

Camera Mouse: Dwell vs. Computer Vision-Based Intentional Click Activation

Rafael Zuniga and John Magee^(✉)

Math and Computer Science Department,
Clark University, 950 Main Street, Worcester, MA 01610, USA
jmagee@clarku.edu

Abstract. People with severe motion impairments may face challenges using assistive interface devices for common point-and-click tasks. A motion tracking interface, the Camera Mouse, allows users to control a mouse pointer with their head and click by dwelling the pointer over a target. Previous studies evaluated the use of an attached sensor (ClickerAID) as an alternative to the dwell-time clicking. However, the sensor's proprietary hardware is a barrier to adaptation. Here, we present a computer-vision based alternative that can be used to actuate mouse clicks. We conducted a preliminary evaluation of our interface and compare to previous results. Although quantitative evaluation did not achieve the same speed and accuracy as the other measures, the non-contact approach to intentional click activation demonstrates benefits compared to the other techniques.

Keywords: Mouse-replacement interfaces · Camera Mouse · Dwelling · Intentional muscle contractions

1 Introduction

The *Camera Mouse*¹ [1, 8] system has been developed to provide computer access for people with severe disabilities. The system tracks the computer users movements with a video camera and translates them into the movements of the mouse pointer on the screen. This system also provides a clicking feature with *dwell-time* selection. This involves hovering over a button for a certain period of time in order to generate a click. While this clicking approach is intuitive and easy to use for some people, it has several disadvantages for other users and for use in certain applications. Anytime the mouse stops moving, a click can be generated, potentially causing unintended selection of whatever happens to be under the link. It is hard to click small buttons or links because users have problems keeping the pointer on top of the button for the time required. Other clicking interfaces such as the *ClickerAID* [2, 7] solve the problem of inadvertent clicking but do so with an attached sensor in order to detect a single intentional muscle

¹ The Camera Mouse is freely available as a download at <http://www.cameramouse.org/>.

contraction. We present a computer vision based approach to detect intentional muscle contractions such as an “eyebrow shrug” (as in [3, 5]), an upward motion followed by a downward motion.

This paper is a follow-up to a previous study [7] that compared dwell-time selections against intentional muscle selections using an evaluation conforming to ISO 9241-9, conducted as an empirical investigation using 2D Fitts law. The method for click activation was a sensor worn in a headband by the users. In the prior study, dwell-time resulted in higher communication throughput, but intentional muscle selections were qualitatively preferred by the participants. The major downside of the intentional muscle selection was that it required specialized hardware, and that the device must be attached physically to the user, causing some discomfort. The contribution of the study now presented are (1) the development of a computer-vision based gesture clicker, and (2) an empirical investigation to compare the new computer-vision based clicker against the prior studys results.

2 Alternative Point and Click Interfaces

Users of mouse replacement interfaces perform two different tasks when using a graphical user interface. These tasks involve first positioning the mouse pointer (“pointing”) followed by selecting the user interface element under the pointer (“clicking”). Here we investigate an alternative hardware-free mouse selection technique: muscle-shrug selection. We then compare it against two other selection techniques: Dwell-Time and a single intentional muscle contraction with an attached sensor.

Our investigation is targeted for selection techniques that can be used with the Camera Mouse. The Camera Mouse provides a time-based selection technique called Dwell-Time. This technique involves hovering the pointer over a user interface element for a specified period of time in order to actuate a click (Fig. 1). Because of the time-based nature of this selection technique, there exist several issues such as the “Midas Touch” [4] problem and selecting small user interface elements.

The “Midas Touch” problem refers to the unintentional selection of any user interface element. The dwell-time technique relies on checking whether the Camera Mouse should actuate a click or not at all times. This means that even if the user is merely reading text on screen without the intention of clicking, but happens to stay still while the pointer is on top of a button, the Camera Mouse will actuate an unwanted click.

Another common problem involves trying to click small user interface elements. For the dwell-time technique to be responsive a shorter dwell-time configuration should be chosen, one to two seconds is usually best. The problem is that users might have problems maintaining the pointer on top of a user interface element long enough to actuate a click. Therefore, there are drawbacks regardless of what dwell-time configuration the user chooses. If the dwell-time configuration is too long, there is less inadvertent clicking but harder to select

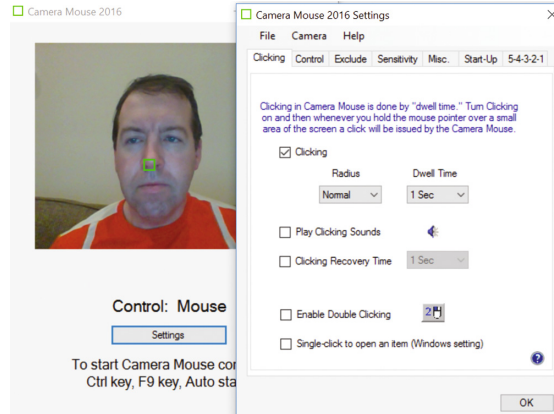


Fig. 1. Camera Mouse - tracking of a selected feature and menu system for dwell-time click configuration.

small user interface elements. If the dwell-time configuration is too short, the technique is more responsive but causes more inadvertent clicking. For other users with involuntary motions, holding the mouse still may be impossible for any period of time.

ClickerAID offers an alternative selection technique. It uses an attached sensor to detect intentional muscle contractions and actuates a mouse click when a contraction is recognized. This technique can be flexible because the user can decide what muscle group works best for him or her (e.g., eyebrow, jaw, forearm, ankle).

ClickerAID uses a Piezoelectric sensor in direct contact with the skin to measure small muscle movements. The user can choose any small muscle group that they can intentionally control. The sensor can be held in place with some elastic tape. The prior ClickAID studies tended to use a headband to hold the sensor over the brow muscle. Therefore, an eyebrow raise was used to control the clicking. The system is customizable by modifying a configurable threshold to determine when a mouse click should be simulated. The configuration screen is shown in Fig. 2. Since the system requires specialized hardware, accessibility is drastically reduced (i.e. the number of people who could easily adopt the interface).

In the next section we introduce the Muscle-Shrug selection technique that has capabilities similar to that of the ClickerAID but is completely software based.

3 Muscle-Shrug Technique

The Muscle-Shrug selection technique is a computer vision approach to a clicking in a mouse-replacement interface. This technique allows the user to select

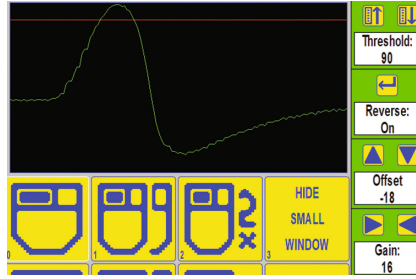


Fig. 2. ClickerAID configuration window. The signal from the piezoelectric sensor is displayed along with controls for configuring the threshold, offsets, and gains. The user can also select different types of clicking modes. Image credit Felzer and Rinderknecht [2].

two features (eyebrow, eye, jaw, chin, etc.) and actuate a click by making a “shrugging” motion with the muscle group that belongs to one of the features. Muscle-shrug selection also allows the same flexibility the ClickerAID does; the user can choose which ever pair of features work best for him or her. Furthermore, muscle-shrug selection can adapt to the user’s range of movement and to the speed of the shrug and because of this it can also adapt to the user’s distance from the camera.

Similar to the ClickerAID, the Muscle-Shrug selection technique solves the Midas Touch problem by actuating a click through an intentional muscle-gesture instead of a time based technique like dwell-time. Muscle-Shrug selection also has the advantage that performing double clicks is possible as compared against the dwell-time selection technique.

3.1 Computer Vision Clicking

Muscle-shrug selection takes advantage of the same tracking algorithm that the Camera Mouse implements, in order to keep track of the position of two features (eyebrow, eye, jaw, chin, etc.). We then define a shrug (a click actuation) as an increase in the distance between the two features followed by a decrease. This way we can detect the upward and downward motion of an eyebrow shrug or the downward and upward motion of opening and closing the users jaw. See Fig. 3.

With the users visual input, we calculate the change in distance between the two selected features across a specified number of frames. At every frame, our goal is to process N frames and calculate the average change in distance in terms of pixels of the two features being tracked across the first $N/2$ frames and the last $N/2$ frames. Where N is usually a number between eight to twenty depending on the framerate of the camera feed. If one of the features being tracked do a shrug type of motion (upward movement followed by a downward movement) then the average change of the first $N/2$ frames will be a positive number and the last $N/2$ frames will be a negative number. Then we compare these values

to a positive and a negative threshold that can be adjusted to the user. If there is ever a frame where both thresholds are surpassed, a click is actuated.

A problem that we encountered was that depending on the speed of the shrug, more than one click can be actuated from a single shrug. That issue was easily solved by setting a small time delay after the first click recognition in order to not actuate any other recognized shrugs for a small period of time. Note that the delay is not long enough to affect the users ability to double click.

Muscle-Shrug selection gives us the flexibility to adapt to the user in two different ways. It can adapt to the users mobility by adjusting the thresholds either manually or through calibration. It can also adapt to the users movement speed by varying N , the number of frames we use to perform the calculations. A higher N being better to recognize slower shrugs and a lower N being better to recognize faster shrugs.

3.2 Failure Mode

Muscle-Shrug selection has some disadvantages though. Since our algorithm depends on the tracking algorithm of the camera mouse, if the tracking of any of the two features fails, the muscle-shrug selection will not be able to perform the calculations correctly until the features are assigned again. This means that moving out of the camera, moving too quickly, or anything that will hinder the tracking will also affect the muscle-shrug selection performance.

This failure mode is the *same* as that of the Camera Mouse: loss of tracking requires manual initialization. Prior experience with Camera Mouse users “in the wild” have shown that caregivers and assistants can easily understand a basic failure mode of: reset the tracking if it is lost.

4 Preliminary Evaluation

4.1 Participants and Apparatus

We performed an evaluation of the muscle-shrug selection technique using the Camera Mouse, replicating the evaluation conditions from the previous study comparing dwell-time selection versus ClickerAID selection [7]. This is a preliminary evaluation of dwell-time selection our proposed selection mechanism here. The pointing task is done with the Camera Mouse. Five participants, two female and three males, mean age 20, participated in this evaluation.

The interface test was conducted on a laptop screen viewed from a distance of approximately 2.5 ft. The integrated camera of the laptop, with a resolution of 1280×720 , was used. The following Camera Mouse settings were used for all participants: medium horizontal and vertical gain, very low smoothing, and dwell-time click area was set to “Normal” and 1.0s. Our click actuation selection was based on movements of the jaw.

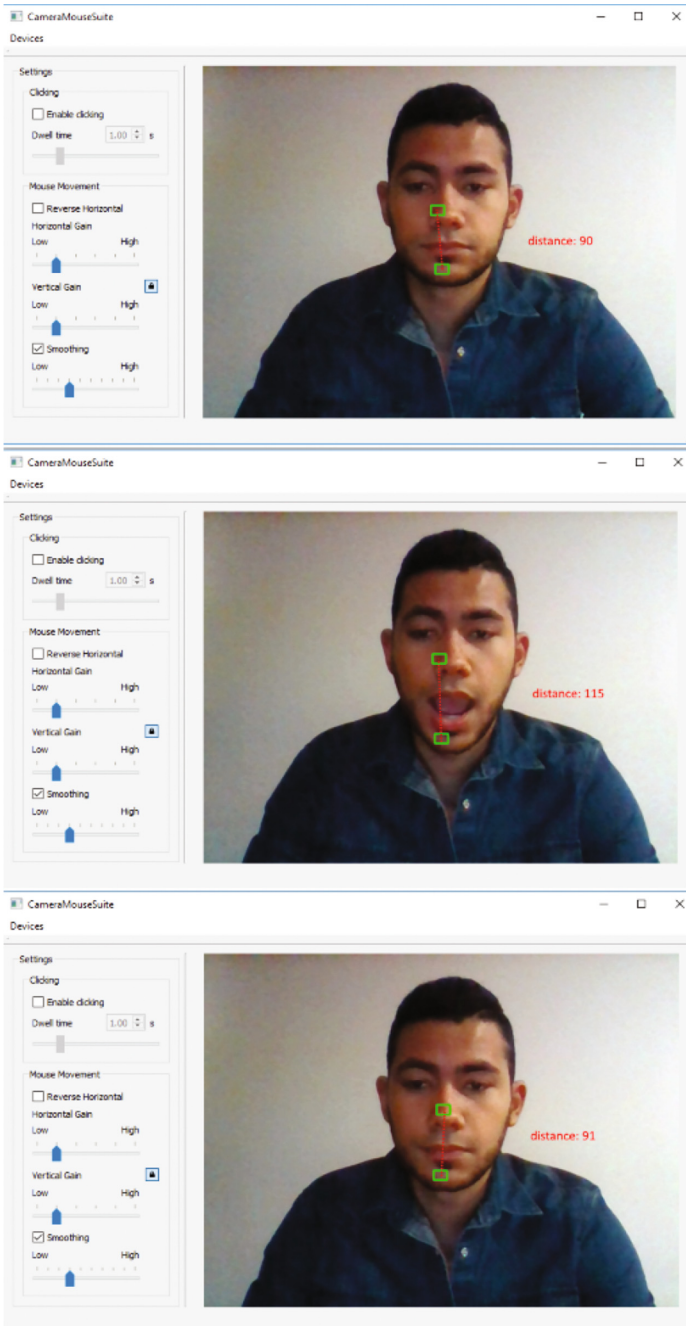


Fig. 3. Muscle-Shrug Detection - Two features are tracked with the Camera Mouse's computer vision tracking. The distance between the two features is monitored for an increase followed by a decrease. In the example above, the points start close together and move further apart as the jaw opens, then return closer together as the jaw closes. This sequence triggers a mouse click.

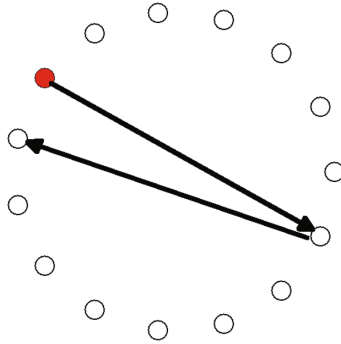


Fig. 4. FittsTaskTwo - Intended targets are highlighted in the a sequence as depicted by the overlaid arrows. Sizes and distances to targets are configurable. The software records and calculates movement time, throughput, error rates, and number of target re-entries. Trajectories of mouse movements are also recorded.

4.2 Procedure and Design

An interactive evaluation tool called FittsTaskTwo² [6] was used to perform the preliminary evaluation. Users performed repeated target selection tasks that involve first positioning the mouse pointer over a target and then selecting it with a click (Fig. 4). Log files from the tool were then analyzed to compare performance between the click modalities. Log files are also used to generate traces of mouse movements during the tests.

Each participant's session contained four sequences of thirteen targets at amplitudes 300 and 600 and widths 50 and 80 pixels. The main independent variable was input method with the following conditions:

- CM_DWELL - Camera Mouse with 1.0s dwell time,
- CM_CA - Camera Mouse with ClickerAID,
- CM_MS - Camera Mouse with Muscle Shrug.

The dependent variables were movement time (speed), throughput (speed and accuracy - bits/s), error rate (%), and target re-entries.

4.3 Results and Discussion

We report our average measurements for the CM_MS condition and compare against CM_CA and CM_DWELL previously reported. The mean movement time for CM_MS was 4284 ms versus 2226 for CM_CA and 2609 for CM_DWELL.

For throughput (speed and accuracy), the CM_MS fared worse (0.67 bits/s) compared to CM_CA (1.43 bits/s) and CM_DWELL (1.28 bits/s).

² The software is freely available as a download at <http://www.yorku.ca/mack/HCIbook/>.

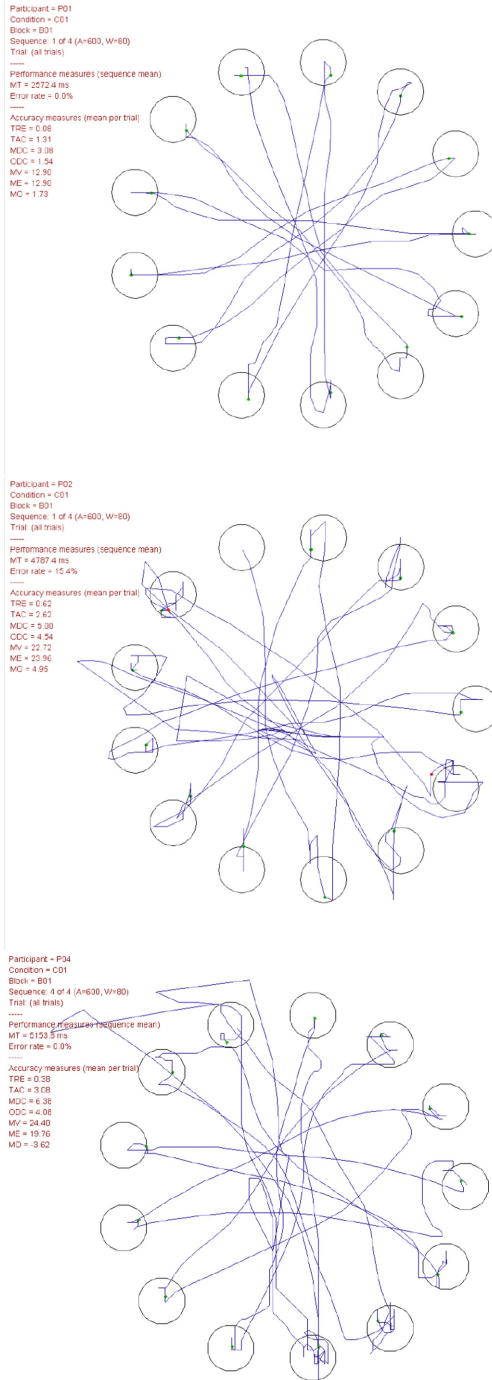


Fig. 5. Traces of mouse trajectories in target selection task.

Error rate demonstrated larger differences with means of 19.6% for CM_MS, 8.1% for CM_DWELL, and 10.8% for CM_CA.

Traces of mouse movements from three participants on the same target amplitude and width are shown in Fig. 5. The first user had more experience with the interface and his trace demonstrates more-or-less direct movements between targets and their selections. The other users were not as familiar with Camera Mouse or our selection interface - their traces show that the mouse pointer deviates significantly from the intended target trajectories. A longer study may show a learning effect and bring the performance of our system more in line with the other approaches.

In our subjective observation of the participants, we noted that many participants performed well for part of the experiment, but the tracking of one of the features drifting away from their original positions caused degraded performance. Sometimes the features would be lost completely and the tracking would have to be manually reset. This additional time was a factor in the averages reported above.

5 Conclusion and Future Direction

Our approach gives the user more control as to when the user wants to click, helping to address the Midas Touch problem. It is also more accessible for users because it does not require any hardware such as the sensor in the ClickerAID. Also, our algorithm is not limited to using nose and eyebrow. Nose and jaw actually seemed to perform better because the tracking algorithm worked better on them. Unfortunately, if the tracking algorithm fails, muscle-shrug selection will not work. At the same time though, this means that the performance of muscle-shrug selection will continue to improve as tracking algorithms get more accurate.

The muscle-shrug selection technique has room for improvements. A future direction can be to automatically recover the features being tracked if the user ever moves them out of the camera or moves too quickly.

References

1. Betke, M., Gips, J., Fleming, P.: The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Trans. Neural Syst. Rehabil. Eng.* **10**(1), 1–10 (2002)
2. Felzer, T., Rinderknecht, S.: ClickerAID: a tool for efficient clicking using intentional muscle contractions. In: *Proceedings of ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2012)*, pp. 257–258. ACM (2012)
3. Grauman, K., Betke, M., Lombardi, J., Gips, J., Bradski, G.: Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. *UAIS* **2**(4), 359–373 (2003)
4. Jacob, R.J.K.: What you look at is what you get: eye movement-based interaction techniques. In: *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI 1990)*, pp. 11–18. ACM (1990)

5. Lombardi, J., Betke, M.: A camera-based eyebrow tracker for hands-free computer control via a binary switch. In: 7th ERCIM Workshop User Interfaces for All, UI4ALL 2002, pp. 199–200 (2002)
6. MacKenzie, I.S.: *Human-Computer Interaction: An Empirical Research Perspective*. Elsevier, New Delhi (2013)
7. Magee, J., Felzer, T., MacKenzie, I.S.: Camera mouse + ClickerAID: dwell vs. single-muscle click actuation in mouse-replacement interfaces. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2015. LNCS, vol. 9175, pp. 74–84. Springer, Cham (2015). doi:[10.1007/978-3-319-20678-3_8](https://doi.org/10.1007/978-3-319-20678-3_8)
8. Magee, J.J., Epstein, S., Missimer, E.S., Kwan, C., Betke, M.: Adaptive mouse-replacement interface control functions for users with disabilities. In: Stephanidis, C. (ed.) UAHCI 2011. LNCS, vol. 6766, pp. 332–341. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-21663-3_36](https://doi.org/10.1007/978-3-642-21663-3_36)