Journal of Eye Movement Research
2(4):4,1-7

On Object Selection in Gaze Controlled Environments

Anke Huckauf
Bauhaus-Universität Weimar

Mario H. Urbina
Bauhaus-Universität Weimar

In the past twenty years, gaze control has become a reliable alternative input method not only for handicapped users. The selection of objects, however, which is of highest importance and of highest frequency in computer control, requires explicit control not inherent in eye movements. Objects have been therefore usually selected via prolonged fixations (dwell times). Dwell times seemed to be for many years the unique reliable method for selection. In this paper, we review pros and cons of classical selection methods and novel metaphors, which are based on pies and gestures. The focus is on the effectiveness and efficiency of selections. In order to discover the real potential of current suggestions for selection, a basic empirical comparison is recommended.

Keywords: Gaze control, Usability, Human–Computer–Interaction

The fascination of gaze control

Controlling computers by gaze is a fascinating undertaking, for users as well as for researchers. The user's fascination is two-fold; on the one hand, system control seems to be somehow magical, which might be due to it's intangibility. On the other hand, one gets feedback about one's own eye movements, which one is usually unaware of.

For researchers, this fact is also of interest: Since the gaze moves towards objects and areas of interest, it provides valuable information about the focus of attention and the users intentions. With intelligent algorithms, the cursor position can thus not only been adjusted on the basis of the current gaze position, but even be predicted. In such a way, gaze control can be used to pre-activate certain areas of interest, objects, or actions. This comprises the term of "*attentive input*" (Zhai, 2003). The vision of attentive input is that, by gaining information about a user's gazing behaviour, a system can react or, even better, predict his intentions and can thus provide suitable assistance.

Another important feature of gaze control is its velocity: An object can be looked at much faster than it can be reached by hand or any tool. This speed advantage might contribute also to the user's fascination. Moreover, it is clearly one advantage over any other interaction technique known hitherto.

Despite these useful qualities, the implicit nature of gazing bears also several disadvantages. The most obvious and probably also most disturbing one is the inadequateness of implicit movements for performing explicit actions, and thus for achieving explicit goals. But of course, the most important tasks in computer control require explicit control.

In this paper we provide an overview of selection metaphors which have been developed and used in the past twenty years and describe the efforts, qualities and challenges of using gaze for explicit computer control.

Task requirements: Selecting and confirming

The majority of subtasks during computer control can be divided into two main classes; namely navigation tasks and object selection tasks. Examples for navigation are

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

orienting within a desktop surface or scrolling a text. Tasks like eliciting commands or handling objects can be subscribed as object selection tasks. In current manual control (e.g., mouse, touchpad, keyboard input), deictic movements (e.g., pointing with the mouse) are mainly used for navigation, whereas arbitrary key presses are used for object selection. With other words, manually controlling a computer is based on relatively automatic deictic movements for navigation, and explicit movements like finger presses for object selection. In gaze control, navigation and object selection tasks have to be performed using the eyes as input organ.

When our eyes move over a scene, times of relatively stable eye positions (fixations) interchange with rapid movements (saccades). Information intake takes place during fixations. They last, depending on visual, cognitive, and situational settings, between 200 ms and 1000 ms. Factors determining the duration of fixations also affect their positions. The movements of our eyes are implicitly controlled. Therefore, the information they offer about the orientation of attention can hardly be absolutely controlled by the observer. This valuable attentive information is also the foundation of terming gaze control *attentive input*. The respective information includes mainly information about fixation locations and gaze paths. On this basis, one can predict future fixation locations and thus, areas of interest within the scene.

As becomes clear, the benefits of gaze input trace mainly back to performance in positional orientation, that is, in navigation tasks. For these tasks, performance can benefit a lot from the fast and reliable information provided by gaze. However, in order to select objects, explicit control not inherent in gaze control is required. With other words, the search for the optimal selection method in gaze control equals the search for the most explicit movements or gestures possible to be executed with the eyes.

Blinking for clicking

One immediately obvious solution to select objects via gaze is using blinks. Blinks, however, happen about 10 times per minute (Doughty, 2002). These frequent automatic blinks must be distinguished from the intentional blinks for object selection. In order to distinguish these types, one might use winks for selection. However, not everybody is able to wink. In addition, with head-mounted cameras, blinks and especially winks leads to movements of the face and forehead, which in turn displace the cameras and thus destroy calibration. Another solution to differentiate between intentional and automatic blinks might be the time the eye is closed: Whereas during blinks the eye is closed for about 300-400 ms (Moses, 1981), one might require longer blink durations for selection. One argument against this is that vergence changes with closed eyes so that the current fixation position is lost after a blink. Nevertheless, there are few applications working with blinks (e.g., Murphy & Basili, 1993). Unfortunately, a detailed comparison between blinking and other selection techniques is still missing. Due to the formerly described disadvantages, other gazing behaviors were tried for their suitability for object selection tasks.

Gazing for object selection: Dwell times

The standard solution for object selection in gaze-controlled systems is to use fixations on objects for their selection. But, obviously, the fixations mainly serve for inspection. This leads to the so-called Midas touch problem (Jakob, 1991), resembling King Midas, who turned everything to gold by touching. In gaze control, a certain feature of fixations, namely their duration, is used for eliciting the selection. That is, setting a certain threshold for fixation duration, the critical dwell time, is the standard solution to implement object selection in current gaze-based systems.

Dwell times can be found in the majority of gaze controlled applications. In eye typing applications, they are used to select the letters (e.g., Majaranta et al. 2004, Hansen et al. 2004). In drawing, the starting point, as well as various formatting tools, is selected using dwell times (e.g., Hornhof & Cavender, 2005). With prolonged fixations, applications are started, files are managed (e.g., MyTobii (2008)), and so on. It seems as if dwell times are the best solution for selection via gaze control.

However, dwell times bear a considerable amount of disadvantages. First of all, the Midas Touch problem is of importance. Nevertheless, it decreases in significance the more familiar the application becomes (i.e., the less inspection is required). But still, the duration of the dwell time is crucial to achieve optimal performance: Dwell times which are too short will increase the amount of unintended selections, whereas those too long will ham-

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

per users performance, since fixating the gaze at one position slows down the intake of new information as well as the execution of new actions. This let to the development of adaptation algorithms (Spakov & Miniotas, 2004; Huckauf & Urbina, submitted). However, adaptive dwell times require distinguishing intended from unintended selections, immediately and in real-time. This can only be done after lot of training of all; the user, the application, and the algorithm.

Another disadvantage of dwell times is the achievable speed of system control: Whereas eye movements and thus gaze control can be extremely fast, remaining on a certain object reduces or even destroys this advantage. Moreover, gazing at an object for a long time requires effort, and, in case of having lost the object, refixations.

From an ergonomic perspective, dwell times require a constant behavior (namely fixation) in order to change system parameters (namely, eliciting the selection of an object). Such incompatible correspondence between a behavior and its consequence should, of course, be omitted. Whereas dwell times might be easy to use for novice users because the gaze can rest on an object, for experienced users working in a familiar surround, dwelling can be annoying. Although the transition from novice to experienced usage can be handled by decreasing the threshold, behavior cannot become fluent when using dwell times.

To sum up, there are several disadvantages of selection by dwell times. But, within the last decade, the standard dwell time solution is increasingly substituted by new concepts of selection methods. Our goal is to review some of these selection metaphors. Unfortunately, empirical comparisons among various methods for object selection are still missing. Nevertheless, the knowledge about alternative metaphors may stimulate the development of more new concepts and the research about suitable eye movements.

The usage of other modalities

As alternative method for object selection, the addition of other modalities has been suggested. For example, Zhai and co-workers (1999) combined gaze control with manual reactions (key strokes), or Surakka and co-workers (2003) suggested frowning to assist gaze control. Kaur and co-workers (2003) as well as Miniotas and co-workers (2006) complement gaze control with speech. Due to the explicit nature of these other modalities, these systems were shown to provide the possibility of fast and efficient control. However, they afford an additional device. This reduces some of the advantages of gaze control; namely the idea of saving channel capacity and having ones hands free. Hence, using additional modalities must be restricted to certain settings of tasks and users.

Alternative methods for gaze controlled object selection

On- and off-screen buttons

Ware and Mikaelian (1987) suggested using buttons, which were placed on or off the screen in order to select objects. An object is selected by a fixation and a subsequent saccade towards the button. In their user study, these keys were used effectively for selection. Nevertheless, screen buttons require saccading through various objects on the screen, which are situated between the targeted object and the button. This may hamper orientation. It may also interfere with focussing on the relevant object and distracts attention from the area of interest. Taken together, Ware and Mikaelian (1987) have shown that on- and off-screen buttons require additional dwell times on the objects in order to work effectively. A similar selection method was also developed for eye typing (Urbina & Huckauf, 2007; Iwrite). This time the letters were positioned along the border of the display. The area between the letters and the border served as selection button. By this placement, the previously mentioned unintended selections were avoided. However, this eye typing layout was inferior relative to one based on dwell time selection and also relative to one based on pEYEs (see below).

Gaze gestures

Gaze paths can also be used as method for selecting objects, and are then mostly referred to as gaze gestures. Gaze gestures are sequences of fixation locations, which are not necessarily coupled to dwell times. As an easy example, one may go from the object of interest to a certain location and back to the object. This resembles the screen buttons discussed above. But, more complex paths can also be used. Here, a certain gaze gesture requiring several steps (saccades towards various locations on and off the screen to elicit and confirm differential com-

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

mands) is suggested for selection. But, given that a scan path means that various saccades (and in consequence fixations) have to be carried out, and that each fixation costs from 150 to 600 ms of time (Duchovski, 2003), it is unclear whether such a scan path can indeed be seen as a better alternative relative to dwell times. That is, a gaze path may consist of maximum four fixation locations to be able to compete with the dwell time approach. The main pro of gesture-based selection is its insensitivity for tracking accuracy and its spatial resolution independency.

Wobbrock and co-workers (2007) suggested a text entry application, Eye Write, which is based on gaze gestures, inspired by the work of Isokoski (2000) who introduced a gesture like text entry system based on off-screen targets. In Eye Write, each letter is defined by certain points within a standard letter box. If, for example, a user wants to type an A, the gaze has to follow the corners of an A in a certain order. Such a gaze path defines an object and does not need a further selection. It thus can be run without dwell times. As the data of Wobbrock and co-workers show, this method functions effectively. Of course, it doesn't belong to the fastest selection methods.

Drewes and Schmidt (2007) investigated which gestures are appropriate for gaze based computer interaction. They reported times over 3600 ms to perform a simple gesture over a square, starting on the top-left, to the bottom-right, top-right, bottom-left, finishing on the stating point using helping lines. For the same gesture over 4400 ms on a blank background, taking in mean 557 ms for each segment. Only five of nine participants where able to perform the most difficult gesture, which consisted on six segments. That is, gestures involve not only a high cost of time, but they also require large training until users are able to perform certain gestures.

Anti-saccades

Huckauf and Urbina (under review) examined the usability of anti-saccades. The underlying idea is that anti-saccades are the only eye movement, which can be considered as an explicitly controlled movement. Usually, saccades target a certain stimulus. For anti-saccades, however, the targeted area is directly opposed to the stimulus. That is, a stimulus in three degrees to the left of a current fixation requires a saccade of three degrees to the right. As their user study showed, although efficiently working for some users, the effectiveness of anti-saccades was below that of dwell times. And although some experienced users favored selection by anti-saccades, its application will be restricted to certain situations: Several novices did not like this method. Moreover, it requires space on the screen since the stimulus as well as the target area have to be unoccupied by further stimuli.

pEYEs

For mouse and stylus control, marking menus have been shown to provide a useful alternative to pull down-menus. Marking menus are circular menus, which have the advantage of comparable distance from the centre of attention. In addition, due to their circular arrangement, they increase in size towards their outer border. This is especially useful for gaze controlled applications in which spatial accuracy is always a critical factor. Due to the form, marking menus are referred to also as pie menus. In mouse or stylus control objects, which are contained in the pie slices, are selected by key presses. In gaze control, one might simply use dwell times in order to select a certain slice (e.g., Kammerer et al., 2008). Besides, Urbina and Huckauf (2007, s.a. Huckauf & Urbina, 2008) suggested a certain area of the slice, namely its outer boundary, for selection. This can speed up selection since waiting times become negligible.

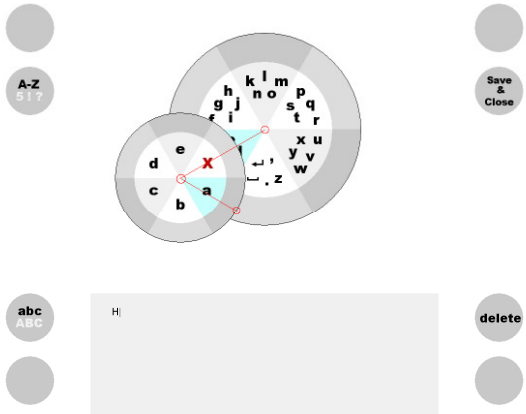


Figure 1: pEYEdit. Selection of letter "a".

This selection type have some similarity to gaze paths or gestures since the enable users to mark ahead (Kurtenbach & Buxton, 1993) the path towards the intended action (typing a letter, copying a file). However, the gaze path is independent from a fixed location on the screen but relies on relative movements from the current viewpoint. Combining gesture selection with pEYEs can lead to an accurate, fast, tracking resolution and spatial inde-

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

pendent gaze control modality and finally a reliable alternative to dwell time selection.

For novice users, pEYEs provide enough time that can be spent for searching and investigation. As has been shown, this pEYE concept seems to work well within a text entry application (Urbina & Huckauf, 2007). However, a direct comparison between pie menus to be handled by dwell times or by selection areas is still missing.

Special selection methods

In eye typing, there is one application offering a very interesting selection method: In Dasher (Ward et al., 2000), letters are presented vertically in alphabetical order on the right side of the screen. In order to select a letter, the user points (gazes) to the desired letter. The interface zooms in, and the letter moves towards the centre. The letter is selected when it crosses a vertical line placed in the middle of the application's window. Gazing to the right from the vertical line leads to a selection of the target. By gazing left, leads deleting written text, causing the opposite visual effect. Writing and deleting speed is controlled by the distance between the gazing point and the vertical line. This means looking to the middle of the vertical line ceases any action. The cooperation of vertical navigation and horizontal selection leads to a very smooth motion.

A word completion algorithm increases probable items and brings them into the centre of the screen. When both, the user and the word completion algorithm, are sufficiently trained, typing with Dasher becomes fast and fluid, feels comfortable and even fun (Tuisku et al. 2008). Nevertheless, transferring Dasher onto other applications – especially those requiring more objects and more different selections - seems to be nearly impossible.

Towards an optimal selection method in gaze control

Theoretically, a decision amongst the different kinds of selection methods in gaze control seems scarcely possible. This is mainly due to the unknown effect sizes of disturbances, disadvantages and advantages. But also, the task affect the usability of selection methods. This can be seen in Dasher (Ward et al., 2000) with its uncommon selection method. Moreover, as has been pointed out by Majaranta and co-workers (2006), the feedback given after eliciting a command is of importance for the efficiency of selection. The optimal kind of feedback may depend on the task. In addition, it may vary with the selection method used. Therefore, not only the method, but also the circumstances (task, feedback provided, experience of users) should be taken into account when evaluating certain methods.

As the current state of work shows, selection by dwell times works in various application and settings. Although it is certainly of disadvantage in several aspects, dwell times have been demonstrated to be an effective method for object selection in gaze controlled systems. This holds especially when dwell times could be individually and task-dependently adapted – what is still difficult to achieve. One reason for the superiority of dwell times might be that dwell time selection is based on fixations. Since the fixated object is directly projected into the fovea and can thus be discriminated and identified without effort, controlling the fixation location is somehow possible despite the implicit nature of gaze movements. This factor might also be one severe disadvantage for gaze paths, which are visualized with small objects only: Since saccades cannot be controlled, the correct gaze path has to be validated at each fixation. Hence, gaze paths, even if they work effectively, might be assumed to function not as efficient as dwell times. Anti-saccades, which are assumed to be explicitly controlled eye movements, bear the same problem: Since there is no object to validate the correctness of a saccade, their efficiency should not be large.

A detailed empirical comparison between various types of selection is still missing. The crux of such a user study is, however, that some of the most interesting methods summarized above can be used effectively and efficiently only within a certain application. Therefore, for future designers, one should strongly recommend to take a user-centred approach in order to develop applications, which result in fluent oculo-motor behaviour. One challenging criterion, not only for selection methods in gaze control, will be the usability for novice as well as for experienced users. In this respect, adaptive procedures and pEYE menus seem to be promising tools.

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

References

Doughty, M.J. (2002). Further assessment of gender- and blink pattern-related differences in the spontaneous eyeblink activity in primary gaze in young adult humans. *Optom Vis Sci*, 79 (7), 439-447.

Drewes, H. & Schmidt, A. (2007). Interacting with the Computer using Gaze Gestures. *Proceedings of Human-Computer Interaction - INTERACT 2007*, 11th IFIP TC 13 Int. Conference, Part II, Springer LNCS 4663, September 2007, 475-488.

Duchowski, A. T. (2003). Eye Tracking Methodology - Theory and Practice. *Springer*, 2003, p 49.

Hansen, J. P., Torning, K., Johansen, A. S., Itoh, K. & Aoki, H. (2004). Gaze typing compared with input by head and hand. *Proceedings of the 2004 symposium on Eye tracking research & applications*, ACM ETRA 04, 131-138.

Hornof, A. J. & Cavender, A. (2005). EyeDraw: enabling children with severe motor impairments to draw with their eyes. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM CHI '05 161-170.

Huckauf, A., Urbina, M. H. (2008). Gazing with pEYEs: Towards a universal input for various applications. In *Proceedings of the symposium on Eyetracking research & applications*, ACM ETRA 08, 51-54.

Huckauf, A., Urbina, M. H. (under revision). Improving object selection in gaze controlled systems.

Isokoski, P. (2000) Text input methods for eye trackers using off-screen targets. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA'00)*, pp. 15–22.

Jacob, R.J.K. (1991). The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. *ACM Transactions on Information Systems,* 9, 152-169.

Jacob, R. J. K. (1993) Interaction Techniques: Toward Non-Command Interfaces In H.R. Hartson and D. Hix (Eds.) *Advances in Human-Computer Interaction*, Ablex Publishing Co., 1993, 4, 151-190.

Kammerer, Y., Scheiter, K. & Beinhauer, W. (2008). Looking my Way through the Menu: The Impact of Menu Design and Multimodal Input on Gaze-based Menu Selection. *Proceedings of the symposium on Eyetracking research & applications*, ACM ETRA 08, 213-220.

Kaur, M., Tremaine, M., Huang, N., Wilder, J., Gacovski, Z., Flippo, F. and Mantravadi.C. S. Where is "it"? event synchronization in gaze-speech input systems. In ICMI '03: *Proceedings of the 5th international conference on Multimodal interfaces*, 151-158, New York, NY, USA, 2003. ACM Press.

Kurtenbach, G. & Buxton, W. (1993). The limits of expert performance using hierarchic marking menus *CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM SIGCHI*, 482-487.

Majaranta, P. & Räihia, K.-J. (2002). Twenty years of Eye Typing: Systems and Design Issues. *Proceedings of the symposium on Eyetracking research & applications, ACM ETRA 02*, 15-22.

Majaranta, P., MacKenzie, S., Aula, A., and Räihä, K. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Univers. Access Inf. Soc.* 5, 2 (Jul. 2006), 199-208.

Miniotas, D., Špakov, O., Tugoy, I., and MacKenzie, I. S. 2006. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications* (San Diego, California, March 27 - 29, 2006). ETRA '06. ACM, New York, NY, 67-72.

Moses, R. A. (1981). In Adler's Physiology of the Eye: Clinical Application; Paul L. Kaufman and Albert Alm, Editors. Mosby, St. Louis, MO.

Murphy, R.A. and Basili, A. (1993). Developing the user-system interface for a communications system for ALS patients and others with severe neurological impairments. Designing for Diversity; *Proceedings of the Human Factors and Ergonomics Society*, 2, 854-858.

MyTobbi (2008). http://www.tobii.com/assistive_technology/products/mytobii_d10.aspx. Last visited on Mai 2008.

Sibert, L. and Jacob, R.J.K. (2000). Evaluation of Eye Gaze Interaction. *Proceedings of the SIGCHI conference on Human factors in computing systems, ACM CHI*, 281-288.

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Spakov, O. and Miniotas, D. (2004). On-line adjustment of dwell time for target selection by gaze. *ACM NordiCHI*, 203-206.

Surakka, V. Illi, M., and Isokoski, P. (2003). Gazing and frowning as a new human--computer interaction technique. *Transactions on Applied Perception*, 1, 40-56.

Tuisku, O., Majaranta, P., Isokoski, P., and Räihä, K. 2008. Now Dasher! Dash away!: longitudinal study of fast text entry by Eye Gaze. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia, March 26 - 28, 2008). ETRA '08. ACM, New York, NY, 19-26.

Urbina M. & Huckauf A. (2007). Dwell time free eye typing approaches. *Communication by Gaze Interaction COGAIN 2007*, 65-70.

Ward, D. J., Blackwell, A. F. & MacKay, D. J. C. (2000). Dasher: a data entry interface using continuous gestures and language models. *Proceedings of the 13th annual ACM symposium on User interface software and technology*, ACM Press, 129-137.

Ware, C. and Mikaelian, H. (1987). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM CHI, 183–188.

Wobbrock, J. O., Rubinstein, J., Sawyer, M., & Duchowski, A. T. (2007) Not Typing but Writing: *Eye-based Text Entry Using Letter-like Gestures. Communication by Gaze Interaction COGAIN 2007*, 61-64.

Zhai, S. (2003). What´s in the eyes for attentive input. *ACM Communications*, 46, 3, 34-39.

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze Input cascaded (MAGIC) pointing. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM CHI, 246-253.

DOI 10.16910/jemr.2.4.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692