example, shifts may occur when the speechreader has selected critical information, become fatigued, or alerted to unexpected facial movement of the talker. It would be interesting to know whether a blink interferes with recognition or comes at a time when the information in the speech signal is low. Additionally, patterns of eye gaze may differ for varying talker characteristics or demands of the speechreading task. These hypotheses may be tested with a variety of speechreading stimuli and tasks using eye tracking and measures of performance.

CONCLUSION

We have illustrated the feasibility and application of eye tracking in studying speechreading. The technology is available to accurately link the direction, duration, and sequences of eye-gaze patterns to spatial and temporal patterns of speech events. Such data will contribute to an understanding of how a speechreader selects and extracts information from the complex environment of the face and may help indicate the basis for proficiency in speechreading. Knowledge about patterns of eye gaze may also be useful in developing hypotheses about attentional and visual processes in speechreading.

Although the results described are based on unrelated sentence stimuli produced by a single talker, the techniques are applicable to a variety of experimental materials and paradigms. The combination of eye monitoring and perfor-

mance in visual speech perception should provide sensitive real-time, on-line measures of attentional processes in visual speech perception. Knowledge about visual processes is needed in order to design research-based intervention protocols, to optimize sensory aids to augment speechreading, and to enhance the design of human computer interfaces.

ACKNOWLEDGEMENTS

This research was supported by NIDCD Grant DC001600 to the University of Illinois. Portions of this research were presented at the 1993 Summer Institute of the Academy of Rehabilitative Audiology. The authors wish to thank Ying Kong for her assistance in software development.

REFERENCES

American National Standards Institute. (1973). *American National Standards specifications for the safe use of lasers* (Z136.1-1973). New York: Author.

Bartlett, R. (1949). Attention in speech reading. *Hearing News, 17* (3), 1-2, 18, 20, 22.

Berger, K.W. (1972). *Speechreading principles and methods.* Baltimore: National Educational Press.

Bernstein, L.E., Demorest, M.E., Coulter, D.C., & O'Connell, M.P. (1991). Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America, 90,* 2971-2984.

Bernstein, L.E., & Eberhardt, S.P. (1986). *Johns-Hopkins lipreading corpus I-IV, discs I & II* [Laser Videodiscs]. Baltimore: The Johns-Hopkins University.

Davis, H., & Silverman, S.R. (Eds.). (1970). *Hearing and deafness* (3rd ed.). New York: Holt, Rinehart and Winston.

De Filippo, C.L. (1990). Speechreading training: Believe it or not! *Asha, 32* (4), 46-48.

Erber, N. (1972). Speech-envelope cues as an acoustic aid to lipreading for profoundly hearing-impaired children. *Journal of the Acoustical Society of America, 51,* 1224-1227.

Gailey, L. (1987). Psychological parameters of lipreading skill. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 115-141). London: Lawrence Erlbaum Associates.

Hall, R.J., & Cusack, B.L. (1972). *The measurement of eye behavior: Critical and selected reviews of voluntary eye movement and blinking* (Technical Memorandum 18-77, AMCMS Code 501B.11.84100.). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (NTIS No. AD-752 904)

Hallett, P.E. (1986). Eye movements. In K.B. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance I: Sensory processes and perception* (pp. 10-1 - 10-102). New York: Wiley.

Jeffers, J., & Barley, M. (1971). *Speechreading (lipreading).* Springfield, IL: Charles C. Thomas.

Lesner, S.A., & Hardick, E.J. (1982). An investigation of spontaneous eye blinks during lipreading. *Journal of Speech and Hearing Research, 25,* 517-520.

McConkie, G.W. (1981). Evaluating and reporting data quality in eye movement research. *Behavior Research Methods & Instrumentation, 13,* 97-106.

McConkie, G.W., Scouten, C.W., Bryant, P.K., & Wilson, J. (1988). A microcomputer-based software package for eye monitoring research. *Behavior Research Methods, Instruments & Computers, 20,* 142-149.

O'Neill, J.J., & Oyer, H.J. (1981). *Visual communication for the hard of hearing* (pp. 66-89). Englewood Cliffs, NJ: Prentice-Hall.

O'Regan, J.K., & Lévy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 363-383). Hillsdale, NJ: Lawrence Erlbaum Associates.

Posner, M.I. (1980). Orienting attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

Rayner, K. (1984). Visual selection in reading, picture perception and visual search: A tutorial review. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 67-96). London: Lawrence Erlbaum Associates.

Rönnberg, J., Arlinger, S., Lyxell, B., & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *Journal of Speech and Hearing Research, 32,* 725-735.

Samar, V.J., & Sims, D.G. (1984). Visual-evoked response components related to speechreading and spatial skills in hearing and hearing-impaired adults. *Journal of Speech and Hearing Research, 27,* 162-172.

Shepherd, D.C., DeLavergne, R., Frueh, F., & Clobridge, C. (1977). Visual-neural correlate of speechreading ability in normal-hearing adults. *Journal of Speech and Hearing Research, 20,* 752-765.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 3-51). London: Lawrence Erlbaum Associates.

Summerfield, Q. (1991). Visual perception of phonetic gestures. In I.G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 117-138). Hillsdale, NJ: Lawrence Erlbaum Associates.

Walker-Smith, G.J., Gale, A.G., & Findlay, J.M. (1977). Eye movement strategies involved in face perception. *Perception, 6,* 313-326.

Yarbus, A.L. (1967). *Eye movements and vision.* New York: Plenum Press.

Young, L.R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation, 7,* 394-429.