

# Wink Detection on the Eye Image as a Control Tool in Multimodal Interaction

Piotr Kowalczyk and Dariusz Sawicki<sup>(✉)</sup>

Warsaw University of Technology, Warsaw, Poland  
p.kowalczyk@iem.pw.edu.pl,  
Dariusz.Sawicki@ee.pw.edu.pl

**Abstract.** The selected problem of multimodal interaction for multimedia is discussed. The application for visual analysis of eye winking is presented. The proposed algorithm allows interpreting “eye gestures” as appropriate system events, which corresponds to executing system commands. The main aim of this paper is to develop replacing the usage of mouse keys by eye winking. Such a solution can work in multimodal interaction for many different applications. The work of surgeon in the operating room is a good example of situation, where busy hands do not allow controlling the computer system by traditional devices. Introduced interaction based on eye winking can help to communicate with system. The prototype has been built and set of experiments have been carried out in a group of 30 people. Tests have shown high acceptance of the proposed solution.

**Keywords:** Multimedia · Multimodal interaction · Eye winking recognition

## 1 Introduction

Multimedia covers a wide range of techniques from computer vision, machine learning, human-machine communications and artificial intelligence [1]. Computer vision and pattern recognition are treated as a tools which allow to develop devices providing multimedia entertainment. Today consumer electronics is commonly present in everyday life. However, multimedia is not just entertainment. Multimedia is also a multimodal interaction [2, 3], which is important in intelligent multimodal presentation as well as a working tool needed in the operating room. This last medical example is a specific sign of our time. On the one hand, it shows the need of using advanced graphical computer technology in a way that should be extremely simple – even intuitive. In this context effective control of a computer in the operating room [4] is one of the most difficult and also most interesting tasks of modern information technology. On the other hand this example, like others examples from multimodal interaction, breaks the tradition of combining functions in a pointing device. Since Ivan Sutherland and his sketchpad [5], the pointing device allows selecting the appropriate object in space and simultaneously generating the corresponding system event. In this way, the popular computer mouse, in addition to the ability to move the screen cursor, is equipped with keys (1 to 3). In multimodal interaction such functions can be controlled independently. This is important not only for the player, but also for the surgeon in the

operating room, who has a busy hands and looking at the computer screen would like to run a specific system event.

The main aim of this paper is to develop the simple method that allows generating system events using multimodal interaction. We focused on elements of body language which can be recognized using visual analysis.

## 2 The Task Analysis and the Main Assumptions of the Project

The typical (and known for the user) methods of generating system events is using the keys of mouse. If we assume that the hands are busy (e.g., in case of the surgeon), the control of the mouse keys can be replaced for example by movement of the head or facial expression.

The general survey of multimodal human computer interaction was presented by Jaimes and Sebe in [6], where many different methods of communications were analyzed. The image of face is used in many interactive solutions for multimedia. In the work of Mandal et al. [7] the video frames are analyzed. In the first step, the skin color is recognized, in the second the face/head pose and its movement is determined. Strumiłło and Pajor proposed a similar system based on face recognition [8]. The simple analysis of the closing and opening eyes is used in this solution. Surgeon's face analysis can be difficult due to the protective clothing required in the operating room. On the other hand, in such analysis there is a need to direct the face toward the camera or to provide a number of cameras so that in any head position one of the cameras can record the head/face properly. The analysis of the closure of the eye seems to be a good option, provided, however, that it will be made from a close distance. Using eye tracking (oculography), we could isolate also eye blinks. Al-Rahayfeh, Faezipour [9] presented a survey of methods used in eye tracking and head position detections. A survey of eye tracking methods in different methods related to multimodal interaction was presented by Singh and Singh in [10]. Such a solution could be effective and elegant in considered task. However, the use of oculography in this task seems to be not only too complicated but also unnecessary.

In some multimedia applications infrared radiation is used. Kapoor and Picard [11] described a vision based system that detects head nods and head shakes using an IR camera equipped with IR LEDs. This idea allows analyzing the image independently of the room lighting, and more importantly, regardless of the head position.

After analyzing previous work on multimedia and multimodal interaction, we decided to use winking to generate events. The proposed solution should be able to correctly replace the keystrokes of a computer mouse by winks, on the following assumptions:

- Any position of the head of the user and direction of the eyesight are possible.
- Head and face of the user can be obscured (partially or almost completely).
- The solution should be simple, fast and effective. This applies particularly to the used algorithms for image processing.
- Analysis of the stage of eye should be performed in real time in order to measure eye closure time and distinguish natural blinking from intentional winking.

Simplifying the issue, we can use a camera to “observe” the face, extract the eye image from the face image, and after that identify the pupil and iris. Then, based on the corresponding image, we can recognize winking. In order to fulfill proposed requirements and to give the ability to work under any conditions and any head position, we decided to use an IR sensor and additionally we decided to mount sensors as close to the face/eyes as possible. This way the wearable technology [12] would be applied in our solution. Traditionally always accepted element (even in the operating room) are glasses. Therefore, wink detector can be “attached” to the glasses as a wearable sensor. It should ensure the correct operation in any eyesight direction.

### 3 Introduced Technical Solution

In the proposed solution, two modules that record reflection from the surface of the eye are used. Each module consists of micro camera recording infrared radiation and IR LED. The use of infrared radiation (instead of visible light) makes the analysis independent of external lighting conditions. In addition, it prevents the occurrence of glare, which could interfere with the user’s work. Processing and analysis of images taken from a micro camera allows independently determining the state of the left and right eye. By the state of the eye to be understood one of two options – the eye is open or closed.

Cameras should be placed in such a way that they do not obscure the field of view and that they do record only image of eye – no other objects.

The quality of recorded images is a decisive parameter for the entire project, however small, low resolution cameras were also taken into consideration. We have experimentally determined that the micro camera should capture images with resolutions ranging from 4.8 thousand pixels ( $80 \times 60$ ) to 0.3 megapixels ( $640 \times 480$ ). Increasing the resolution improves the effectiveness recognition of the winking, while slightly increasing the computational cost. However, too large resolution does not improve effectiveness, only increases computational cost.

### 4 Influence of Infrared Radiation on the Eyeball

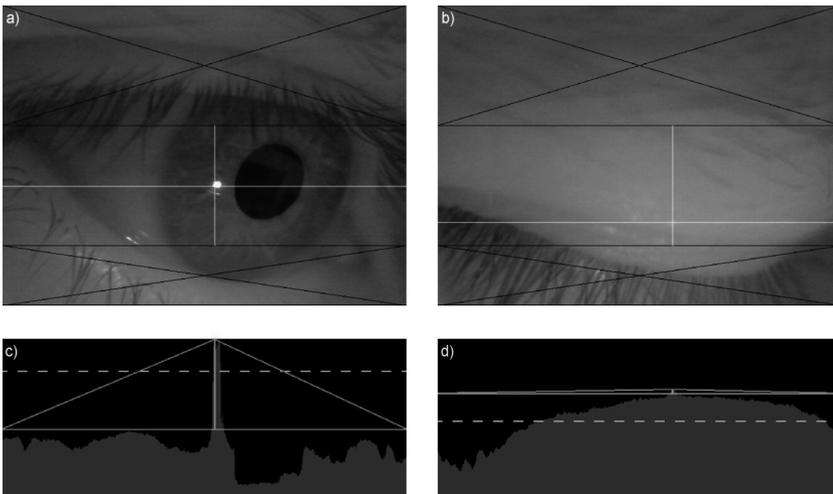
The influence of infrared radiation on the human body, especially on the eyes may be harmful. It is therefore necessary to select futures of IR LEDs that illuminate the eyes in such a way as to ensure work safety. It is a crucial condition for practical applications.

It is documented that the wavelength of the infrared radiation that is safe to the human eye should be greater than 1400 nm. Such waves do not penetrate the retina of the eye. It is further assumed that at the retinal level the emission should not exceed  $100 \text{ W/m}^2$  [13]. We have used IR LEDs with a wavelength of 1550 nm and rated power that does not exceed the permissible level. In addition, experimentally, we have repeatedly reduced the power relative to the nominal value so that it is the smallest power at which the winking detection algorithm works correctly.

## 5 Algorithm of the Eye Winking Recognition

The proposed method of determining the state of the eye is based on the analysis of the degree of diffusion in the reflection of the radiation. This method allows performing accurate analysis while maintaining low computational cost. Usage of camera in this case is a typical application of camera as a sensor for registering and analyzing reflection [14].

The light that falls on any surface, including the surface of the body, is partially reflected [14, 15]. Depending on the type of surface, this reflection may be specular (directional) or diffuse (scattered). Infrared radiation emitted by the IR LED can be reflected from the eyeball when the eye is open or from the eyelid if the eye is closed. The eyeball is characterized by a smooth, glassy surface. From such a surface the light is reflected, primarily, directionally. On the images recorded by the micro cameras in this case, IR reflection emitted by the diode, is clearly visible. (Fig 1a). The reflection has small size but is very bright. Human skin, on the other hand, has very strong diffusing properties (Fig. 1b). The eyelid is illuminated, but there is no visible bright point of reflection.



**Fig. 1.** The captured image of the right eye: (a) open (b) closed. The distribution of brightness in the horizontal line passing through the brightest pixel (c) open eye, (d) closed eye

It is therefore easy to determine the state of the eye based on the analysis of reflection. Strong diffusion means reflection from the eyelid: the eye is closed. The directional reflection from the eyeball means that the eye is open.

The algorithm for analysis of images recorded by micro cameras is implemented in three steps:

1. **Determination of the point of light reflection in the image.** This point is rarely one pixel in size – this is only possible with a very low resolution camera. Most often the point of reflection is represented as a group of bright, neighboring pixels. We analyze the difference in brightness between neighboring pixels and this way we build a group of pixels that creates a reflection area. Then, by specifying the center of gravity, we determine the exact point/pixel of the reflection. For several groups (what is possible while we notice diffuse reflection from wrinkled skin), we select the brightest group. This solution is accurate and, theoretically, allows eliminating other bright areas that could be mistakenly recognized as a reflection point. However, after conducting experiments on a set of about 5000 opened eye images and about 5000 closed eye images, we simplified the method. It turned out that simple finding the brightest pixel in the eye image, is effective enough and also much easier. The result of determining the brightest point of the image for the open eye is shown in Fig. 1c, and for the closed eye in Fig. 1d.

The position of the reflection in the image is determined by the location of the IR LED and the camera. There is no need to look for reflection in the whole image. After experiments, in order to speed up the analysis, we have limited the search area in the image. The excluded areas have been marked (crossed) in Figs. 1a and b.

2. **Determination of the brightness profile for the image.** Through the point selected in the first step, a horizontal line (cross-section) of one pixel width is carried out. For each point of this line the brightness of the image is determined. Then we generate a brightness graph on the section line. The pixel is brighter, the higher value it has and is represented in the brightness graph by a longer line. Because we analyze the image in one byte grayscale, so for each point of the section, a segment of up to 255 pixels high may appear. Examples of defined brightness profiles for the open and closed eye are illustrated respectively in Figs. 1c and d.
3. **Determination of the eye stage.** In this stage we analyze the brightness profile. The graph of specular reflection (Fig. 1c) is characterized by large differences in levels of luminance and rapid increase and decrease in luminance around the point of the highest level. In this case the eye is open. The graph for diffuse reflection (Fig. 1c) is characterized by a small differences in luminance levels and slow increase and decrease in luminance around the highest point. In this case the eye is close. The purpose of the analysis of brightness profile is only to isolate these two cases, rather than determining the reflection properties. Therefore, it is enough to test the differences in levels experimentally for many lighting environments and to specify luminance threshold. After that, using this threshold we look for local changing of luminance level around the point of maximum. Thus, the search starts from the maximum point (moving in the graph left or right) and finds the largest decrease in the neighborhood, taking into account local level differences. As the slope decreases, the process stops and the luminance level is determined for this point of the graph.

The proposed algorithm does not require complicated operations. After going through three stages, it is possible to unambiguously determine whether the user's eye is currently open or closed. The conducted experiments have shown practically 100% efficiency of determining the state of the eye in real time.

## 6 Analysis of Signals Generated from Eye Winks

In traditional control devices such as a mouse or keyboard, practically only single click (keyboard, mouse) or double click (mouse) is used. In both cases with very short time of individual click. Performing such activities using hands is not a problem, whereas analogous action with appropriate eye winking could be difficult for the user. The signals generated on the basis of winking, should be classified by the duration of the closure of the eye, and what is very important, the times of winking should be matched to the human capabilities. This will ensure the comfort of work. An additional problem is the filtering of natural eye blinks. To correctly identify winking, we have experimentally determined the longest time of natural blinking ( $t_0$ ) and the shortest time of intentional winking ( $t_1$ ).

We have separated two independent channels: for left and right eye. In each channel, the eye closure signal is recorded and its duration ( $t_x$ ) is measured. We have proposed a simple algorithm for recognizing “eye gestures” and assigning the corresponding system event to them.

- **Simultaneous winking of both eyes – higher priority event.** Natural eye blinking is characterized by simultaneous occurrence of very short signal on both channels ( $t_x < t_0$ ) – gesture  $B_0$  in Fig. 2. This case is not further analyzed. When the user closes simultaneously his both eyes intentionally for a certain period of time ( $t_x > t_0$ ), the signal  $B_1$  is shown on both channels, as shown in Fig. 2. In this case, we have assumed that the  $B_1$  gesture is used to enable/disable the detection system of winking.

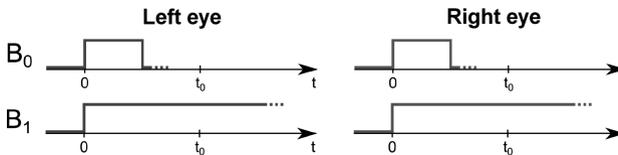
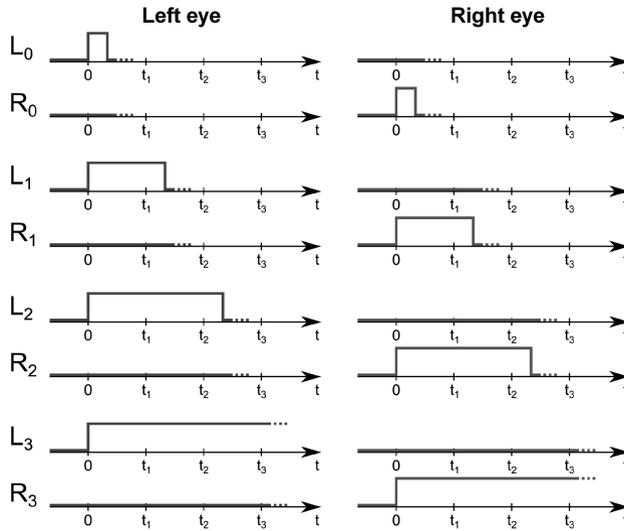


Fig. 2. Eye gestures:  $B_0$  and  $B_1$

- **Winking with one eye (the other is open) – event with lower priority.** Taking into account the time  $t_1$ , we have experimentally determined two additional times  $t_2$  and  $t_3$ , such that  $t_1 < t_2 < t_3$ . If  $t_x < t_1$  then the wink time is shorter than the shortest time of intentional wink and this case is not further analyzed. If  $t_1 < t_x < t_2$ , then there is a shorter intentional wink. If  $t_2 < t_x < t_3$ , then there is a longer intentional wink. If  $t_3 < t_x$ , then the eye is closed in intentional way (and it is not a wink!). The list of gestures for left and right eye is shown in Fig. 3.

The proposed method for time analyzing of the eye winking is the simplest possible, which gives the ability to separate up to 4 independent gestures for each eye.



**Fig. 3.** Eye gestures:  $L_0$  and  $R_0$ ,  $L_1$  and  $R_1$ ,  $L_2$  and  $R_2$ ,  $L_3$  and  $R_3$

Experiments have shown that the proposed division into gestures is acceptable by the user and allows for quick and comfortable work. We have proposed how to assign system events to recognized gestures (Table 1.); based on known system commands generated using the mouse. Of course, the gestures described, generated on the basis of one or both eyes winking, can be arbitrarily translated into commands. This may be application-dependent and may be defined at the stage of software installation.

**Table 1.** Recognized eye gestures and corresponding events

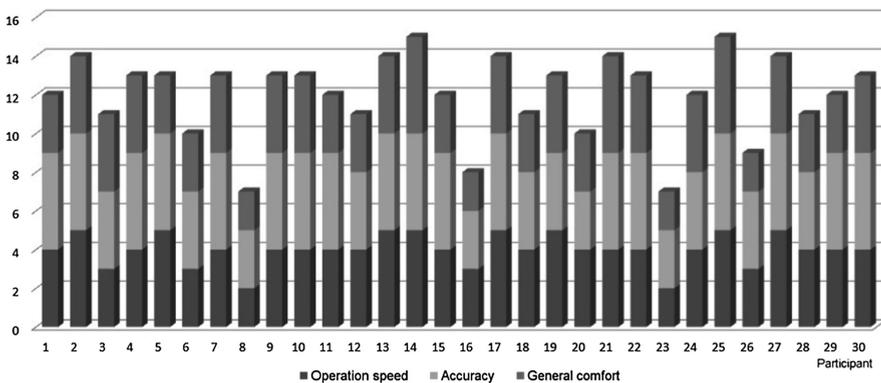
Eye gesture	Closing time ( $t_x$ )		System event
	Left eye	Right eye	
$B_0$	$0 < t_x \leq t_0$	$0 < t_x \leq t_0$	None (this prevent blink recognition)
$B_1$	$t_0 < t_x$	$t_0 < t_x$	Switch on and off (of the interaction)
$L_0$	$0 < t \leq t_1$	$t_x = 0$	None
$L_1$	$t_1 < t_x \leq t_2$	$t_x = 0$	Single click of the left mouse button
$L_2$	$t_2 < t_x \leq t_3$	$t_x = 0$	Double click of the left mouse button
$L_3$	$t_3 < t_x$	$t_x = 0$	Hold down of the left mouse button
$R_0$	$t_x = 0$	$0 < t_x \leq t_1$	None
$R_1$	$t_x = 0$	$t_1 < t_x \leq t_2$	Single click of the right mouse button
$R_2$	$t_x = 0$	$t_2 < t_x \leq t_3$	Single click of the central mouse button or scroll (double click of the right mouse button is not popular)
$R_3$	$t_x = 0$	$t_3 < t_x$	Hold down of the right mouse button

## 7 The Prototype and Tests

We have developed and manufactured the prototype. It made it possible to carry out the necessary studies on the step of analyzing problem and also at the stage of developing algorithm for winking recognition. The prototype allows also testing the whole introduced solution. We have built the multimodal interaction: as a pointing device we have used an engine for recognition of head movements. The introduced winking recognition was used as a tool for simulating clicks of mouse buttons – this way user can create proper events in operating system.

Two modules of IR LED and camera are fixed rigidly to the frame of glasses worn on the head. The modules control behavior of the user’s eyes (independently left and right eye). However frame of glasses were used only as a tool for mounting cameras and LEDs in proper places before the eyes. Cameras were connected by USB cables to computer where our algorithm allows interpreting the image of eye and turned winking into system events.

The conducted experiments have shown practically 100% correctness of the proposed solution. We have not noticed a case when a closed or open eye would be incorrectly identified. From technical point of view, the aim has been achieved. However, a much bigger problem is the acceptance of users – we deal with a completely new kind of interface. We have tested our eye control tool in a group of 30 participants. All participants used computer at work or home, but the proposed solution was new to all. Participants had to perform a set of simple tasks that are normally performed with the mouse: moving or selecting elements in the screen but using winking instead of mouse clicking. After the tests, participants were asked to evaluate the new solution. In the assessment we used methods consistent with the Standard ISO 9241-411:2012 [16]. The participants evaluated our solution using a 5-step subjective scale, in three independent subjects: Operation speed, Accuracy, General comfort (Fig. 4).



**Fig. 4.** The participants evaluation: Operation speed (from 1 = unacceptable to 5 = acceptable), the average result was 4.0 ( $\delta = 0.83$ ); Accuracy (from 1 = very inaccurate to 5 = very accurate), the average result was 4.47 ( $\delta = 0.73$ ); General comfort (from 1 = very uncomfortable to 5 = very comfortable), the average result was 3.5 ( $\delta = 0.86$ )

In the discussion, participants drew attention to the need of getting used to the new standard of communication. The high value of Operation speed (4.0) and Accuracy (4.7) shows the acceptance and correctness of our solution. Interesting, in this context, was lower level of General comfort (3.5), but the participants emphasized that this was the result of novelty of our solution and habits related to the standard mouse. One of the main problem for participants was to replace double clicking by long period of closed eye. The proposed algorithm allow for individual tuning the time conditions for each participant. The general opinion was that the solution seems to be slower, however the overall evaluation of all was positive. The participants pointed also out that there are people for whom winking at the right moment can be a very difficult task. This limitation seems to be the most serious problem of our solution.

On the other hand, the proposed method for identifying eye wink works properly in all cases. The proposed simple algorithm allowed interpreting the image of the eye and recognizing winking correctly. We have also tested the algorithm in different lighting environments. Additionally we have changed timing of events creation dependently of participant needs and we have noticed no problem in eye image interpretation. In all cases the replacement of mouse buttons by winking worked correctly.

## 8 Summary

The aim of the study was to present a new simple method for recognizing eye winking. We have used IR LEDs and cameras and built simple but effective algorithm for analysis of eye image. Initially, it was dedicated to control by head movement. However, as an independent module of interaction our solution can replace mouse buttons for creating proper system events in other situations. The introduced solution can be applicable in multimodal interaction for multimedia. As an additional control element for game players, as well as in professional systems for operating room or in control of the production line. In every situation, where operator hands are busy and cannot be used for keys click.

We have also built the prototype in which the proposed method was used. The solution was tested in different lighting environment. Simple algorithm of eye image interpretation allows selecting proper system events in any environment and with any head positions. The application shows the high usefulness of very simple method. The performed test on a group of participants showed correctness of proposed solution. The main advantage of the new solution was the ability to adjust the timing to the individual requirements of the users. For the users it is also important that the proposed solution will replace standard mouse clicks, and thus generate system events that are familiar to users.

In the future we plan, first of all, to change the connection to the computer – the wireless method will be used. In the future also, it is worth considering the possibility of automatically selecting the timing to the individual needs of users. This could be realized for example on the basis of a short test in which the user would be required to perform the winking in a predetermined manner in order to distinguish it from the standard blinking. More advanced methods (statistical model) will be also considered.

## References

1. Shao, L., Shan, C., Luo, J., Etoh, M. (eds.) *Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications* (Advances in Computer Vision and Pattern Recognition). Springer (2010)
2. Turk, M.: A multimodal interaction: a review. *Pattern Recogn. Lett.* **36**, 189–195 (2014). doi:[10.1016/j.patrec.2013.07.003](https://doi.org/10.1016/j.patrec.2013.07.003)
3. Muda, Z., Mohamed, R.E.K.: Adaptive user interface design in multimedia courseware. In: *Proceeding of 2nd International Conference on Information and Communication Technologies*. vol. 2, pp. 196–199 (2006) doi:[10.1109/ICTTA.2006.1684369](https://doi.org/10.1109/ICTTA.2006.1684369)
4. O'Hara, K., et al.: Touchless interaction in surgery. *Commun. ACM* **57**, 70–77 (2014). doi:[10.1145/2541883.2541899](https://doi.org/10.1145/2541883.2541899)
5. Sutherland, I.E.: *Sketchpad, a man-machine graphical communication system*. Ph.D. thesis MIT, January 1963. Reprinted in Technical Report Number 574. University of Cambridge, Computer Laboratory. UCAM-CL-TR-574. September (2003)
6. Jaimes, A., Sebe, N.: Multimodal human computer interaction: a survey. *Comput. Vis. Image Underst.* **108**(1–2), 116–134 (2007). doi:[10.1016/j.cviu.2006.10.019](https://doi.org/10.1016/j.cviu.2006.10.019). (Special Issue on Vision for Human-Computer Interaction)
7. Mandal, B., Eng, H.-L., Lu, H., Chan, D.W.S., Ng, Y.-L.: Non-intrusive head movement analysis of videotaped seizures of epileptic origin. In: *Proceeding of 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 28 August–1 September, pp. 6060–6063 (2012) doi:[10.1109/EMBC.2012.6347376](https://doi.org/10.1109/EMBC.2012.6347376)
8. Strumiłło, P., Pajor, T.: A vision-based head movement tracking system for human-computer interfacing. In: *Proceeding of New Trends in Audio and Video/Signal Processing Algorithms, Architectures, Arrangements and Applications (NTAV/SPA)*, 28–29 September 2012, pp. 143–147 (2012)
9. Al-Rahayfeh, A., Faezipour, M.: Eye tracking and head movement detection: a state-of-art survey. *IEEE J. Trans. Eng. Health Med.* **1**, 2100212 (2013). doi:[10.1109/JTEHM.2013.2289879](https://doi.org/10.1109/JTEHM.2013.2289879)
10. Singh, H., Singh, J.: Human eye tracking and related issues: a review. *Int. J. Sci. Res. Publ.* **2** (9), 1–9 (2012)
11. Kapoor, A., Picard, R.W.: A real-time head nod and shake detector. In: *Proceeding of 2001 Workshop on Perceptive User Interfaces*. November 2001, pp. 1–5 (2001) doi:[10.1145/971478.971509](https://doi.org/10.1145/971478.971509)
12. Calvo, A.A., Perugini, S.: Pointing devices for wearable computers. *Adv. Hum. Comput. interact.* **2014**(527320), 10 (2014). doi:[10.1155/2014/527320](https://doi.org/10.1155/2014/527320)
13. Wolska, A.: *Artificial Optical Radiation - Principles of Occupational Risk Assessment* (in Polish). Central Institute for Labour Protection - National Research Institute, Warsaw, Poland (2013)
14. Dorsey, J., Rushmeier, H., Sillion, F.: *Digital Modeling of Material Appearance*. Morgan Kaufmann Publishers, Burlington (2007)
15. Baranoski, G.V.G., Krishnaswamy, A.: *Light and Skin Interactions: Simulations for Computer Graphics Applications*. Morgan Kaufmann, Burlington (2010)
16. Ergonomics of human-system interaction—Part 411: evaluation methods for the design of physical input devices. ISO Standard: ISO 9241–411:2012 (2012)