Experiments with a Camera-Based Human-Computer Interface System

Robyn Cloud*, Margrit Betke**, and James Gips***

* Computer Science Department, Boston University, 111 Cummington Street, Boston, MA 02215, USA, robyn_l_cloud@yahoo.com

** Computer Science Department, Boston University, 111 Cummington Street, Boston, MA 02215, USA, betke@cs.bu.edu, http://www.cs.bu.edu/faculty/betke

*** Computer Science Department, CSOM, Fulton Hall 460, Boston College, Chestnut Hill, MA 02467, USA, gips@bc.edu, http://www.cs.bc.edu/~gips

ABSTRACT

A human-computer interface (HCI) system called "The Camera Mouse" is evaluated. It tracks a user's movements with a video camera and translates them to movements of the mouse pointer on the screen [1]. The main objectives for the experimentation were to quantitatively define the performance of the system for different users, features, and applications, to determine the optimal settings for different kinds of users, and to compare measurements over all users in order to suggest enhancements to a future system of this type. The experiments were conducted with 11 participants including a subject with severe physical disabilities. Each subject repeatedly performed a number of tasks. During each trial, a different feature was tracked and the elapsed time and mouse movement trajectories were measured. These measurements were used to quantify the system's performance.

1. INTRODUCTION

Human-computer interface (HCI) systems are key components in allowing persons with severe disabilities to communicate with family, friends and caregivers. In many cases, persons with severe disabilities are non-verbal and have limited voluntary movement that greatly hinders their ability to share their needs and desires with others. For example, persons afflicted with severe cerebral palsy may be quadriplegic, spastic and athetiodic whereby muscles in all parts of their body are unable to relax and they are unable to control the movement of many muscles in any part of their body [2]. An assistive HCI device like the "Camera Mouse" [1,3] not only allows severely disabled persons to communicate their wants, but also allows them to use the computer for educational and recreational purposes. With an effective assistive HCI system, communication via a computer can be accomplished without the use of traditional devices such as a keyboard or mouse and can instead be tailored to the user's own motor abilities. These qualities of an assistive HCI system make it an invaluable asset to those living with severe disabilities.

2. THE CAMERA MOUSE HCI SYSTEM

"The Camera Mouse" system is comprised of a single computer, a camera, and a driver program. The driver program tracks the user's movements with the camera and translates them

into movements of the mouse pointer on the screen. Any body feature such as the nose, finger, or lip can be tracked to control mouse movements. The feature to be tracked is initialized by the user (or caregiver), when it is selected with a mouse click in the live video box displayed on the screen. The field-of-view of the tracked feature in the video box can be changed by zooming in or out of the scene with the camera controls (see Figure 1).

The user is able to move the mouse pointer by moving his/her selected feature in space. When the feature's field-of-view is small in the video box (i.e. the feature is zoomed-in on), small movements of the feature in space will translate to larger movements of the mouse pointer, while a large field-of-view will translate to smaller mouse movements.
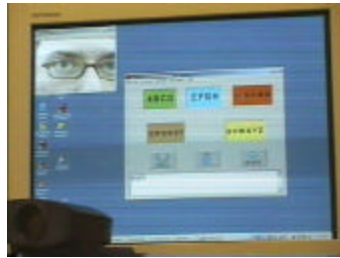

Figure 1: Participant using Camera Mouse for SpeechStaggered application where inner left eye contour is zoomed-in on (video box is in upper left corner of screen)

In order to select items on the screen, the system provides click-event functionality such as right clicks, left clicks, and double clicks. The clicks are generated based on "dwell time." Specifically, when a user wishes to select an item on the screen, he/she must hold the mouse over that item (within a 30-pixel radius) for a specified period of time (between 0.5 seconds and 1.5 seconds) then the click-event selected will execute and the item will be selected .

The tracking algorithm that governs this system utilizes a technique called "template matching." When the feature to be tracked is selected, a square is placed over the feature and the subimage within this square is saved. The saved subimage becomes a template that will be used to determine the location of the feature in the next frame. The tracking algorithm attempts to match the template to a search window in the new frame that is centered at the position of the feature in the previous frame. All subimages within the search window are correlated against the template and the subimage with the highest correlation becomes the new template while its center coordinates become new pointer coordinates for the mouse on the screen. If, at any time, the correlation value computed is low and the original feature cannot be found in the next frame (due to quick movements by the user for example), the feature can be re-selected by the user and the template will thereby be manually updated.

3. EXPERIMENTATION PARAMETERS AND METHODS

Three application programs were used by the participants to perform tasks from which data was captured. These programs are "EaglePaint" and "SpeechStaggered," both developed at Boston College, and an educational program titled "Richard Scary's Best Reading Program Ever" (see Figure1 above for SpeechStaggered screen shot, see Figures 2 and 3 below for EaglePaint and Reading Program screen shots).

Figure 2: Participant using Camera Mouse for EaglePaint application (tracking lower lip)

Figure 3: Participant using Camera Mouse for Reading Program: Kite Game (tracking nose)

The EaglePaint program is a simple painting program that allows the user to draw multi-colored shapes in the application window using only mouse pointer movement (see Figure 2). For experimentation purposes, each participant was asked to draw a horizontal straight line from one marker on the left side of the application window to another marker on the right side (for each feature tracked).

The SpeechStaggered program allows the user to spell out words and phrases from five boxes together containing all letters of the English alphabet. The boxes are organized such that there are three centered at the top of the application window containing clusters of the letters "ABCD" "EFGH" and "IJKLMN", and two centered just below those with clusters of the letters "OPQRST" and "UVWXYZ." The program also includes buttons to add a space between words, delete letters, and give a computer-generated pronunciation of the word or phrase. In order to spell a word using SpeechStaggered, the user positions the mouse pointer over a particular letter cluster then left clicks to select the cluster. Upon clicking, a new window appears showing each letter of the cluster in an individual box. The user then left clicks on his/her desired letter which is then shown in a dialog box in the application window. During experimentation, each user was asked to spell three words (for each feature tracked), "RAINING", "MINIMAL", and "POOR." These words were chosen on the basis of the location of their letters in the application window which influenced the differentiation that their x- and y-mouse-coordinate trajectories would have.

The Richard Scary program is a children's educational program which contains five "worlds" where the user can play games associated with spelling and reading. For the purposes of experimentation, each subject was asked to play a matching game in the "beach world." The game involved having the user match words and pictures on kites in order to make them fly. Two sets of four kites appear on the screen. The first set, which appears in the upper portion of the screen, contains kites with words describing objects. The second set appears below the first set and contains pictures of the objects described in the first set. The user was asked to match either a kite with a word to one with a picture or vice-versa by selecting one kite then its matching partner. The user played this game four times (once per feature) and selected kites in random order in each iteration of the game.

Four features were tracked for each subject, the nose, lower lip, interior of left eye, and thumb. Figures 4-7 below show the field-of-view of the camera mouse for each feature tracked Table 1 shows the experimentation participants.
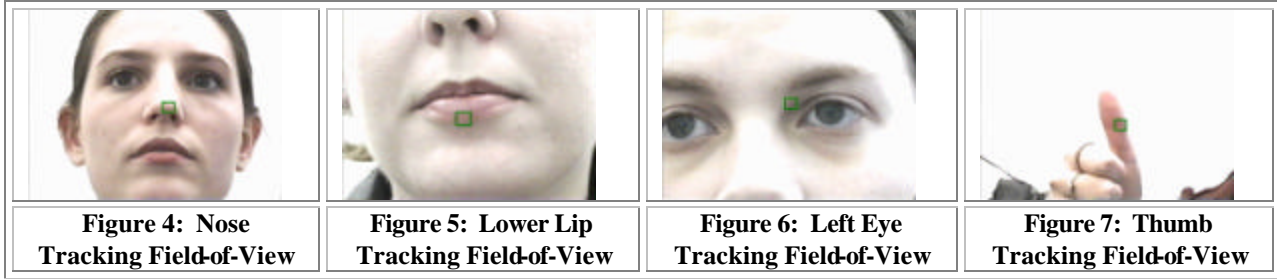
| | Figure 4: Nose Tracking Field-of-View | Figure 5: Lower Lip Tracking Field-of-View | Figure 6: Left Eye Tracking Field-of-View | Figure 7: Thumb Tracking Field-of-View |

Figure 4: Nose Tracking Field-of-View

Figure 5: Lower Lip Tracking Field-of-View

Figure 6: Left Eye Tracking Field-of-View

Figure 7: Thumb Tracking Field-of-View

Table 1: Participants

4. EXPERIMENTATION RESULTS

4.1 Time Results

The following table summarizes the average elapsed time (in seconds) taken by all non-disabled subjects to perform each task by feature. The average was taken over 4 trials.

	All Subjects			
	Nose	Lip	Eye	Thumb
EaglePaint	7.1	7.1	5.9	9.3
SpeechStaggered: Raining	46.0	53.6	53.4	58.5
SpeechStaggered: Minimal	47.9	52.5	51.7	52.1
SpeechStaggered: Poor	22.9	26.0	24.0	31.0
KiteGame	56.2	61.6	50.2	59.8
	Nose	Lip	Eye	Thumb
Overall Avg. Time per Feature	36.0	40.2	37.0	42.1

Table 2: Average elapsed time in seconds per task and feature

The following table summarizes the average elapsed time taken by people wearing glasses versus people not wearing glasses to complete each task while the eye was being tracked:

	Glasses Wearing Subjects	Non-Glasses Wearing Subjects
	Eye	Eye
EaglePaint	5.0	6.1
SpeechStaggered: Raining	65.0	50.5
SpeechStaggered: Minimal	53.0	51.4
SpeechStaggered: Poor	25.5	23.6
KiteGame	49.0	50.5
	Eye	Eye
Overall Avg. Time for Eye	39.5	36.4

Table 3: Average elapsed time in seconds per task while tracking the eye of glasses and non-glasses wearers

4.2 EaglePaint Results

This section describes the data from the EaglePaint application that were used to measure the smoothness of mouse movement over the different features.

For each subject and feature, a graph was generated showing the x mouse coordinate trajectory for the user drawn horizontal line. An example graph is shown in Figure 8 below:
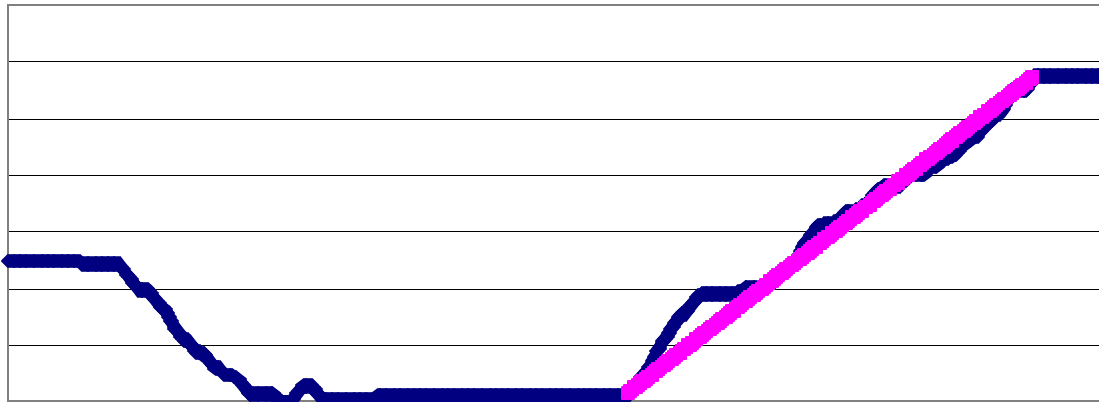

**Figure 8: X trajectory (blue line) with Ground Truth Straight Line (pink line)**

Figure 8 above shows the x-coordinate movement of the mouse for subject 1 using EaglePaint to draw a straight line while the nose was tracked. In the figure, the x-axis represents time while the y-axis represents the mouse x-coordinate. The blue line shows the movement of the mouse by the user in the horizontal direction. A stationary segment of the blue line indicates that the user was not moving the mouse in the horizontal direction either left or right, a downward movement in the line indicates that the user was moving to the left, and an upward movement indicates that the user was moving to the right. The pink line in Figure 8 represents a straight line from the x-coordinate where the user started to draw the line to the coordinate where the user finished drawing his/her line. The blue line trajectory up to where the pink line begins represents the adjustment period of the user to move the mouse from its position on the screen when tracking begins to the left marker in the application window.

The results measured the smoothness of movement for each feature and show that subjects were able to draw the smoothest lines using their thumbs and the least smooth lines using their eyes. For the thumb feature, the smallest Euclidean distance between the drawn line and the ground truth straight line was measured, while for the eye, the largest Euclidean distance

between the drawn and straight lines occurs. The results also show that glasses wearing subjects were able to draw a significantly smoother line while their glasses were being tracked as opposed to the non-glasses wearing subjects whose left eye interior was tracked.

4.3 Results for Subject with Severe Physical Disability

A user with severe disabilities was the subject of the final experiment (see Figure 10 below). This user is a 40-year-old woman who is afflicted with cerebral palsy and has participated in six Camera Mouse sessions since December of 2001. She is non-verbal except to make sounds indicating agreement or disagreement with a question being posed. Figure 9 below is a picture of the Camera Mouse setup for users with disabilities in the Human-Computer Interface Laboratory at Boston College.

Figure 9: Disabled User Setup at Boston College	Figure 10: Disabled User
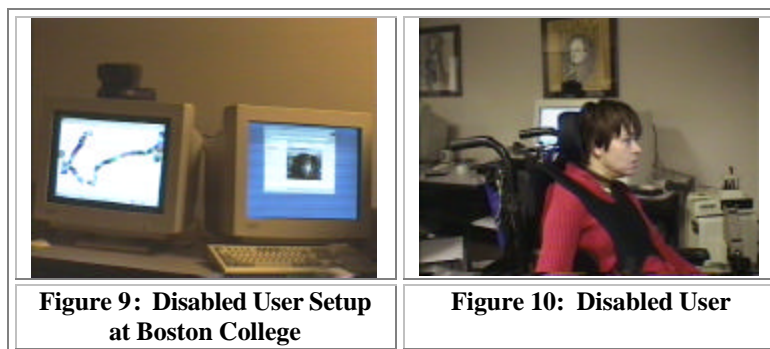
The subject views the monitor on the left in figure 9 while she is working with the software applications. The monitor on the right is used by the caregiver to observe the Camera Mouse video box and manually update tracking of the subject if necessary.

During the user's experimentation session, she was able to use the EaglePaint and SpeechStaggered programs. The EaglePaint program was the first application used. Initially, the disabled user's nose was tracked with the same field-of-view as the non-disabled users and she was asked to draw a horizontal line from the left side of the application window to the right side. The field-of-view parameters proved unsatisfactory for the user since her spastic movements would cause her face to move entirely out of the Camera Mouse video box. As a result of this, the field-of-view was changed to include her full face in approximately the center 30% of the screen. This new setting alleviated the problem of her face moving out of the screen, however, the nose proved problematic as a tracked feature. Throughout this trial, the tracker lost the position of the nose due to the irregular and quick movements of the user and the template had to be manually updated several times. In addition, the user did not have enough fine motor control to draw a horizontal line and found it especially difficult to reach the first marker on the left of the screen. After also experimenting with lower lip tracking and eye tracking, it was observed that the tracker worked best for this subject when tracking the patch of skin just above the bridge of the nose and between the eyes. This was the tracking feature that was used in subsequent experiments with the user.

In addition to having the user draw a horizontal line with the EaglePaint program, she was also asked to draw a vertical line from the top of the screen to the bottom. This task proved easier for the user as she seemed to have better motor control of her head in a vertical direction than in a horizontal direction. She performed this task four times with a progressive increase in

elapsed time to perform the task. As the user continued to perform the task, she seemed to become tired and somewhat frustrated that she was unable to make the mouse function as she wanted it to. This most likely accounted for the increase in task performance time. Figure 11 below describes the time taken for the disabled subject to draw lines with the EaglePaint application and shows that the eye was the fastest feature to be tracked on average.

	User with Disabilities		
	Nose	Lip	Eye
EaglePaint - Horizontal Line Trial 1	9		
EaglePaint - Horizontal Line Trial 2		21	
EaglePaint - Horizontal Line Trial 3			20
EaglePaint - Horizontal Line Trial 4			4
EaglePaint - Horizontal Line Trial 5			17
EaglePaint - Vertical Line Trial 1			11
EaglePaint - Vertical Line Trial 2			6
EaglePaint - Vertical Line Trial 3	28		
EaglePaint - Vertical Line Trial 4	55		
Average	31	21	12

Figure 11: Elapsed time in seconds for subject with disabilities using the EaglePaint Program

The other task performed by the disabled user was to spell the word "POOR" with the SpeechStaggered program. This task was extremely difficult for the user for a number of reasons. One significant reason is that the size of both the letter cluster boxes and the individual letter boxes in the SpeechStaggered application window seemed to be too small for the user. Another is that the individual letter boxes were too close together and awkwardly placed such that the user was not able to select the letters she wanted. For instance, the user was able to effectively select the letter cluster box containing the letters for "POOR" and was even able to select the letter 'P' after an initial attempt, however, she had great difficulty selecting the letter 'R' because of both its position in the window and its size. The positions of the letters in the individual window that are contained in the word "POOR" are roughly like that in figure 12:

Figure 12: Positions of Letters in Individual Letter Boxes of EaglePaint Program

When the cluster box 'OPQRST' is selected, a window like the table in Figure 12 appears and the mouse is initially contained over the letter 'O'. As a result, in order to select the letter 'R' for example, movement in both the horizontal and vertical directions is required. As movement in the horizontal direction in addition to the fine motor control required to hold her head in a fixed position was problematic for the user, when she attempted to select the letter 'R,' she would inadvertently select the letter 'T' instead. This limitation in her movement caused her to spell the following: "OPOTOTTR" instead of "POOR."

An observation directly resulting from this disabled user's experience with the SpeechStaggered program would be one of the need for more effective application appearance design for spelling programs of this type. For instance, keeping in mind that a severely disabled user may have a very limited range of movement as well as a limited ability to hold his or her

head in a fixed position, spelling applications should be designed so that letters are in easy to reach locations on the screen and are large enough so that the user can waver in movement without fear of failing to select the desired letter. A communicative application such as this could also incorporate functionality for quick phrase generation so that a user is not forced to painstakingly spell out every letter in the phrase.

5. Conclusion

This paper explained the results and evaluation of experimentation with an HCI system called "The Camera Mouse." The experiments were conducted with ten non-disabled users and one disabled user who were asked to perform various tasks through mouse movement and selection while a facial feature or the thumb was tracked. The results found through these experiments could help to augment and customize future HCI systems for different types of users. Overall, the results show that the nose is the most reliable tracking feature for non-disabled users. This feature was the fastest feature over all users and the second smoothest feature. In future HCI systems, it would be possible to use this information in order to customize tracking settings for non-disabled users. The results of the SpeechStaggered program for non-disabled users show that different words can be recognized based on their trajectories. For future HCI systems, this could mean employing a technique to recognize similar trajectories in order to identify and spell words quickly based on their initial trajectory observations. For disabled users, this could dramatically speed up their rate of communication. The results obtained from the experimentation session with the disabled user highlighted the importance of not only user customization for feature tracking and other settings, but also effective application design that takes into account user disabilities. The disabled user described in this paper could greatly benefit from an HCI system that automatically tracks the easiest feature for her to use as well as a spelling application with larger letters that are ergonomically positioned for most efficient use.

Acknowledgements

References

[1] M. Betke, J. Gips, and P. Fleming. The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People with Severe Disabilities. IEEE Transactions of Neural Systems and Rehabilitation Engineering, in press, June 2002.
[2] F. Miller, M.D., and S. Bachrach, M.D. "Cerebral Palsy: A Complete Guide for Caregiving", 1998. JHB: A Johns Hopkins Press Health Book.
[3] J. Gips, M. Betke, P. Fleming. The Camera Mouse: Preliminary Investigation of Automated Visual Tracking for Computer Access. Proceedings of the Rehabilitation Engineering and Assistive Technology Society of America 2000 Annual Conference (RESNA 2000), pp. 98-100, Orlando, FL, July 2000.