Proceedings of

COGAIN 2006
'Gazing into the Future'

4th-5th September 2006, Turin, Italy

COGAIN NoE is funded by the EU IST 6th framework program.

(Blank)

Welcome

Welcome to COGAIN 2006! - the second conference of the COGAIN research network of excellence, supported by the European Commission, and the very first public event of its kind. COGAIN is coming to the end of its second, very successful year. One of our intentions is to create a permanent conference in the field of gaze-based communication to support disabled users, and we hope this is the first of many such events.

We have been delighted with the response to the call for contributions to the conference. We have received 32 submissions from 26 institutions, 12 of which are from outside the COGAIN network. We have received submissions from 16 countries in Europe, the US and in Asia. This conference seeks to be different from other events in the following ways - the emphasis in COGAIN and its conference is on gaze-based communication and supporting technology primarily (but not exclusively) for disabled users. In this way, we are different from events like ETRA, which covers a range of eye tracking applications beyond communication for disabled users. There are conferences that cover communication and applications for disabled users, such RENSA and AAATE, which do not specifically deal with gaze-based techniques. Another defining element of the COGAIN conference is its active involvement of user communities. This provides a valuable opportunity for industry and academia to meet and listen to the people they are building systems for. It enables users and user organisations the opportunity at first hand to see and try out the latest systems developments and to meet members of the academic research groups involved in COGAIN – and to tell us where we are doing well, and where we need to do better!.

In this years conference, we are looking to the future. What do we want to achieve in the next 5 to 10 years in the field of gaze-based communication? What are the issues we need to resolve in order to achieve these objectives? The first day of the conference will address these two questions in relation to four themes around which the paper sessions are organised. Each session will contain an informal workshop style discussion to attempt to map out a future agenda and research direction for COGAIN in these areas. We hope you will want to join in and contribute as much as you can to making this a milestone event in defining the way ahead in this important field. So let's 'Gaze to the Future' and enjoy a great conference - 'the more you put in, the more you get out…'

Enjoy the conference and enjoy Turin!

Howell Istance & Laura Farinetti
COGAIN 2006 Conference Chairs

COGAIN 2006 Program

Monday, 4 September, 2006

Politecnico di Torino - Sede "Lingotto"
Aula Magna, via Nizza 230, 10126 Torino

8:30-9:00	Registration

9:00-10:30

Session 1: Eye-tracking Systems

Chair: (to be confirmed)

Remote Eye Tracking: State of the Art and Directions for Future Development
Martin Böhme, André Meyer, Thomas Martinetz, Erhardt Barth
Institute for Neuro- and Bioinformatics, Germany

Open-Source Software for Real-Time Visible-Spectrum Eye Tracking
Dongheng Li, Derrick Parkhurst
The Virtual Reality Application Center, Iowa State University, USA

Progress on an Eye Tracking System Using Multiple Near-Infrared Emitter/Detector Pairs with Special Application to Efficient Eye-Gaze Communication
Dale Grover
Electrical and Computer Engineering Dept, Michigan State University, USA

Basics of Gaze Estimation
Arantxa Villanueva[1], Dan Witzner Hansen[2], Javier San Agustin[2], Rafael Cabeza[2]
[1]Public University of Navarra, Spain; [2]IT University, Copenhagen, Denmark

GoldenGaze: an inexpensive real-time gaze tracking system
Detlev Droege, Fabian Fritzer, Dietrich Paulus
Institute for Computational Visualistics, University of Koblenz-Landau, Germany

10:30-11:00	Coffee Break

11:00-12:30

Session 2: Multimodal Interaction with Real-World Objects

Chair: (to be confirmed)

Control Application for Smart Houses through Gaze interaction
Dario Bonino, Emiliano Castellina, Fulvio Corno, Alessandro Garbo
Politecnico di Torino, Dipartimento di Automatica ed Informatica, Italy

Direct Gaze-Based Environmental Controls
Fangmin Shi, Alastair Gale, Kevin Purdy
Applied Vision Research Centre, Loughborough University, UK

Gaze Pointing and Facial EMG clicking
John Paulin Hansen
IT University of Copenhagen, Denmark

When can eyes make up for hands?
Marcela Fejtová, Petr Novák, Olga Štepánková
Czech Technical University, Department of Cybernetics, Czech Republic

Multi-modal interface: Gaze-EEG-based system
Carmen Vidaurre[1], Arantxa Villanueva[1], Reinhold Scherer[2], Markus Joos[3], Alois Schloegl[2], Sebastian Pannasch[3], Rafael Cabeza[1]
[1]Public University of Navarra, Spain; [2]Graz University of Technology, Austria; [3]Technische Universität Dresden, Germany

12:30-14:00	Lunch (offered to registered participants)

14:00-15:30

Session 3: Attentive Interfaces and Reacting to Intent

Chair: John Paulin Hansen

Gaze-based Attentive User Interfaces (AUIs) to support disabled users: towards a research agenda
Howell Istance[1], Aulikki Hyrskykari[2], Daniel Koskinen[2], Richard Bates[1]
[1]De Montfort University, UK; [2]University of Tampere, Finland

Automatic Preference Detection by Analyzing the Gaze 'Cascade Effect'

Nikolaus Bee[1], Helmut Prendinger[2], Elisabeth André[1], Mitsuru Ishizuka[3]
[1]University of Augsburg, Germany; [2]National Institute of Informatics, Japan; [3]The University of Tokyo, Japan

An Estimation of Certainty for Multiple Choice Responses using Eye-movement

Minoru Nakayama, Yosiyuki Takahasi
Tokyo Institute of Technology, Japan

Gaze tracking for robotic control in intelligent teleoperation and prosthetics

Claudio Castellini, Giulio Sandini
University of Genova, Italy

Eye Tracker Input in First Person Shooter Games

Poika Isokoski[1,2], Benoît Martin[2]
[1]University of Tampere, Finland; [2]University Paul Verlaine - Metz, France

15:30-16:00	Coffee Break

16:00-17:30

Session 4: Users and Usability

Chair: Mick Donegan

The 'KEE' Concept for Eye-Control and Complex Disabilities: Knowledge-based, Enduser-focused and Evolutionary

Mick Donegan, Lisa Oosthuizen
The ACE Centre, UK

Gaze Path Playback Supporting Retrospective Think-Aloud in Usability Tests

Merja Lehtinen, Aulikki Hyrskykari, Kari-Jouko Räihä
University of Tampere, Finland

Recent experiences, using eye-controlled communication systems with "locked-in" patients

Jeffrey D. Morris, Gavin Cooper
Cardiff and Vale NHS Trust, Rookwood Hospital, UK

Towards Remote Evaluation of Gaze Typing Systems
John Paulin Hansen[1], Kenji Itoh[2], Hirotaka Aoki[1,2]
[1]University of Copenhagen, Denmark; [2]Tokyo Institute of Technology, Japan

Magic Key
Luis Filipe Figueiredo[1], Ana Isabel Gomes[2], João Bento Raimundo[3]
[1]Escola Superior de Tecnologia e Gestão, Portugal; [2]Agrupamento de Escolas de Penalva do Castelo, Portugal; [3]Associação Augusto Gil, Portugal

Tuesday, 5 September, 2006

Città di Torino, Servizio Passepartout
via San Marino, 10, 10134 Torino

8:30-9:00	Registration
9:00	**Welcome** from the Rector of Politecnico di Torino, from Città di Torino and from Regione Piemonte
9:30	**Opening** from Kari-Jouko Räihä, coordinator of the COGAIN Network of Excellence
9:40	**Keynote Address: When Less is More: Applying Attention to the Design of Context-Aware Computers** Roel Vertegaal – Queen´s University's School of Computing, Canada
10:25	**Technology and Disability: Industry and Research in Piemonte**, Raffaele Meo – Politecnico di Torino, Italy
10:40	Coffee Break
11:25	**Living with Eyegaze,** Angela Jansen – actress and author, Germany
11:50	**Eye Control helping People with a Disability in Torino,** Adriano Chiò – ALS Centre of Molinette Hospital in Torino, Italy
12:10	**Who can Benefit from Eye Control and how?,** Michael Donegan – ACE Centre Advisory Trust, UK

12:30	**Gaze Interactive Communication,** John Paulin Hansen – IT University of Copenhagen, Denmark
12:50	**Concluding Remarks**
13:00- 14:30	Lunch (offered to registered participants)
14:30- 18:00	**Exhibition** Users and Industry (+ a coffee break)

COGAIN 2006:
'Gazing into the Future'
4-5 September 2006
Turin, Italy
www.cogain.org/cogain2006

COGAIN 2006 Papers Index

Session 1: Eye-tracking Systems

Remote Eye Tracking: State of the Art and Directions for Future Development

Martin Böhme, André Meyer, Thomas Martinetz, and Erhardt Barth
Institute for Neuro- and Bioinformatics
Ratzeburger Allee 160
D-23538 Lübeck, Germany
{boehme, meyera, martinetz, barth}@inb.uni-luebeck.de

Keywords
Remote eye tracking, free head motion, eye model

Introduction

Recent years have seen rapid developments in the field of remote eye tracking. Whereas only a few years ago the standard in eye tracking was for systems to be intrusive, i.e. they either required the user's head to be fixated or equipment to be mounted on the user's head, systems have now evolved to the point where the user can move freely in front of the system (within certain limits), and good accuracy (1 degree or better) is achieved throughout the whole working range. This has been demonstrated by a number of commercial and academic systems, both multi-camera (Beymer and Flickner, 2003; LC Technologies, 2006) and single-camera (Tobii, 2002; SMI, 2006; Hennessey et al., 2006; Guestrin and Eizenman, 2006; Meyer et al., 2006). To clarify terms, we will use the term "remote eye tracking" here to mean a system that operates without contact with the user *and* permits free head movement within reasonable limits without losing tracking.

In this paper we give an overview of our own work in this field and give our view on where worthwhile opportunities for future research lie.

State of the Art

The first remote eye tracking systems that appeared in the literature used multiple cameras (Shih et al., 2000; Beymer and Flickner, 2003; Ohno and Mukawa, 2004; Brolly and Mulligan, 2004; Yoo and Chung, 2005), usually in some kind of stereo setup. Morimoto et al. (2002) describe a single-camera eye tracker with an accuracy of about 3 degrees. The first single-camera remote eye tracker with high accuracy (0.5 to 1 degree) and good tolerance to user movement was a commercial system (Tobii, 2002), but implementation details have not been made available. Recently, several academic groups have built similar single-camera systems (Hennessey et al., 2006; Guestrin and Eizenman, 2006; Meyer et al., 2006). (Guestrin and Eizenman's system allows only small head movements, but it appears that their well-founded approach would allow greater head movements with a higher-resolution camera.) The main additional difficulty in the single-camera setting is determining the distance of the user from the camera, since a triangulation as in the multi-camera case can not be carried out. The advantage of a single-camera system is of course the reduced cost and smaller size.

Figure 1. Remote eye tracker system setup. The eye tracking hardware consists of a single high-resolution camera below the display and two infrared LEDs to either side.

The setup of our own single-camera system (Meyer et al., 2006) is shown in Figure 1. It consists of a high-resolution camera (1280x1024 pixels) and two infrared LEDs mounted to either side of the camera. The LEDs provide general illumination and generate reflexes on the surface of the cornea. These corneal reflexes (CRs) are used to find the eye in the camera image and determine the location of the centre of corneal curvature in space. The system is shown here mounted below an LCD display.

The software consists of two main components: The image processing algorithms that are used to determine the position of the CRs and pupils in the image, and the gaze estimation algorithm, which uses this data to compute the direction of gaze.

The image processing component is based on the Starburst algorithm (Li et al., 2005), which was reimplemented and modified to fit the needs of the remote eye tracking setting. The gaze estimation component is based on a physical model of the eye (see Figure 2), which models the optical properties of the cornea (reflection and refraction), the location of the pupil centre (PC) and centre of corneal curvature (CC), and the offset of the fovea (and hence the line of sight) from the optical axis. The model contains three user-dependent parameters: the curvature radius of the cornea (r_{cornea}), the distance between PC and CC (r_{pc}), and the offset of the direction of gaze from the optical axis (α_{fovea}) (only the horizontal component of this offset is currently modelled). Given the observed positions of the CRs and of the pupil centre in the camera image, there is only one possible position and orientation of the eyeball that could have given rise to these observations. The gaze estimation algorithm deduces this position and orientation, then intersects the direction of gaze with the display plane to determine the location the user is fixating.
Preliminary measurements on the system have shown an average accuracy of 1.2 degrees; with additional fine-tuning, we expect to improve the accuracy to better than 1 degree. The system allows head movements of 20 cm between the extremes of the working range on all three spatial axes.

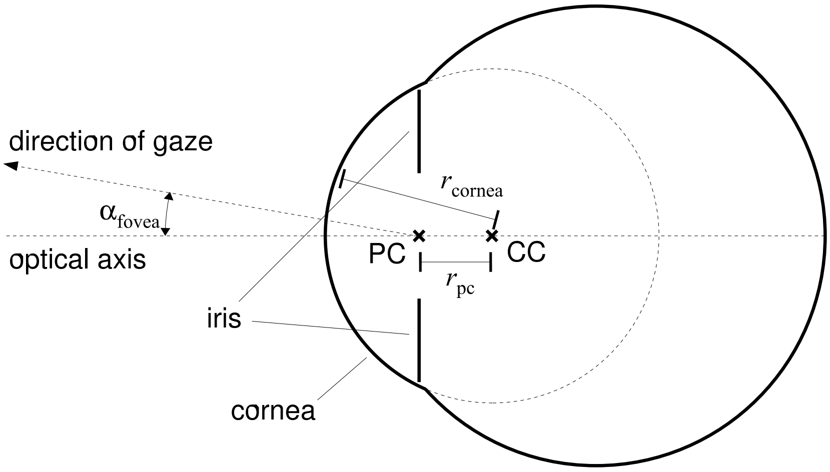


Figure 2. Eye model used in the remote gaze estimation algorithm. PC: Pupil centre. CC: Cornea centre. r_{cornea}, r_{pc}, α_{fovea}: User-dependent parameters (see text for explanation).

The image processing component is based on the Starburst algorithm (Li et al., 2005), which was reimplemented and modified to fit the needs of the remote eye tracking setting. The gaze estimation component is based on a physical model of the eye (see Figure 2), which models the optical properties of the cornea (reflection and refraction), the location of the pupil centre (PC) and centre of corneal curvature (CC), and the offset of the fovea (and hence the line of sight) from the optical axis. The model contains three user-dependent parameters: the curvature radius of the cornea (r_{cornea}), the distance between PC and CC (r_{pc}), and the offset of the direction of gaze from the optical axis (α_{fovea}) (only the horizontal component of this offset is currently modelled).

Given the observed positions of the CRs and of the pupil centre in the camera image, there is only one possible position and orientation of the eyeball that could have given rise to these observations. The gaze estimation algorithm deduces this position and orientation, then intersects the direction of gaze with the display plane to determine the location the user is fixating.

Preliminary measurements on the system have shown an average accuracy of 1.2 degrees; with additional fine-tuning, we expect to improve the accuracy to better than 1 degree. The system allows head movements of 20 cm between the extremes of the working range on all three spatial axes.

Directions for Future Development

Despite the advances in remote eye tracking systems in recent years, there are still quite a number of areas in which improvements have to be made if these systems are to see widespread use in human-computer interfaces (including, but not limited to, AAC applications). The aspects that require further progress include robustness, accuracy, ease of setup and use, and price.

In the following, we will present ideas that we intend to pursue in the future to achieve advances in the areas mentioned above:

- Tolerance towards glasses. Systems that use infrared illumination often do not work well for users who wear glasses because of reflections on the surface of the glasses. The existing systems can usually be made to work with glass wearers to a certain extent, but only for some head orientations where no interfering reflections occur. For other head orientations, the reflections can obscure the user's eyes completely, making eye tracking impossible. One way of dealing with this problem might be to use more than two infrared illuminators. At any given time, the system would use two of the illuminators. If the system detected that the user's pupils were being obscured by reflections, it would switch to a different set of illuminators at a different angle relative to the user and the camera. In this way, the reflections should shift off the eyes or even be eliminated entirely.

  To achieve high accuracy in the presence of glasses, the eye model may have to be augmented with a model of the glasses to account for their effect on the image of the eye. However, preliminary tests indicate that the accuracy is still tolerable even if the effect of the glasses is not modelled.

- Ease of setup / use. Remote eye tracking systems are typically based on a physical model of the eye, the eye tracking system (camera and illuminators), and the monitor. Because of this, they require the spatial relationship between the camera, the illuminators, and the monitor plane to be known. These measurements are usually obtained by hand, a process that is time-consuming, error-prone, and difficult for an end user to carry out. Beymer and Flickner (2003) calibrated the orientation of the monitor plane automatically using a mirror to reflect the image of a checkerboard pattern taped to the monitor back into the camera. We intend to implement a similar automatic calibration in our system.

- Price. Existing remote eye trackers typically use high-resolution industrial cameras with relatively high-grade lenses. This makes the systems quite expensive, even before labour costs for assembly are taken into account. For example, the camera and lens used in our eye tracker have a combined price of around 1000 USD. This puts the system out of reach of many potential users. An alternative would be to use webcams, but we are sceptical if their typical resolution of 640x480 pixels can deliver satisfactory results. Instead, we are confident that advances in sensor hardware will solve this particular problem and that sensors with the required resolution will soon reach consumer price points.

  Another obvious idea for cutting cost is to eliminate the infrared illuminators and use natural illumination (see e.g. Hansen and Pece (2005)), though this makes the image processing task significantly more difficult.

- 3D cameras. Recent years have seen the development of so-called 3D time-of-flight (TOF) cameras (CSEM, 2006). In addition to providing an intensity image like a conventional camera, these cameras also provide a depth image that gives the distance of the object in the scene at each pixel. This allows the three-dimensional shape of the scene, e.g. the user's head, to be reconstructed. Many recognition and tracking tasks can be implemented more robustly on 3D range data than on intensity images, and so this technology has the potential to be used for robust head and eye tracking. Two participants in COGAIN, together with other European partners, will be working on TOF-based eye, head and gesture tracking within an EU project.

Robust, affordable eye tracking technology would have a broad range of potential applications. It would of course be invaluable for AAC applications, but beyond that, eye tracking has the potential to become a new general-purpose interaction medium. Eye tracking may change the way we interact with technology and how visual information is communicated – our work on gaze guidance (Barth et al., 2006) has the goal of augmenting a video or visual display with a recommendation of how to view the information, of what is to be seen.

With the advances currently being made in eye tracking hardware and software, widespread low-cost eye tracking may finally become a reality.

Acknowledgments

Research was supported by the German Ministry of Education and Research (BMBF) under grant number 01IBC01 with acronym ModKog and by the European Commission within the COGAIN Network of Excellence. We thank SensoMotoric Instruments GmbH, Germany, for their support.

References

Barth, E., Dorr, M., Böhme, M., Gegenfurtner, K.R., and Martinetz, T. (2006) Guiding the mind's eye: improving communication and vision by external control of the scanpath, *Human Vision and Electronic Imaging*, Volume 6057 of *Proceedings of the SPIE*, pp. 116–123.

Beymer, D. and Flickner, M. (2003) Eye gaze tracking using an active stereo head, *Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition (CVPR)*, Volume 2, pp. 451–458.

Brolly, X.L.C. and Mulligan, J.B. (2004) Implicit calibration of a remote gaze tracker, *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '04)*, Volume 8, p. 134.

CSEM (2006). SwissRanger SR-3000, CSEM SA, Zurich, Switzerland.

Guestrin, E.D. and Eizenman, M. (2006) General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections, *IEEE Transactions on Biomedical Engineering*, Volume 53, Number 6, pp. 1124–1133.

Hansen, D.W. and Pece, A.E.C. (2005) Eye tracking in the wild, *Computer Vision and Image Understanding*, Volume 98, Issue 1, Elsevier, pp. 155–181.

Hennessey, C., Noureddin, B., and Lawrence, P. (2006) A Single Camera Eye-Gaze Tracking System with Free Head Motion, *Eye Tracking Research & Applications 2006*, pp. 87-94.

LC Technologies (2006) Eyegaze Eyefollower, LC Technologies Inc., Fairfax, VA, USA. `http://www.eyegaze.com`

Li, D.,Winfield, D., and Parkhurst, D.J. (2005) Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches, *Proceedings of the IEEE Vision for Human-Computer Interaction Workshop at CVPR 2005*, pp. 1–8.

Meyer, A., Böhme, M., Martinetz, T., and Barth, E. (2006) A single-camera remote eye tracker, *Perception and Interactive Technologies*, Volume 4021 of *Lecture Notes in Artificial Intelligence*, Springer, pp. 208–211.

Morimoto, C.H., Amir, A., and Flickner, M. (2002) Detecting Eye Position and Gaze from a Single Camera and 2 Light Sources, *Proceedings of the International Conference on Pattern Recognition 2002*, pp. 314–317.

Ohno, T. and Mukawa, N. (2004) A free-head, simple calibration, gaze tracking system that enables gaze-based interaction, *Eye Tracking Research & Applications 2004*, pp. 115–122.

Shih, S.W., Wu, Y.T., and Liu, J. (2000) A calibration-free gaze tracking technique, *Proceedings of the 15th International Conference on Pattern Recognition*, pp. 201–204.

SMI (2006) Remote eye tracking system (announced), SensoMotoric Instruments GmbH, Teltow, Germany. `http://www.smi.de`

Tobii (2002) Tobii 1750 eye tracker, Tobii technology AB, Stockholm, Sweden. `http://www.tobii.se`

Yoo, D.H. and Chung, M.J. (2005) A novel non-instrusive eye gaze estimation using crossratio under large head motion, *Computer Vision and Image Understanding*, Volume 98, Issue 1, Elsevier, pp. 25–51.

Open-Source Software for Real-Time Visible-Spectrum Eye Tracking

Dongheng Li
Human Computer Interaction Program
Iowa State University, USA
dhli@iastate.edu

Derrick Parkhurst
Human Computer Interaction Program
Iowa State University, USA
derrick@iastate.edu

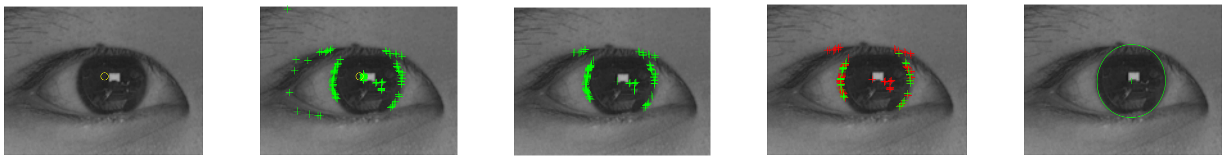
Keywords
Eye typing, limbus, RANSAC

Introduction

Video-based eye-tracking techniques frequently rely upon infrared-spectrum imaging because lighting and image exposure levels are precisely controllable through active illumination. However, there are a number of limitations to infrared-based eye tracking approaches. For example, infrared-spectrum systems perform poorly outdoors due to the presence of ambient infrared light. Moreover, performance of infrared-spectrum systems can vary significantly due to the individual differences in the physiological properties of the eye. For a more in depth review of these problems, see Hansen and Hansen (2005). We have developed a real-time eye-tracking system that uses visible-spectrum imaging in order to address these problems. We make this software freely available for download over the Internet as an open-source software package (see, http://hcvl.hci.iastate.edu/openEyes)

Algorithm

The most notable feature in visible-spectrum images of the eye is the limbus, i.e., the contour between the iris and the sclera (see Figure 1a). The position of the limbus is fixed with respect to the direction of gaze. The shape of the limbus in the image can be modeled as an ellipse. We adapted the Starburst algorithm (Li, Winfield, Parkhurst, 2005), originally designed to track the pupil in infrared-spectrum images of the eye, to track the limbus. In visible-spectrum eye tracking, light from ambient sources is relied upon illuminate the eye. Unfortunately, this can lead to the presence of uncontrolled specular reflections in the image of the eye. Fortunately, as the Starburst algorithm was originally designed to be robust to image noise, it is also well suited to handle the presence of extraneous specular reflections.



(a) (b) (c) (d) (e)

Figure 1. Eye-tracking algorithm (a) An eye image with the starting point shown in yellow circle. (b) detected features (green crosses). (c) Remaining features after distance filtering. (d) Inliers (green crosses) and outliers (red crosses) differentiated by RANSAC (e) Best fitting ellipse using only inliers.

Limbus Feature Detection

The algorithm begins at a starting point that is a best guess of the limbus center (see Figure 1a). This point is derived from the limbus center from the previous frame, and in the case of the first frame, is initialized as the center of the image. The limbus feature points are found by computing the derivatives along rays extending radially away from a starting point, until a threshold is exceeded. For each ray we detect two features before halting. An example set of detected features is shown in Figure 1b. Because the limbus is likely to be occluded by the eyelids and eyelashes, we restrict the directions of the rays. This range of angles is an adjustable to accommodate different users, but is initially taken to include -45^{o} to 45^{o} and 135^{o} to 225^{o}. One ray per degree of angle is traced resulting in at most 360 candidate limbus feature points.

Distance Filtering

Relying on the elliptical shape of the limbus in the images, a distance filter is applied to remove the features that are outliers. The features whose distance from the starting point is greater than 1.5 standard deviations from the mean are removed. The starting point is then replaced with the geometric center of the remaining features and the filtering is repeated again. The features remaining after filtering are shown in Figure 1c.

Ellipse Fitting

An ellipse is fitted to the candidate feature points using the Random Sample Consensus (RANSAC) paradigm (Fischler and Bolles, 1981). The candidate limbus feature points may still contain false alarms even after distance filtering, which would strongly influence the accuracy of the results if a least-squares fitting approach was used. RANSAC is an effective model estimation method in the presence of a large but unknown percentage of outliers in a measurement sample. We introduce two restrictions on the RANSAC fitting process to increase the robustness of the inlier selection process. First, only candidate ellipses with a radius ratio (major radius / minor radius) greater than 0.75 are considered. Second, only candidate ellipses of an area within plus or minus 1.5 standard deviations of the mean ellipse radius determined over the course of the tracking. The inliers are shown as green crosses and the outliers are shown as red crosses in Figure 1d. The final ellipse fit is shown in Figure 1e.

Application

We developed a low-cost remote eye tracker to test the algorithm. The remote eye tracker uses an inexpensive webcam, Unibrain Fire-i camera ($95 US dollars). The camera is mounted on the extended arm of a chin rest. To obtain a full-frame image of the eye we replace the original lens with a 12mm zoom lens ($70 US dollars) which required the addition of a CS lens mount ($10 US dollars). The system requires the user to places his/her head in the chin rest to assure proper alignment of the camera and use the remote eye tracker for desktop applications. The distance between the eye and the display was 26 inches and the visual angle of the screen was 32 degrees.

To calculate the point of gaze in the scene image, a mapping between the limbus center and the point of gaze must be determined. The user is therefore required to look at a 9-point grid, for which the scene locations are known. We use then can estimate a second-order polynomial mapping. After calibration, the user's point of gaze in the scene for any frame can then be established from the limbus center using this mapping. The average error in terms of visual angle is approximately 1 degree of visual angle after calibration. A significant limitation of this remote eye tracker is that it requires the user to hold very still to avoid introducing error into the estimated point of gaze. With a head-mounted system, the user would not be similarly restricted.

Conclusion

We developed a visible-spectrum eye-tracking algorithm and tested this algorithm with a remote eye-tracking system. We have made this algorithm freely downloadable and open source so that it can be easily integrated into human computer interaction applications, for example, gaze-based communication. We believe that in order for applications of eye tracking to become widespread, low-cost eye tracking solutions must be developed. With the decreasing cost of hardware and now the availability of free, open-source eye-tracking software, we expect that gaze-based interfaces will become more prevalent.

References

Hansen, D. and Hansen, J. (2005), Review of Current Camera-based Eye Trackers, *First Conference on Communication by Gaze Interaction*, 2005, pp 7-9.

Fischler, M. and Bolles, R., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, vol. 24, 1981, pp. 381-395.

Li, D., Winfield, D., Parkhurst, D. J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. Proceedings of the IEEE Vision for Human-Computer Interaction Workshop at CVPR, 1-8.

Progress on an Eye Tracking System Using Multiple Near-Infrared Emitter/Detector Pairs with Special Application to Efficient Eye-Gaze Communication

Dale Grover
Electrical and Computer Engineering Department
Michigan State University
2120 Engineering Building
East Lansing, MI 48824, USA
groverda@egr.msu.edu

Keywords
Eye tracking, eye-gaze communication, word-level disambiguation

Introduction

The "Owl" eye tracker uses a small sensor (Figure 1) mounted close to the user's eye, bouncing multiple channels of unfocused near-infrared light on and around the user's eye to determine direction of gaze. The hardware makes no assumption about what structures reflect light from any particular emitter/detector combination - it is thus not subject to the setup and adjustment constraints of techniques like limbus reflectometry. Specifically, specular and diffuse reflections may come from the iris/cornea, sclera, eye lids, and surrounding tissue, rather than from specific regions of the eye. Whereas conventional gaze-trackers use accurately aimed and focussed hardware along with detailed assumptions about the geometry of the eye, the Owl's approach is to use software to extract useful information from the raw unfocused data.

The Owl was developed in Ithaca, New York by Martin King in the early 1980's for augmentative and alternative communication users. Originally conceived as a direct-select device, it had a fixed display of approximately 36 hexagonal cells to allow typing English text, one letter per dwell. This arrangement required steady physical mounting due to sensor mass, good gaze detection accuracy, and resulted in a less than cosmetic appearance for the user, as well as occluding one eye. The later development of efficient word-level disambiguation (described below) allowed the display to be eliminated, leaving only eight peripheral LEDs as both targets and feedback, and producing a system that generates unlimited English text with one dwell per letter using only eight or nine cells. The reduced size of the sensor, and proximity to the eye (out of focus), means the users can still see out of both eyes, rather than having one occluded purely for communication needs.

Hardware

The present Owl sensor is an annulus 2.4 cm in outside dimension, with a 1.7 cm interior hole, and weighs less than 5 gm (<0.2 ounces). Eight LEDs and eight phototransistors are arrayed uniformly around the sensor, and a total of 64 different measurements of reflection are taken 50 or 60 times a second. The 700 nm "near-infrared" wavelength LEDs are just visible to the human eye, but also match well to the phototransistors and thus serve two purposes on the compact sensor. A lightweight, flexible cable with just three conductors suffices for signalling between the sensor and an interface board which converts the signal using a 3-bit variable gain amplifier and a 12-bit ADC with quite low sampling requirements (<100k samples/second). Bidirectional communication with the host computer is via a serial or USB connection at less than 100k bps.

Data Processing

In operation, the 8 LEDs of the sensor are turned on one at a time, and reflected light measurements from the 8 phototransistors are taken. The entire cycle is repeated 50 or 60 times a second, ideally locked to the AC line frequency since ambient interior light may have significant line frequency components. At this frequency, the user perceives the LEDs to be glowing steadily. The resulting 64 measurement channels include 8 from adjacent emitter/detector pairs, which are discarded due to excessive direct coupling. Noise is reduced using a filter that can accommodate discontinuities without excessive smoothing, such as the Savitzky-Golay filter (also known under various names such as least-squares polynomial filter). A third-order Savitzky-Golay filter of length 11 is currently used for 60 Hz frame rates.

The filtered channels are reduced by projection from 56 channels to 2 orthogonal components. The projection matrix is produced using a technique such as Principal Component Analysis (PCA). Given data from a brief calibration, the arbitrary orientation of the PCA-derived projection can be corrected via a rotation/flip matrix and the results used with parametric classifiers, though non-parametric classifiers would not need the geometric correction.

Once projected, each frame can be classified as fixation, saccade, blink. etc., with varying degrees of delay based on the method used. At present, fixations are identified using a variation of the method described by Sibert, et al (Sibert 2000), wherein eye positions within approximately 0.5 degree for 0.1 seconds are taken as the start of a fixation, further positions within a degree are taken as continuations, and 0.05 seconds of positions greater than a degree are taken as the start of a saccade. Another dwell-time fixation detection algorithm such as "N of M samples within distance D of the mean μ" (Duchowski, 2003) is implemented for future comparison.

It should also be noted that unlike image-based processing, it is possible to use alternative projections (e.g., based on Multiple Discriminant Analysis) for different classification tasks - for example, an alternative projection might produce superior results for blink classification versus PCA.

Once a group of contiguous measurements are classified as a fixation, the fixation can be classified according to target. The reduced number of targets (i.e., 8 or 9) and low processing devoted to extracting position data thus far leaves a great deal of flexibility in choosing one or more algorithms for target classification. Simple but reasonably robust non-parametric methods such as variations on the "K-Nearest-Neighbor" algorithm are useful early in a user's session when little data is available. After a short period of use, sufficient statistics will be available for a wider range of classifiers, especially parametric classifiers.

The processing requirements for the Owl are quite low. Filtering incoming data (60 Hz) requires on the order of 67,000 floating (or fixed-point) operations a second, and projection of the data to 2 components 13,000 operations a second. Thus on the order of 80,000 floating (or fixed-point) calculations per second comprise the main processing burden for generating "X/Y" eye tracking data.

During calibration, the projection matrices must be found. Calculation of the required covariance matrix for 4 seconds of raw data requires approximately 750,000 operations (once the data has been normalized and bias removed). PCA requires fewer operations depending on exact algorithms used, but a common implementation using optimized LAPACK routines on a 1 GHz PowerPC took 3 msec. For comparison, a frame at 60 Hz is 16.7 msec.

The data communication and processing requirements can thus be met without undue burden on a modern computer, reducing the cost of the hardware which needs only to make the measurements and convey them to the user's personal computer.

The Owl and Word Level Disambiguation

The dramatic reduction in price due to both high volume and advancing technology for small, easily interfaced video imaging devices (e.g., CMOS USB web cams) largely removes the historical benefits of the Owl technology as a general purpose eye tracking technology. The need to mount the Owl on the user's glasses and the fact that it provides only relative gaze information are further reasons why the Owl is not a compelling eye tracker for general purpose use.

The present research is motivated however by combining the sensor with word-level disambiguation, similar to methods now found on mobile phones. In word-level disambiguation, each of a small number of targets has multiple meanings - as for example a telephone keypad with multiple letters per key. The user selects a sequence of targets, and in the preferred implementations, the system presents the user with the most likely interpretations of those sequences. The characteristics of most languages are such that efficiencies that approach one letter generated per target selected are easily achieved - even when the letters are arranged inefficiently on keys (e.g., alphabetically). Since spelling is typically used, no special encoding method need be learned beyond spelling (though this could be an issue for some users), and vocabulary is unlimited.

Of particular importance is that visual feedback to the user can be limited to the LEDs in the sensor during target selection (target meaning is static within a sequence), and disambiguation at the end of sequences is often not required or minimal, and can be potentially replaced by auditory or other non-visual feedback. In many situations, the user thus has a cognitive load more akin to touch-typing than a feedback-intensive operation such as scanning. Since the interpretation of sequences remains static in preferred implementations, words can be generated by rote. A dynamic display, aside from the feedback offered by the Owl LEDs, can be consulted as desired, rather than being a necessity. Nor need the user maintain a relatively fixed relationship to a display, as eye position is significant relative to head position only. This provides more opportunity for eye contact with communication partners and less potential for physiological problems associated with a constrained position.

In addition, the reduction in the number of targets relaxes the accuracy requirements for the eye tracker, which should lead to more robust performance of the system.

Present Status

Offline analysis of data (see Figure 2) using the filtering and projection techniques discussed reveals good separation of targets. These techniques are being incorporated into a real time demonstration program hosted on Mac OS X, which continues to also support logging raw data for offline algorithm development (e.g., in Octave) as well.

A central issue with the Owl is how to address issues such as changes in sensor position (e.g., due to slippage of glasses down the user's nose) or ambient light. The program is written to keep time-tagged data histories to support exploration of classification algorithms that can robustly, and gracefully, handle these changes without resorting to a disruptive full calibration.

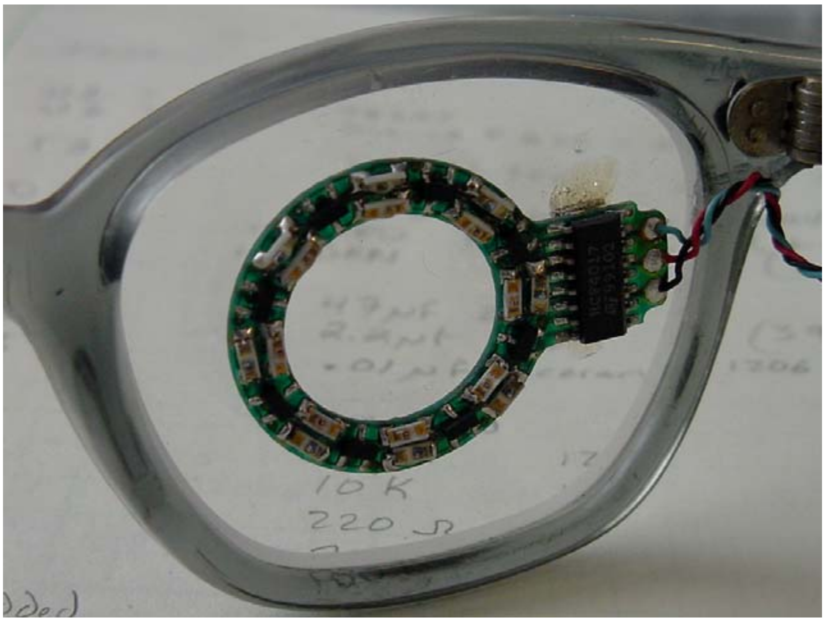


Figure 1. Owl sensor mounted on eyeglasses

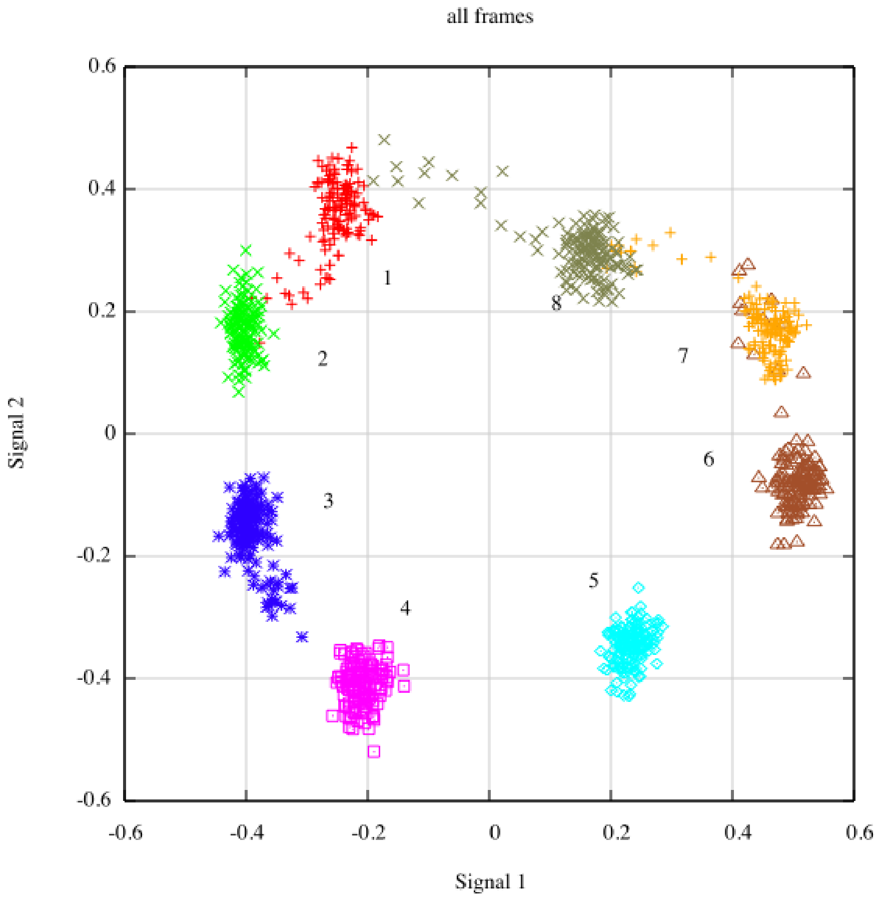


Figure 2. Frame data projected using Principal Component Analysis

References

Duchowski, A. (2003) *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, London, p. 115.

Sibert, L., Jacob, R., and Templeman, J. (2000) *Evaluation and Analysis of Eye Gaze Interaction*, NRL report. Naval Research Laboratory, Washington, D.C., p5.

Basics of Gaze Estimation

Arantxa Villanueva
Public University of Navarra
Campus Arrosadia sn
31006 Pamplona (Navarra)
Spain
avilla@unavarra.es

Dan Witzner Hansen
IT University, Copenhagen
Address Rued Langaardsvej 7
2300 Copenhagen S
Denmark
witzner@itu.dk

Javier San Agustin
Public University of Navarra
Campus Arrosadia sn
31006 Pamplona (Navarra)
Spain
javier.sanagustin@unavarra.es

Rafael Cabeza
Public University of Navarra
Campus Arrosadia sn
31006 Pamplona (Navarra)
Spain
rcabeza@unavarra.es

Keywords
Gaze estimation, calibration, eye tracker geometry

Introduction

During the last decade, a tremendous effort has been made on developing robust and cheap eye tracking systems for various human computer interaction applications. Robust non-intrusive eye tracking and gaze estimation is crucial for human computer interaction with attentive user interfaces, for diagnostic applications such as understanding human affective states, and is gaining importance outside laboratory experiments (Duchowski 2003).

Several camera-based eye trackers have been proposed in research and as commercial systems. The proposed methods have shown that the use of several cameras and light sources seems to make gaze estimation more robust, but stereo setups require calibration and are more expensive. Single camera systems may on the other hand reduce costs. The use of single camera system without pan-and-tilt faces the trade-off between large head movements and high accuracy.

The main objective of eye tracking is to provide an accurate estimation of eye movements. Based on the image of the eye and possibly additional data, gaze information can be determined. Gaze determination has previously been denoted for both: describing the task of determining the location on a 2 dimensional (2D) surface the user is looking at, as well as the 3 dimensional (3D) direction of gaze. We denote the 2D location as the *Point of Regard* (PoR) and the 3D directional vector as the *gaze direction or line of gaze* (LoG). Obviously, the point of regard can be determined by the intersection of the 2D surface and the gaze direction. In the following discussion note that using gaze direction for on-screen applications needs an additional model for finding the intersection, which in turn may require additional calibration. The success of the gaze estimator depends largely on the technology (cameras and light sources), prior knowledge of the setup and other system parameters, as well as the algorithms employed to gather information from the images. To a certain extent the accuracy of gaze estimation will improve as the image quality improves.

The objective of this paper is to discuss the problem of gaze estimation from a formal (geometric) point of view. We intend to propose a mathematical framework for geometrically determining the relationship between the information provided by the technology and gaze (PoR / LoG) under varying system information. This work is mainly theoretical and practical issues are omitted. Figure 1 illustrates the topic of analysis. It shows a schematic description of the matter under study. The video oculographic system obtains an image of the eye from which system specific features are extracted. Based on the features the connection between the PoR/LoG and the image information is sought.

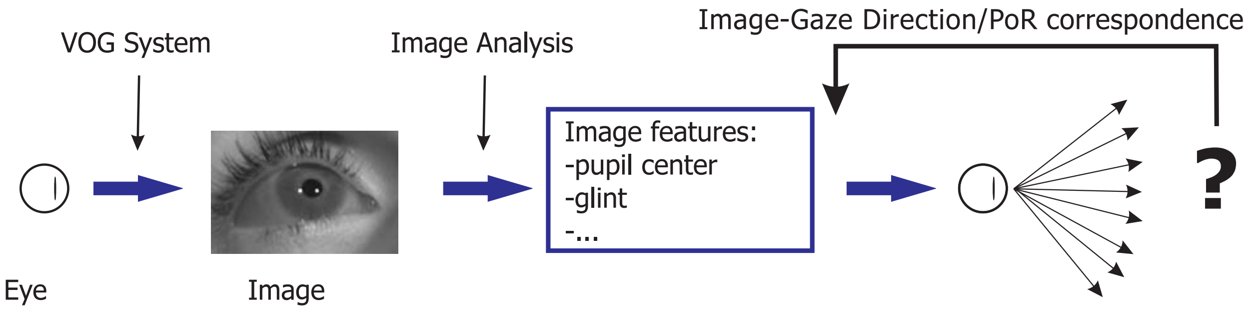


Figure 1.Image-gaze direction connection problem. The VOG system obtains an image from the eye. The geometric connection between image and eye position is pursued.

Basics of Gaze Estimation

The purpose of any gaze estimation method (PoR or LoG) is to find a mapping from the image to the output space (either 2D position or a 3D direction). That is,

$$P_g = \Phi_c(X) \quad (1)$$

where Φ is the gaze function with parameters c, X is the features vector from the image and P_g denotes the output vector (PoR or LoG). In principle the input vector X may contain the entire image, features extracted from the image or parameters from other systems (such as an external head pose system). Different models for the function Φ have been proposed during the years to relate the observations X and gaze P_g. In equation (1) a set of parameters may need to be inferred, c. The process of finding these values is called calibration and is usually performed by asking the user to look at N predefined points on the screen.

The objective of this paper is to examine the gaze function (equation 1) based on geometrical principles. In this paper we will divide the methods for gaze estimation into two types. (i) Point feature-based geometric methods, that is methods that rely on extracting point features such as reflections (glints) and center of pupil. Where possible we relate our findings with existing methods. (ii) Shapes-based methods in which the ellipse of the pupil is considered.

In previous work the lower bound on the number calibration points varies with prior knowledge of the system geometry. We provide lower bounds on gaze estimation methods when varying the prior knowledge and the number of point features.

With regard to methods based on points, we will show as first approach that gaze estimation can be performed using solely the center of the pupil needing only 4 points of calibration (Witzner and Pece 2005) (Villanueva et al. 2005). The method is not head pose invariant and therefore less viable for desktop-based eye trackers. On the other hand it may be applicable in head mounted or head fixed eye trackers. Among others we will show that even by adding one light source, it does not make eye tracking invariant to head pose changes. We will show that, geometrically, using two or more glints can make gaze estimation invariant to head pose changes. Such systems can be calibrated with only one calibration point.

Several methods consider the shape of the pupil as a working feature. In our paper we will analyze this feature and its importance when dealing with refraction. The effects of the corneal refraction affect considerably the pupil image, making the analysis of the system geometry more difficult. We will describe how refraction affects the different methods and provide techniques to compensate for the error.

Many systems effectively only use one camera and one eye for gaze estimation. Using either both eyes or several cameras have previously been used for adding robustness to the systems, but do they actually provide additional information? Many systems use a stereo system to locate the head in 3D space. We want to demonstrate how many cameras are really necessary to estimate the gaze based on purely geometrical criteria. This will provide firstly a mathematical reasoning to infer the minimum number of cameras needed, and secondly increase the possibilities of systems based on more cameras. It is clear that more cameras make the system more robust but having a deeper knowledge about the geometry is surely going to extend trackers possibilities to estimate gaze and to reduce calibration. With regards to binocular eye trackers we would like to explore the possibility of using both eyes information in order to estimate the gaze point. As the PoR is considered as the intersection of the visual axes, the study can provide interesting conclusions. We will, through geometric reasoning, elaborate on these issues.

Conclusions

A study of the geometry of gaze estimation using an eye tracking system is proposed. The objectives of the work are:

- to provide a mathematical review of different methods that combine image information and previous system data in order to estimate gaze.
- to propose new methods to estimate gaze.
- to identify the minimum hardware requirements, and the lower bound on the number of calibration points using purely geometrical criteria.

To this end it is necessary to make a thorough review of the geometry of the system, i.e. a mathematical description of the elements of the system: camera(s), lighting, screen, eye, and its geometrical connections.

References

Duchowski, A.T.(2003), Eye Tracking Methodology. Theory and Practice. Great Britain: Springer-Verlag London Ltd.

Hansen, D.W. and Pece, A.E.C. (2005), Eye Tracking in the Wild, *Computer Vision and Image Understanding*, 98r, Elsevier, pp. 182-210.

Villanueva, A., Cabeza, R., Porta, S., Eye tracking: Pupil orientation geometrical modelling, *Image and Vision Computing* (in press).

GoldenGaze:
An inexpensive real-time gaze tracking system

Detlev Droege
University of Koblenz
56068 Koblenz
droege@uni-koblenz.de
++49 261 287 2769

Fabian Fritzer
Koblenz
weinschwaermer@web.de

Dietrich Paulus
University of Koblenz
56068 Koblenz
paulus@uni-koblenz.de
++49 261 287 2788

Keywords
Gaze tracking, low cost

Introduction

We present an improved version of a simple gaze tracking system based on an inexpensive, yet highly sensitive camera, equipped with a low-cost IR-filter and near infra red LEDs. Using the corneal-reflection method and common image processing algorithms we easily achieve the accuracy required to control reduced keyboard gaze typing systems such as UKO-II.

Overview

By further improving the inexpensive gaze tracking system presented last year at COGAIN 2005 we can achieve promising results. The hardware setup consists of a single, highly sensitive monochrome mini-camera, a group of infra red LEDs and a frame grabber card to connect the camera.

The main algorithms still follow the well established layout:

- preprocessing and ROI detection.
- glint detection.
- pupil detection.
- pupil center determination.
- sub-pixel accurate iris and glint position determination.
- viewing direction / gaze point estimation.

Different methods for eye detection have been considered and, if appropriate for the given setup, have been evaluated. With respect to the real time constraints, the missing of color in the (grayscale) input image, the low resolution and the small size of the ROI (roughly 60 by 60 pixel at best) a rather simple approach performed best. Using template matching around detected highlights (glints) pupils were found with high accuracy. Switching from on-axis to off-axis illumination solved the problems introduced by the bright pupil effect. However, this causes a less compact hardware setup for camera and illumination.

The corneal reflection method gives considerably better results when using sub-pixel accurate input values. Refining initial results to sub-pixel accuracy proved inexpensive in computing time while greatly improving the results. Several algorithms for pupil center determination were evaluated. The circle Hough transformation gives very accurate results, but is limited to sufficiently round circles. When degraded to an ellipse due to perspective the detection quality lowers. Extending the algorithm to ellipsis introduces considerable additional computational complexity and is not suitable for a real time application. Similarly, ellipsis fitting did not fulfill the real time constraints given by the problem. Further investigation on detection algorithms will be done in a following diploma project.

While users wearing glasses introduced severe problems for the system in its initial version, some simple heuristics meanwhile solved this problem in almost all situations. Provided the additional highlights introduced by the lenses and the frame are not too close to the pupil, the correct glints are chosen for gaze determination.

Results

While the system performed well under lab conditions the first user test suffered from technical problems which have to be investigated further. The intended user of the system was able to type some sentences, however some tuning to reflect her usual working environment still have to be done. Coupling the gaze tracker to her customized input software ERIC, based on UKO-II, performed well.

Further work will be performed on the optimal selection of camera and IR illumination positions, relative to each other as well as to the monitor.

References

Fritzer, F.. (2006) Texteingabe per Augenbewegung. *Diploma Thesis*, University of Koblenz-Landau, Campus Koblenz.

Session 2: Multimodal Interaction with Real-World Objects

Control Application for Smart House through Gaze interaction

Dario Bonino
Politecnico di Torino
dario.bonino@polito.it

Emiliano Castellina
Politecnico di Torino
emiliano.castellina@polito.it

Fulvio Corno
Politecnico di Torino
fulvio.corno@polito.it

Alessandro Garbo
Politecnico di Torino
alessandro.garbo@polito.it

Keywords
Domotics, eye tracking, head tracking, multiple layout application, usability

Introduction

The research community has to face many challenges in smart house system development: interoperability, communication, security, hardware, interface, etc. The biggest challenge probably is to maintain the focus on the user as the ultimate target of this intense effort. One of the main aspects in the definition of the quality of life of a human being is autonomy, which in people with disabilities is often strongly compromised. This paper proposes an environmental control application that works with domotics and eye tracking systems, for enabling people with severe motor diseases to interact independently with the home environment that surrounds them.

System Architecture

In order to completely control a domestic environment through gaze, some facilities are necessary for transforming user glances into commands and for physically actuating such commands. Two main technological components are needed: a gaze control system and a domotic house. The gaze tracker follows the user movements and maps these movements to proper actions, by means of a control application. The domotic house executes the actions, closing the forward link of the user-home interaction loop. In the opposite direction, the house senses the environment and communicates its state to the control application which manages such information, possibly warning the user and requiring actions when necessary. The main feature of our system is the possibility of combining smart home and gaze control devices from different manufacturers and platforms.The system architecture (see Figure 1) is divided in four sections:

- Handling smart home appliances (low level).
- Handling smart home systems (high level).
- Handling Gaze Control devices.
- Environmental Control Application.

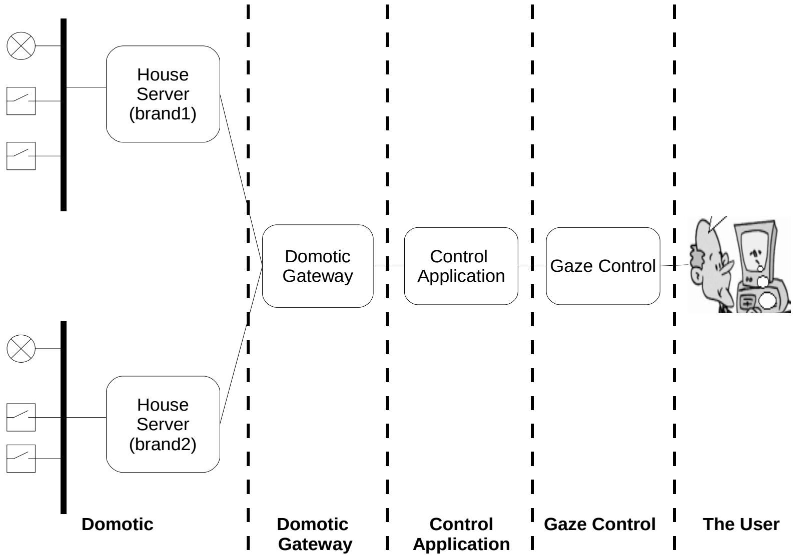


Figure 1. System architecture

Domotic System and Domotic House Gateway

There are many domotic system brands on the market: most of them adopt different standards for the communication buses, either proprietary, like the SCS by BTicino, or freely available (for association members) and widely supported like the Konnex standard, which is the result of a joint effort of more than twenty international partners.

A software component like a "gateway" that abstracts the physical configuration of the house and gets uniform access to various appliances is necessary to achieve interoperability amongst different systems. The house gateway has to provide three important functions for a house automation architecture:

- independence from domotic brands
- abstraction of appliances
- home intelligence

Unfortunately there is not yet a standard for the representation of domotic appliances and for communication protocols between control applications and house gateways. This standardization is one of the tasks included in the COGAIN project. In our system we have employed a Domotic House Gateway (DHG) previously developed within our research group and described in detail in (Pellegrino et al., 2006). The DHG provides access to the description and to the control of the smart home system through a simple XML-RPC interface.

Gaze control

There are many gaze control systems either commercially available or developed by academic researchers. It is only recently, thanks to the COGAIN network, that researchers and producers of different commercial systems are working together to define a new universal standard for eye control applications. Unfortunately nowadays few devices support the first standard release (Bates and Spakov, 2006), yet, an element that is conventional in most systems is the capability to control mouse cursor movements and events. Our environment control application uses information about mouse cursor movements and events in order to be compatible with most existing systems.

Environmental control application

The purpose of this software is to present to the users an interface where they can check the state of the various appliances and send control commands to the house.

The environmental control application communicates on one side with the DHG to obtain information about the state of the domotic home and to deliver commands, and on the other side with the gaze control system through "point and click" interaction.

The main feature of the control application is the dynamically resizable layout (see Figure 2). This characteristic meets two requirements: to support eye control systems with different performance and precision and to accommodate possible decreases in users' capabilities as the disease they have progresses.

When the user adopts a different application layout due to the disease evolution or just to use a different gaze control system, he or she shall not be compelled to learn a different way of interacting with the application. In other words, the way commands are issued shall stay the same even if the layout, the calibration or the tracking mode changes. For domotic control applications, the commands issued with such a gaze-based interface shall be similar to the ones issued by means of the standard interface, i.e., buttons, switches, etc. However, since the layout is constrained by the accuracy of the tracking system, which may change depending on the disease, the interface elements have been grouped by house room. This means when the user wants to switch on the kitchen light, he or she shall first select the room, e.g.. the kitchen, then select the light object and finally confirm the desired change of the object state by clicking on a confirmation button. These three steps remain the same while the application layout varies.

Experimentation

Our test system uses Sandra as gaze control software, which is based on a low cost webcam and can work both as head tracker and as eye tracker. Each of these two modalities offers a different precision and performance. The platform is described in more detail in (Garbo and Corno, 2005). We have used the aforementioned DHG in a smart home that is part of a scientific and technological park maintained by C.E.T.A.D. and dedicated to promotion, development and diffusion of technologies and innovative services for rehabilitation and social integration of elderly and disabled people.

Conclusion and further research

This paper has presented a control application that uses eye and head tracking as a means to allow not only communication but also interaction by severely impaired people with the environment that surrounds them.

The results are still preliminary. However they provide some positive feedback on the feasibility of the approach and on the capability of the application to improve the current interaction between disabled users and houses. Several issues need further improvement, as this is still work-in-progress. Future works on the Control Application will adhere to the to COGAIN standard and will improve the user-application interaction in terms of responsiveness and will result in a platform ready to be tested by end users. To set-up this more extensive experimentation campaign the authors are already collaborating with care-givers institute Molinette, in the context of the COGAIN European network of excellence.

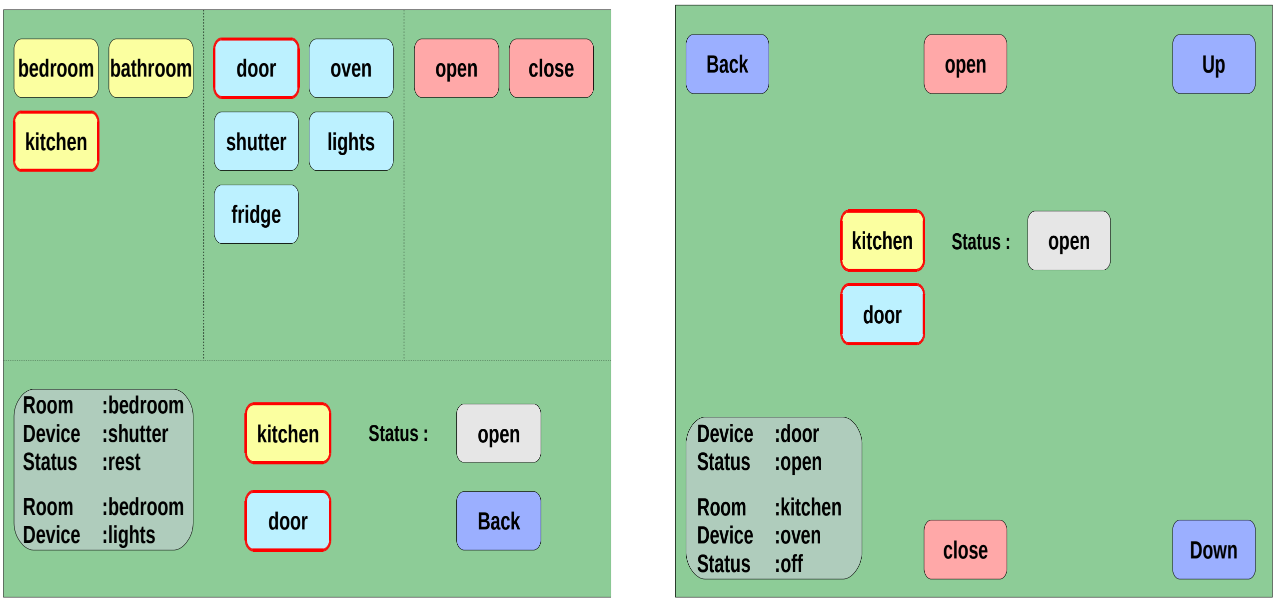


Figure 2. Differt layouts of control application

References

Bates and Spakov (2006) D2.3 Implementation of COGAIN Gaze Tracking Standards. *Communication by Gaze Interaction (COGAIN)*, IST-2003-511598: Deliverable 2.3.

Garbo and Corno (2005) Multiple Low-cost Cameras for Effective Head and Gaze Tracking. *11th International Conference on Human-Computer Interaction.*

Pellegrino, Bonino and Corno (2006) Domotic House Gateway. *ACM Symposium on Applied Computing*, ACM.

Direct Gaze-Based Environmental Controls

Fangmin Shi
Applied Vision Research Centre, Loughborough University, UK
f.shi@lboro.ac.uk

Alastair Gale
Applied Vision Research Centre, Loughborough University, UK
a.gale@lboro.ac.uk

Kevin Purdy
Applied Vision Research Centre, Loughborough University, UK
k.purdy@lboro.ac.uk

Keywords
Eye-tracking, environmental control, computer vision

Introduction

Living in modern times, people at home are greatly enjoying the convenience brought about by advanced technologies. With steadily increased home automation applications, it is becoming more and more popular for individuals to use one central control interface to set up and operate of all the audio, video and many household appliances in a home. However, such interfaces often are too complicated for people with a disability to operate. However, the technology has long been available to achieve Environmental Control (EC) for disabled people with limited mobility, which then helps them live with more independence.

This paper presents a specially designed EC system for use by people who have lost significant mobility but who have good control of their eye movements. Through attention responsive technology, a user will be able to perform either simple or complex operations of any electrical household appliance by directly gazing at it.

Current Environmental Control Technologies

As part of home automation, the Environmental Control system allows the remote operation of electrical devices in the home surroundings. Although such system can be used regardless of a users' physical ability, it is primarily designed to achieve more independent living for people with disability. Using an appropriate interface, a disabled user can independently control the room lighting, home entertainment system, air conditioning, or even open/close curtains, doors and operate intercoms.

An EC system should address the following two technological issues:

1. The input/output interface for users to operate

   The simplest input is by pressing buttons on a remote control such as for a typical TV control. Other input devices operated by hand can be a touch screen, switch, or a joystick. More advanced means are voice recognition or even eye tracking. The latter is one of the main challenges in the current research.

2. What wireless technology to adopt

   There exists three different types: ultrasound, infrared (IR) and radio frequency (RF). Each type uses the resource known from their names. An ultrasound system is not cost effective. IR and RF are currently very popular, even though conventional RF controlling requires additional wiring. The increasingly developed type is RF communication transmitted over the AC home power supply. This makes use of existing electric outlets in the environment but can also switch on/off the power supply through a special RF communication protocol. It is also the type that is employed in this research.

A control unit, an input device, together with electric appliances form the three main components of an EC system. A typical example of such system can be illustrated by the SRS series (http://www.srstechnology.co.uk/). These provide a range of solutions for use with either hand-held devices or integration onto a wheelchair. Both IR and new RF control are available. Furthermore, there are a variety of input choices such as integral keys, keypad or joystick; all of which have to be referred to using either a key display or a LCD screen for a confirmative selection.

R&D of Attention Responsive Technology for Environmental Control

Such assistive systems can be very useful for many mobility-restricted people by allowing a simple selection operation. This can be as straightforward as keeping the joystick moving in order to select from a complicated menu. However, for people suffering from severe disability due to diseases such as Amyotrophic Lateral Sclerosis or Cerebral Palsy, then their ability to operate a simple interface with even a few items can be very restricted.

In recent years, eye movement research has made significant progress in developing eye gaze based input techniques for interacting with computers. Instead of using a normal keyboard and a mouse, systems such as Dasher (Ward 2002) have been successful in achieving eye gaze based word typing. Eye tracking research also has applications in Environmental Control of which the Eyegaze System (Cleveland 1994) is an example. However, the collected eye gaze data of such systems are usually relative to a computer monitor. In operation, a user selects a device by looking at its representative icon on the screen. Therefore, the selection is indirect. In addition, it can be very tiring for a severely disabled user to go through such a content-rich menu.

This motivates the investigation of a new alternative EC input technique for severely disabled users - hence the current research project – Attention Responsive Technology (ART).

The aim of the project is to establish an EC approach by using an individual's eye gaze as a direct input method. In use, a user gazes at a real household appliance and his/her gaze attention will result in an action for the device. For instance, if a user wants to turn on a light, it can be automatically achieved currently by looking at the light in three-dimensional space for a second or less. This is achieved by applying the following methodologies:

- Computer vision: the user's living environment and eye movements are monitored in real time.

- Object recognition: based upon the user's gaze direction, any controllable devices are detected.
- Device control: X10 technology enables the remote control by plugging in simple modules to outlets.
- Selection interface: simple ON/OFF menu on a touch sensitive screen or via simple switch selection.

Computer vision and Object recognition

A head mounted eye tracking system is firstly employed in the project. With the user wearing a head band, two fixed cameras (an eye camera tracing eye movements and a scene camera monitoring the front view) work together to give the output of eye pupil co-ordinates relative to their positions in the scene image. This needs the system calibration in advance. The prototype system setup is described in Shi et al. (2006-1).

Certain criteria are also established to detect an eye gaze from a set of eye movement data obtained at the frequency of 50Hz (Shi et al., 2006-2). Upon any gaze, object recognition using the SIFT approach (Lowe, 2004) is activated with a view to finding a match in the real time scene image with any pre-known device. Images of these devices are captured in advance and saved in the image database. Figure 1 (a.& b.) show the resultant images of a lamp when it is in the OFF and ON status, respectively. The highlighted squares are the SIFT matches with reference to one of the database images. The '+' signs are the detected eye gaze co-ordinates from the gaze analysis.

The advantages of the SIFT approach to object matching are that it is not only invariant to scale and illumination but also stable in the case of occlusion. Figure 1 (c) is a correctly recognized result of a captured image when a visitor was partly blocking the fan.

a. b. c.

Figure 1. Objects captured by directly looking at them

For severely disabled users, the continual use of a head mounted eye tracker can be problematic. Consequently, the use of a remote eye tracker - the Smart Eye system (http://www.smarteye.se) is being investigated. This consists of two or three cameras which form a computer vision system and this measures the user's head position and orientation as well as gaze direction. A research challenge is how best to place the scene camera in order to relate the eye gaze data to where they are looking in the recorded scene image.

Device control

X10 technology (http://www.simplyautomate.co.uk/GuideToX10.asp?Cat1=82) is employed in the ART remote control. X10 is a wireless communication protocol which allows short digital messages to be sent and received over existing wiring. Only simple and cheap X10 modules need to be plugged between each electric device and its mains power supply. Each device is assigned a unique address. Each command then addresses one or more devices with an action such as ON and OFF and the module will operate the device accordingly. Although over the past decade many other home automation technologies have come into use, X10 has been expanding because of its low cost and because it uses existing electrical wiring, it's key feature.

Selection interface

Although the ART project aims to develop a direct eye-operated control interface, the word 'direct' must be understood in terms of the eye tracking with reference to real devices placed in 3D space. As a matter of fact, a gaze at a real appliance will be confirmed in ART by introducing a simple pop-up menu with say only two items, ON or OFF, on a touch sensitive screen. This is also the main feature that distinguishes the ART system with other 'always-on' eye tracking control systems, which can trigger numerous wrong actions for random gazes. Depending on a user's requirement, the touch screen can be replaced by any type of switch input.

Results

Currently the ART system is PC-based. It has a simulation interface as shown in Figure 2. The system has integrated the whole process including the system calibration, eye gaze data analysis, objection recognition, X10 control and selection, and intermediate result display with a list of gaze control history. An eye gaze can be obtained stably from 50 eye movement data points, taking 1second (less time can be easily used). It takes less than 2 seconds to determine whether a gaze falls on any found object with the current Matlab-driven software.

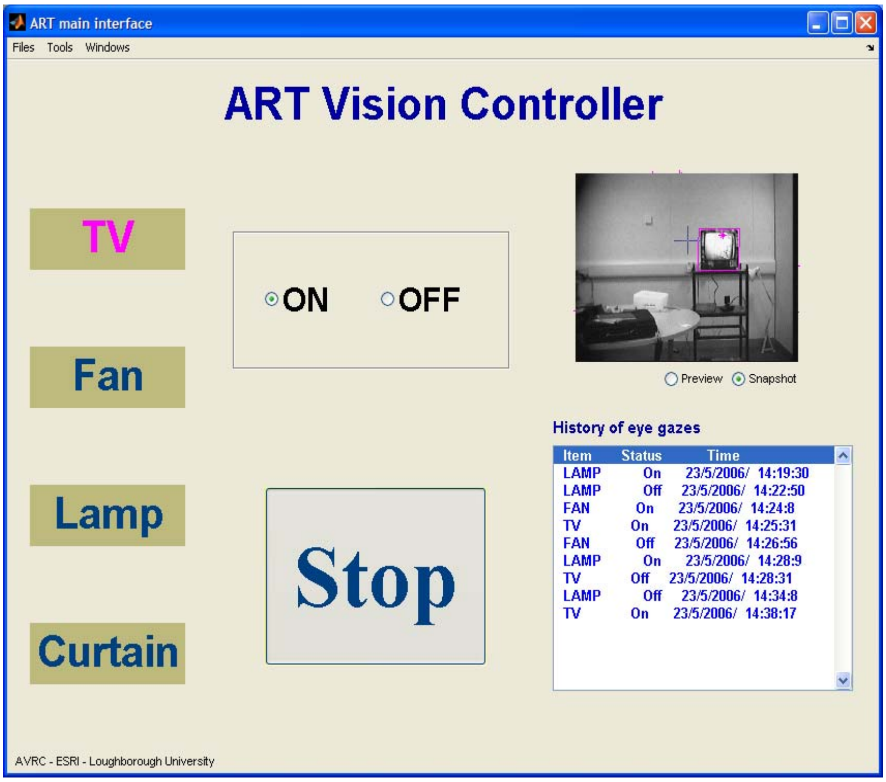


Figure 2. Current ART user interface

Conclusion and Future Plan

The ART system based on the head mounted eye tracker has demonstrated the successful process of using eye gaze to address a control directly. The current tested household appliances are limited but can be easily extended to as many as a user requires. As a PC-based system with no need for extra electrical wiring, it can be very handy for the ART components to integrate into a wheelchair Environmental Control system using other control technologies such as Infra Red.

We are currently progressing to a remote eye system, which has low cost and does not need any attachment to the user. However, the other modules discussed above will remain the same as the head mounted one. Although the SIFT approach to object recognition has shown its powerfulness, to apply a more stable 3D object recognition in a real environment with a complicated background and ambient light will still be the core next stage of the research programme.

Acknowledgements

This research is funded by the UK ESRC PACCIT Programme and also part of the COGAIN network.

References

Cleveland N.R. (1994). Eyegaze human-computer interface for people with disabilities. *First Automation Technology and Human Performance Conference*, Washington, DC, April 7-8, 1994.

Lowe D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2, pp. 91-110.

Shi, F., Gale, A.G. & Purdy, K.J. (2006). Eye-centric ICT control. In Bust P.D. & McCabe P.T. (Eds.) *Contemporary Ergonomics 2006*, pp. 215-218.

Shi, F., Gale, A.G. & Purdy, K.J. (2006). Helping people with ICT device control by eye gaze. *10th International Conference on Computers for Helping People with Special Needs*, Austria, July 2006. pp. 480-487.

Ward, D. J. & MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, 418, 838

Gaze Pointing and facial EMG clicking

Andrew Michael Junker
Brain Actuated Technologies, Inc.
1350 President Street,
Yellow Springs, Ohio 45387 USA
andrew@brainfingers.com

John Paulin Hansen
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S Denmark
paulin@itu.dk

Keywords
Multimodal interaction, gaze and brain integration, user performance and preference studies, eye-typing systems, usability issues with gaze-based systems

Introduction

For a user to control an eye gaze system they must be able to control their eyes and keep their eyes open. In most cases where this is possible the user will have some control of their facial muscle activity as well. Facial muscle activity as indicated by the electomyographic (EMG) signal may be used for clicking to enhance a gaze communication system (Surraka et. al., 2004). The EMG signal can provide a fast click that would overcome problems associated with dwell or eye blink click. The EMG click may serve as a back-up for gaze control when the portability of a user's eye gaze system is a problem or if the user should loose the ability to control the eyes.

The need for a durable switch

Gaze interaction appears to be one of the most desirable alternatives for individuals who cannot direct cursor movement with their hands. Pointing the eyes is a natural action, it takes seemingly no effort, and the visual system along with some facial motor control, often remain functional when degenerative diseases or traumatic injuries affect most other motor systems.

Once a cursor has been directed over an item for selection, some click action must be made to complete the selection process. Most gaze communication systems incorporate a dwell function for simulating a simple click. Some systems also include the ability to detect an eye blink as a means of creating a simple click. Both methods of clicking have inherent problems that can make eye gaze as a communication method more difficult to utilize.

In the case of dwell, the user must hold their gaze in a pre-defined range of motion for a pre-defined duration of time. This results in a delay in responsiveness of the eye gaze system (Hansen et al., 2003). Another problem is the selection of false targets from holding one's gaze too long in the wrong location. Most users need to practice for a couple of hours before they master dwell selection (Itoh et. al., 2006).

In order to employ intentional eye blinks for clicking the user must close and hold closed at least one eye for a pre-defined length of time. This time must be longer than that of involuntary eye blinks to prevent false selections from involuntary eye blinks. Thus the use of eye blinks for selection will delay the responsiveness of the eye gaze system. Further, some users report that the muscle effort needed to blink becomes tiring with constant use.

At times the portability of a user's gaze communication system may present a problem. In some cases it is not possible to take the eye gaze system along in the car or to a hospital for example. There may be times when the user's eyes are tired but the user still needs an easy way to communicate simple messages. There may be users who will eventually need a click to substitute for a gaze pointer that they can no longer control. A male ALS patient described his concern about a potential gradual loss of his eye control in an e-mail: "This is an important question, since many final stage people with ALS (PALS) experience that their eye muscles become weaker and weaker. This is contrary to what is written in most textbooks on ALS, but it is actually happening. I can mention that some Japanese PALS, who have survived 20 years using ventilator, can only stare straight out. Secondly, many final stage PALS take drugs (Scopoderm, Atropin etc) against saliva. Unfortunately, these drugs interfere with the sight leading to loss of precision and accurateness."

An EMG click can compensate for noise on the eye gaze tracker. An EMG click can be combined with a dwell click. EMG clicks can be strung together as sequences or held for predefined durations to have other meanings, like mode shift, double click etc.

EMG-switches may be faster than a finger button

There are some indications that EMG-switches may also be attractive because they can become very fast. In a study performed with able-bodied participants by the USAF in which a forehead EMG switch was compared to a finger switch, response accuracy was found to be extremely high, approximately 98%, and reaction times fell between 180- 200 ms, a range considered to be the limit of simple reaction time. Several participants achieved 15-20% faster reaction times with the forehead EMG switch than with a manual switch (Nelson et. al., 1996). Surraka et.al. (2004) found indications that gaze pointing in combination with EMG-clicking would be faster than hand controlled mouse pointing and clicking for longer movements.

Recently, three student subjects at IT University of Copenhagen conducted an experiment to compare gaze typing on the GazeTalk system in combination with dwell-, mouse- or EMG-switch activation. They used a Tobii-1750 system to point with their gaze. Clicking was done by either dwell, a standard mouse or by the use of a Brainfinger system to control an EMG-switch by activation of the forehead muscle, i.e. *corrugator supercilii*. In average (across 60 sentences typed), mouse clicking produced the fastest typing (10.3 wpm, s.d. = 0.8), the EMG-switch came second (8.1 wpm, s.d. = 2.4), a dwell time setting at 500 ms produced 7.7 wpm, (s.d. = 0.7) while a dwell time at 1000 ms produced 5.9 wpm (s.d.= 0.6). The fastest session average (12.2 wpm) was made by a subject using the EMG-switch to click. (Bech et. al., 2005).

In another experiment, conducted at Wright State University in Ohio USA, 3 subjects typed 3 blocks of 10 sentences on the GazeTalk system in scanning mode (without gaze tracker or any other pointer), with a fast setting for the step-time (300 ms). They typed an average of 2.66 wpm (s.d.= 0.86) when they used a finger on the mouse button to click, they typed 3.17 wpm (s.d.=1.02) when they used a forehead mounted EMG-switch activated by their forehead muscle, and they typed 4.25 wpm (s.d.= 1.80) when they used their jaw-(bite)-muscle to activate the forehead EMG-switch. The keystroke per character was lower for jaw activations (KSPC= 0.98, s.d. = 0.28) than for finger activations (KSPC=1.34, s.d.=0.45). This is attributed to the fact that the subjects often clicked too late on a target button with their finger; while with jaw activations they did not.

In another study, Surraka et.al. (2005), found that EMG-clicking by a smiling technique was significantly faster and less erroneous than EMG-clicking by a frowning technique (i.e. using *corrugator supercilii).*

The above observations point to the need for more research in ways to optimize the measurement and use of facial EMG signals.

Considerations when using facial EMG switches with gaze pointing

The majority of people who currently use gaze communication equipment usually have a degenerative disease such as ALS. In the case of ALS, motor neurons die off reducing the number of motor units and leaving muscle fibers that have lost their nerve supply. These orphaned fibers reattach to other motor neurons. This results in a decrease of functional motor units with an increase in action potentials of these motor units (Stashuk, 2001). We hypothesize that a decrease in the number of motor units results in a perceived need to produce more effort to achieve a muscle contraction. We have observed that users with ALS approach control tasks with a tendency to over control and work from a high muscle tension level. This tendency should be taken into consideration when implementing an EMG switch algorithm. We hypothesize that it would be beneficial for individuals with ALS to operate at a lower muscle tension baseline level and command smaller muscle contractions to create a trigger. In this way less muscle fibers would be recruited, the overall response could be lower, the rise and fall could be faster, and less effort would be needed. This would result in a faster trigger requiring less effort. We feel that EMG-switch software should take into consideration the above findings by facilitating operation at lower baseline levels and triggering at lower levels.

It is possible to measure brainwave resonance (EEG) and lateral eye movements (EOG) as well as EMG at the forehead. We are currently experimenting with a composite measure of these three signals to produce a metric of effort. It is hypothesized that this metric can be fed back to the ALS user to help them monitor the level of stress at which they are operating.

Conclusion

Facial muscle activity, the electomyographic (EMG) signal, may be used for clicking to enhance a users eye gaze system. If used correctly, the EMG signal can provide a fast click that would overcome the problems associated with using a dwell or eye blink click. The EMG click could also be used with an on-screen keyboard in a switch scanning mode when the portability of a user's eye gaze system is a problem. When introducing EMG-switches to individuals with ALS, one should consider that the individual may have a tendency to over control and work from a high muscle tension level and that the EMG signal may contain excessive individual motor unit firing. The EMG switch should be designed to cope with user tension levels and high levels of motor unit firing.

References

Bech, C., Dam-Andersen, M., Hansen, J. E. & Schifter, L. R. (2005) Cyberlink & Eyetracking – multimodale systemer som fremtidens computernavigation. (in Danish, English title: Cyberlink & Eyetracking – the future of multimodal systems for computer navigation.) Student paper, IT University of Copenhagen, 65 pages.

Hansen, J. P., Hansen, D. W., Johansen, A. S., Itoh, K. and Mashino, S (2003). Command without a click: Dwell time typing by mouse and gaze selections. *Proceeding of the 9th IFIP TC 13 International Conference on Human-Computer Interaction*, pp 121-128.

Itoh, K., Aoki, H., and Hansen, J. P. 2006. A comparative usability study of two Japanese gaze typing systems. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications* (San Diego, California, March 27 - 29, 2006). ETRA '06. ACM Press, New York, NY, 59-66. DOI= http://doi.acm.org/10.1145/1117309.1117344

Nelson, W., Hettinger, L., Cunningham, J., Roe, M., Hass, M., Dennis, L., Pick, L., Junker, A., Berg, C. (1996) Brain-body-actuated control: assessment of an alternative control technology for virtual environments. *Proceedings of the 1996 IMAGE CONFERENCE*, 225-232

Surakka, V., Illi, M. & Isokoski, P. (2004) Gazing and Frowning As a New Technique for Human-Computer Interaction. *ACM Transactions on Applied Perception*, 1, 40-56.

Surakka1, V. Isokoski1, P., Illi1, M. and Salminen, K. (2005): Is it better to gaze and frown or gaze and smile when controlling user interfaces? *Proceedings of Universal Access in Human-Computer Interaction* (UAHCI 2005), Las Vegas. LEA, cd-rom.

Stashuk, D. (2001): EMG signal decomposition: how can it be accomplished and used? *Journal of Electromyography and Kinesiology* 11 (2001) 151-173

When can eyes make up for hands?

Marcela Fejtová
Czech Technical University
Department of Cybernetics
Technická 2, Prague 6
Czech Republic
fejtovam@k333.felk.cvut.cz

Petr Novák
Czech Technical University
Department of Cybernetics
Technická 2, Prague 6
Czech Republic
novakpe@labe.felk.cvut.cz

Jan Fejt
Czech Technical University
Department of Cybernetics
Technická 2, Prague 6
Czech Republic
fejtj@k333.felk.cvut.cz

Olga Štěpánková
Czech Technical University
Department of Cybernetics
Technická 2, Prague 6
Czech Republic
step@labe.felk.cvut.cz

Keywords
Eye trackers, mouse emulation, gaze PC control, gaze-controlled wheelchair

Introduction

There are a number of situations where a person cannot use their hand as an actuator for handling simple everyday tools (e.g. to write with pencil), to control his/her environment (e.g. to switch a light on or off) or to take advantage of different sophisticated assistive technologies. In Europe, there are more than 500,000 people who need alternative access to computer due to their impairment. For some this may be a temporary problem (e.g. after a severe operation) while for others this is a long lasting condition. Gaze-based interaction can be a reasonable solution for a significant part of this population. COGAIN tries to set technical standards for eye tracking (Bates et al. 2005). When searching for an adequate gaze-based solution for a specific user it is important to take into account a number of criteria – *the user's physical condition* and *the price* of the used system appear to be the first candidates for this purpose. These are however far from sufficient. A number of features make a difference when searching for the best choice for a specific person. Let us consider at least three of them, namely *flexibility* (= ability to be used in combination with various software tools which are not created by the gaze-control producer), *takeup time* (= time a person needs to learn to utilize the system) and *setup time* (= time needed to install the system). All of these features represent strong points of I4Control® (Fejtova et al. 2005), an original low cost solution, which emulates computer mouse and thus ensures hands-free computer interface. Moreover, the same principle can be applied to control various toys or tools, e.g. a wheelchair. This paper first reviews several applications designed and developed by the CTU team to prove these claims. Finally, some challenges related to gaze control of a wheelchair are considered.

The system I4Control®

The I4Control® system (Fejtova et al. 2005) uses a small camera mounted on a spectacles frame to monitor the user's eye position and movements.. Data provided by a videooculogram (VOG) is evaluated to estimate deviation of the user's eyes from their normal position: left-right, up-down. This information is interpreted as the control signal driving the computer cursor in the same direction. On the other hand, the absence of the signal for a certain pre-defined period (when eye is closed) is used to express a simple mouse click. In this way I4Control® emulates the computer mouse and thus provides direct means for operating a computer through gaze interaction in the most *flexible* way – it is ready to cooperate with any mouse/controlled software system. Installation of the I4Control® system is as easy as that of a standard PC mouse – its *setup time* is very low. It's user can utilize the software keyboard included in a common operating system and she/he can write a message, send an e-mail or browse web pages without a single touch of any physical object. What the user needs to master is the ability to use eye movements to navigate the cursor sufficiently precisely to the selected place on the computer screen. As there are more than 30 keys on the software keyboard, this requires precise positioning and consequently it results in a necessary increase of *take-up* time. Unfortunately, for some persons the required precision can become an inconquerable obstruction. To overcome this problem we suggest substituting the software keyboard with a simple dedicated application **IPad** (Kšára 2006). Here the computer screen is divided into 13 boxes and the user gets access to the required letter of an alphabet by gradual division of the selected interval of ordered letters. She/he makes the choice by clicking anywhere in the left or right upper box – see Fig. 1. Here, the take-up time and requirement for precise positioning is significantly reduced at the expense of speed of writing – 5 clicks are needed to reach the required letter. Another feasible option is to combine I4Control® with the Dasher system (The Dasher Project) – to do so the proper language model has to be created first and the user has to get used to this specific form of writing.

Figure 1. IPad

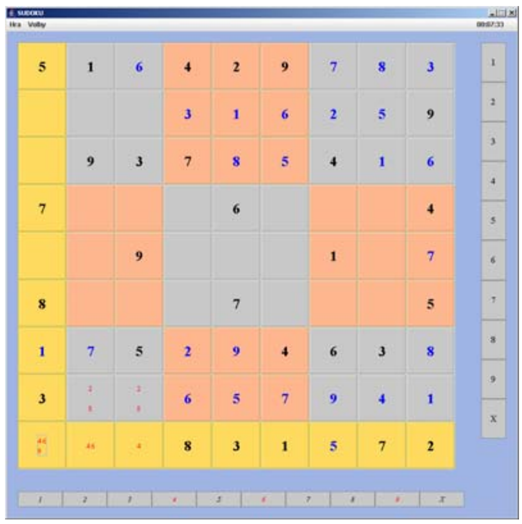


Figure 2. Eye controlled SUDOKU puzzle

Control signals produced by the I4Control® system can provide the user with direct access to many e-learning applications and provide an intellectually stimulating environment, see for example our simple mouse-controlled environment for solving **SUDOKU puzzles** (Hodač 2006) on the screen of a computer (http://sudoku.euweb.cz/sudoku.jnlp), Fig. 2. The same signals can even drive physical objects. To support this claim we have designed a special toy-car called **Gertie** which is controlled directly by means of I4Control®, too. It is easy to specify the goal Gertie should reach – the goal can be e.g. "to move Gertie into the soccer goal". Gertie1, the first version of our toy, was assembled from LEGO® and it used following basic commands. If the driver (=I4Control® user) looks up, the car moves straight forward, and goes backwards if he/she looks down. It turns to the side the user looks to, while eye blink sets the rotation of the locator. It reacts promptly to any change of user's eye position and one can easily estimate how far he/she is from the intended goal, thus it provides a well interpretable feedback to most types of users. Wireless communication with I4Control® is ensured through a bluetooth module in the most recent model (Gertie2). The successful experiments with Gertie have proved that the potential application of I4Control® is not restricted to a computer. On the contrary, our system can step out from the virtual reality of a PC into the physical world.

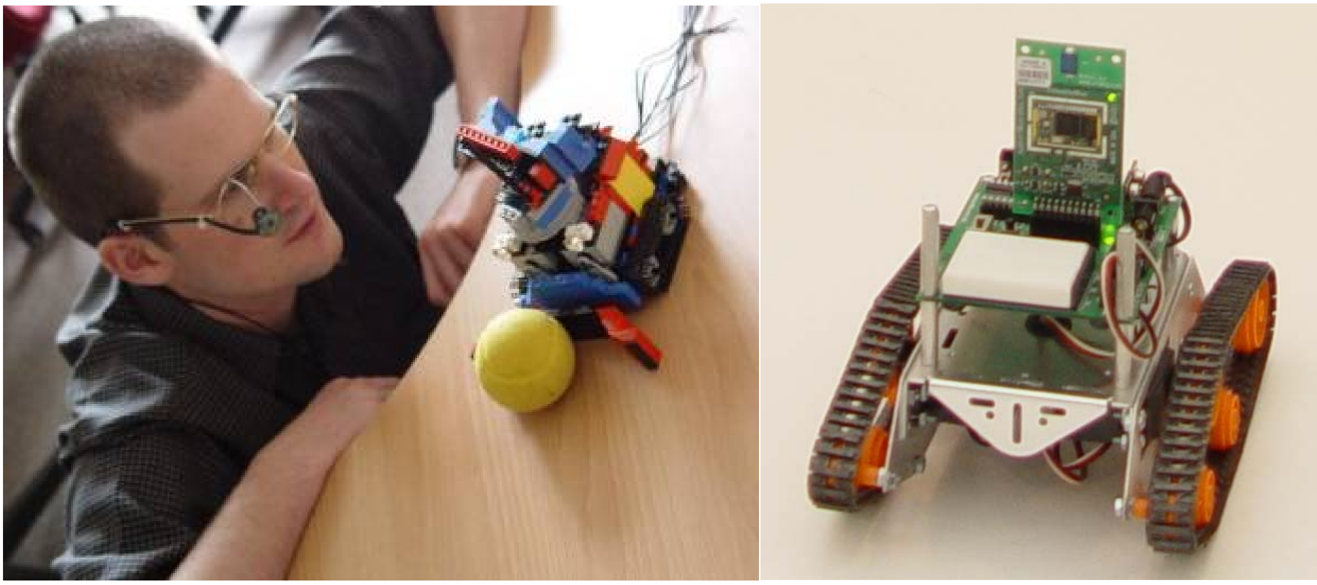


Figure 3. Gertie – an educational toy

Challenges of the gaze controlled wheelchair

The functionality offered by the I4Control® seem to provide a good starting point for an ambitious goal to design a gaze-based control system for an electric wheelchair. Seemingly, this task is closely related to that of driving Gertie the toy. This is true as far as the navigation is concerned. Numerous new problems arise however in connection with the question of how to guarantee the safety of the human user. Gaze control can create number of false or unintended signals caused by various reasons, for example, when the user:

- suddenly looks in a different way (towards the source of a suspicious noise),
- moves into a place with bad light conditions (and the system lacks the control signal),
- is forced to close his or her eyes due to irritation (dust, strong light, etc.)

It is clear that in real life conditions the gaze controlled wheelchair cannot fully rely on its user only. One way of proceeding is to require confirmation of commands conveyed by gaze through an independent channel provided by an alternative approach, e.g. (Tanaka et al. 2005). The other approach is to consider a wheelchair equipped by a certain degree of autonomy and offers the user only a limited possibility to interfering through gaze control. One can imagine scenarios of various complexity which have to be designed and tested – they can range from the case when mostly the user is in charge of the system equipped by some collision avoidance solution up to the situation when the chair moves autonomously to the destination described using gaze interaction. Given so many options, it seems that the appropriate combination of techniques will have to be customized according to the needs and constraints of each individual user. This task represents additional dimensions of the considered task. To build such complex systems, artificial intelligence techniques will have to be applied and enhanced. Even though we hope that AI can provide significant support for this purpose, the most important part of the development will still rely on close cooperation with end users and their communities.

Acknowledgement

The presented research and development has been partially supported by the Czech grant NR8261 Therapy of paresis n. abducens by means of biological feedback and IST Network of Excellence IST-2003-511598 COGAIN (Communication by Gaze Interaction).

References

Bates, R., Istance, H., Oosthuizen,L., Majaranta,P. (2005) Survey of De-Facto Standards in Eye Tracking, D2,1, Project COGAIN

Fejtová, M., Fejt, J. (2005) System I4Control: The Eye As a New Computer Periphery. *In The 3rd European Medical and Biological Engineering Conference - EMBEC´05* [CD-ROM]. Praha: Společnost biomedicínského inženýrství a lékařské informatiky ČLS JEP, vol. 11, ISSN 1727-1983.

Kšára,M. (2006) SW tool for writing messages using I4Control, Bc. Thesis, FEE, Czech Technical University

The Dasher Project [online]. http://www.inference.phy.cam.ac.uk/dasher/, Cavendish Laboratory, Cambridge

Hodač, J. (2006) SUDOKU – implementation of an interactive environment, Bc. Thesis, FEE, Czech Technical University

Tanaka, K., Matsunaga, K., Wang, H. O. (2005) Electroencephalogram – Based Control of an Electric Wheelchair. *IEEE Transactions on Robotics*, Vol. 21, No. 4, IEEE, pp. 762-766.

Multi-modal interface: Gaze-EEG-based system

Carmen Vidaurre(1), Arantxa Villanueva(1), Reinhold Scherer(2), Markus Joos(3), Alois Schloegl(2), Sebastian Pannasch(3) and Rafael Cabeza(1)

(1) Electric Engineering and Electronics Department Public University of Navarra Campus Arrosadia sn 31006 Pamplona (Navarra) Spain. carmen.vidaurre@unavarra.es
(2) Laboratory of Brain-Computer Interfaces (BCI-Lab) Institute for Knowledge Discovery Graz University of Technology A-8010 Graz, Krenngasse 37, Austria. Reinhold.Scherer@TUGraz.at
(3) Applied Cognitive Research / Psychology III, Technische Universität Dresden, Mommsenstrasse 13, 01062 Dresden, Germany. joos@applied-cognition.org

Keywords
Multi modal interfaces, electroencephalogram (EEG), midas touch, eye tracking, Brain-Computer Interface (BCI)

Introduction

Future Human Computer Interaction (HCI) interfaces for severely disabled people should make use of the remaining capabilities of each person and be designed to compensate for his/her non-existing abilities. In this respect, all kinds of physiological signals should be combined to develop a so-called multi-modal interface. We have long experience in eye movement tracking and electroencephalogram-based (EEG) Brain Computer Interfaces (BCI). In this paper, we present a preliminary design of a mixed interface based on video oculography (VOG) and EEG.

Current Trends

In recent years, the rapid development of eye tracking techniques has provided new avenues for the exploration of eye movements in different environments and research is no longer restricted to artificial laboratory experiments. New research paradigms can now be developed and applied to more natural tasks and environments (Hayhoe and Ballard, 2005). In the field of HCI especially, the analysis of eye movements becomes more important for the design of interfaces and the exploration of usability. However, one of the major problems of eye tracking HCI is the differentiation between eye movements serving visual orientation and the subject's intent to select an interface object. This is known as the *Midas Touch* problem by the eye tracking community (e.g. Jacobs, 1990; Velichkovsky et al., 1997).

Over the last years interesting approaches have been proposed to overcome this effect. The solutions can be divided into two main groups. Firstly, there are systems that try to recognize selection using only eye information through fixation dwell time or blinking (Velichkovsky and Hansen, 1996; Majaranta et al. 2004). Secondly, there are systems that combine eye movement with other physiological signals, such as voice control or electromiographic (EMG) signals among others to activate selection, i.e. multi-modal interfaces (Starker and Bolt, 1990; Surakka et al., 2004).

Another idea is to combine eye tracking and EEG, the feasibility of which has recently been demonstrated by Baccino and Manunta (2005) in reading tasks, and by Graupner et al. (2005) during the free exploration of visual scenes. During recent years we pursued two main research lines of approach: (i) the development of an eye tracking system and (ii) EEG-based BCIs. By combining both, we are able to introduce a new approach of multi-modal interfaces based on eye tracking and EEG, that will be used for the selection procedure of the element gazed at on the interface.

Ad (i) The eye tracking system we developed is based on dark pupil technique and two infrared lateral lighting sources. The subject needs to conduct a typical 4x4 grid calibration. The system presents high robustness to lighting variation and exhibits mean accuracies of about 0.35º. At the moment dwell time and blinking based selections can be used.

Ad (ii) BCIs are systems capable of accepting commands directly from the human brain without the use of any muscle activity. For severely paralyzed people, or people in a locked-in state, such systems can reestablish a communication channel. We have developed several on-line adaptive cue-based BCI systems based on electroencephalogram (EEG) and motor imagery (the subject imagines performing a movement; e.g. left or right hand). Operating a BCI is a skill which users have to acquire. Therefore on-line feedback training is needed in which system and subject are adapting to each other. Usually cue-based training is performed. A BCI is called cue-based or synchronous when the user has to wait for a cue from the BCI before switching to the next mental state (following a fixed repetitive time scheme). We conducted on-line feedback experiments with a large number of subjects. We were able to train them in 3 sessions of 1.5 hours each and most of the subjects reached a high classification accuracy. For the training, we used on-line adaptive systems in both the feature extraction and the classification modules. This adaptation is very important for inexperienced subjects because they are rarely able to generate stable and stationary EEG patterns. (Kaplan, Lim, Jin, Byeon, & Tarasova, 2003).

After the training period subjects with an acceptable control of the BCI are ready to use an asynchronous or self-paced system in which the subject generates mental commands without system supervision. The BCI has the task of detecting the voluntarily induced changes in the ongoing brain activity (intentional control). For the user this operation mode represents a more natural way of communication. An eye tracking based interface works, in BCI terms, asynchronously. As the next step in our BCI research is towards real world applications, combining it with our eye tracking experience is a challenging opportunity.

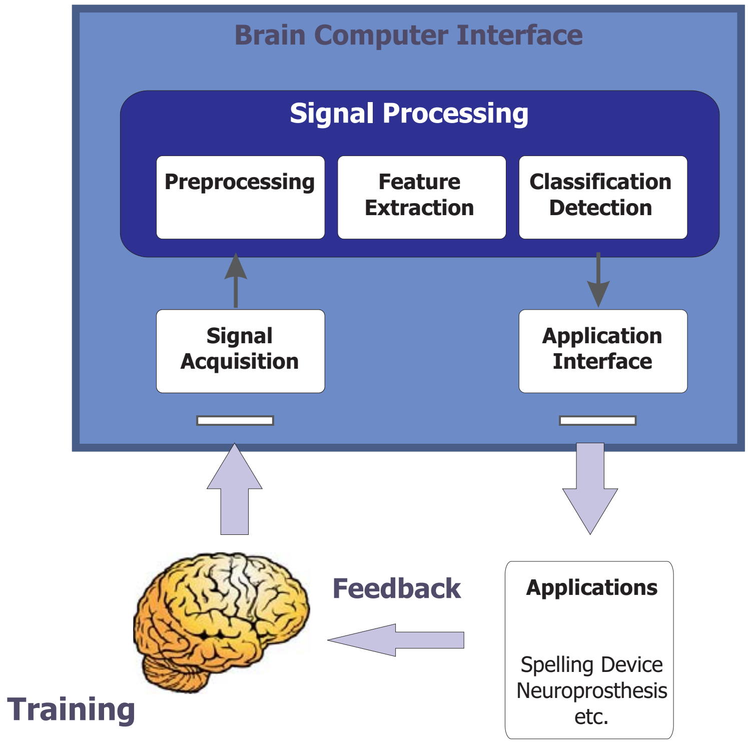


Figure 2.Brain Computer Interface modules

Suggested Design

The system presents two main areas, i.e. eye tracking and BCI. Although the algorithms involved could be improved the eye tracking system design will be identical to the one already developed, except for the inclusion of the BCI classifier output. The tracker would use the BCI classification output in two different ways:

1. The selection/activation could be purely activated by the BCI classifier again in two ways.
    a. Activation using motor imagery. The subject will imagine a specific movement when he/she wants to make a selection of a gazed element. This will be helpful for subjects who have problems blinking and will be faster than dwell selection. We would like to achieve selection intervals of 1 second or less.
    b. Activation detecting the subject's intent to select. The classifier should recognize the EEG pattern associated to this will when the subject is fixating an element in the interface and wishes to activate it.

    It is clear that this type of multi modal interfaces with EEG signals is more intrusive than a pure eye tracking system. However, it could be a solution for some people with blinking difficulties and could speed up the selection comparing to dwell time.

2. Another possibility for the BCI classifier output is to support the recognition of tracking. The loss of calibration generates tracking problems that cause frustration. We would like to study EEG patterns related to frustration during an eye tracking session. This information could be used to correct the gaze estimation system resulting in a very innovative contribution to the research.

The first objective, i.e. selection recognition, will be quite straightforward using motor imagery. The main part of the work will be devoted to study appropriate feature extraction methods which allow the fastest recognition of the selection. To accomplish this goal, the subjects must be previously trained like in every BCI application.

Studying EEG patterns of selection intent and frustration require further research which will involve psychologists and neurophysiologists. At this stage we are able to record EEG signals related to both mental states as a start to research in this area. This interface is primarily based on eye movement, therefore an automated artefact processing module will be also necessary. We plan to use a fully automated EOG reduction method based on regression analysis and validated by two independent scorers on a large database (Schlögl et al 2006).

Immediate Steps

The hardware for the system consists of infrared lighting sources, camera, lenses, infrared band pass filter and a computer for the eye tracker. With regards to BCI system we have a low noise portable bio-signal amplifier, electrodes, an acquisition card and computer.

We consider that we are able to accomplish two main steps of the work once the hardware step is completed.

- A number of subjects will be trained with a BCI to learn to distinguish between two classes, i.e. selection and no selection in which selection would be associated with the imagination of a movement such as “move the arm” or “touch the screen”. Once the user is trained the integration of the BCI and the eye tracker is straightforward and therefore, could be easily tested.
- EEG patterns of intent to select and frustration will be recorded during a tracking session for a number of selected subjects. The objective is to create a database for further studies that contains signals recorded during an eye tracking session. The main goal is to recognize electroencephalographic patterns produced by the user which represent selection of the gazed element in the interface and frustration when he/she detects that the tracking is not working well and the focused element does not coincide with his gaze.

  For this step once the subject calibrates the tracking system he will be asked to accomplish a specific task such as “write the word hello” using a known virtual keyboard. The session will be conducted in such a way that the tracking will perform satisfactorily for the first part of the word leading the subject to “selection” state when he gazes ‘h’, ‘e’ and ‘l’. Then the tracking will be forced to fail for the rest of the word making the cursor to appear near but not exactly over the right letters creating “frustration” in the subject. The subject will not be informed of this change but he will perceive it as a tracking loss. In this manner, we could label the recorded signals as “selection” or “frustration” according to the part of the word he is working on. In addition a third class should be also included in the database indicating no selection and no frustration state.

References

Baccino, T., and Manunta, Y. (2005). Eye-Fixation-Related Potentials: Insight into Parafoveal Processing. *Journal of Psychophysiology, 19*(3), 204-215.

Graupner, S. T., Marx, J., Pannasch, S., and Velichkovsky, B. M. (submitted). Surprise, surprise: Two distinct components in the visually evoked distractor effect. *Vision Research*.

Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*(4), 188-194.

Jacob, R.J.K. (1990). What you look at is what you get: Eye movement-based interaction techniques. *Proceedings ACM CHI'90 Human Factors in Computing Systems*, pp 11-18.

Kaplan, A., Lim, J.J., Jin, K.S., Byeon, J.G. and Tarasova, S.U. (2003). The conception of intellectual brain computer interface: fundamentals and experiment. *1st Indian Int. Conf. Artif. Intelligence* (IICAI-03). (Ed. by B.Prasad.) IICAI, pp. 712-721.

Majaranta, P., Aula, A., and Räihä, K.-J. (2004). Effects of Feedback on Eye Typing with a Short Dwell Time. *Proceedings of Eye Tracking Research & Applications, ETRA2004*, San Antonio, TX, March 2004, 139-146.

Surakka, V., Illi, M. and Isokoski, P. (2004) Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception, 1*, 1, 40-56.

Schlögl A., Keinrath C., Zimmermann D., Scherer R., Leeb R. and Pfurtscheller G. (2006e, in revision) A fully automated correction method of EOG artifacts in EEG recordings. *Clin. Neurophys*.

Staker, I. & Bolt, R.A. (1990). A gaze -responsive Self -Disclosing Display. *ACM CHI'90*, 3-9

Velichkovsky, B.M. and Hansen, J.P. (1996). New technological windows into mind: There is more in eyes and brains for human-computer interaction. In *CHI-96: Human factors in computing systems*. NY: ACM Press.

Velichkovsky, B.M., Sprenger, A., and Unema, P.J.A. (1997). Toward gaze-mediated interaction: Collecting solution of the "Midas touch problem". In S. Howard, J. Hammond & G. Lindgaard (Eds.), *Human-Computer Interaction: INTERACT'97*. London: Chapman & Hall.

Session 3: Attentive Interfaces and Reacting to Intent

Gaze-based Attentive User Interfaces (AUIs) to support disabled users: towards a research agenda

Howell Istance
School of Computing
De Montfort University
Leicester, UK
hoi@dmu.ac.uk

Aulikki Hyrskykari
Dept. of Computer Sciences
University of Tampere
Finland
ah@cs.uta.fi

Daniel Koskinen
Dept. of Computer Sciences
University of Tampere
Finland
daniel.koskinen@cs.uta.fi

Richard Bates
School of Computing
De Montfort University
Leicester, UK
rbates@dmu.ac.uk

Keywords
Attentive user interfaces, usability evaluation, research agenda

Introduction to current achievements

Gaze on the Desktop

The last twenty years have seen marked improvements in the usability of eye-based communication systems. Many limitations in earlier systems that have had an impact on usability have been overcome (Istance, 2006). There is now a much greater tolerance to head movement during calibration and use, which means that eye-based communication for groups of users with involuntary movements has become feasible. In addition the introduction of gaze-responsive screen objects[1] that respond directly to gaze means that the on-screen pointer can be dispensed with.

This leads to sense of direct interaction with objects rather than indirect interaction via the mouse pointer. This reduction in the need to control a cursor can be assumed to lead to a reduction in effort and workload. The rate of text entry has also increased with the advent of improved soft text entry and control devices such as Dasher and Gazetalk, which has lead to improved efficiency in the context of the usability of gaze-based communication systems. Finally, the advent of wearable gaze tracking devices will mean that more flexible and usable solutions to gaze-based communication problems can be found away from the desktop.

Away from the Desktop and into the Real World

One area integral to COGAINs view is the use of gaze-based communication for mobility and for environmental control. “One of the future challenges should be to make a computer so that you can drive the wheelchair (safely!) using only an eye tracker” is the view of a user with ALS who uses gaze-based communication regularly (Larsen, 2006). It is useful to consider distributing the focus of gaze-based communication from the desktop to the real-world. This implies interacting with objects in the real world by looking at them directly rather than only looking at an on-screen representation of them, and in turn this implies making real-world

[1] e.g. by Tobii Technology (http://www.tobii.se/).

objects gaze aware. Shi, Gale and Purdy (2006) introduce a technology (ART, Attention-Responsive Technology) aiming to allow objects in the environment to be controlled by gaze. The idea is to recognize the objects attended to by combining information of both the user's eye movements and the computer vision system identifying the objects in the user's field of view.

An essential caveat is that such technology should offer a real benefit to the disabled user group. This benefit is likely to come from more efficient interaction or reduced effort and workload during interaction. For some users moving to look at the object in the real world may be more effortful than looking at an on-screen representation of the same object. "If you can only use your eyes (as I can), you are not very mobile and it will be inconvenient to turn the wheelchair to look at the door, or whatever" (Larsen, 2006). For other users the ease of interacting directly with objects in the real world will doubtlessly bring considerable benefits. Further work will be required to understand the trade-offs between types of disability, workload and benefits of direct interaction with real-world objects. This area of research is related to activities in the field of ambient assisted living (AAL), where a persons activities are intelligently monitored in a domestic environment, for example. Successful application of these systems can enable people to continue living in their own surroundings rather than being obliged to move to specialist care environments.

Areas for improvement

Donegan (2005) identifies a number of areas for improvement based on an analysis of user requirements for gaze-based communications: (1) reducing the price of eye tracking systems; (2) improving the reliability and robustness of tracking; (3) enabling gaze selections to be accelerated and thereby reducing user errors; (4) allowing for multimodal input and output on gaze-controlled systems, and (5) increasing mobility. The likelihood of a low cost eyetracker being realised has increased by the establishment at ETRA2006 of the IPRIZE (IPRIZE, 2006), a Grand Challenge. The challenge is to produce a eyetracking system for $100 that meets or exceeds the performance specifications of currently commercially available (and far more expensive systems). Reliability and robustness is continually increasing with improvements in processing power, algorithms and camera and tracking technologies. These are important avenues to follow to enable command-based communication to be achieved by a much larger population of disabled users than is currently the case, particularly in the context of using desktop applications.

However even with a low-cost tracker and improved tracking there remain problems with accelerating selection and control actions. These problems are associated with the type of 'command-based' interface, where the interface requires deliberate selection and commands from the user, often by dwell click or other means.

Research directions for non-command based interaction

Looking at a broader set of objectives, there has been relatively little research into non-command-based interaction for disabled users. Jacob (1991) advocated using gaze for this type of interaction long ago, instead of deliberate, command-based interaction. Non-command-based interaction has re-emerged in recent times in the form of 'attentive' interfaces, although primarily for able-bodied users and applications.

Context for Action

An attentive user interface (AUI, Vertegaal, 2002) embodies the notion of monitoring what the user is attending to during interaction with a system (in a broad sense of the word). The system is able to offer the user commands based on what it thinks the user wants to do in the current context of interaction. This reduces the need for the user's attention to be distracted away from the primary task to first find and then issue a command. It also reduces the workload and effort associated with giving that command. Gaze position gives a good indication of which objects the user is visually attending to and these give important clues as to intent.

An overview of gaze-based attentive systems is given in (Hyrskykari, Majaranta and Raiha, 2005). The scope of attentive systems is not limited to desktop interaction. Vertegaal describes a means of instrumenting real world objects with Eye Contact Sensors (ECS) such that, within limits, these are aware when the user is looking at them. (Dickie, Vertegaal, Shell, et al., 2004). A headmounted system (Viewpointer) allows that the ECS devices are replaced with simple IR emitters and all of the sensing and computation takes place via the headmounted device (Smith, Vertegaal and Sohn, 2005). The outcome of inferences made by the system about the user's intent or interest can take different forms. The system can prepare suggestions for commands, which the user can ignore or accept, a principle that Smith, Vertegaal and Sohn (2005) refer to as 'context for action'. These are intelligent default actions. An important issue here is the way in which the commands are presented and the workload inherent in attending to, or ignoring, the suggested commands. For disabled users, the primary motivation for context-based actions is the quest to reduce workload and increase efficiency of interaction with applications, mobility devices and the local environment.

Automatic Action

An alternative mechanism to 'context for action' is that of 'automatic action', where the system acts on its assumption of user interest or need without further reference to the user. A good example to illustrate this approach is embodied in the iDict system (Hyrskykari, 2006), which is a system that aids the reading of foreign language texts. The readers' 'index of difficulty' in comprehending words in a piece of text is computed based on metrics associated with gaze behaviour. When the difficulty threshold is exceeded, a brief translation, or gloss, in the native language is displayed automatically over the word considered to cause difficulty. The more complete translation information can be retrieved by glancing into a panel on the right hand side of the screen (see Figure 1).

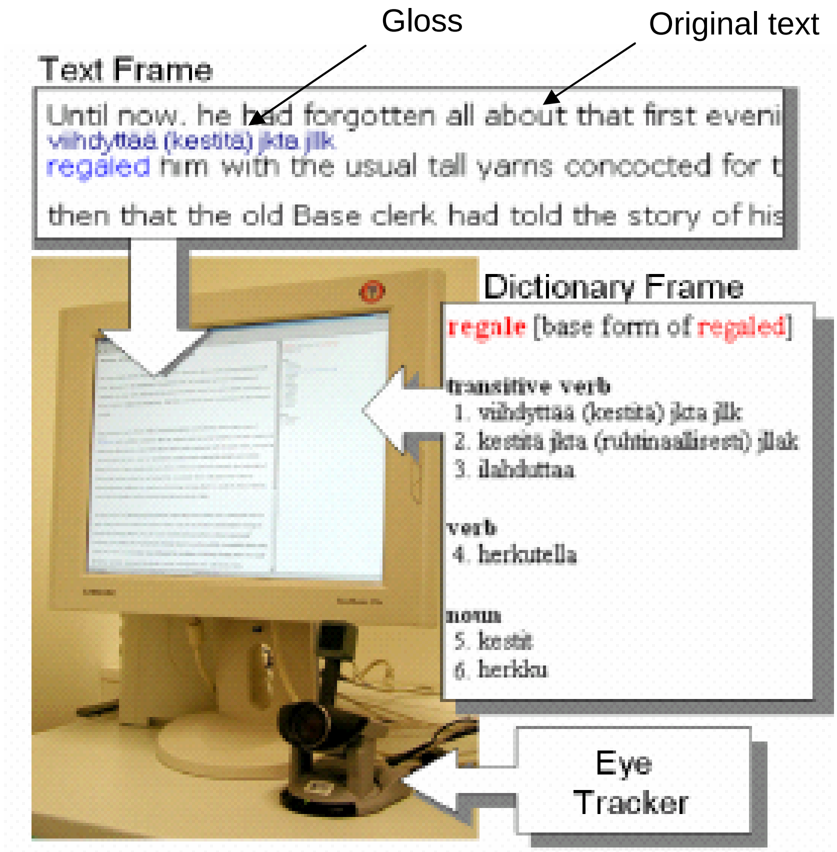


Figure 1: the iDict system showing a gloss presented within the text passage, with the fuller dictionary entry for the word on the right

iDict – an Automatic and Context-based Interface

As stated earlier, a prime motivating factor in providing attentive user interfaces for disabled users is the opportunity that they represent to reduce effort and workload in communicating the users' intent, needs and commands to the system. In order to investigate the potential workload reduction associated with the different paradigms for action, a usability study was made of the iDict system operating in three modes. The normal mode of operation 'automatic action' has the system deduce which words give rise to comprehension difficulties, and presents glosses for these automatically. This was compared with a 'context for action' mode. In this mode, a simple activating command, (pressing the space bar) was used instead of automatic action to present the gloss over the word considered to cause comprehension difficulties. Visual feedback (word became grey) was given after a delay to show that a gloss was available. A third manual condition (analogous to a conventional interface) was included in the study where the user selected which words they wished the system to present a gloss for directly by means of the mouse and without any reference to gaze behaviour. 18 able-bodied subjects, who were all native Finnish speakers took part and read 6 passages of English text taken from the same narrative. The passages were blocked into 3 blocks of 2, the three conditions (mouse only 'conventional', gaze and key 'context for action', and gaze only 'automatic') were presented one per block, and the order of the presentation of conditions was counter balanced across subjects. The first passage in each block was used as a practice session. The System Usability Scale (Brooke, 1996) was used to obtain Likert scale ratings of 10 statements relating to overall usability. Ratings of overall preferences were obtained after the trials. Furthermore, subjects were asked after the trials to indicate which words in the test passages they would have wished to be explained. This list of words per passage was then compared with the words for which the subject either requested the gloss for (context-based mode), or, in the gaze only condition (automatic mode), for which the system correctly anticipated the need for the gloss to be shown. Full details of the study can be found in Koskinen (2006).

Comparing Context-based and Automatic Interfaces

Here a subset of results will be presented. Two of the statements in the System Usability Scale are particularly relevant to the aims of reducing workload and effort, these are 'ease of use' and 'confidence in the system' (Tables 1a and 1b).

Tables 1a and 1b: Frequencies of responses to two statements from the System Usability Scale

I thought the system was easy to use	*Strongly disagree*	*disagree*	*No opinion*	*Agree*	*Strongly agree*
Automatic action (gaze only)	0	0	3	8	7
Context for action (gaze and key)	0	1	3	8	6
Manual (mouse only, no gaze)	0	1	0	6	11

I felt very confident using the system	*Strongly disagree*	*disagree*	*No opinion*	*Agree*	*Strongly agree*
Automatic action (gaze only)	0	1	7	8	2
Context for action (gaze and key)	0	4	2	9	3
Manual (mouse only, no gaze)	0	0	2	8	8

The data related to these two individual statements suggests that subjects prefer the conventional manual option for requesting presentations of glosses compared with the two attentive action options. This was not surprising given the prior level of familiarity subjects had with the different modes of operation, with all subjects being familiar with conventional mouse operation but not with the context and automatic modes as they had not used the iDict system previously. The overall preference ratings however do not show an overwhelming bias towards the familiar manual operation mode (Table 2).

Table 2: Frequencies of responses of overall device preferences

Overall preference	*Count*
Automatic action (gaze only)	5
Context for action (gaze and key)	5
Manual (mouse only, no gaze)	8

More subjects recorded a preference for one of the attentive interface modes than for the manual non-attentive option (Table 2). There is no clear evidence suggesting one attentive mode is better than the other. The number of actual glosses presented by the system compared with the number of those words that the subject considered to be sufficiently difficult to require a gloss when asked after the test were compared between conditions. This gives the 'hit rate' or level of correct determinations made by the system that a gloss was required by the subject. In the table below (Table 3), 86% of the words subsequently judged to be problematic by the subject after the test were glossed in the *automatic action* mode, and 14% were missed. 2.0% of the glosses were presented when the words were subsequently not considered sufficiently difficult to warrant this. In contrast, subjects only choose to invoke the gloss suggested by the system 74% of the time in the *context for action* condition.

Table 3: Percentages of correct and incorrect glosses

Number of glosses	*% correct hits*	*% false alarms*
Automatic action (gaze only)	86	2.0
Context for action (gaze and key)	74	0.3
Manual (mouse only, no gaze)	87	0.6

The results viewed collectively for this particular attentive application show no clear preference for either attentive mode of operation from a usability point of view, although there is some indication that automatic action is more efficient in providing assistance than the context for action mode, but it also produces more false alarms.

Designing attentive user interfaces for disabled users

The iDict case study has indicated that attentive interfaces have much to offer. These interfaces may offer an alternative and perhaps more suitable and effective method of communication and control for people with communication disabilities. Attentive interfaces present the user with suggested options, much as a human helper would do. Hence one approach to designing an attentive user interface would be to emulate a human helper, where the system attempts to guess what the user wishes to do in much the same way that a human helper proactively guesses the need of a user. The helper may list suggestions in what areas help may be required (much as the iDict interface offered glosses when required), and the user may indicate their agreement in different ways. Whether this is the best way to approach the issue of efficient communication remains to be seen. The user requirements produced by COGAIN (Donegan, 2005) form the platform for the investigation of those tasks within the domains of application usage, mobility and environment control for which attentive interface support may well be appropriate.

Attentive user interfaces have the possibility of reducing interaction and control effort for a wide range of users. They can assist users who would not normally need gaze control of applications, as well as those who do. Work is required to understand how best to use the opportunities that attentive interfaces offer for both of these groups of users, and in the case of those who do use gaze control we need to understand how best to switch the use of gaze between deliberate control mode and passive attentive monitoring mode. Developing attentive interface support for disabled users brings the benefits of gaze-based communication to a much larger population than currently is the case.

References

Brooke, J. (1996). SUS: A 'Quick and Dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, I. L. (Eds.) Usability Evaluation in Industry. London: Taylor & Francis, 189-194.

Dickie, C., Vertegaal, R., Shell, J. S., Sohn, C., Cheng, D., & Aoudeh, O. (2004): Eye contact sensing glasses for attention-sensitive wearable video blogging. In Proceedings of on Human Factors in Computing Sys-tems (CHI '04). ACM Press, 769-770.

Donegan, M. (2005): User requirements report with observations of difficulties users are experiencing, IST-2003-511598 (NoE) report, available through COGAIN pages at http://www.cogain.org/

Hyrskykari, A., Majaranta, P., & Räihä, K-J. (2005): From gaze control to attentive interfaces. In Proceedings of HCI International 2005. Erlbaum.

Hyrskykari A. (2006): Eyes in Attentive Interfaces: Experiences from creating iDict, a gaze-aware reading aid. Dissertations in Interactive technology, Number 4, Department of Computer Sciences, University of Tampere.

Istance H. (2006): Communication through Eye-gaze: Where we have been, where we are now and where we can go from here. Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA 2006). ACM Press.

IPRIZE (2006): http://hcvl.hci.iastate.edu/IPRIZE/ (accessed 16th June, 2006)

Jacob, R. J. K. (1991): The use of eye movements in human-computer in-teraction techniques: What you look at is what you get. ACM Trans-actions on Information Systems, 9, 152-169.

Koskinen, D. (2006): Comparison of three different input modes in a gaze aware reading aid environment (in Finnish). Unpublished manuscript for Master's Thesis. Department of Computer Sciences, University of Tampere.

Larsen, A (2006): Personal Communication

Shi F., Gale A.G. & Purdy K.J.(2006): Vision Responsive Technology to Assist People with Limited Mobility. In Proceedings of 3rd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT´04).

Smith, John D., Vertegaal, R., & Sohn, . (2005): ViewPointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. in Proceedings of the 18th annual ACM symposium on User interface software and technology (UIST '05:). ACM Press, New York, 2005, 53-61.

Vertegaal, R. (2002): Designing attentive interfaces. In Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA 2002). ACM Press: NewYork, 22-30.

Automatic Preference Detection by Analyzing the Gaze 'Cascade Effect'

Nikolaus Bee
University of Augsburg
Eichleitnerstr. 30
86135 Augsburg,
Germany
nikolaus.bee@gmail.com

Helmut Prendinger
National Institute of Informatics
2-1-2 Hitotsubashi,
Chiyoda-ku,
Tokyo 101-8430, Japan
helmut@nii.ac.jp

Elisabeth André
University of Augsburg
Eichleitnerstr. 30
86135 Augsburg,
Germany
andre@informatik.uni-augsburg.de

Mitsuru Ishizuka
The University of Tokyo
7-3-1 Hongo,
Bunkyo-ku,
Tokyo 113-8656, Japan
ishizuka@i.u-tokyo.ac.jp

Keywords
Preference detection, gaze 'cascade effect', two-alternative forced choice task

Introduction

The purpose of our work is to predict (visual) preference decisions of users in real-time, with the overall goal of designing systems that may recognize a user's choice of a particular visually presented stimulus in the presence of other stimuli, and respond accordingly. Our system, called *AutoSelect*, may automatically detect a user's visual preference solely based on eye movement data in a two-alternative forced choice (2AFC) setting. In a pilot study involving the selection of neckties, the system could correctly classify subjects' choice with an accuracy of 81%.

We believe that visual attention based interactive technology is of high relevance to various applications, including e-learning, future interfaces, as well as devices for handicapped people. In fact, many decisions of our daily life can be reduced to choices between several items, and cannot be easily explained in terms of overt reasoning on premises. In a restaurant, for instance, we choose between different types of dishes. Unless price or dietary considerations are of primary importance, our decision for a particular dish might be based on our taste, our expectation of a specific (eating) experience, or even our current mood.

The analysis of eye gaze patterns may provide an effective means to unveil non-conscious preference decisions of people. This paper describes our AutoSelect system that exploits the gaze 'cascade effect' and a recently conducted pilot study.

Gaze 'Cascade Effect'

When presenting pairs of human faces to subjects and giving the instruction to decide on their attractiveness, (Shimojo et al., 2003) observed a phenomenon they called gaze 'cascade effect'. This phenomenon involves the gradual gaze shift toward the face that was eventually chosen (as more attractive), while gaze bias was initially distributed evenly between the two presented faces. The results of the 2AFC task used in their study demonstrated a progressive bias in subjects' gaze toward the chosen stimulus (preference formation), which was measured by the gaze time spent on the selected stimulus. However, the strong correlation between choice and gaze duration occurred only in the last one and half seconds before the decision was made. A finding that (Shimojo et al., 2003) declared as surprising relates to the result that a larger cascade effect was found in the 'difficult' task, where the comparison between the attractiveness of faces was difficult, while intuitively, subjects were expected to more evenly distribute their gaze between stimuli in this case, in order to compare stimuli in as much detail as possible. The result was be explained by a theory claiming that gaze would significantly contribute to decision-making when cognitive bias is weak. The importance of this result for our research derives from the fact that a large number daily choices, e.g. regarding consumer products, are also deficient of a strong cognitive bias, and hence contributes to the importance of investigating non-conscious human decisions.

Pilot Study

A system that is able to automatically detect users' choices seems to break new ground. We therefore conducted an exploratory study using the AutoSelect system. Our first application is an automatic necktie selector, where subjects are shown a pair of ties and the AutoSelect system tries to detect the preferred tie. Subjects were given no instruction other than having to choose a tie for themselves or their friend for a graduation party.

We used faceLAB™ v4 from (Seeing Machines, 2005), a non-contact vision-based system with a sampling rate of 60 Hz. We implemented an algorithm based on the findings of (Shimojo et al., 2003), which detects visual preference in real-time.

Eight subjects (4 female, 4 male), all students or researchers from our institute, participated in our study. Subjects entered the experimental room individually and were provided written instructions about their task. Subjects were seated in front of a 20.1 inch display with attached infrared lights and their head and eyes were calibrated. This procedure had to be performed for each individual once, and took approximately 5 minutes. A session was initialized by subjects pressing a 'start' button in a web page based interface (see Fig. 1).

The following procedure was then iterated for 62 pairs of ties. First, a center located 'dot' was shown on the screen for 2.5 s in order to eliminate any initial gaze bias. Next, a pair of ties was presented, located to the left and to the right on the screen. In order to guarantee that subjects actually compare the ties, automatic selection was suppressed within the first 2.5 s. This value was based on the empirically determined decision time of 4 s in (Shimojo et al., 2003). After the system decision, the selected tie was presented and subjects were asked to indicate whether the system choice is correct by clicking on a 'yes' or 'no' button. Then the next iteration started with the initial view of a center dot. One initial set of 32 tie pairs was prepared, and the chosen ties were put back into the tie pool, which was used to create the subsequent set of 16 pairs, and so on. Eventually, subjects were shown a single pair of ties they presumably liked best. Hence, subjects were exposed to 63 pairs and performed 62 decisions in total. In the initial set of tie pairs, two partitions were created with 13 pairs each. One partition contained pairs of 'different' type ties, i.e. formal (decent) vs. 'entertainment' (adventurous) style ties, whereas the other partition contained 'similar' type ties that differed only in color or had a slightly different pattern but the same color. The motivation of this grouping was to investigate differences in subjects' decision behavior for presumably 'easy' vs. 'hard' decisions. All sessions were logged and lasted for about 10 minutes.

Results

The primary result concerns the classification accuracy of the AutoSelect system. In our study, the system was able to detect subjects' choices correctly in 81% of the cases. The worst recognition rate was 68%. Given a chance level of 50%, the system performed very well. (One subject was excluded from the analysis because of distorted values due to starting a conversation during the experiment.) We wanted to investigate the users' interactive experience with a running system, which can reveal e.g. issues related to the latency between user decision and system decision. Informal comments on the system indeed indicated that subjects were surprised about the system's reliability to timely identify which tie they liked more. Some of the misclassifications were related to a design problem, i.e. when subjects moved their face out of the camera range. The next version of AutoSelect should alert subjects in those situations. Furthermore we were particularly interested in results comparable to the 'difficult' vs. 'easy' choice finding reported in (Shimojo et al., 2003).

We hence compared recognition rates and decision times for 'different' vs. 'similar' tie pairs. Recognition rates were 75% (different ties) and 81% (similar ties); decision times were 6.8 s (different ties) and 7.65 s (similar ties) In line with (Shimojo et al., 2003), the decision time for similar ties was significantly longer than for different ties ($t(180) = -1.66$; $p = 0.049$). A one-tailed t-test assuming unequal variances was used in our analysis. This result supports the hypothesis that a choice between unlike items relies on (time consuming) cognitive processing, whereas similar items might be chosen based on non-conscious ('intuitive') preference. We also note that the system calculated the choice between similar ties more accurately.

Conclusion

We conducted a pilot study to test whether the AutoSelect system can correctly predict the choice of a user. The accuracy of the system with a limited number of subjects (seven) is reasonably high (81%). Future research needs in our case are two-fold: (1) We plan to improve the experimental design. In the current study, subjects are asked to confirm (or not confirm) the system choice, which involves the risk to receive 'polite' rather than 'truthful' answers (whether the system chose the preferrered tie correctly). Although this goes along with a situation that occurs i.e. in sales talks with vendors, we will obtain preference decisions predicted by AutoSelect and subjects' decisions independently in an upcoming study. In this way, the accuracy of the AutoSelect system can be estimated on a more solid basis. (2) A larger number of subjects will be included in the follow-up study.

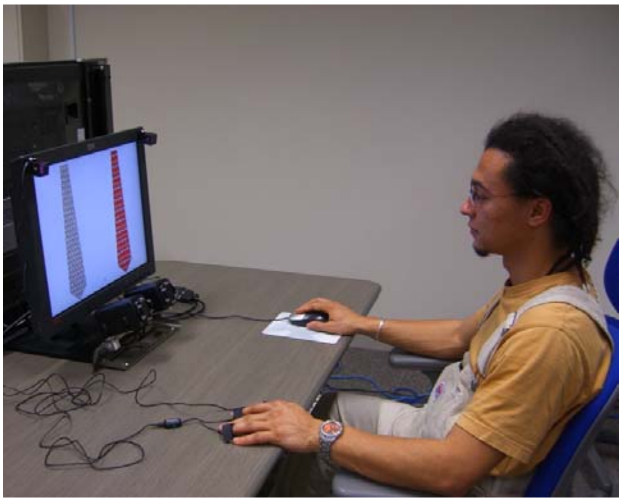


Figure 1: Experimental setup

References

Seeing Machines (2005) URL: http://www.seeingmachines.com/.

Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003) Gaze bias both reflects and influences preference. *Nature Neuroscience,* 6(12), pp. 1317-1322.

An Estimation of Certainty for Multiple Choice Responses using Eye-movement

Minoru Nakayama
CRADLE, Tokyo Institute of Technology
Ookayama, Meguro, Tokyo 152-8552 Japan
nakayama@cradle.titech.ac.jp
+81-3-5734-3234

Yosiyuki Takahasi
Interdisciplinary Graduate School of Science and Technology, Tokyo Institute of Technology
Nagatsuta, Midori, Yokohama 226-8503 Japan

Keywords
Strength of Belief, Multiple choice response, Estimation

Introduction

Communication is the sharing of a common belief about a topic between humans, or between humans and computers or robots [Iwahashi 2003]. During a dialogue consisting of human conversation or human-computer interaction, a person feels certainty in the sharing of a common belief if there is good communication. Otherwise a person feels uncertainness. This suggests that the performance of memory and learning are influenced by certainty and belief as with communication. We often try to voluntarily find further information when there is poor communication, and a typical human reaction is the unstable eye-movement associated with uncertainness [Underwood 2005]. This shows the possibility of making subjective evaluations of viewers using eye-movements. The phenomenon has already been applied to detecting relevant text during an information retrieval experiment [Puolamaeki et al. 2005]. However, the feasibility of measuring certainty using eye-movement patterns is not clear. In this paper we examine whether eye-movement based certainty can be an index of the "strength of belief (SOB)" for answers to a multiple choice task. The behavioral characteristics of eye-movement are also investigated.

Experiment

A multiple choice task, arrayed in a 4 by 4 matrix with 4 answer choices, was prepared as a full screen web page displayed on a 17 *inch* LCD monitor positioned 65 *cm* from the subject. Five subjects were asked to answer questions by selecting their choices using a mouse for the first minute (Figure 1a), and were then asked to review their own answers without making corrections during the second minute (Figure 1b). After the review, subjects noted their own subjective certainty as SOB on a scale between 0 to 100 for each answer. Each subject reviewed three sets of tasks so that in total 48 SOBs were reported.

During the experiment, subject's eye-movements were observed using a video-based eye tracker (nac:EMR-8NL). The subject rested his or her head on a chin rest and a small infra-red camera was positioned between the subject and the monitor, 40 *cm* from the subject. Eye-movement was tracked on a 800 by 600 pixel screen at 60 *Hz*. Eye-movements were divided into saccades and gazes. A gaze is defined as eye-movement staying within a 0.3 *degree* visual angle and at a velocity of 3 *degree/sec.* or less.

According to the definition, all gaze points were classified into 16 cells on a grid consisting of multiple choice responses to the questions, bordered by question items on the horizontal and vertical axes. The scan-paths between gaze points were extracted by analyzing the video record. In other words, the reviewing process is illustrated in a state transition diagram, where gazing cells are defined as states. To determine whether any eye-movement due to uncertainness appeared in this phase, two scan-path series between gaze points, namely the state transitions, were analyzed. If there was a scan-path between a multiple choice cell and a question item, the certainty for the answer was defined as low SOB. Otherwise the certainty for the answer was defined as high SOB. These two-class classifications were automatically calculated from the scan-path data.

Results

The "strength of belief (SOB)" was freely reported by subjects as an analogue value for each question. To examine the relationship between subjective reports of certainty and answer correctness for questions, the average reported "strength of belief (SOB)" was summarized as hit or false in Figure 2. There is a significant difference in SOB between hit and fault ($t(8)=7.3$, $p<0.001$). This suggests that subject's reports show the correctness of their answers.

Since the level of certainty estimated from eye-movement was high or low, subject's reports were also divided into two levels using a threshold as the overlap point of two normal distributions. These distributions were estimated from mean and standard deviations (SD) for hit and false responses by each subject, as shown in Figure 3.

Subject's SOB reports and eye-movement estimations were summarized in a contingency table (Table 1). When the estimation coincided with a subject's report, it showed that eye-movement can estimate SOB correctly. According to the analysis, the total rate of correct responses was 65.4%, and was significantly higher than chance ($p<0.05$).

The correct rate of estimation may depend on the threshold, which is the overlap point between hit and false responses. Therefore the rate was investigated in accordance with the threshold. When the threshold was adjusted from 0 to 100, the percentage correct changed with the SOB threshold, as illustrated in Figure 4. The correct rate decreases with the threshold for SOB. Figure 4 shows that a significant correct rate is obtained when the threshold is lower than 60%. This result suggests that a scan-path between a question item and an answer area appears when a SOB report is low.

Recall and precision rates are often used as evaluation metrics of discrimation performance (Jackson & Moulinier, 2002). A recall rate is defined as the percentage of correct hits, meaning that the estimation of SOB coincided with the subject’s report, based on the number of correct reports. The precision rate is defined as the percentage of correct hits per correct estimation. According to Table 1, there are two correct decisions, therefore recall and precision rates are calculated as rates for high SOB and low SOB respectively. The rates at the threshold are summarized in Table 2. These results show that precision and recall rates for high SOB are higher than those for low SOB. The algorithm, which detects low SOB using eye-movement, may affect the difference of performance between high and low SOB.

Signal detection theory is often used to evaluate discrimination performance and to discuss discrimination characteristics using *D'* [Palmer 2000]. In this paper, *D'* shows discriminability for high and low SOB. *D'*s are derived from both the hit rates (P_h) and the false alarm rates (P_{fa}) for high SOB, using subject's reports and estimations of eye-movements. Table 2 summarizes *D'*s for each subject across the subject's reports and eye-movement estimations. The results show that average discriminability of eye-movement estimation for high and low SOB is 34% of all subject's reports (0.35/1.05).

The performance of eye-movement estimation may increase when the accuracy of definition of the gazing point, or the algorithm for the detection of irregular eye-movement is improved. Those improvements will be topics of our further study.

Summary

To examine the feasibility of estimating the degree of "strength of belief (SOB)" of responses using eye-movement, the scan-path of eye-movements were analyzed while subjects reviewed their own responses to multiple choice tasks. If there is a scan-path between an answer area and the question item area, SOB is estimated to be low. In other cases, it is estimated to be high. Comparing subject's reports of high and low SOB and eye-movement estimations, a significant correct rate of discrimination was observed. When the threshold of discrimination was controlled, a high correct rate was obtained when the threshold was set to a low level.

These results provide evidence that "strength of belief" can be estimated using eye-movement. This technique can be applied to Web communication systems such as e-learning, or to detecting the willingness of subjects in evaluations or tests.

References

Iwahashi, N (2003) A method for the coupling of belief systems through human-robot language interaction. *Proc. of IEEE ROMAN* 2003, pp.385-390

Underwood, G. (2005) Eye fixations on pictures of natural scenes: Getting the gist and identifying the components, *in Underwood Ed. Cognitive processes in Eye Guidance*, Oxford University Press, UK.

Puolamaeki, K., Salojaervi, J., Savia, E., Simola, J., Kaski, S. (2005) Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval, *Proc. of SIGIR 2005*, pp.145-153, ACM Press, US.

Jackson, P. and Moulinier, L. (2002) *Natural Language Processing for Online Applications*, John Benjamins Publishing, The Netherland.

Palmer, S.E. (2000) *Vision Science*, The MIT Press, US.

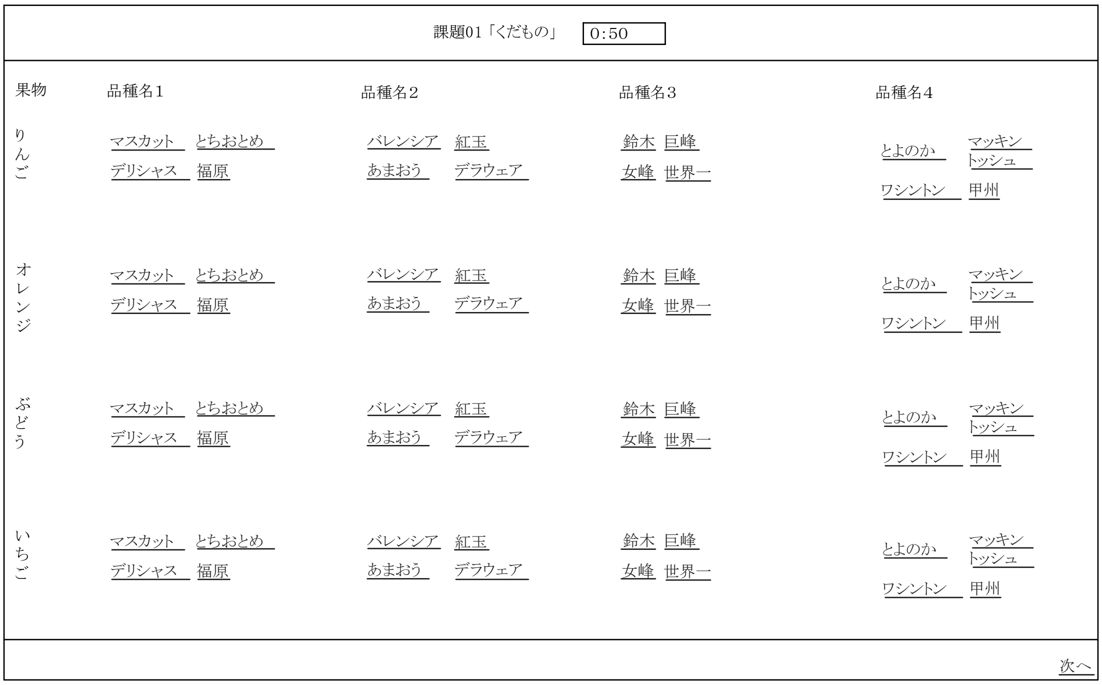


Figure 1(a). Example of a multiple choice task: multiple choice questionnaire.

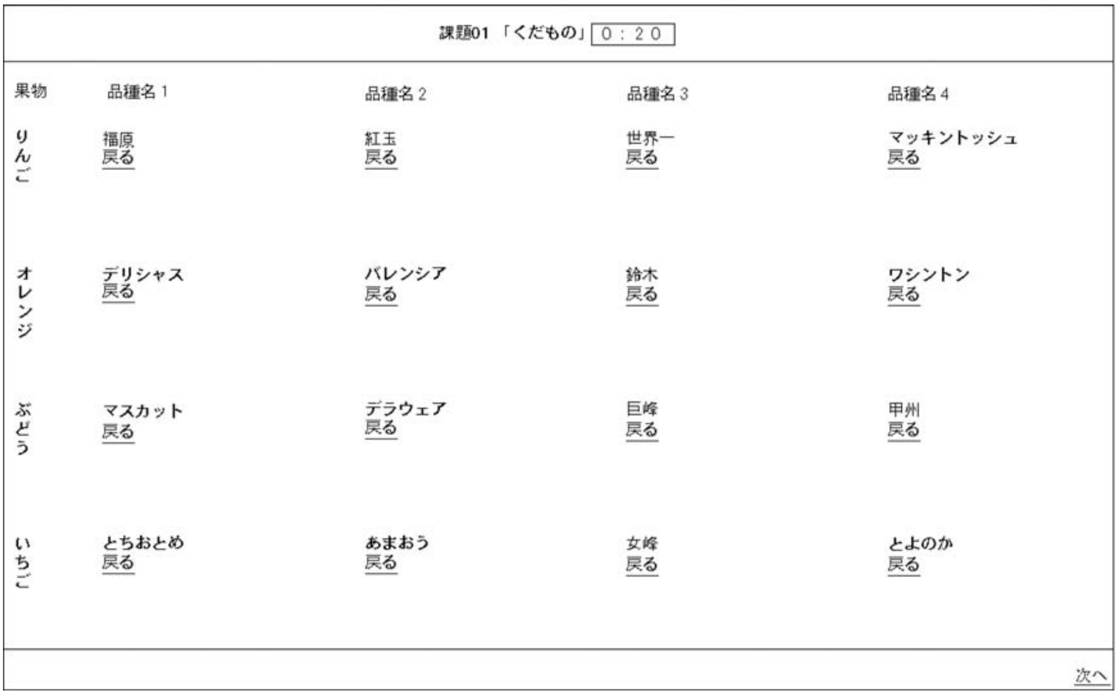


Figure 1(b). Example of a multiple choice task: answer selections.

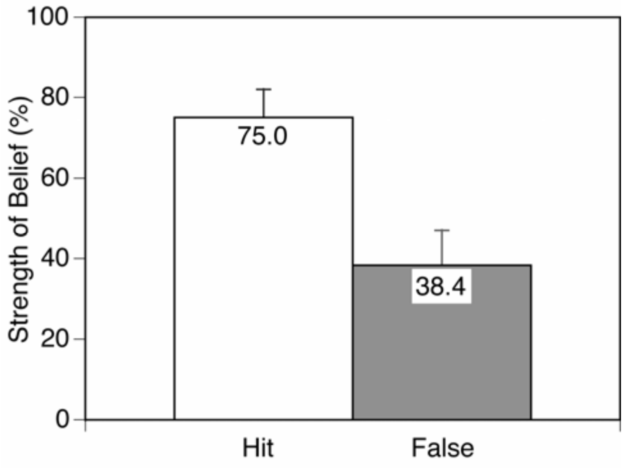


Figure 2. Mean Strength of Belief across Hit and False.

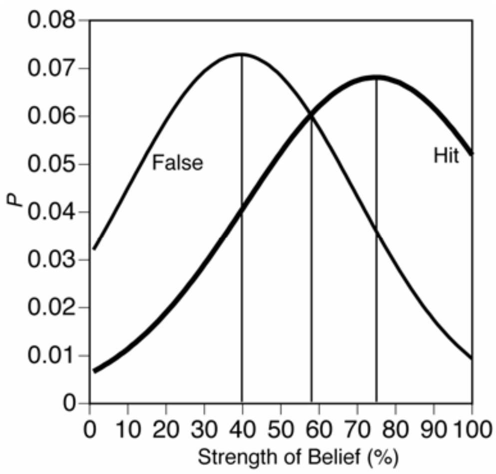


Figure 3. Mean Strength of Belief across Hit and False.

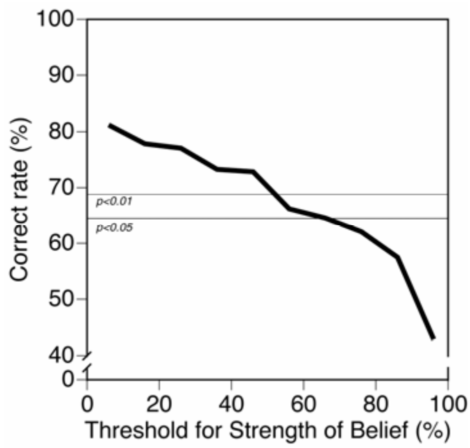


Figure 4. Correct rate of discrimination against threshold for SOB.

Table 1. Contingency table for the estimation.

	Subject's report	
Eye-Movement estimation	SOB [High]	SOB[Low]
SOB[High]	correct	miss
SOB[Low]	miss	correct

Table 2. Result of recall and precision rates.

	Precision	Recall
SOB[High]	63.5	79.2
SOB[Low]	55.5	35.6

Table 3. D' comparison between eye-movement estimation and subject's report.

Subject	*D'* [Eye-movement]	*D'* [Subject's report]
A	0.12	0.77
B	0.49	0.80
C	0.24	0.92
D	0.36	1.01
E	0.52	1.64
Mean	0.35	1.03

Gaze tracking for robotic control in intelligent teleoperation and prosthetics

Claudio Castellini
(corresponding author)
LIRA-Lab, University of Genova
viale F. Causa 13, 16145 Genova (Italy)
claudio.castellini@unige.it
phone +39 010 353 2946
fax +39 010 353 2948

Giulio Sandini
Italian Institute of Technology, Genova (Italy)
and LIRA-Lab, University of Genova (Italy)
giulio.sandini@unige.it
phone +39 010 353 2779
fax +39 010 353 2948

Keywords
Gaze tracking, robotic control, teleoperation, prosthesis, machine learning

Introduction

In the rehabilitation of amputees it is crucial to be able to *control* a prosthesis in such a way to reduce as much as possible the gap between the user's intentions and the response of the prosthesis. In particular, as far as prosthetic *hands* are concerned, the situation is rather poor. Recent advances in building sophisticated artificial hands have made it possible to give the user an artifact which, although still far from a real human hand, could potentially allow for complex grasping actions to be performed. Moreover, advanced tactile sensors can be fitted on the hand, so that a great deal of sensory feedback can be sent to the user's Peripheric Nervous System (PNS in short). An example is the *CyberHand* project [1].

However, it is hard for the user to effectively *control* the prosthesis. Direct connection to the PNS will result in poor control abilities and sensory feedback – the technology of invasive PNS interfaces still does not allow for severed nerves to be connected one-to-one to artificial sensors and actuators. Other solutions, such as non-invasive interfaces (e.g., monitoring the electro-myographic signal) and high-level commands issued via voice or buttons are of no help in reducing the above mentioned gap. Good artifacts can be built, but there is no way for the disabled person to control them effectively.

In order to overcome this gap, we believe one must put *intelligence* in prostheses. The computer system controlling the prosthesis must be able to *learn* and *adapt* to the user's needs, capabilities and feelings. Picture, for instance, grabbing and holding a pen in order to hand-write: this type of grasping is extremely delicate and precise, but still the grip must be strong enough to allow the pen to be held against the paper. Moreover, although the shape of a pen is similar to that of the handle of a hammer, for example, what one can do with it is completely different.

In the framework of the Neurobotics project [2] we are trying to improve hybrid bionic systems this way by employing machine learning algorithms. As a prototypical experiment, we are working on a *teleoperation* setup which, after a period of training, will eventually *guess* the user's intentions and correctly grasp a series of different objects placed in front of the slave robot; the technology acquired will then be transferred to a prosthetic hand with high dexterity (for instance, that being developed at the German Aerospace Center, see http://www.dlr.de/rm/en/). Intelligent teleoperation bears more than a casual resemblance to intelligent prosthetics: driving a prosthesis is like teleoperating a robot, only the slave lies in the very same place as the master. We envision that the prosthesis will gather sensory data from cameras[2], tactile and pressure sensors, in order to gain insight on the shape and affordance [3] of objects lying around, and in order to guess what it is supposed to grasp.

In this paper we describe how the *gaze* of the user also can be used in order to direct the robot toward the correct object on a table and guess whether the user is actually willing to grasp that particular object.

Teleoperation

A basic teleoperation setup consists of a *master*, by means of which movement/sensory data are gathered off a human user, and a *slave*, a robot acting according to the intentions of the user. Figure 3 shows the setup. The master consists of an Immersion *DataGlove* with 22 sensors, describing in real-time the position of the fingers and wrist of the user; an Ascension *Flock-Of-Birds* magnetic tracker, which tells us where the wrist is, in absolute coordinates; and, lastly, an ASL *E504* gaze tracker, telling us where the user is looking at.

Since the objects the master wants to grasp are in front of the slave setup (which is exactly the situation one would be presented with in intelligent prosthetics), we place a monitor in front of the user, showing the robot's point of view. The slave consists of two colour cameras mounted on a five degrees-of-freedom (DOFs) robotic head, a *PUMA200* six DOFs robotic arm, and a custom built dexterous humanoid hand, also having six DOFs. The communication is realised via YARP [4], a modular, abstract robotic control environment, which allows distributed computation and fast transmission of data through a standard network.

The control loop starts with the data gathered off the master: hand position (using data the magnetic tracker), shape of the grasp (using data from the DataGlove) and direction of gaze, coming from the gaze tracker. The user looks at the monitor and contextually moves his/her arm and hand, performing reaching and grasping actions. The master's data is sent to the slave, where it is interpreted according to the robot's geometry and kinematics, and immediately executed. (A simple inverse kinematics algorithm is applied to evaluate the robot joints positions.) The loop is then closed showing the user what the slave is doing in real time on the monitor.

The robotic control is, at this stage, still performed *in position*, that is, position data is translated to position commands to the robotic joints. This means the user has no control over the *speed* of the robot movements, which in some case can be awkward. However, thanks to a smart evaluation of the velocity profiles onboard the robot, the delay (of about half a second) is tolerable, giving the user a reasonable feeling of tele-presence.

[2] Light microcameras can be embedded in a pair of glasses, which would make the disabled person able to carry the system along.

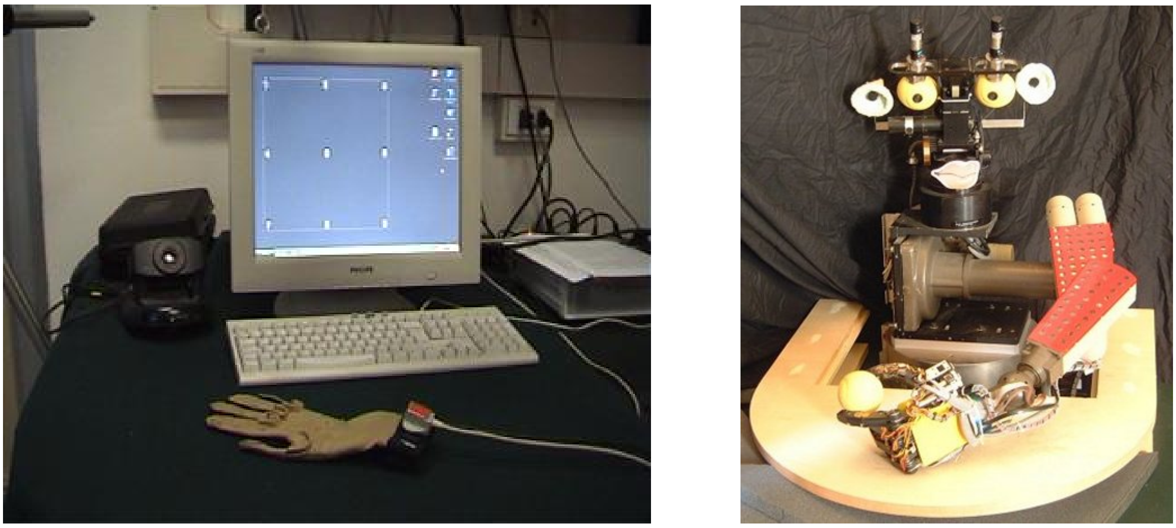


Figure 3: the master (left) and the slave (right) parts of the teleoperation setup.

Intelligent teleoperation

Basic teleoperation (what the master commands, the slave blindly executes) has several problems: mainly, it requires a high bandwidth, and it impossible for the user to correct every problem which might occur in the slave's setup. As an example, imagine the slave holding a mine while carrying it to a safe place, when suddenly a hole in the ground is met. In this case the user cannot react fast enough to prevent the mine to shake and possibly blow up.

In order to overcome these problems, we envision a learning machine to be put in the control system, able to react in place fast enough to compensate for such problems[3]. Even more interesting in this context is the ability of the slave to learn, from the real-time data coming from the user, *what the user wants to do*. For instance, the system could monitor a reasonable time-frame of data coming from the hand position and the gaze of the master; in most cases, *gazing* at an object and *moving the hand* towards that object means: *I want to grasp that object*. After a supervised training phase, the system would then learn to associate a certain speed of reaching, associated with the gaze fixation upon an object, with the *action* of grasping that object. In a second phase then, the system would be instructed, upon recognising a "grasp-that-thing" sequence, to de-activate basic teleoperation, grasp the object and then release the control to the user. The user would then have the feeling of the machine "having read his/her thoughts".

Notice that this schema overcomes the bandwidth limitation detailed above, since all calculations and learning would happen onboard the slave. The schema can also be extended toward more elaborate forms of learning, e.g., learning to grasp by recording the shape of the master's hand during the grasp, and associating it with the visual appearance of the object. This would aid the slave in grasping the object the right way automatically. Not incidentally, this is exactly what is needed in intelligent prosthetics.

[3] By *learning machine* we mean a software artifact based upon a machine learning algorithm, such as, e.g., Support Vector Machines [5].

Gaze tracking for saccades control

Gaze is a quintessentially voluntarily driven motion, and can be actively used to infer cognitive processes (see, e.g., [8] for a survey). Our slave's robotic head is currently able to *saccade* to a certain point in its viewfield. We have linked the gaze tracking device to the saccading mechanism in order to control the head's position by gazing. Having the slave gaze accordingly to the master has the beneficial effect of placing the objects of interest at the center of the viewfield: this enables us to employ *log-polar* vision [6], which greatly reduces the bandwidth needed by the transmission of images – in most cases, the bottleneck of the system. The system works in real time: we evaluate the mean and variance of the gaze signal over a carefully chosen time-window; once we find that the variance has remained "small" for the whole duration of the time frame, we guess that the master is *fixating* a point in the slave's view field, and therefore command a saccade toward that point. The coordinates of the saccade are gathered by considering the mean of the gaze signal, essentially the *center* of the "cloud" of the gaze data.

The above mentioned time-window is currently set at 400 milliseconds: in an extensive series of experiments on human adults, Johansson et al. [7] have shown that, during ordinary reaching and grasping tasks, (a) we always gaze at the *reaching and grasping points*, and never at our own hand; (b) we fixate the objects to be grasped for about 350-450 milliseconds, and then direct our hand toward the object. Therefore, it seems reasonable to instruct our system to do the same. The data reported in [7] is actually an average over nine human subjects; indeed, the time-window necessary for deciding when to saccade varies from person to person. Therefore, we plan to extend the learning mechanism in this sense too: to adapt the time-window to the needs and will of the user.

Conclusion

The Neurobotics project is due at the end of 2007. By that time, we plan to have accomplished the two phases sketched above: (a) setting up and testing the described learning machine on the intelligent teleoperation setup, and then (b) migrating the system to a dexterous robotic hand, which will then be worn by an amputee. Gaze will therefore be extensively used to understand how a disabled person can realise a better control of his/her artificial hand, and reduce the frustration gap induced by the poor chances of a direct control.

Acknowledgments

This work is supported by the Neurobotics project, FP6-IST-001917. We wish to thank Claes Von Hofsten for lending us the gaze tracker, and Giorgio Metta and Francesco Orabona for invaluable discussions and support.

References

[1] The CyberHand project, IST-FET-2001-35094. See *http://www.cyberhand.org*
[2] The Neurobotics project, FP6-IST-001917. See *http://www.neurobotics.org*
[3] Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston.
[4] The YARP project. See *http://yarp0.sourceforge.net*.
[5] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[6] *Anthropomorphic Visual Sensors*. F. Berton, G. Sandini, G. Metta. In the Encyclopedia of Sensors. Ed. C.A. Grimes, E.C. Dickey, M. V. Pishko, American Scientific Publishers. Volume X, pp. 1-16.

[7] Roland S. Johansson, Göran Westling, Anders Bäckström, and J. Randall Flanagan, *Eye-Hand Coordination in Object Manipulation*, The Journal of Neuroscience, September 1, 2001, 21(17):6917-6932

[8] Mary Hayhoe and Dana Ballard, *Eye movements in natural behavior*, TRENDS in Cognitive Sciences, Vol. 9 No. 4, 2005.

Eye Tracker Input in First Person Shooter Games

Poika Isokoski
TAUCHI
Department of Computer Sciences
University of Tampere
FIN-33014 University of Tampere, Finland
poika@cs.uta.fi

Benoît Martin
LITA
University Paul Verlaine - Metz
Ile du Saulcy
57045 METZ CEDEX 1, France
benoit.martin@univ-metz.fr

Keywords
Eye tracker, game, input device, first person shooter, aiming

Introduction

We report ongoing work on using an eye tracker as an input device in first person shooter (FPS) games. In these games player moves in a three-dimensional virtual world that is rendered from the player's point of view. The player interacts with the objects he or she encounters mainly by shooting at them. Typical game storylines reward killing and punish other forms of interaction.

The reported work is a part of an effort to evaluate a range of input devices in this context. Our results on the other devices in the same game allow us to compare the efficiency of eye trackers as game controllers against more conventional devices. Our goal regarding eye trackers is to see whether they can help players perform better. Some FPS games are played competitively over the Internet. If using an eye tracker gives an edge in competitive play, players may want to acquire eye tracking equipment. Eye trackers as input devices in FPS games have been investigated before (Jönsson, 2005), but that investigation focused on user impressions rather than on the efficiency and effectiveness of eye trackers in this domain. However, Jönsson's results on eye tracker efficiency in a non-FPS game were encouraging.

The Game

Rather than using an existing game engine for our experiment, we contracted a student group to build a new one. A new game was necessary, because we wanted the source code to be very simple and compact so that it would be easy to modify. It was possible to avoid most of the complicated code because we did not need network or mulpiplayer capability, artificial intelligence for game creatures, or special techniques to speed up the graphics rendering. For input device experiments, unintelligent targets are better because they make the experimental situation more controlled. In experiments we can use powerful hardware instead of clever coding to keep the frame rate high enough.

The game world consisted of a square area covered by randomly generated hills and valleys with randomly placed trees and tufts of grass. The randomness helps to avoid map-specific bias in the results.

A screenshot from the game is shown in Figure 1. The task of the player was to move in the world and shoot as many targets as possible. The targets were round plaques with a portrait of a penguin on them. The targets moved slowly along the terrain to make hitting them at least moderately difficult. In our experiments the targets did not shoot back. Our purpose was to focus on the efficiency of moving and aiming. The efficiency of aiming while evading enemy projectiles was left for further work. Whenever a target was hit, it disappeared, and another was generated at a random location in the world.

Figure 1. A scene from the penguin hunting game.

When planning the use of an eye tracker in a game like this, one has to find a way that does not interfere with the use of gaze for acquiring information about the world. This excludes the use of gaze as a simple pointer replacement for controlling the player's movement on the map. It may also be important to be able to move to all directons while freely observing the scene. Another possibility for eye-tracker use is to contol the direction where the player is facing. However, we found that the need to do swift 180 and even 360 degree turns is frequent because one needs to survey the environment to find targets. We did not find a natural way of doing this with the gaze. Also, a direct mapping of the gaze position to camera orientation would lead to either a jittery display or slow movements due to averaging in order to avoid the jitter (Jönsson, 2005).

Finally we decided to use the gaze for aiming the weapon within the scene shown on the display. The mouse in the right hand was used for controlling the camera angle, and the left hand operated the arrow keys on the keyboard for moving the player around in the world. The white crosshairs in the center of the display showed where the player was facing and acted as the only aiming device when the eye tracker was not used. When the eye tracker was used, the red reticle (pointed at by the white arrow for the benefit of grayscale printout readers) showed where the player was looking. Shooting to the position of the white crosshairs was possible by pressing the left mouse button, and shooting to the position of the gaze-controlled red reticle was possible by pressing the right mouse button.

The advantage that we envisioned this setup might have over the conventional keyboard and mouse setup was that aiming with the gaze should be faster than aiming with the mouse. We thought that this would be a significant advantage in situations where the player reaches a top of a hill or steps out from behind a tree so that several targets are revealed. It should be possible to shoot the targets rapidly with the combination of gazing and pressing the mouse button.

Note that performance in this kind of scenario is very important in FPS games. The player is often thrown into rooms full of unfriendly creatures and the only way to survive is to aim and kill fast in order to survive.

The disadvantage of aiming with the gaze is that aiming over long distances is difficult because of the accuracy issues with eye trackers. Because there are advantages and disadvantages in the use of eye trackers in FPS games, and theoretical answers to player performance were hard to come by, we decided to approach the problem empirically.

Results

Our experimentation is in early stages. So far we have data for only one of the authors playing 10 five-minute sessions using a Tobii 1750 eye tracker. The results are shown in Figure 2. For comparison we show the last 10 sessions of a 30-session trial completed by the same player with other input devices. These devices are keyboard and mouse without the eye tracker, and the Xbox 360 controller. With the Xbox 360 controller the left stick was used for moving, and the right stick for aiming. The shoulder buttons were used to shoot.

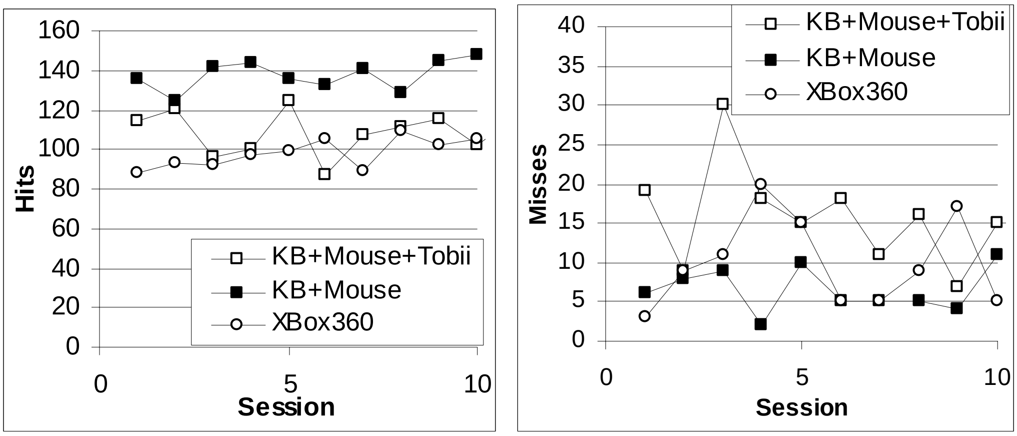


Figure 2. The number of hits and misses per session for three input device configurations for one player.

It appears that using the eye tracker does not improve performance for this player in comparison to the keyboard+mouse condition. However, the keyboard+mouse+tobii condition performs on the same level with the Xbox 360 controller. We should notice that the player had more training with the conditions without the eye tracker. Therefore, with continued training the performance with the eye tracker might improve more than the performance with the other devices. While fairly flat, the data in Figure 2 do not prove that no learning happens. Games are often played for longer than the 50-minute period shown in Figure 2. Even slow improvement adds up in the course of hundreds of hours of training.

Conclusions and Further Work

While it appears that adding eye-tracker support to FPS games will not always improve player performance, we find our preliminary results promising. For example, it may be that eye-tracker input can improve performance with input device configurations other than the keyboard+mouse combination. By the time of the conference, we hope to have more data to report on combining eye tracker with other input devices. It may also be possible to design an eye tracker based input device configuration that allows disabled users without the manual dexterity to aim with a mouse or a gamepad have a satisfying gaming experience.

Acknowledgments

The game was developed by Laurent Gomilla and Maurice Svay, who we thank for their excellent work.

References

Jönsson, E. (2005) *If Looks Could Kill – An evaluation of Eye Tracking in Computer Games*. Master's Thesis, Royal Institute of Technology, Stockholm, Sweden.

Session 4: Users and Usability

The 'KEE' Concept for Eye-Control and Complex Disabilities: Knowledge-based, End user-focused and Evolutionary

Mick Donegan
The ACE Centre
92 Windmill Road, Headington, Oxford, OX3 7DR
donegan@ace-centre.org.uk

Lisa Oosthuizen
The ACE Centre
92 Windmill Road, Headington, Oxford, OX3 7DR
oosthuizen@ace-centre.org.uk

Keywords
User Perspectives; User Trials; Usability; Case Studies

Introduction

The ACE Centre is using a case-study based action-research methodology to investigate conditions for successful use of eye control technology by users with complex visual and physical difficulties. By comparative analysis of the progress of a group of participants, a number of conditions are emerging which can enhance even the most complex users' chances of successful access to this technology and optimise its usability. Eye control is a not only a relatively new technology, but also a completely new skill for anyone trying it for the first time, whether that person has a disability or not. It is important to remember this when attempting an eye control trial if one is to make it as successful and positive an experience as possible for both the participant and the researcher. This paper will discuss some of the results of our user trials and provide recommendations to enhance the chances of complex users achieving success from the outset.

The Calibration Process

A range of factors need to be considered when starting a user trial including the expectations involved, the environment and the calibration process. Starting the trial with an explanation and, if necessary a demonstration of the process the user is about to experience can help all parties involved feel more at ease and also provides an opportunity for concerns and questions to be dealt with. There are likely to be many people involved in the care/education of complex users, many of whom would be interested in attending a user trial. However, the more people who attend, the greater the potential pressure on the participant and this can create a more stressful environment.

A range of environmental and situational factors, such as anxiety, sound, background noise, lighting conditions, can make calibrating a far more difficult process than for people without complex difficulties. Nonetheless, it has been our experience that, with good preparation and planning, a customised trial – designed with the particular user in mind – can increase the chances of a successful calibration. Under such conditions, the chances of achieving a successful calibration for the user are much greater and this provides a good foundation for the rest of the trial. Through customised planning with the needs of specific users in mind, we have achieved 'usable' calibrations with even the most complex of users. This paper describes some of the ways in which success has been achieved through creating a relaxed, yet focused atmosphere in order to achieve a calibration that we are confident will see the user through their trial.

Progression is Key

There are various important factors to consider when deciding on which activities to present to a user during an eye control trial. It has been our experience so far that many of the physically complex users involved in our trials also have complex visual/perceptual difficulties which may not be fully diagnosed or understood. For this reason, the trials begin with a range of very undemanding activities designed simply to gain an insight into what our participants are able to see and understand. Even for those users who might have been ready to start eye-typing immediately, there is no harm in starting with a less demanding activity initially, making the participant feel confident and relaxed. It must be remembered that, even for fully literate, visually and cognitively able users, controlling a cursor with their eyes is still a new skill. At the other extreme, many users involved in our trials have complex difficulties that could be exposed by their use of eye control. The concept of dwelling on an item in order to make a selection is unfamiliar to most people and it can be very disconcerting to make unexpected and/or unwanted selections. Not being familiar with software or knowing how to navigate from area to area can also be confusing and/or frustrating. For many of the people we work with, increased emotional anxiety is very closely linked to increased involuntary physical movement, which then affects eye control and that in turn becomes frustrating and less successful – clearly a vicious circle.

It has become apparent during many of our trials is that eye control is often suggested when almost all other options have failed or have become impractical. Because of this, expectations can be high and, even when it is made clear that any work being done is project-related and of an exploratory nature (rather than an assessment of any kind), there can still be very high expectations of the trial. If the trial is unsuccesful, therefore, it is important to make it clear to those involved that it is not the fault of the user. Rather, it is the fault of the currently available technology and/or software at this moment in time and <u>not</u> the fault of the user. If this view is understood, there is a greater chance of the user retaining a positive attitude for any future trials with this technology. At the same time, as members of a key European project, we are in an excellent position to report calibration difficulties to researchers and developers and to collaborate in order to overcome them. This is a <u>key</u> benefit of the opportunity offered by COGAIN of involving clinicians to evaluate and represent User Requirements.

A carefully planned, progression of personalised activities can enhance the chances of a successful and positive experience for the user during their trial. During our initial trials, we regard it as our role to explore whether or not eye control can be achieved and, if not, adjust or modify the software and/or hardware available us to enhance the chances of success in subsequent trials. If necessary, this might involve working in partnership with manufacturers and developers. As a result, changes can be made with the aim of making the system more accessible to specific users. In the best traditions of action research methodology, this is an 'evolutionary', cyclical process where each modification or set of modifications is re-trialed with the user involved with the ultimate aim of acquiring a personalised eye-control system for the user involved.

We refer to this process as the 'KEE' approach to trialing and implementing eye control technology:

- Knowledge-based - founded on what is known of the user's physical and cognitive abilities.
- End-user focused - designed to meet the end-users' interests and needs.
- Evolutionary – ready to change in relation to the end-users response to eye-control technology and software provided.

As a result of this 'KEE' approach, a different way of unlocking the door to eye control technology can be found for each user. One user, for example, might prefer to use an on-screen cursor/pointer and to access Windows directly, whilst another might need or prefer to control the computer via a grid. The examples below, for example, show two sharply contrasting 'template screens'.

One user, for example, might prefer to access the computer by direct control over the cursor/pointer, ie by using their eye(s) to carry out the same function as a hand controlling a mouse. Instead of a real keyboard, he/she might prefer to use a small on-screen keyboard. In this way, he/she can access and use the computer in a very similar way to everyone else.

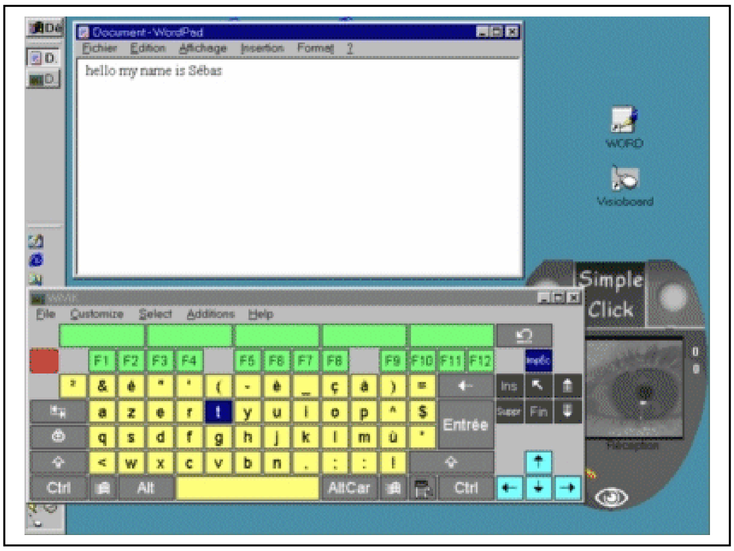


Figure 1. Metrovision on-screen keyboard

On the other hand, other another enduser might require a a completely different 'KEE' approach. A user who has a visual difficulty, for example, might not be able to control an on-screen pointer satisfactorily using a standard on-screen keyboard, so he/she might needs to use an interface in the form of an on-screen grid with large cells and prediction. This is a sytem developed over a period of time for Michael (see Case Studies below).

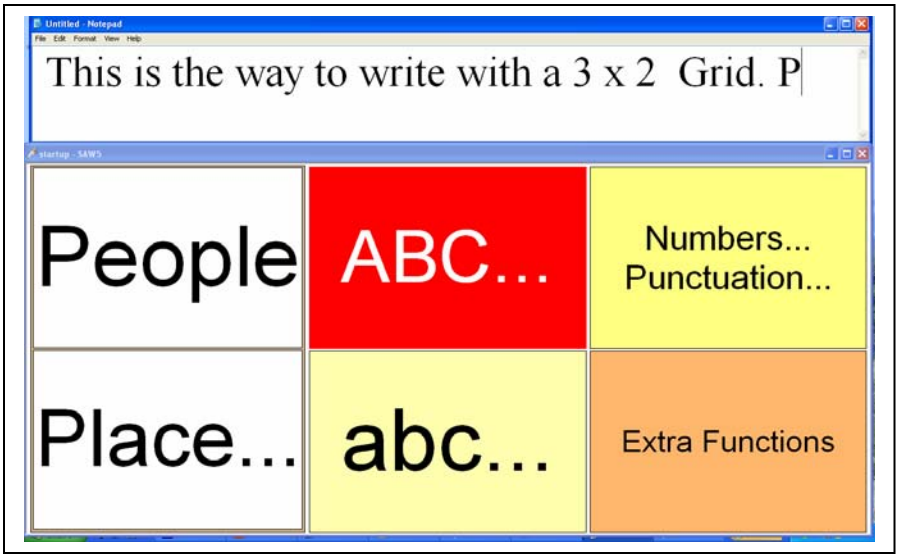


Figure 2. Michael's writing grid

Considerations when taking the 'KEE' approach include the following:

- An appropriate eye control system(s) that accommodates the users physical and visual/perceptual needs.
  - *ie. a system that is appropriate for the user. For some, a system that is able to accommodate involuntary head movements might be required*

- Appropriate mounting and positioning of the system in relation to the users needs.
  - *ie. the sytem must be positioned in a position for optimal comfort, function and visibility for the specific user.*

- Appropriate on-screen visual representation (pictures, symbols, text, foreground/background colours, etc.)
  - *ie. how should visual images be presented in a way that is clearly visible and comprehensible to the user?*

- Appropriate organisation of the images on the screen in relation to the visual abilites of the users (eg. visual scanning ability, range/direction of eye movement).
  - *How should the visual images be arranged in a way that is most appropriate for the users*

- Appropriate auditory support and feedback.
  - *What kind of additional support should be provided to provide feedback to the user? (eg. Should the system speak out the symbols, letters, words or pictures?)*

In some cases, this can take months and even years, depending on the levels of complexity involved and the speed with which the available hardware and software can be adapted or modified. Initial trials with eye control for both Helen and Michael, for example (see below), were unsuccessful due to the levels of complexity of their physical/visual difficulties. Under the COGAIN project, it is *not* regarded as our role to establish users' levels of cognitive and/or visual perceptual disability. However, when designing/planning activities for a user trial, it is necessary to gather any information that might enhance the chances for a successful initial trial.

Results of the 'KEE' approach – Case Studies: Michael and Helen

Michael (43) first tried eye control about two years ago. Due to his nystagmus and involuntary head movement, calibration was extremely difficult. However, by working closely with him and adapting some existing software to suit his abilities, he now has his own eye control system that he uses to write with successfully. (See Michael's writing grid above.)

Figure 3. COGAIN Participants, Helen and Michael

Helen (9) had never been able to successfully access any technology independently. When an opportunity to use eye control was initially offered to her, she found it impossible to use because it was not sufficiently personalised to her needs and interests. However, as a result of taking a KEE approach over an extended period of time that involved adapting software to match her specific needs and interests, she has progressed from not being able to use eye control at all to managing to write emails independently.

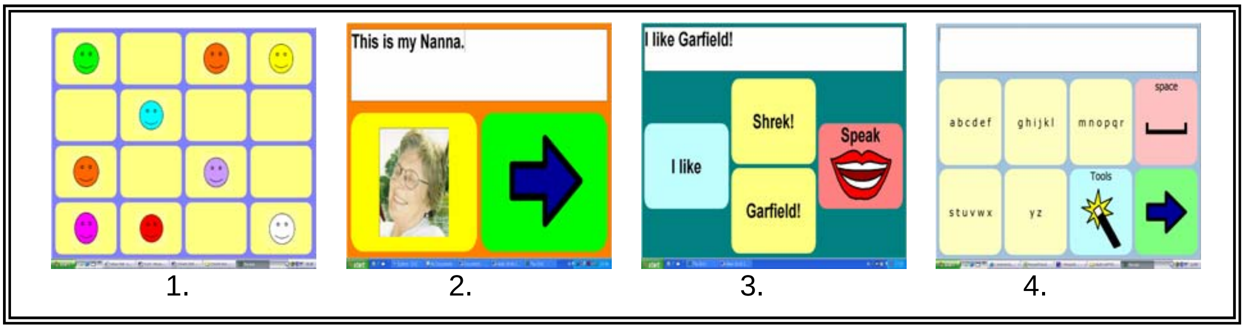


Figure 4. Helen's software progression using the KEE approach.

Gaze Path Playback Supporting Retrospective Think-Aloud in Usability Tests

Merja Lehtinen
University of Tampere
FIN-33014 Tampere
merja.inkeri.lehtinen@uta.fi

Aulikki Hyrskykari
University of Tampere
FIN-33014 Tampere
ah@cs.uta.fi

Kari-Jouko Räihä
University of Tampere
FIN-33014 Tampere
kjr@cs.uta.fi

Keywords: Eye tracking, usability, think-aloud, retrospective think-aloud.

Introduction

The think-aloud method is the most common technique to collect information in usability tests (Nielsen, 1993; Boren and Ramey, 2000). It is a way to get insight of the user's cognitive processes during the use of a product. Even though think-aloud still is the predominant usability testing method, its shortcomings are well known. Many users find thinking aloud difficult and it makes them feel uncomfortable (Nielsen, 1993). Since we think much faster than we are able to verbalize our thoughts, "thinking aloud" is actually an unreasonable demand. Consequently, the verbalized functions are imperfect. In addition, thinking aloud probably affects the user's task performance (Nielsen, Clemmensen & Yssing, 2002; van Someren, Barnard & Sandberg, 1994; Guan, Lee, Cuddihy & Ramey, 2006). An obligation to verbalize the performed processes may slow normal behaviour with the product and even change the steps of execution from those the user would take in a normal situation.

One of the suggested ways to avoid these problems is to use the think-aloud method retrospectively. The user is permitted to carry out the given tasks without an obligation to think aloud, which hopefully makes the interaction with the product more natural. After the task the user gives a verbal report of the task session. The usual way is to present the user a playback of the task session during which the user explains what s/he was thinking at the time. The playback includes a video of the screen, with possibly an inserted video of the user. When the retrospective report is given right after the task session, the user still has part of the information in short-term memory, and the supporting playback helps the retrieval of information from long-term memory (Ericsson and Simon, 1993).

So far, the use of eye tracking for assessing the usability of a product has not been as fruitful as one could expect (Jacob & Karn, 2003). Interpreting the data is intricate. Even though the data reveals how the user's focus of visual attention varies during a task session, it is difficult to deduce the reasons for the user's behaviour. For example, a prolonged gaze to some widget does not necessarily mean that user doesn't understand the meaning of the widget. The user may just be pondering some aspect of the given task unrelated to the role of the widget on which the gaze happens to dwell during the time. Thus, gaze path information seems to call for the user's interpretation.

Hansen (1991) presented the idea of showing the gaze path to the user to assist in retracing the cognitive processes during the task session. He compared the quality of retrospective verbalization in the case the user was presented a video of the task session with and without an overlaid gaze path. Hansen found that the users were able to recall their thoughts during the task execution more precisely when a playback of the task session was supplemented with the overlaid gaze path.

It is not surprising that the shown gaze path helps the user to remember the steps and the thoughts during the task session. First, we were interested in whether it was possible to substitute the think-aloud method with retrospective think-aloud when the user is provided with the gaze path during the playback? Second, we thought that it would also be interesting to know if the usability problems found with traditional think-aloud differ from the ones found with retrospective think-aloud (when the task session playback is supplemented with the user's gaze path).

Experiment set-up

We performed a traditional usability test of a Finnish car brokerage web site with eight test users. We gave the test users eight tasks, varying from maintaining their own profile in the service to searching cars with certain properties. Half of the users were asked to think aloud. The other half performed the assigned tasks without verbalising their thoughts. The gaze paths of both user groups were recorded and both of the user groups were asked to give a retrospective verbalization of their recalled thoughts during the playback of the task session. The gaze path was overlaid on the video replay.

Data analysis

Consequently, we got verbal data on three different conditions: (1) from a traditional think-aloud session, (2) from a retrospective think aloud session, when think-aloud was used during the original task session and (3) from a retrospective think aloud session, when the original task session was carried out without obligation to think aloud. First we computed the word count of the verbal data recorded in each of the conditions. Then the *operational comments* (Hansen, 1991), i.e., the user's verbal expressions on behaviour or operations, were searched for in the data. After that we used the coding presented by Hansen to compare the quality of the think aloud data in the three different conditions. Hansen categorised the given comments to *manipulative*, *visual*, or *cognitive operations*.[4]

Manipulative operations are the ones expressing performance. Some examples of manipulative operations in our data are:

"I *write* the name into this field",
"Of course, I could have *clicked* all of those…", or
"Oh, I *gave* an erroneous input…".
Visual operations reflect perceptual operations, like
"I *saw* it here somewhere…",
"Then I *look* for a picture of the car…", or
"I *read* it from the previous page …".

[4] More elaborate categorizations appear in the literature; see, for instance, Guan et al. (2006). However, the simpler categorization points out the differences more clearly.

Cognitive operations reflect interpretations, evaluations, expectations and specifications, like
"I *remember* seeing it before",
"Now I finally *understand* that there is a scroll bar on the right", or
"I *found out* that I can't make a search from this page".

The results of the analysis are shown in Table 1.

Table 1. Analysis of verbal data in the three different conditions (four subjects in each group).

	Words	Comments	Manipulative	Visual	Cognitive
Think aloud	1148	66	82%	14%	4%
Think aloud + retrospection	3309	214	53%	14%	33%
No think aloud + retrospection	4136	267	42%	15%	43%

Discussion

We were interested in studying how the think-aloud method affects the observations and the gaze paths of participants in a usability test. By presenting the playback with an overlaid gaze path, we hoped to get the same information from the user that we would get with the think-aloud method. The preliminary analysis shows that retrospective think-aloud, when it is supported with gaze path playback, produced distinctively more verbal data than the original think-aloud method. The think-aloud session produced 1148 words and 66 operational comments. The retrospective think-aloud data after the original think-aloud session contained 3309 words and 214 operational comments. The retrospective think-aloud after a task session without think-aloud produced even more verbal data: 4136 words and 267 operational comments.

Already the increased amount of data received from retrospective think-aloud simulation encourages us to assume that retrospective think-aloud works better than conventional think-aloud. Because in this case the original task session can be performed in a more natural way without interruptions, we assume that the data corresponds to more genuine behaviour than the data obtained from the traditional think-aloud method.

It is also interesting to compare the quality of operational comments in each condition. When testing usability each operational comment is valuable, but especially the cognitive comments give the testers information which usually cannot be deduced from the user's observed interaction with the product. In the think-aloud condition the proportion of those comments is considerably low, while in the retrospective conditions the proportion is clearly higher.

Already these preliminary results justify assuming that when using gaze path playback, the retrospective think-aloud method may produce more and better quality data than the traditional think-aloud method. Next we are going to study the received data in more detail to get better understanding what are the typical differences of the comments received in each of the conditions.

Future work

Users with disabilities are not able to verbalize their thoughts in a manner needed for applying the think-aloud method. For them, it would be crucial that the usability problems could be, as much as possible, detected from the gaze path alone – without the need for think-aloud. We would like to work with the data we have collected and give it to some usability experts for analysis, to see what kind of usability problems they can detect from the gaze data alone. If successful, this could leave to a method for carrying out some form of usability analysis without the need of think-aloud at all: neither during the task session nor retrospectively.

References

Boren, M. T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication, 43 (3), 261-278.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data.* Revised edition. Cambridge, MA: The MIT Press.

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06).* ACM Press, 1253-1262.

Hansen, J. P. (1991). The use of eye mark recordings to support verbal retrospection in software testing. *Acta Psychologica,* 76, 31-49.

Jacob, R. J. K., & Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Radach, R., Hyönä, J., & Deubel, H. (Eds.) *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Oxford: Elsevier Science, 573-605.

Nielsen, J. (1993). *Usability Engineering*. Cambridge, MA: AP Professional.

Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In Proceedings of the Second Nordic Conference on Human-Computer Interaction (NordiCHI 2002), 101-110.

Van Someren, M., Barnard, Y., & Sandberg, J. (1994). *The Think Aloud Method. A Practical Guide Modelling Cognitive Processes.* London: Academic Press.

Recent experiences, using eye-controlled communication systems with "locked-in" patients

Dr. Jeffrey D. Morris
Cardiff an Vale NHS Trust
RE Unit, Rookwood Hospital, Cardiff UK
jeff.morris@data2file.com
+44-(0)7770-971526

Mr. Gavin Cooper
Cardiff an Vale NHS Trust
RE Unit, Rookwood Hospital, Cardiff UK
Gavin.cooper@rehabeng.uhw-tr.wales.nhs.uk

Keywords

Eye movement control, locked in syndrome, communication

Abstract

The authors describe the assessment for and functional use of eye controlled computer systems with disabled subjects exhibiting "locked-in" syndromes, of multiple origins. In particular, the paper deals with the difficulties encountered in matching this high technology to severe and profound physical disability. It is stressed that the paper reflects upon the results of a collection of individual case studies with common goals and measurement systems, as opposed to a prefabricated, equipment trials process.

Introduction

It is often the case that the more profound the disability the more likely it is that high technology environmental control and communication will be based upon indirect access methods, often via a scanning interface. The two widely available exceptions to this "rule" often involve the employment of head and eye movement.

Until 1991 direct access to a graphical user interface (GUI) required that the user had the ability to access a touch-screen, to move a mouse, or to control a joystick. With the advent of head tracking systems, GUIs became accessible directly to those people whose disability had little, or no affect on their head control. This technological advance increased significantly the communication rate of people who had hitherto been dependent upon scanning or switch activated interface systems. However, this paradigm shift in access technology still excluded those whose disability compromised controlled functional head movement. In 1995 EyeGaze, an eye controlled system was introduced into the field of assistive technology. The system allowed eye movement bounded by limited head movement to be employed as a 'point and click' device within a customised software environment. In more recent years, this technology has expanded into multiple systems employing both open and closed software architecture, allowing generalised GUI access.

The contra-indications for provision of this technology can be ranked according to their level of negative contribution for each individual but will include; influences of the particular medical condition/disability, the environment in which the system is to be used, the specific eye movement measurement device, associated communication/control software and the available technical/carer support.

Materials

To date, four different eye movement measurement systems were used within the study. These comprised; Quickglance 1 & 2, VisioBoard and EyeGaze. These devices, theoretically, span the range of systems that are either limited by, or are independent of head/body movement.

In addition, one of the devices (EyeGaze) was modified to allow one user with injury related presbyopia the opportunity to operate the device. This involved the replacement of the system monitor with a data projector and screen at a distance of > 3 m from the user (the image roughly occupying the same visual field area as a monitor and thus the equivalent subtended angle of eye movement). All of the systems illuminate the subjects' pupil(s) with infra red radiation and after correctly focusing (achieved automatically in the case of the VisioBoard system) measure the change in pupil shape or position with reference to IR reflection corneal landmarks and translate this into a gaze direction vector, which is then mapped onto the computer screen, according to the functional translation formula:

$$f(x,y) = g(D.\tan ø + C, D.\tan ß+E)$$

where:
D is the distance from the eye to the camera
C/E are the vertical and horizontal offsets
ø is the vertical angular displacement
ß is the horizontal angular displacement

Subjects

Clients whose individual circumstances precluded the use of switch and scanning technology were considered for the trial process. These included four long term and ten short term trial assessments. Their primary diagnoses, all resulting in 'Locked in Syndrome' included; motor neurone disease (MND/ALS), brainstem cerebro vascular accident (CVA), and traumatic brainstem/spinal C1:C2 injury.

Each of the subjects had 'normal' levels of cognition, no apparent visual field disturbances (as reported by their optometrist) and were all literate. Two of the subjects required spectacles. In addition to the four disabled users, four “normal” subjects were included in the study so as to set a benchmark communication rate using the same software packages as the disabled subjects.

Method(s)

Prior to using the devices many problems with movement related artifacts had to be overcome. These unintentional movements can be catagorised into two sets: patient generated and life support generated.

Patient generated movement included; spasms, tremors, coughs, bruxim, low tonal displacements and transient or pathological myoclonic/opsoclonic activity.

Life support generated movements derive from; assisted/forced ventilation, mobility aids and carer intervention.

Further complications in the setup and use of eye controlled systems result from spurious specular reflections caused by tear drops, dry eyes, perspiration, spectacles and a range of background reflective surfaces. Eyelid droop due pathological/traumatic effects on the nerve supply and other medical complications can also reduce the efficacy of eye-movement measurement systems.

The gaze direction was measured for each user with a standard (commercially available) software that allowed the grouping of gaze, saccadic movements and dwell time to be assessed. As small variations in gaze direction are to be expected, the important parameter of functional gaze location was assessed by repeatedly adjusting the time threshold for activation for a mouse click. This provided not only a figure for activation for each subject but for those subjects with occulomotor control deficit, a measure of the maximum fixed gaze time.

A standardised test for communication rate was developed requiring that the users generate a short text passage (provided to them audibly), over which, either their time to completion was recorded or the time at which fatigue ended the session.

To date, with so few subjects the quality of expression was difficult to measure, particularly as the novelty of the devices allowed the users to 'suddenly' express basic needs and wants, however, the time to fatigue was measured over both the standardised test and free expression sessions, and each parameter is reported individually.

Results

It was seen that the “locked in” subjects were able to generate the text passage audibly provided, at a significantly higher rate than using the most appropriate switch and scanning system tailored for each individual, with a mean completion time of 16.47 ± 5.43 mins for the eye control as opposed to 34.1 ± 13.6 mins for switch and scanning.

Additionally, the eye-control systems appeared to cause less fatigue over the same text generation trial with all subjects completing the task.

Conclusions

A significant range of inter-related variables need to be considered when employing eye movement as an access and control system for people with profound physical disability.

The results obtained in this study show promise for the adoption of this technology (these technologies) in the communication/control arena. Indeed, when functioning correctly, these systems offer significant improvements in both communication rate and accuracy. However it was evident from the studies so far, that due to setup, calibration, adjustment and software problems, the level of technical support required to provide the desired level of communication and control is currently prohibitive for the most severe levels of disability.

In conclusion, whilst this technology shows great promise, greater emphasis will need to be paid on the reliability and ease of use of these systems, if those with the most profound levels of disability are to derive significant benefit from this technology.

Towards Remote Evaluation of Gaze Typing Systems

Hirotaka Aoki
IT Univ. of Copenhagen
Rued Langgaards Vej 7, DK 2300
Copenhagen S, Denmark
hirotaka.aoki@cogain.org

John Paulin Hansen
IT Univ. of Copenhagen
Rued Langgaards Vej 7, DK 2300
Copenhagen S, Denmark
john.paulin.hansen@cogain.org

Kenji Itoh
Tokyo Institute of Technology
2-12-1, Oh-okayama, Meguro-ku,
Tokyo 152-8552, Japan
kenji.itoh@cogain.org

Keywords
Eye-typing systems, user performance and preference studies, usability issues with gaze-based systems.

Introduction

Remote usability evaluations have been suggested as a new way to collect larger amounts of quantitative usability data in an efficient and ecologically valid way that allows the user to remain in their normal environment (c.f. Dray & Siegel, 2004). Remote data collection seems particular relevant for design of Augmented and Alternative Communication (AAC) systems for individuals with a mobility problem and for special user groups, like ALS patients that are relative few in numbers and thus spread over large areas.

Automated collection of usability data has been recommended by Lecerof & Paterno (1998). Ivory & Hearst (2001) provides an overview of methods and benefits. These data will be valuable because they may point to possible design improvements and they may serve as a window to the progress of a particular user. When data has been automatically collected, it may be sent off to the designers when the user is online. Monitoring user progress could also help communication specialists timing their advice and/or visits. For instance, it may be recognized when a user has reached a stable, but sub-optimal level of performance, and possible ways to improve could then be suggested to him.

User data from large populations would enable us to benchmark gaze typing interfaces or eye tracking systems against each other in actual use. Data collection from a significant number of users of a particular system reduces the influence from a few individuals with exceptional slow performance. On the other hand, if we were able to identify those few users with exceptionally fast typing, we could get very useful inspiration from them about how to help others.

Text data from people’s personal communication is confidential information. We therefore recommend calculating the performance metrics locally on the users PC and only submitting the metric values, not the text itself, in order to protect the privacy of the user. This paper will present new metrics for gaze typing systems that can be collected remotely and distributed without revealing the content of the users communication.

Measuring Typing Performance at a Distance

In remote real-life conditions we cannot control users' behaviour during typing. This makes it impossible to obtain metrics such as "Overproduction Rate" (e.g. Hansen et. al. 2004), that requires comparisons between the actual typed text and the optimal input stream for the target text. The user may type idiosyncratic words and abbreviations that will only be understandable to the people who know him well. Since we cannot compare this kind of personal text to any general dictionary standard, we cannot judge the quality of the productions by counting, for instance, the number of spelling mistakes. Even if we could, users differ enormously in their ability to spell. Whatever text the user produces, we will have to accept as the target text. On the other hand, the user will remove some of the errors that he recognizes, so the corrective action will be at least weakly correlated to the numbers of errors actually committed, we presume. But again, the carefulness by which people would actually correct errors is likely to be highly individual. Therefore, we should probably measure the changes in frequencies of corrections for a particular user instead of just the raw number of corrections.

Gaze typing metrics

Experimental evaluations of gaze typing are often based on metrics like word per minute (WPM), Keystroke per character, (KSPC) and error rates (e.g. Majaranta et. al, 2004, Hansen et. al. 2004, Ward & MacKay, 2002). If we would like to collect user data from larger populations outside experimental rooms, we need performance metrics that can be derived while the user types their own free text in real-life situations without the strict control of typing conditions that some of the traditional metrics require. Fortunately, gaze typing systems have a unique advantage when collecting data "at home" since it is known where the user is looking.

We suggest logging how many input keys the user looks at but never selects, termed *"Attended Keys Not Selected"* (AKNS), and to report this as the AKNS-rate, i.e. the number of AKNS per character entered. The definition of when an input character has actually been attended to can be based on the 125 ms minimum fixation time commonly used in the eye tracking research community (c.f. Jacob and Karn, 2003). AKNS does not include the character looked at for the final activation. It measures the attention overhead associated with production of a character, not just the overhead of every single keystroke, since it may take several keystrokes to produce a character – cf. the KSPC-factor. By adding the KSPC and the AKNS of a gaze typing system we then get the total number of keys attended per character produced.

The idea behind the Attended-Keys-Not-Selected-metric is inspired by the "Principles of Motion Economy" from Barnes (1949). Barnes provided guidelines for efficient manual work that conserve human energy. We suggest that this also applies to eye movements during skilled gaze typing. Although the "energy cost" of an eye movement is negligible, we believe that the cumulated cognitive cost of all unnecessary fixations during routine operations are very big. We expect this cost to be a strong predictor of long-term user satisfaction and perceived workload. We expect that AKNS will be related to error frequencies for gaze typing systems, since every key that is unnecessarily fixated, also risks to become a wrong selection, c.f. the Midas touch problem.

Calculating the AKNS- rate for Dasher is probably rather difficult because of the high information density in the input field. So for this particular system it may be more relevant to analyse saccade patterns of all input sequences and compare them with the optimal (i.e. shortest) scan path possible for the text that the user has entered. This is similar to the “Cumulative Deviation from the most Efficient Scan-Path” index, suggested by Aoki et al (2006).

The number of *Deleted Characters* can be measured without knowing what the user actually intended to type. It is a simple, but useful index to measure frequency of error correction per character. This index is closely related to the Number of Backspace Activations used by e.g. Itoh et al. (2006). Since it is not possible to measure the number of backspace selections in a gaze typing system like Dasher, which is operated by continuous navigation and not by single selections, we prefer to avoid the term “backspace activations” and just focus on any kind of deletions. Some systems provide an editing function that allows deletion of a full word. In this case we also suggest counting the numbers of characters in the full word deleted.

Outside the laboratory, the user is likely to be distracted during typing. Frequently, he may sit for a while and think deeply on what to write or he may just be waiting for others to reply on a question before typing the next sentence. Consequently, if we would like to include time-based measures in the remote metrics we need to filter out those breaks. WPM may be measured for all the rather short intervals when the user is actively looking in the entry field. Once the user looks at the text field or away from the on-screen keyboard, this indicates a break and the WPM for the recent active period can then be calculated. We expect that these *micro-WPM´*s will be somewhat higher than the traditional WPM measure, since reading of typed text is now excluded. The duration of the micro-WPM´s are most likely to be short burst in the beginning, but if the system is well designed, it will most likely increase with training. Supplementing the micro-WPM, we could also report *the average duration of uninterrupted text-input,* which would be the time periods the user were actively looking at input characters. Both the micro-WPM´s and the average duration of uninterrupted text-input would most likely correlate inversely with the metric “Number of Read Text Events per Character” suggested by Majaranta et al. (2004), which measures how often people look at the text they have produced. Majaranta et al (2004) defined “Read Text” as an event in which a user switches his/her point of gaze from the virtual keyboard to the typed text field to review the text written so far (Majaranta et al, 2004). However, we expect that the “Read text events” in real life situations may include cases where the user is just thinking on what to write, or just looking in the text field to take a break at a location that will not activate something.

Table 1 lists the indexes mentioned previously in four groups according to whether they can be collected remotely (outside the lab) and whether they require a gaze tracking system or not.

Table 1: Measures of gaze typing performance

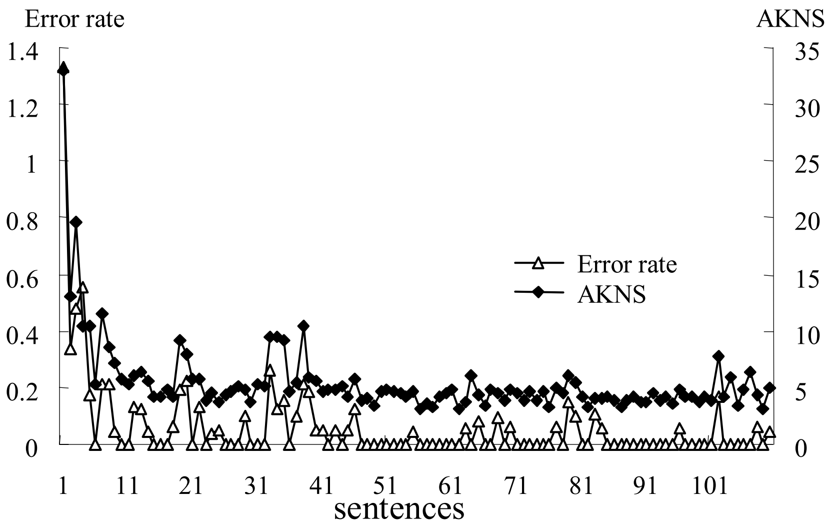
	Measures that require gaze tracking	Measures that do not require gaze tracking
Measures that can only be collected in controlled experiments	Rate of Premature Movement Errors (Aoki et al, 2005) Cumulative Deviation from the Most Efficient Scan-Path (Aoki et al, 2006)	Minimum String Distance (e.g., Soukoreff and MacKenzie, 2001) Cost per Correction (Gong and Tarasewich, 2006) Overproduction Rate (Hansen et al, 2004) Error Rate (e.g., MacKenzie, 2002)
Measures that can also be collected in real-life conditions	Number of Read Text Event Per Character (Majaranta et al, 2004) Attended-Keys-Not-Selected Rate (this paper) Micro- WPM (this paper) Average duration of uninterrupted text-input (this paper)	Words Per Minute (e.g., MacKenzie, 2002) Key Strokes Per Character (e.g., MacKenzie, 2002) Rate of Backspace Activations (Itoh et al, 2006) Deleted characters (this paper)

Experimental validation

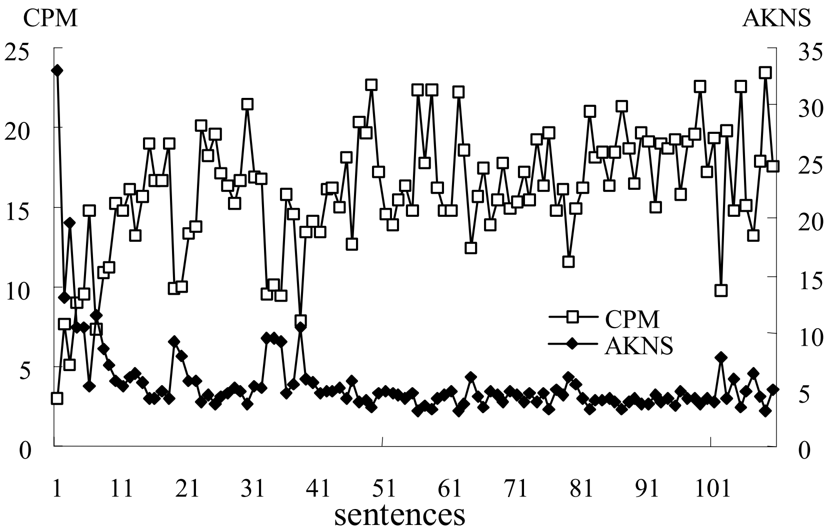
All of the new metrics suggested should be validated by experiments. So far, we have only tested the validity of AKNS by calculating the AKNS and compared it with two traditional metrics, Character Per Minute (CPM) and *error rate*. This was on data collected in a learning experiment with the Japanese version of GazeTalk (Aoki, Itoh & Hansen, 2006). Six Japanese students, who had never used any kind of gaze typing system before, typed 110 sentences in 22 experimental blocks. The average KSPC was 3.4, the average CPM was 15.9 and the total number of characters typed by each subject was 2075. The error rate (mean 0.09, S.D. 0.21) was calculated as the number of errors divided by the number of characters typed. Each error was identified by close examination of all key activations, deciding if every single keystroke made were in fact appropriate at the time of execution. This definition of an error not only requires knowledge of the target input stream but also a great deal of time for analysis, since it cannot be automated. It may be regarded as "ground truth" for comparisons with the other metrics, but the traditional error rate measure will not be a likely candidate for remote collection of performance data.

Figure 1 shows the transitions of AKNS, CPM and error rate for Subject 1. The learning pattern of AKNS seems closely related to CPM and the error rate. Figure 1 (1) shows that the AKNS for Subject 1 became almost flat after just 20 sentences. Then the AKNS data indicates problems in the period between sentence 31 and 35 and also around sentence 81 and 101. This suggests an interesting possibility for immediate detection of typing problems, which may be caused by e.g. a temporary decrease in the precision of the eye tracker. If a well-trained user deviates from his base-line AKNS for a period, we could then automatically suggest him to do a re-calibration or maybe to take a rest. It is also worth noting that, even when Subject 1 had got some practice, the average AKNS was 5.5. With practice, AKNS of a completely predictable keyboard should become close to zero. AKNS are high for the Japanese version of GazeTalk (mean value for all subjects was 5.52) because it requires shifts between different typing system layouts (Kanji, Hiragana, Katakana, and Roman alphabet etc) and it requires selections among several Kana-

Kanji conversion candidates. The monitoring of these different states introduces an attention overhead that seems to persist at least throughout the experiment. It is a design goal for us to reduce the AKNS, and we will need AKNS measures from several users to tell us if a new design actually brings it down.



(1) AKNS and Error rate



(2) AKNS and CPM

Figure 1: Comparison of AKNS and traditional metrics (S1)

Table 2: Correlation coefficients among AKNS, CPM, and Error rate

	S1	S2	S3	S4	S5	S6	Average
AKNS - Error rate	0.93**	0.45**	0.87**	0.92**	0.76**	0.91**	0.81
CPM - Error rate	-0.66**	-0.44**	-0.61**	-0.61**	-0.58**	-0.47**	-0.56
KSPC - Error rate	0.91**	0.38**	0.67**	0.88**	0.68**	0.64**	0.70
AKNS - CPM	-0.77**	-0.85**	-0.80**	-0.69**	-0.67**	-0.77**	-0.76

**: $p<0.01$

We calculated the correlation coefficients for the three metrics for all the subjects, c.f. Table 2. In general, the coefficients were high, and the correlation between AKNS and error rate were the highest (mean r = 0.81). In addition, the AKNS - error rate coefficients were higher than the inverse correlations between CPMs and error rates for all subjects and higher than KSPC – error rate correlation, also. These suggest that AKNS may become an attractive substitute for the error rate since it can be calculated automatically, remotely and without knowledge of the target input stream. We also calculated the correlation between attended keys *per activation* (mean 2.56), but this measure was not so well correlated with the error rate (mean r = 0.66) as the AKNS. It is worth noting, though, that one subject (S2) showed a rather weak AKNS-error rate correlation. This subject had the slowest performance (CPM= 14.4) and the highest KSPC (3.77) of all subjects, but also a very low error rate (0.06). So AKNS may not be so good to predict the error rate of subjects with a cautious typing strategy.

Future Work

We have suggested some measures that could be candidates for remote evaluation of gaze typing performance. Most of them would work for gaze operated on-screen QWERTY keyboards and for special gaze typing interfaces like Dasher and GazeTalk, and most of them would work for Roman alphabets as well as other character-based typing systems.

A metric like WPM/CPM can be measured on any text entry system and it is device independent, while some of the metrics suggested in this paper requires an eye tracker device. At first, this may restrict them to be used in tests of gaze typing systems only. However, if the metrics turn out to be as informative for keyboard design research, as AKNS seems to be, then gaze recording could become a standard procedure when designing typing systems for all kinds of input devices.

Besides from the validation of AKNS, no empirical evaluations of the metrics were presented in this paper. Future research is required to clarify how the proposed metrics correlate with each other and how well measures from the laboratory reflect actual typing performance at home.

Acknowledgements

This research was partly supported by Grant-in-Aid for Scientific Research (A), No. 14208040, the Japan Society for the Promotion of Science(JSPS), by the European Network of Excellence COGAIN, Communication by Gaze Interaction, funded under the FP6/IST programme of the

European Commission, and by Bilateral Program for Scientist Exchanges under Memoranda of Understandings between JSPS and Danish Rectors' Conference. Thanks to Paivi Majaranta for a very inspiring discussion that lead to the idea of micro-WPM´s, and thanks to the anonymous reviewers for useful comments to this paper.

References

Aoki, H., Itoh, K., Sumitomo, N. and Hansen, J. P. (2003). Usability of gaze interaction compared to mouse and head-tracking in typing Japanese texts on a restricted on-screen keyboard for disabled people. *Proceedings of the 15th Triennial Congress of the International Ergonomics Association*, 1, Seoul, Korea, August, pp. 267-270.

Aoki, H., Itoh, K. and Hansen, J. P. (2005). Learning to type Japanese text by gaze interaction in six hours. *Proceedings of the 11th International Conference on Human-Computer Interaction*, 7, Las Vegas, USA, July (CD-ROM).

Aoki, H., Itoh, K. and Hansen, J. P. (2006). Usability evaluation of gaze interface based on scan-path analysis. *Proceedings of the 16th World Congress on Ergonomics*, Maastricht, Netherland, July (CD-ROM).

Barnes, R.M. (1949). *Motion and Time Study*, John Wiley and Sons.

Drey, S. & Siegel, D. (2004): Remote Possibilities? International Usability Testing at a Distance. *Interactions*, March+April 2004, p. 10 – 17

Gong, J and Tarasewich, P. (2006). A new metric for text entry method evaluation. CHI 2006 Proceedings, Montreal, Canada, pp. 471-474.

Hansen, J. P., Hansen, D. W., Johansen, A. S., Itoh, K. and Mashino, S (2003). Command without a click: Dwell time typing by mouse and gaze selections. *Proceeding of the 9th IFIP TC 13 International Conference on Human-Computer Interaction*, pp 121-128.

Hansen, J. P., Tørning, K., Johansen, A. S., Itoh, K. and Aoki, H. (2004). Gaze typing compared with input by head and hand. *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications,* pp.131-138.

Jacob, R. J. K., & Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Radach, R., Hyönä, J., & Deubel, H. (Eds.) *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Oxford: Elsevier Science, 573-605.

Itoh, K., Aoki, H., and Hansen, J. P. (2006). A comparative usability study of two Japanese gaze typing systems. In Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (San Diego, California, March 27 - 29, 2006). ETRA '06. ACM Press, New York, NY, 59-66. DOI= http://doi.acm.org/10.1145/1117309.1117344

Ivory, M. Y. and Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv*. 33(4), pp. 470-516. DOI= http://doi.acm.org/10.1145/503112.503114

Lecerof, A. and Paterno, F. (1998). Automatic support for usability evaluation. *IEEE Transactions on Software Engineering*, 24(10), pp. 863-888.

MacKenzie, I. S. (2002). KSPC (keystrokes per character) as a characteristic of text entry techniques. *Proceedings of the Fourth International Symposium on Human Computer Interaction with Mobile Devices*, Heidelberg, Germany, pp. 195-210.

Majaranta, P., Aula, A., and Räihä, K.-J. (2004). Effects of feedback on eye typing with a short dwell time. *Proceedings of the 2004 Symposium on Eye Tracking Research and Application*, San Antonio, USA, March, pp.139-146.

Soukoreff, R. W. and MacKenzie, I. S. (2001). Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2001*, New York, USA, pp. 319-320.

Ward, D. J. and MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, 418(6900), pp. 838.

Magic Key

Luis Filipe Figueiredo
Escola Superior de Tecnologia e Gestão
Av. Francisco Sá Carneiro, nº 50,
6300-Guarda Portugal
luis.figueiredo@ipg.pt

Ana Isabel Gomes
Agrupamento de Escolas de Penalva do Castelo
Rua da Ramalhosa, Lote A,
3500-688 Repeses Portugal
ana.isabel.gomes@netvisao.pt

João Bento Raimundo
Associação Augusto Gil
Av. Alexandre Herculano
6300-659 - Guarda
Portugal
joao.raimundo@netvisao.pt

Keywords
Quadraplegic, mouse control, eye tracking.

Introduction

Based on the real needs felt by many people with limitations in the use of their arms, and on the different solutions that already exist, a set of requirements was defined which a new system should obey in order to allow these people to easily use any computer application. These are some of the requirements:

- A simple system to be used by anyone, even by people who have no arm movement, and who can only control head movement.
- A system that makes using the computer by these people as similar as possible to its use by people who control their arm movements.
- A system that allows the use of any computer application without having to configure it.
- A system that is quickly learned and which needs no memorization.
- A system that is totally independent from its user, which means no electrical or mechanical appliance attached to the user.
- A low-cost system that does not require special equipment for its implementation.

Next, we will explain how these targets were met throughout the development of this project.

System description

Using an ordinary webcam, adapted for this application, installed below the monitor and aiming the user's face, this application uses DirectX technology (Microsoft) to acquire real-time images of the user's face. These images have a 640*480 resolution and a frequency of 30 Hz.

The first processing phase consists of the automatic tracking of the user's nose, which will always be the central reference point in the user's face. Once the coordinates of the nose have been established in the image system, they are mapped onto the monitor coordinate system, where the mouse cursor is then placed. A direct mapping system of coordinates is employed, which always implies the use of the absolute coordinates of the nose position, and not the relative coordinates in relation to the previous image. This means that when the user is turned to the right the cursor is on the right and when he is turned to the left the cursor is on the left. It is important to mention that in spite of using an ordinary webcam with a 640*480 resolution, and considering that the horizontal movements of the user's face are not usually higher than 100 pixels, and that the vertical movements aren't usually higher than 50 pixels, the algorithms developed for mapping the coordinates make it possible to place the cursor in the exact desired pixel, even if the monitor has a graphic resolution of 1280*1024. It is equally important to mention that the cursor always responds to movement of the head in real-time.

The blinking of the eyes signifies pressing the mouse buttons. When the eye is closed a button is pressed, when the eye is open the button is released. It is possible to configure the right or the left eye to activate the main button of the mouse, and use or not the other eye for the secondary button. It is also possible to define the minimal time for an eye to be closed in order to be understood as a click, thus solving the problem of involuntary blinking, which happens more or less frequently to everyone. Dragging can be achieved in two ways. The first way is to close the eye, and keep it closed during the whole dragging period. This is a simple solution, but it leaves the user only one eye to keep track of what he is doing, which, in some cases, may be a problem. The second solution implies keeping an eye closed for a period of time just long enough for the button to remain depressed. When the eye is then opened, the button remains depressed, which allows the user to see what he is doing with both eyes, while the button is depressed.

For eye tracking, a proprietary algorithm based on the Hough Transform (Zahid Hussain, 1991) for the calculation of circles was developed, but with an important group of changes that not only permit a substantial decrease in processing time, but also a considerable increase in eye tracking efficiency, even when the eyes aren't completely open, meaning they are not perfect circles. At this point it is important to mention that poor eye tracking when, for example, the eyes are not completely open, and therefore not in the shape a perfect circle, would lead to numerous false mouse clicks, and therefore to a malfunction of the application.

Practical results

The computer used to perform these tests was equipped with a Pentium 4 CPU, running at 3.0 GHz. The use of advanced digital image processing techniques alongside other simpler and well-known ones (Gonzalez and Woods, 2001) has delivered excellent results, thus diminishing the need for more expensive and complex equipment, such as high-definition cameras, whose market price is higher than the webcam used. On the other hand, the efficient manner in which the proprietary algorithms were developed and implemented, making use of, in the most critical parts of the code, a lower level implementation in order to better take advantage of the existing SIMD technologies from Pentium 4 processors (Intel, 2003), prevented the application from compromising the overall performance of the computer, thereby almost completely maintaining the performance of the remaining applications. The average percentage of CPU usage in the process of image acquisition and processing of the 30 images per second remain around 11% of total CPU time.

Although this is a generic application, adaptable to any user, it is necessary to undergo an initial personalization process, which will remain in memory for future use, thus establishing a profile for each user. The system is very tolerant of fluctuations in the levels of illumination. The tests performed under real conditions, with two very different users, show the success and adaptability of this application to individualised conditions.

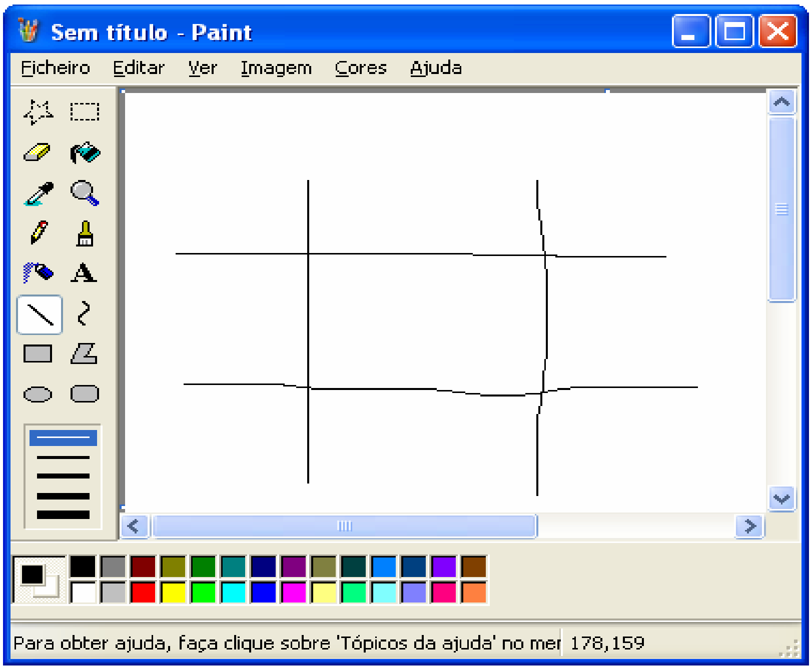


Figure 1. Paint Application

An image of the well-known Paint application is shown in Figure 1, which consists of two horizontal and two vertical lines. One of the horizontal and one of the vertical lines were drawn with the traditional mouse, by keeping the button pressed throughout the drawing. The remaining two lines were drawn using this application, by keeping the eye closed throughout its drawing. The left line and the top line were drawn with the Magic Key.

It is perfectly possible to make a Power Point presentation using this mouse control system to navigate through the several slides, which has been done in every public presentation of this application.

Conclusion

Making use of existing technology - ordinary low-cost equipment - it was possible to develop a “magic key” that allows handicapped people to control the mouse, and thereby any other associated computer application.
There is total mouse control, as it combines the high precision of the cursor movement, at the pixel level, with the high movement speed that relates to real-time movement of the head, and, above all, with the simple and absolute control of all types of clicks by eye control.

As a result of the efficient elaboration and implementation of its algorithms, this application doesn’t interfere with the performance of the other computer applications, since it uses about 11% of CPU in a Pentium 4 at 3.0GHz.

The numbers of tests performed, as well as its practical utilization by target users, have already shown the success of this "magic key" which undeniably opens the doors of the World to many people.

Effectively, for those many people who see themselves confined to a wheelchair, totally dependent on others, having an application that allows them - in an independent, simple, low-cost way - to use the computer and access the Internet is, no doubt, a huge contribution to the quality of their lives.

As for future work, many projects are already being developed, which emerged from this application. We are looking for communication solutions for people who suffer from strong spasms, or solutions to solve the mobility problem of quadriplegic individuals, or even specific solutions to meet the specific needs of each user. For all these solutions we adhere to the principles that guided the development of this application: the search for simple, low-cost but functional systems.

References

Gonzalez, R. and Woods, R (2001): *Digital Image Processing*, Prentice Hall.

http://www.microsoft.com/windows/directx/default.aspx

Intel (2003). *IA-32 Intel Architecture Software Developer's Manual*, Intel Corporation.

Zahid Hussain (1991). *Digital Image Processing-Practical Applications of Parallel Processing Techniques*, Ellis Horwood.

COGAIN 2006 Keynote by Dr. Roel Vertegaal

When Less is More: Applying Attention to the Design of Context-Aware Computers

Ubiquitous Computing promised to provide users with many simple computing appliances, each appliance suitable for a singular task. While users own more computers than ever before, devices have not become easier to use. Each individual interface is still designed as if the user had only one computer. Instead, featuritis has become a primary marketing mechanism, with cell phones now duplicating functionality of early desktop computers. Users are faced with ever smaller yet ever more demanding user interfaces. How can we design computing appliances that work in synchrony with the user and with each other? How can we reduce complexity through combined functionality of many individual computers?

One approach is to design computing interfaces such that they share common resources, as well as users, by embedding them in the user's social networks. Attentive User Interfaces allow devices to observe human social cues that are used to manage group conversations. By observing the attention of users, devices may determine the user's task focus and their preferred channels and moments of communications. By modeling the user's attention, devices may understand when to await their turn and leave the floor to others. By observing human social networks, devices may share context between many communications. I will illustrate our approach through several prototypes developed at Queen's University's Human Media Lab. These include eye contact sensing phones; appliances that contextualize speech interactions by observing eye contact with users; robot eyes that communicate attention; attentive video conferencing systems that optimize turn taking, attentive wearables as well as attentive architecture.

Biography

Roel Vertegaal, Human Media Lab, Queen's University, Canada.

Dr. Roel Vertegaal is Associate Professor in Human-Computer Interaction at Queen's University's School of Computing in Canada, where he leads the Human Media Lab, Canada's premier media laboratory. He is also CEO of Xuuk, Inc, a attentive sensor company. Dr. Vertegaal's first degree was in Music at Utrecht Conservatory, and he spent time as a visual artist and photographer at the Vrije Academie in The Hague. Roel holds an MSc in Computing and a PhD in Human-Computer Interaction, from Twente University in The Netherlands. Roel co-chaired the ACM Eye Tracking Research and Applications conference (ETRA), the world's premier eye tracking conference. He co-founded and chaired alt.chi, an alternative papers venue at the prestigious ACM CHI conference, for which he served as associate program chair. His work was awarded with the Premier of Ontario's Research Excellence Award, and was featured on

ABC Good Morning America, Discovery Channel's Science Daily and Scientific American, amongst others. Roel's current interest lies in the design of completely transparent computers. www.roelvertegaal.com