Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International conference on Intelligent Human Computer Interaction, IHCI 2014

# New solution to the Midas Touch Problem: Identification of visual commands via extraction of focal fixations

Boris B. Velichkovsky*, Mikhail A. Rumyantsev, Mikhail A. Morozov

*Department of Psychology, Moscow State University, Mokhovaya 11/9, Moscow 125009, Russia*

**Abstract**

Reliable identification of intentional visual commands is a major problem in the development of eye-movements based user interfaces. This work suggests that the presence of focal visual fixations is indicative of visual commands. Two experiments are described which assessed the effectiveness of this approach in a simple gaze-control interface. Identification accuracy was shown to match that of the commonly used dwell time method. Using focal fixations led to less visual fatigue and higher speed of work. Perspectives of using focal fixations for identification of visual commands in various kinds of eye-movements based interfaces are discussed.
© 2014 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Scientific Committee of IHCI 2014.

*Keywords:* visual commands; gaze control; Midas touch problem; focal fixations; attention depth; visual fatigue

## 1. Introduction

Eye-gaze interfaces are human-computer interfaces in which interaction is mediated by the direction of user's gaze. The development of such interfaces is made possible by the rapid developments in the field of contactless eye movements' registration technology. Today, such interfaces begin to be used in industry, healthcare, and science. Given that eye-movements are an important element of human perceptual-action cycle using them as a human-computer interaction technique has the potential to increase the naturalness of interaction thus making it easier for the users to concentrate on achieving their goals.

* Corresponding author. Tel.: +7-495-629-5795; fax: +7-495-629-5810.
*E-mail address:* velitchk@mail.ru

1877-0509 © 2014 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Scientific Committee of IHCI 2014.

Eye-gaze based interfaces can be divided in two classes. Gaze-control interfaces use the gaze as a pointing device. Gaze direction is used to select an interface element and to activate the function related to it. In such interfaces the command paradigm is used in which the user issues explicit commands. Attention-sensitive interfaces implement the non-command interaction paradigm. Such interfaces trace the direction of user's gaze and other characteristics of user's eye activity in order to detect and to support user's intentions. Both types of eye-gaze based interfaces face the problem of detecting user's voluntary commands and intentions with the help of formal rules.

### *1.1. Midas touch problem*

The central problem in the development of eye-gaze interfaces is the so called Midas touch problem (Jacob, 1995). If a gaze-directed interface is realized straightforwardly then each fixation on an interface element will lead to its activation even when the user has no such intention. Two approaches to the solution of this problem have been proposed. One is to use an explicit movement as an indicator of user's intention to issue a command. Gaze is used to select but not to activate an interface control. A typical example is to use voluntary blinks as such a command indicator. The drawbacks of using voluntary blinks are obvious. They lead to high discomfort of the user, and involuntary blinks may case false alarms. Other alternatives (facial muscle movements, Tiusku et al., 2012; electroencephalographic correlates of imagined movements, Lee et al., 2010) are even less suitable – the detection of target states is error-prone and requires the placement of sensors on the user.

The other method to solve the Midas touch problem is to measure the total time user's eye rests within an interface element ("dwell time"). If dwell time exceeds some threshold value, the element is activated. The threshold is selected to be larger than the duration of typical fixations. Typical values for the threshold are between 500 ms and 1000 ms, but dwell times of 1500 ms have been reported (Majaranta & Räihä, 2007). This approach requires the user to voluntary fixate the gaze at the intended interface element for a duration which exceeds the duration of natural fixations. This slows user's work and leads to user's discomfort and eye strain. The drawbacks of both approaches to solve the Midas touch problem justify the search for other solutions. An alternative solution may be provided by the distinction between ambient and focal fixations.

### *1.2. Ambient and focal fixations*

The distinction between ambient and focal fixation is based on the distinction between two visual processing mechanisms (processing streams) – the ventral stream and the dorsal stream. The ventral stream goes from the visual cortex to the posterior occipital cortex and is responsible for the localization of objects. The dorsal stream goes from the visual cortex to the inferior temporal cortex and is responsible for the identification of objects. The dorsal stream is the basis of environment perception (ambient perception), which includes rapid object localization without their identification. It is also responsible for programming and control of actions during their execution. The ventral stream is the basis for object identification and subsequent interpretation of visual information and enables the programming of actions while they are not yet executed. This perception modus is also the basis for the "vision for consciousness" – the detailed perception and identification of objects.

The work of each processing stream is characterized by specific activity of the eyes. Ambient perception requires quick estimation of objects' locations. This leads to high amplitude saccades exceeding the parafoveal area (4-5 degrees) and short fixations which do not allow identification of objects. Focal perception allows object identification. For object identification long fixations on the object and low amplitude saccades not leaving the parafoveal area are needed. The analysis of eye movements makes it possible to determine the actually dominating modus of visual perception in real-time.

In a study of road situation analysis by Velichkovsky et al. (2002) three clusters of fixations were found: very short fixations preceded by high amplitude saccades (correcting fixations), fixations with duration between 90 and 140 ms preceded by high amplitude saccades exceeding the size of parafoveal area (ambient fixations), and fixations with durations over 140-200 ms followed by saccades not exceeding the size of parafoveal area (focal fixations). These fixation clusters differ in duration: most of the correcting fixations lasts for about 60 ms, most of the ambient fixations lasts between 100 ms and 250 ms, and most of the focal fixations lasts are longer that 280-300 ms. Similar classification of fixations can be found for the perception of static scenes.

Long fixations can be considered as focal fixations accompanying conscious visual processing. We propose to apply the criteria used to define focal fixations for the identification of visual commands in gaze-control interfaces. This approach allows to get rid of short ambient fixations which are related to spatial orientation, reducing the number of false alarms. On the other hand, the duration of focal fixations is usually shorter than typical dwell times. This will allow to reduce user's discomfort. Because specific duration of focal fixations depends on specific task and user's characteristics, the effectiveness of this solution can be increased if individual criteria of defining focal fixations are used. Below, the suitability of this approach is assessed in two experiments. In these experiments focal fixations are identified on individual basis and used to drive the selection of items in a simple gaze-control interface.

## 2. Experiment 1

### *2.1. Method*

*Subjects*. Ten students of a Department of Psychology, 3 men, 7 female, mean age 22.5 years.

*Apparatus*. The stimuli were presented on a 19-inch LCD-screen. Viewing distance was 60-65 cm. Eye movements were registered on an Eyelink 1000 system with 250 Hz. The experiment was programmed with E-Prime 2.0 software and was synchronized with Eyelink system via a COM interface.

*Procedure*. The experiment consisted of three steps. Step 1 was aimed at getting the individual distributions of fixation durations and saccadic amplitudes in a visual search task with realistic visual stimuli. Step 2 was aimed at determination of individual parameters of fixations characteristic for user's visual commands in an imitated work with a gaze-control interface. Step 3 was aimed at checking the effectiveness of determination of visual commands based on extraction of focal fixations during the work with a gaze-control interface. The effectiveness of this method was compared to that of using a dwell time of 500 ms. The experiment lasted about one hour in total.

Step 1. The subjects were presented with 30 photographs of European cities. Subjects were asked if there was a church on each of the photographs.

Step 2. The subjects were presented with 9 number buttons arranged in a 3x3 matrix. Digits were presented auditory and the subjects had to imitate "pressing" corresponding buttons by gaze. The number of digits varied within each trial (1 or 4 digits). 20 trials were presented.

Step 3. The task used in the previous step was again used. This time the pressing of the number button by the gaze was not imitated but real. If a visual command ("press the button") was extracted, the fixated button was pressed. Two conditions were used. In one condition the visual command was identified if a fixation matched the individual criteria for focal fixation identified at step 2. In the other condition the visual command was identified if the dwell time within a button exceeded 500 ms. The identification of visual commands was accompanied by visual (changing the color of the button frame) and auditory (a click sound) feedback. Such feedback was shown to make gaze-based text typing more efficient (Majaranta et al., 2006). The order of experimental conditions was balanced across subjects. In each condition 30 trial of each type (with one or four buttons to be pressed) were presented. From the 10 subjects, 4 (2 female, 2 men) took part in this last step of Experiment 1.

### *2.2. Results*

Step 1. Individual distributions of fixation durations and saccadic amplitudes were built for each subject. General mean fixation duration was 325 ms (individual means ranged from 270 to 380 ms). General mean saccadic amplitude was 6.7 degrees (individual means ranged from 4.7 to 7.6 degrees). Fixation durations and saccadic amplitudes were z-transformed before further analysis. From the z-transformed duration data, mean amplitudes for saccades preceding and succeeding fixations of various durations were computed. The dependency of saccadic amplitude on fixation duration is presented in Figure 1a. Two types of fixations can be distinguished. Short fixations ($z<0$) are accompanied by low-amplitude preceding saccades and high-amplitude succeeding saccades. Such fixations can be interpreted as ambient fixations. Long fixations ($z>0$) are accompanied by saccades with amplitudes around individual mean values ($z=0$). Such saccades can often lie within the parafoveal area (4 degrees). These fixations can be seen as focal fixations. A simple criterion for focal fixations is therefore the individual mean fixation duration.

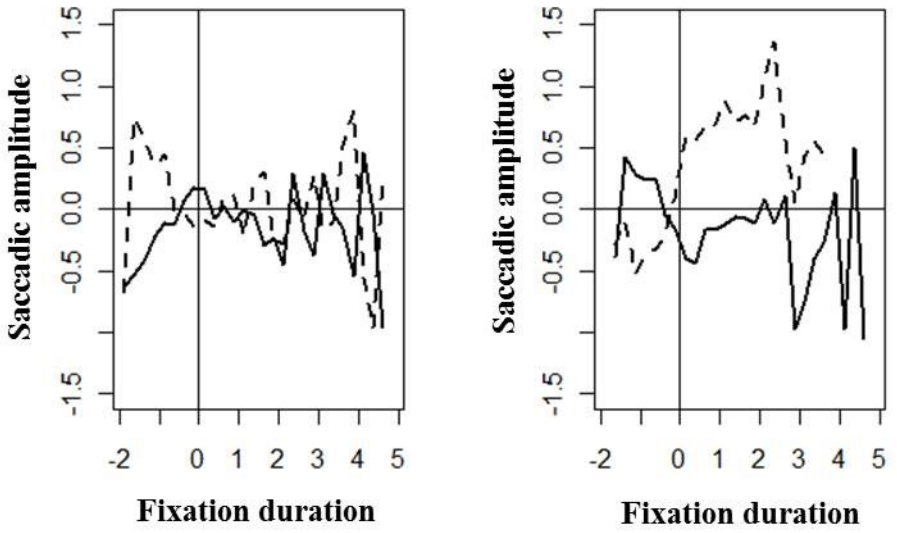


Fig. 1. Dependency between amplitude of preceding (solid line) and succeeding (dashed line) saccades and fixation durations. Z-standardized data from Experiment 1, (a) – step 1, (b) – step 2.

Step 2. During step 2 fixation and saccadic data were collected, and mean amplitudes of preceding and succeeding saccades were computed for fixations of different duration as was done in step 1. The dependency between fixation duration and saccadic amplitude for step 2 is presented in Figure 1b. Two types of fixations are clearly seen on this graph. Short fixations ($z<0$) are preceded by saccades with relatively high amplitude and are succeeded by amplitudes with low amplitude. Long fixations ($z>0$) are preceded by saccades with average amplitude and are succeeded by high-amplitude saccades probably leading outside of the parafoveal area. These fixations can be seen as focal fixations because they are probably related to conscious examining (and "pressing") a button and subsequent redirecting of the gaze towards another button.

The main task of step 2 was to find a possibility to differentiate fixations related to voluntary imitation of button press using only the criteria for identification of focal fixations. To this end all fixations were divided into target fixations (lie within any button which was a target in a given trial) and non-target fixations (lie outside of target buttons). For target and non-target fixations joint distribution densities for fixation durations and saccadic amplitudes were built with function bkde2D from package KernSmooth in the R statistical software.

In figure 2, the difference densities (target fixations minus non-target-fixations) are presented for preceding (Figure 2a) and succeeding (Figure 2b) saccades. Building a difference density makes the characteristics of target fixations more visible. As seen from figure 2a, these fixations are preceded by low-amplitude saccade which presumably lie in the parafoveal area ($z<0$). Such fixations can be short ($z<0$) or long ($z>0$). A possible function of the short fixations is to re-fixate the target button. The long fixations have been identified as focal fixations and their function can be identified with issuing the visual command to "press" the button. As seen from Figure 2b, short target fixations are followed by low-amplitude saccades thus allowing their interpretation as re-fixations. Long target fixations are followed by high-amplitude saccades. It is possible that these saccades serve to direct the gaze onto another button on the screen after a button press has been imitated. Therefore, focal fixations are specific for button press imitation. Such fixations can be objectively identified by their being longer than the individual fixation duration mean.

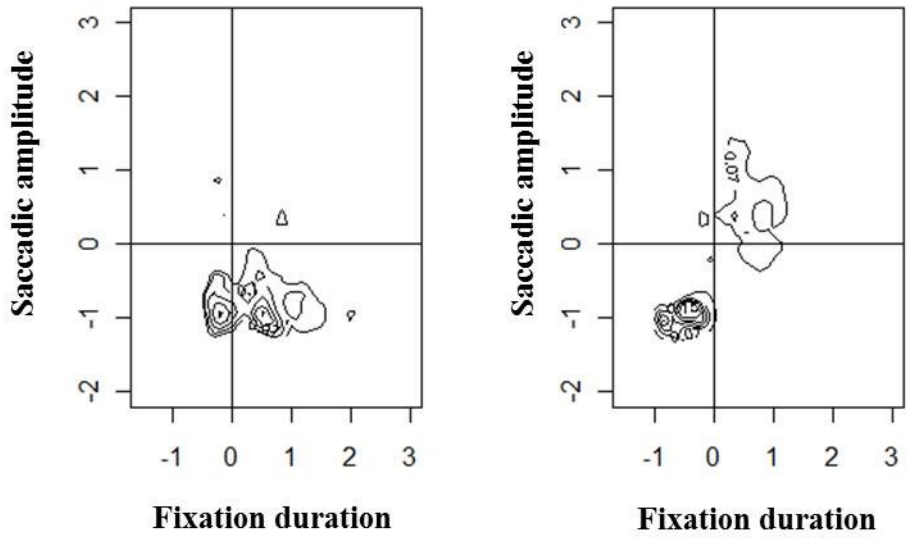


Fig. 2. Difference joint distribution density of fixation durations and saccadic amplitudes (target fixations minus non-target fixations). Only differences over 0.05 are shown. Z-standardized data from Experiment 1 (a) – preceding saccades, (b) – succeeding saccades.

Step 3. For each subject the accuracy of visual command identification was computed as the frequency of correct button presses. Mean accuracy of visual command identification by extraction of focal fixations was 97%. It was almost equal to the accuracy of visual command identification by the method of dwell time which was 98% (difference not significant, $\chi^2$ (1) = 0.21, p>0.1). The use of focal fixations as an equivalent of visual commands does not lead to more errors than when the less error-prone method of dwell-time is used.

## 3. Experiment 2

### *3.1. Method*

*Subjects*. Twenty students (15 female, 5 men, mean age 22 years) from the Department of Psychology who did not took part in the Experiment 1.

*Procedure*. Experiment 2 was aimed checking the results of Experiment 1. Additionally to estimating the effectiveness of visual command identification based on focal fixations, a visual fatigue questionnaire was given to the subjects in order to asses a possible alleviation of visual strain when the focal fixation method was used instead of the dwell time method. The experiment consisted of three steps, conceptually equivalent to the steps of Experiment 1. The experiment lasted about one hour in total

Step 1. Stimuli form a database containing 900 photographs of cities with or without people were used for the visual search task (Ehinger et al., 2009). Fifty photographs with people and fifty without people were selected and presented randomly to the subjects. Subjects had to identify the presence of people on each photographs.

Step 2. The task in this step was equivalent to the one used in step 2 of Experiment 1. Fifty trials with 4 target digits were used.

Step 3. The task in this step was equivalent to the one used in step 3 of Experiment 1. 120 trials with 4 target digits were used. There were two experimental conditions. In one condition visual commands were identified based on focal fixations' identification. In the other, the dwell time method was used for this purpose. The order of

experimental conditions was balanced across subjects (for one half of the subjects the order was focal fixations-dwell time, for another half the order was reversed). Individual means of fixation duration and saccadic amplitude obtained in step 1 were used as individual criteria for focal fixations identification. After the first condition the subjects filled the visual fatigue questionnaire and had rest for 5 minute. After that, the second condition was presented to the subjects, and subsequently the visual fatigue questionnaire was filled.

*Visual fatigue questionnaire*. The questionnaire was used to compare the level of visual fatigue when working with the focal fixation method and the dwell time method. The questionnaire contains 21 items which assess the subjective strength of visual fatigue symptoms (for example, “I clearly perceive the content of the screen”, “I feel pressure and heaviness in the eyes”, etc.). Each item could be scored between 1 and 4, and the general score of visual fatigue could be in the range from 21 to 84. Scores lower than 29 mean low visual fatigue, scores over 44 mean high visual fatigue.

*Stimuli and apparatus*. Stimuli presentation and eye movement registration was done with the same equipment as in Experiment 1.

### *3.2. Results*

Step 1. Mean fixation duration was 274 ms (individual means ranged from 270 to 380 ms). Mean saccadic amplitude was 4.4 degrees (individual means in the range from 3.3 to 5.9). Fixation duration and saccadic amplitude data were z-transformed, and mean amplitudes for saccades preceding and succeeding fixations of various durations were computed (Figure 3a). Long fixations (z>0) are preceded by saccades within the parafoveal area (z=0). They are also succeeded by saccades within the parafoveal area. For preceding saccades there is a tendency of to become lower in amplitude with increasing fixation duration, and for succeeding saccades – to become higher in amplitude. This suggests that fixations with duration exceeding individual mean duration can be seen as focal fixations related to conscious perception of potential visual targets.

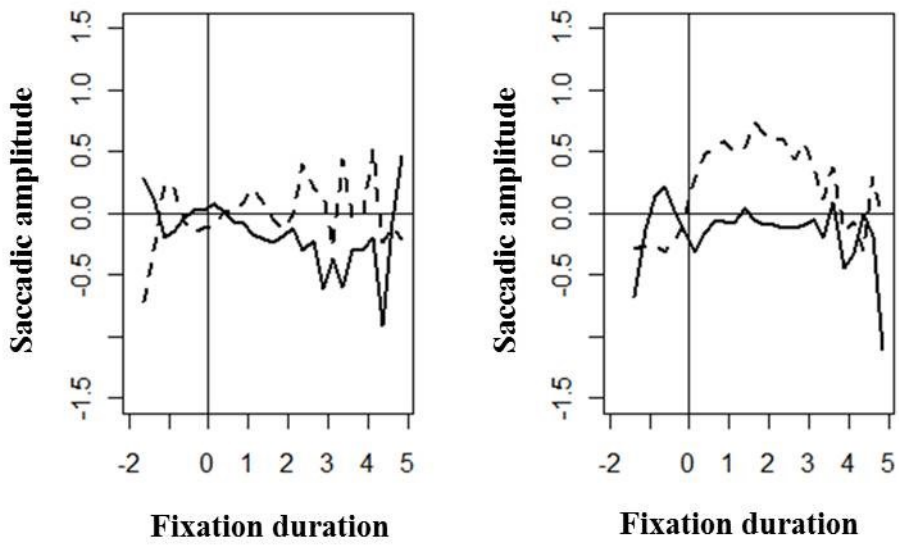


Fig. 3. Dependency between amplitude of preceding (solid line) and succeeding (dashed line) saccades and fixation durations. Z-standardized data from Experiment 2 (a) – step 1, (b) – step 2.

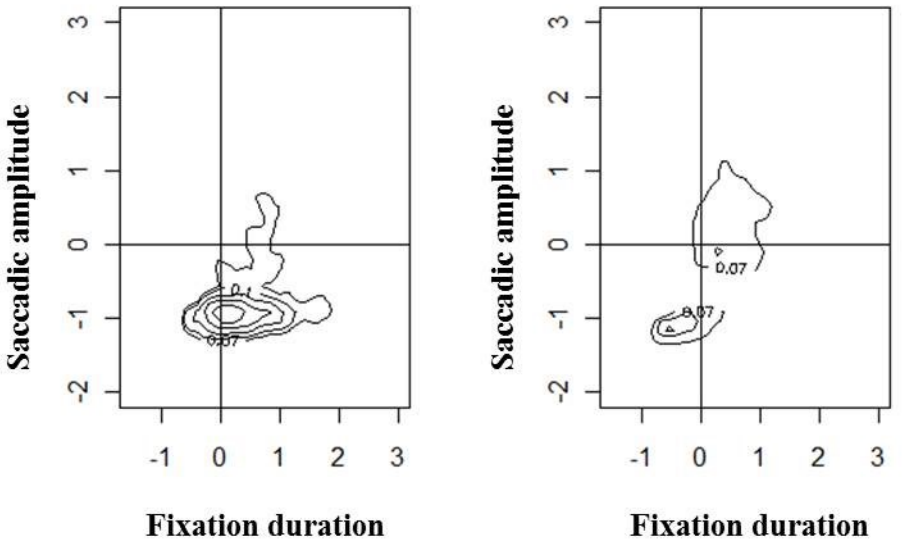


Fig. 4. Difference joint distribution density of fixation durations and saccadic amplitudes (target fixations minus non-target fixations). Only differences over 0.05 are shown. Z-standardized data from Experiment 1 (a – preceding saccades, b – succeeding saccades).

Step 2. In Figure 3b the dependency between fixation duration and saccadic amplitude is plotted. Long fixations ($z>0$) are preceded by saccades within the parafoveal area, while amplitude of succeeding saccades exceeds the size of parafoveal area. Such fixations can, again, be related to attentive (focal) perception of a target button and subsequent re-direction of gaze toward another button. In Figure 4 difference joint distribution densities (target minus non-target fixation) of fixation duration and preceding and succeeding saccadic amplitude are presented. Fixations specific for imitated button presses are preceded by low-amplitude saccade (Figure 4a). These fixations can be short (re-fixations) but also long ($z>0$). Long fixations are followed by high-amplitude saccades possibly related to the re-direction of gaze toward another button after the imitated button press. As in Experiment 1, the imitation of button presses was specifically related to focal fixations. Again, focal fixations can be distinguished by their duration exceeding individual mean duration.

Step 3. Accuracy of visual commands identification was computed as in Experiment 1. Mean accuracy of visual command identification by extraction of focal fixations was 90.7%. Mean accuracy of visual command identification by the dwell time method was 93.5%. The difference in accuracy of visual command identification was not significant statistically ($\chi^2(1) = 2.42$, $p > 0.1$). The level of visual fatigue was lower when focal fixations were used for visual commands identification (40.9±9.5) than when dwell time method was used (43.7±12.2). This difference was significant statistically ($t(19) = 2.28$, $p<0.05$).

## 4. Discussion and conclusions

The experiments showed that it is possible to perform gaze-control in a simple command interface by differentiating focal and ambient fixations. It was shown that the individual mean fixation duration obtained in visual search task is a simple but effective criterion for identification of focal fixations. These results parallel results of the studies of visual fixations distribution in viewing of natural scenes. In these studies, two clusters of fixations – focal and ambient – can also be identified. The proposed method for the solution of the Midas touch problem is thus based on a notion of psychologically sound notion that focal fixations can be associated with conscious processing.

Given the results found in our study, the applicability of visual commands identification by the mean of focal fixations extraction for the use in gaze-control interfaces (and its derivatives like brain-computer interfaces) can be estimated as high. In both experiments the accuracy of visual commands identification via focal fixations extraction was over 90% and did not differ statistically from visual commands identification accuracy obtained for the dwell time method. It was shown that visual fatigue is lower when the method of focal fixations was used meaning less visual discomfort caused in the subjects by this method. The detection of intentional visual commands based on identification of focal fixations is an effective solution to the apparently unsolvable Midas touch problem. On the one hand, the risk of false alarm is relatively low. On the other hand, this reduction in false alarms is not achieved by increasing discomfort of the user.

The proposed solution to the Midas touch problem admittedly bears some similarity to the solution based on estimating the dwell time which it is thought to oppose. However, this similarity is only apparent. As in the proposed solution, the dwell time solution makes the duration of fixations the main marker of user's intention to make a selection in a gaze-control interface. The problem with this approach is that there is no fixed fixation duration which would be indicative of user's intentions and user's state of attention. Although fixation duration is correlated with attention, the exact relationship between fixation duration and voluntary focusing of a given interface element with the intention to activate it may depend on many factors and vary both inter- and intraindividually. In the present study, we were concerned only with the interindividual variation of fixation duration as related to the attentive state of the user. Although this approach is surely also limited, it reflects the characteristics of user's voluntary attention better than any arbitrarily selected dwell time.

The present work was driven by the idea to make gaze-control interfaces more user-friendly by respecting the specifics of user's visual attention. The naturalness and high speed of work are important characteristics of gaze-based interfaces which determine their acceptance by the end users. An important special case is increasing the ease of use of gaze-control interfaces by patients with the locked-in syndrome, for whom they can be the only form of communication with the world. Above was shown that it is possible to detect intentions of the user without disturbing his/her normal eye movement activity. This opens the door for creating effective non-command gaze-based interfaces in the future.

## References

1. 1. Ehinger K, Hidalgo-Sotelo B, Torralba A, Oliva A. Modelling search for people in 900 scenes: A combined source model of eye guidance. Visual Cognition 2009;17:945-78.
2. Jacob R. Eye tracking in advanced interface design. In: Barfield W, Furness TA editors. Virtual Environments and Advanced Interface Design. New York: Oxford University Press; 1995. p. 258-288.
3. Lee E, Woo J, Kim J, Whang M, Park K. A brain–computer interface method combined with eye tracking for 3D interaction. Journal of Neuroscience Methods 2010;190:289-98.
4. Leonova AB, Blinnikova IV, Kapitsa MS. Methodology of work safety and human error research. In: De Keyser V, Leonova AB, editors. Error prevention and well-being at work in Western Europe and Russia. Dordrecht: Kluver Academic Publishers, the Netherlands; 2001. P. 105-133.
5. Majaranta P, Räihä K-J. Text entry by gaze: Utilizing eye-tracking. In: MacKenzie IS, Tanaka-Ishii K, editors. Text entry systems: Mobility, accessibility, universality. San Francisco: Morgan Kaufmann; 2007. p. 175-187.
6. Majaranta P, MacKenzie IS, Aula A, Räihä K-J. Effects of feedback and dwell time on eye typing speed and accuracy. Universal Access in the Information Society 2006;5:199-208.
7. Tiusku O, Surakka V, Vanhala T, Rantanen V, Lekkala J. Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human–computer interaction. Interacting with Computers 2012;24:1-9.
8. Velichkovsky BM, Rothert A, Kopf M, Dornhoefer SM, Joos M. Towards an express diagnostics for level of processing and hazard perception. Transportation Research, Part F 2002;5:145-156.