HMAGIC: Head Movement And Gaze Input Cascaded Pointing

Andrew Kurauchi
Boston University
kurauchi@bu.edu

Wenxin Feng
Boston University
wenxinf@bu.edu

Carlos Morimoto
University of São Paulo
hitoshi@ime.usp.br

Margrit Betke
Boston University
betke@bu.edu

ABSTRACT

Augmentative and alternative communication tools allow people with severe motor disabilities to interact with computers. Two commonly used tools are video-based interfaces and eye trackers. Video-based interfaces map head movements captured by a camera to mouse pointer movements. Alternatively, eye trackers place the mouse pointer at the estimated position of the user's gaze. Eye tracking based interfaces have been shown to even outperform traditional mice in terms of speed, however the accuracy of current eye trackers is not enough for fine mouse pointer placement. In this paper we propose the Head Movement And Gaze Input Cascaded (HMAGIC) pointing technique that combines head movement and gaze-based inputs in a fast and accurate mouse-replacement interface. The interface initially places the pointer at the estimated gaze position and then the user makes fine adjustments with their head movements. We conducted a user experiment to compare HMAGIC with a mouse-replacement interface that uses only head movements to control the pointer. Experimental results indicate that HMAGIC is significantly faster than the head-only interface while still providing accurate mouse pointer positioning.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*input devices and strategies*; K.4.2 [**Computers and Society**]: Social issues—*Assistive technologies for persons with disabilities*

General Terms

Human Factors

Keywords

Assistive technology, video-based mouse-replacement interface, gaze-based interaction

PETRA '15, July 01 - 03, 2015, Corfu, Greece

ACM 978-1-4503-3452-5/15/07$15.00.
DOI: http://dx.doi.org/10.1145/2769493.2769550.

1. INTRODUCTION

Augmentative and alternative communication tools allow people who are not able to use traditional input methods such as mouse and keyboard to interact with a computer. Such interfaces are particularly empowering for people with severe disabilities, for example quadriplegia caused by traumatic brain injuries, cerebral palsy, or stroke, allowing them to communicate using a computer.

Current solutions include the use of speech recognition [3], and gaze-based interfaces. Video-based mouse-replacement interfaces, such as the Camera Mouse [1] and SINA [8], provide an alternative to use the movements of a body part to control the mouse pointer. In the case of people with quadriplegia, it often means using their head. A camera captures videos from the user's head and computer vision algorithms are applied to track the movement of a feature of the head (e.g., nostril or eyebrow corner) that is used to determine the mouse pointer position.

Alternatively, eye trackers place the mouse pointer directly at the user's gaze position on the screen. An eye tracker is either a head-mounted device, or a device, unattached to the user, typically placed near the computer monitor. Eye trackers can therefore be classified as either "head-mounted" or "remote." Most remote eye trackers cannot accommodate large head movements, while head-mounted eye trackers do not restrict the user's head movements.

The accuracy of eye trackers typically varies between 0.5^o and 1^o of visual angle (or approximately 0.5 – 1 cm at a distance of 60 cm) and is not expected to improve much with advances in eye tracking technology, considering the size of the fovea [6]. Such accuracy may not be sufficient for fine mouse pointer positioning. Another issue with using eye trackers as mouse replacement systems comes from the user experience perspective: It can be distracting if the mouse pointer is always following the user's gaze.

Manual And Gaze Input Cascaded (MAGIC) pointing was proposed by Zhai et al. [9] as a combination of manual mouse motion and gaze information to provide faster mouse interaction without depending on high accuracy gaze estimation and distracting users with the mouse pointer constantly following their gaze. MAGIC pointing is based on the idea that a large portion of the manual pointer movement can be eliminated if the interface places the mouse pointer at the point on the screen that a user is looking at. The science-fiction terminology of "warping" the mouse pointer has been used by Zhai et al. for this discontinuous pointer placing. The mouse pointer is “warped” based on a trigger that can be, for example, a manual mouse movement, or the pressing of a button [5]. For fine manipulation, including selection, the MAGIC interface requires manual use of the mouse. Fares et al. [4] suggested an improved MAGIC interface that varies the mouse sensitivity depending on the distance between the pointer and the gaze position.

We here propose the idea that the “warping” of the mouse pointer can also be adapted to mouse-replacement interfaces used by people with severe motor disabilities who cannot use a manual mouse. Head-movement-based mouse-replacement interfaces have been found to be more accurate but slower than gaze-based interfaces [10]. With this in mind, we devised the “Head Movement And Gaze Input Cascaded” (HMAGIC) pointing technique. HMAGIC “warps” the mouse pointer to the user’s gaze point on the screen, and then enables the user to control the pointer via head motion for fine manipulation. HMAGIC combines the advantages of the speed of eye movements with the accuracy of head-movement-based positioning into a single mouse-replacement interface. The contributions of our work are threefold:

1. We propose the HMAGIC pointing technique that combines head-movement-based and gaze-based interaction.
2. We incorporated the HMAGIC pointing technique into the mouse-replacement system “Camera Mouse Suite” as an alternative to the head-only mode.
3. We tested our system in user studies that found that the HMAGIC pointing is significantly faster than the head-only based technique.

2. SYSTEM OVERVIEW

HMAGIC uses both head and eye movements in a cascaded fashion to control the position of the mouse pointer. A rough location is defined by the user’s gaze, while fine adjustments are made by head movements.

The mouse pointer position is primarily controlled by head movements. The system captures a live video of the user and automatically tracks a selected feature on their face. The position of the selected feature on the initial frame is stored as a reference point. For all subsequent frames the difference between the feature position and the reference point is combined with vertical and horizontal “gain” factors to determine the pointer position (Figure 1).

To compute the mouse pointer position $mouse_t$ at time t, we use several variables: The feature position $feat_t$ in the video is compared to the feature reference point $feat_0$ at time $t = 0$. The screen reference point $screen_t$ depends on both the user’s gaze position on the screen and the previous pointer position. At time 0, $screen_0$ is at the center of the screen. We use a diagonal matrix with diagonal values gx and gy to represent the $gain$, where gx and gy are the horizontal and vertical gains, respectively. The gain determines how the difference of the locations of the facial feature (measured in camera coordinates) is scaled to compute the number of screen pixels that the mouse pointer is moved from its current position $screen_t$. The mouse pointer position $mouse_t$ at time t is then given by:

$$mouse_t = screen_t + gain(feat_t - feat_0) \tag{1}$$

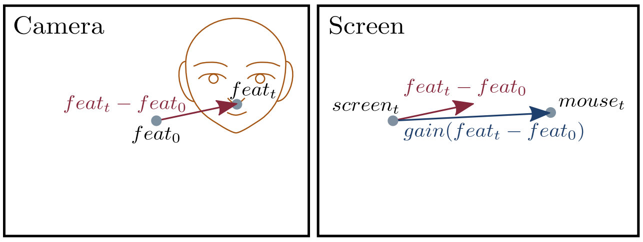

Figure 1: The position of the mouse pointer $mouse_t$ at time t is defined as the product between the $gain$ matrix and the vector $feat_t - feat_0$ dislocated by the screen reference $screen_t$ (that changes depending on the gaze position).

The HMAGIC interface uses an eye tracker to obtain an estimate for the position of the user’s gaze on the screen to set the value of screen reference point $screen_t$. Changing the value of $screen_t$ is equivalent to discontinuously placing (”warping”) the pointer to a new position on the screen. HMAGIC checks if the pointer should be warped when the speed of the head-controlled mouse pointer is beyond a user adjustable threshold (with default value of 30 pixels/second) and the direction from the pointer to the gazed position is within a given angular distance to the movement direction of the pointer (default value of 90^o). We define a region around the current gaze position. If the mouse pointer is already inside this region, it is not warped and the user keeps controlling it by head motion. Otherwise, the mouse pointer is warped, which means, $screen_t$ is updated to the current gaze position.

3. EXPERIMENT

We conducted an experiment to evaluate the performance of HMAGIC pointing and compare differences between its use with remote and head-mounted eye trackers. Three modes were tested in our experiment: mouse pointer control with the (1) head only (Head), (2) head and remote eye tracker (RET), and (3) head and head-mounted eye tracker (HMET). Thus we compared two implementations of HMAGIC (RET and HMET) with the traditional interface based only on head movements (Head).

3.1 Implementation and Apparatus

We have implemented HMAGIC pointing as an extension of the Camera Mouse Suite [2]. An off-the-shelf web camera was used to capture the head movements of the user. To compare the use of head-mounted and remote eye trackers with HMAGIC pointing, we implemented two versions of the software with an EyeTribe eye tracker and a Pupil Pro head-mounted eye tracker from Pupil Labs respectively. The choice for the eye trackers was based mainly on their low price. Since HMAGIC may be useful for people with severe disabilities, price plays an important role, and so we decided not to use more expensive eye trackers.

The experiment was conducted on a Windows 7 PC with a 2.30 GHz CPU, 4.00 GB RAM, and a 19 inch LCD monitor with a 1024 x 768 resolution. We used the Camera Mouse suite and HMAGIC systems described above. The real-time input video was captured by a Logitech QuickCam Pro 9000

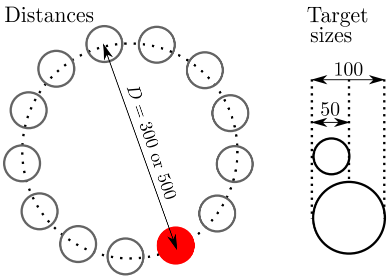

Figure 2: Experimental interface. The four possible combinations of target distances (300 and 500 pixels) and sizes (50 and 100 pixels) shown here, when grouped in a random order, define a block.

webcam with 8 megapixels.

3.2 Experimental procedure

A total of 8 volunteers (5 males, 3 females, between 24 and 31, mean 27) without motor disabilities participated in the experiment. Three participants had less than one year of experience with eye trackers and the other five had never used an eye tracker before. Unfortunately, one of the participants had significant problems with the eye trackers, possibly because of the reflections on his glasses, so his data was not considered in the results.

The experiment involved the standard pointing and selection task commonly used to evaluate the performance of a computer interface with the Fitts' law paradigm [7]. The interface is shown in Figure 2. We used 13 circular targets, arranged in a circle around the center of the screen. Two distances between circles (300 and 500 screen pixels, approximately 12 and 20 cm on the monitor) and two target diameters (50 and 100 pixels, approximately 2 and 4 cm) were tested, yielding 4 possible combinations.

To evaluate HMAGIC, we adapted the standard experimental methodology for the three modes of alternative mouse pointer control. We used the head or eye trackers to move the pointer, and a dwell-time selection mechanism to select a target. This selection mechanism requires the user to hover with the mouse pointer over the target region for a certain amount of time ("dwell time") in order to simulate a left-mouse-button click. As soon as the user selects a target, a new one is highlighted. If the click occurs outside the target, the error is tabulated and the experiment continues by highlighting the next target.

During the experiment, the participants were sitting in a comfortable position at approximately 60 cm from the monitor. For fine motion control of the mouse, we used the traditional Camera Mouse to track the tip of the participant's nose. The dwell time threshold was set as 750 ms. Participants were asked to do the experiment under the three modes described above. The order of the modes was balanced according to a Latin square pattern. Each mode contained two blocks, and each block was defined as the group of the 4 possible combinations in a random order. Every combination had 13 trials, which required the participant to move the mouse pointer from one highlighted target to another and click.

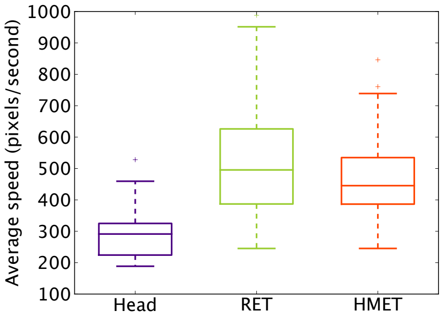

Figure 3: Movement speed averaged over the trials for each experiment participant, mode, target size and distance.

Table 1: Average movement speeds in pixels/s and standard deviations (in parenthesis) over participants for every mode, distance (300 and 500 pixels) and diameter (50 and 100 pixels).

	Head		RET		HMET	
	300	500	300	500	300	500
50	231.12 (37.38)	322.99 (58.42)	407.01 (102.59)	599.98 (186.21)	364.65 (63.08)	548.16 (155.19)
100	247.11 (61.65)	364.71 (69.31)	435.67 (103.31)	649.85 (186.09)	400.09 (69.30)	595.01 (148.18)

For quantitative measurement, we computed the movement speed as $D/(ST - 0.75\text{ s})$, where D is the distance in pixels between the previous and final pointer positions in the current trial, and ST is the time in seconds between presentation and selection of the target. We did not consider the 0.75 s dwell time when computing the speed. After the experiment, the participants were asked about their subjective feedback on speed, comfort, and ease of use of the different interfaces they tested. They were also encouraged to provide additional comments.

3.3 Results

We recorded the movement speed for each mode, target distances, and target diameter and averaged for all experiment participants over their trials. The average movement speed for the three modes is shown in Figure 3.

The mean and standard deviations of the movement speeds over all participants for every combination of mode, distance and diameter are reported in Table 1. Most notable, the average speed for both HMAGIC implementations is higher than the average speed for the head-only mode.

With a three-way ANOVA we found significant main effects of mode ($F(2, 12) = 28.97$, $p < 0.01$), target diameter ($F(1, 6) = 215.9$, $p < 0.01$) and distance ($F(1, 6) = 204.97$, $p < 0.01$) on the average movement speed. We also found a significant interaction of mode and diameter ($F(2, 12) = 27.08$, $p < 0.01$), mode and distance ($F(2, 12) = 26.21$, $p < 0.01$), distance and diameter ($F(1, 6) = 204.72$, $p < 0.01$), and between all three factors ($F(2, 12) = 24.45$, $p < 0.01$).

A Tukey-Kramer's pairwise comparison revealed significant differences between the average movement speeds for the head-only mode and the other two modes with the eye trackers ($p < 0.01$) for both target diameters and distances. No significant difference between the two eye tracker-based modes was found ($p > 0.05$).

Table 2: Results from the questionnaire about the perception of speed, comfort, and ease of use of the three modes. Participants were allowed to choose more than one mode for each question.

	Fastest	Most comfortable	Easiest
Head	1	1	2
RET	6	6	6
HMET	4	1	1

The responses of the 7 participants to the questionnaire regarding the perception of speed, comfort, and ease of use are summarized in Table 2. The remote eye tracker mode (RET) was preferred by most participants.

One of the participants added: "it was sometimes hard for me to use the Camera Mouse [head-only mode] specially for longer paths." Also, during the experiment, one of the participants commented that HMAGIC required much smaller head movements so she didn't have to deviate much from the center of the monitor. Another participant indicated that there might be a learning effect: "Eye tracker methods take some time to get used to, but the speed picks up once you do. However, I felt that they might be slightly less accurate." This accuracy problem was also noted by another participant.

4. DISCUSSION

In this paper we presented the Head Movement And Gaze Input Cascaded (HMAGIC), a novel technique that combines the accuracy of head movements and the speed of gaze movements to position the mouse pointer. The results show a statistically significant improvement on the pointer movement speed using HMAGIC compared to a head-movement-based mouse-replacement interface. This result is reinforced by the overall subjective feedback of improved speed compared to the head-only mode. HMAGIC also requires smaller head movements compared to the head-only mode, which is likely less fatiguing, especially for people with disabilities. Also, it does not require the users to keep their head in uncomfortable positions. We will test these hypotheses in future experiments with people with disabilities.

The statistical analysis indicated a significant effect of target sizes and distances on the movement speed for all modes. So, even if HMAGIC is considerably faster than the head-only mode, the sizes and distances of the targets should still be considered when designing interfaces.

The responses from the questionnaire seem to indicate an overall preference for the remote eye tracker over the head-mounted. We can interpret these results as an indication that the loss of accuracy due to head movements when using the remote eye tracker is not noticeably affecting the overall experience. Considering this, a possible future direction is to use a single off-the-shelf camera to both track the feature of the head and estimate the gaze point, thus eliminating the need for a specialized eye tracker for this application.

A possible reason for why two participants perceived H-MAGIC pointing as slightly less accurate than the head-only mode is the fact that, when the pointer is warped, it sometimes overshoots and the user has to bring the pointer back to the desired position. When using the head-only mode the user is able to just move in a straight line to the target. We will investigate the possibility of using a more conservative warping by stopping before reaching the target and letting the user finish the movement towards the desired position. Another possibility is to, instead of warping the pointer, explore variable *gain* values that depend on the distance of the pointer to the gazed position, similarly to Fares et al. [4].

5. ACKNOWLEDGMENTS

We thank the volunteers who experimented with our system and acknowledge partial support of this work by the National Science Foundation, grants IIS 1337866 and 0910908, and the São Paulo Research Foundation (FAPESP), grants 2014/12048-0 and 2011/00267-1.

6. 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 on Neural Systems and Rehabilitation Engineering*, 10:1–10, 2002.

[2] Camera Mouse Suite - Innovative software for people with disabilities. `http://cameramouse.bu.edu/`. Accessed: 2015-03-05.

[3] Dragon NaturallySpeaking. `http://www.nuance.com/dragon`. Accessed: 2015-03-05.

[4] R. Fares, D. Downing, and O. Komogortsev. Magic-sense: Dynamic cursor sensitivity-based magic pointing. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 2489–2494, New York, NY, USA, 2012. ACM.

[5] J. Hild, D. Gill, and J. Beyerer. Comparing mouse and magic pointing for moving target acquisition. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, pages 131–134, New York, NY, USA, 2014. ACM.

[6] R. J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst.*, 9(2):152–169, Apr. 1991.

[7] S. MacKenzie. Fitts' Law Software. `http://www.yorku.ca/mack/FittsLawSoftware/`. Accessed: 2015-03-17.

[8] C. Manresa-Yee, J. Varona, F. J. Perales, F. Negre, and J. J. Muntaner. Experiences using a hands-free interface. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, pages 261–262, New York, NY, USA, 2008. ACM.

[9] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 246–253, New York, NY, USA, 1999. ACM.

[10] D. Zhu, T. Gedeon, and K. Taylor. Head or gaze?: Controlling remote camera for hands-busy tasks in teleoperation: A comparison. In *Proceedings of the 22Nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, pages 300–303, New York, NY, USA, 2010. ACM.