Apps for Me, Too: Gaze and Natural Interaction Applications for Varying Cognitive Strengths on a Multi-touch Table Network

Cumhur Erkut
Institute of Inclusive Science & Solutions (IISS)
University of Eastern Finland
Joensuu, Finland
Cumhur.Erkut@uef.fi

Roman Bednarik
School of Computing
University of Eastern Finland
Joensuu, Finland
Roman.Bednarik@uef.fi

Susanne Mäkelä
School of Computing, IISS
University of Eastern Finland
Joensuu, Finland
smakela@cs.joensuu.fi

Jari Kleimola
Dept. Media Technology,
School of Science
Aalto University
Espoo, Finland
Jari.Kleimola@aalto.fi

CHI 2013 Workshop on "Gaze Interaction in the Post-WIMP World", April 27, 2013, Paris, France.
ACM 978-1-XXXX-XXXX-X/XX/XX.

Abstract
We present the conceptual and technical framework we are developing in order to design and evaluate suitable applications for varying cognitive strengths. Our concept is based on an interconnected network of multi-touch tables and a hardware-agnostic thin framework that supports natural interaction and proximal connections of tablets and other sensors. While we are at the early stages of integration of eye tracking and application development, we hope to contribute to the CHI'13 workshop research questions on examples, benefits, guidelines, methods and models.

Author Keywords
Eye tracking; gaze; multimodal; interaction; post-WIMP, multi-touch, cognitive impairments

ACM Classification Keywords
H.5.2. **[Information interfaces and presentation]**: User Interfaces– input devices and strategies.

General Terms
Design, Human Factors

Introduction
The combination of eye-tracking with other input modalities such as gesture (natural interaction) and

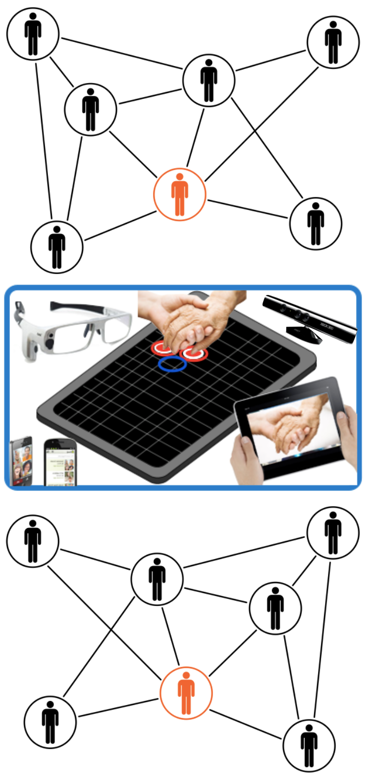

Figure 1: Conceptual illustration of our inclusive system that consists of networked multi-touch tables combined with eye tracking , natural interaction sensors (Kinect, Leap Motion, etc), tablets and smart-phones. Each table transmits interaction data (gaze, multi-touch and gestures) and received adapted content and interface.

sound (sonic interaction [12]), and its integration with multi-touch technologies (surface computers, tablets, and smart phones) provides exciting application and interaction design opportunities in the Post-WIMP world [5]. Like in any emerging technology paradigm [6], the current focus in applications is on early adapters, i.e., physically and cognitively fit young adults, who are able to deal with various demands of the technology, e.g., on cognitive skills, communication skills and persistence.

Various components of the Post-WIMP interfaces were previously used to make interactive technology more accessible for people with special needs. Examples include using eye-tracking for text input [15], multi-touch interfaces targeted especially for older people [14] or special-need children [2]. These efforts have focused on decreasing the negative effects of users' disabilities and lack of skills to allow people with special needs to access the mainstream technology [8,16].

Beyond accessibility, considering the interests and engagement of people with special needs requires shifting the focus from users' lack of abilities to their strengths [21]. The challenge in this shift is the diversity of the user group, which makes it difficult to come up with universal, "one-size-fits-all" solutions. We are currently focusing on *cognitive strengths* that dynamically change in time [16], based on the recent development in eye-tracking [5], user modeling [17], content and interface adaptation [20], and Post-WIMP interfaces [11].

In this position paper, we present the conceptual and technical framework we are developing in order to design and evaluate suitable applications for varying cognitive strengths, with the aim of contributing to the CHI'13 workshop research questions on examples, benefits, and on considerations in form of guidelines, methods and models. A conceptual illustration of our development framework is illustrated on Figure 1.

The structure of this paper is as follows. We first provide a background on our objective and challenges. Next, we provide an overview of eye-tracking research, together with our initial findings on how gaze relates to user experience. Next we focus on the envisioned applications and the technology supporting them. The following section focuses on the conceptual view of interaction while monitoring of interest and engagement.

Background

Our main objective is to develop solutions that provide meaningful content and engaging interaction by combining eye tracking with intelligent adaptive media and suitable post-WIMP interactive technologies. The adaptation will make use of advanced computational techniques such as user modeling and machine learning, enhanced by human-powered computation and informed by user-research. We identify three key intermediate steps to attack the overall challenge: 1) appropriate content generation , 2) engagement and motivation through appropriate interaction design, and 3) evaluation and assessment of the technologies on quality of life. These challenges require a cross-disciplinary approach; therefore we will collaborate with social science and special education experts, creative media technology and design professionals, as well as with industrial partners. This collaboration will allow us to develop culturally, socially, and technically feasible and functional solutions to deal with e-inclusion topics.

Eye tracking

Eye-tracking is currently a well-established interaction method and a way to obtain accurate information about user's interests with a millisecond accuracy. Most of the current eye-tracking techniques rely on high-speed image processing of the images of eyes and on detection of infrared reflections from the cornea. The configuration of the centre of the pupil and the stable reflection is unique for each direction of the sight [9].

The non-invasiveness of eye-tracking is among its main strengths. Compared to other methods, users in fact do not need to be aware of being eye-tracked. Typically, eye-trackers can be placed in a distance from the user, or be mounted into the rim of glasses. Both solutions provide a real-time estimation of the point of gaze. Eye tracking has often been used either as an active interaction technique or as a passive complementary modality. For example, eye-tracking has been previously used in text input systems [15], understanding of reading patterns [18], gaze-input in problem solving [3], and collaboration activities during programming. In our project we will employ eye-tracking for monitoring of the interactions with content in smart environments.

Because visual attention plays a central role in a majority of interactive scenarios, we will employ advanced eye-tracking methods both to construct user models and to estimate the dimensions of user experience. The principal opportunity of eye-tracking and other physiological signals used for user modeling lies in connecting the captured data to the internal processes of the user and to the environment that the user is interacting with. While the general pattern of interest and cognition is apparent from the eye-tracking data, the links to the higher-level processes, such as the user experience for cognitively-varying users are yet to be identified. Based on our initial findings [4], the methods developed in this research will focus on identification and modeling of four dimensions of user-experience relevant to the target group and applications: *engagement, immersion, interest* and *intention*. In addition, we will investigate other behavioral correlates of the four primary dimensions, within the interactive environment described next.

In particular, we plan to collect a database of user behaviors, conduct annotations, and train classifiers by machine learning to recognize the patterns typical for certain levels of the user experience dimensions. For example, we have previously demonstrated that engagement can be inferred from gaze data during multiparty conversation [25]. The system then can act upon the outcomes of the prediction and provide content or interaction that has been observed as influencing the levels of engagement.

At the moment, we aim to employ wearable eye-tracking to monitor the gaze of the users. However, a remote solution is preferable in this context and thus are looking into ways to integrate remote gaze tracking with multitouch surfaces to allow natural interaction.

Applications and technical environment

In order to develop and evaluate user modeling and content production techniques as well as their impact on the user engagement and satisfaction, working prototypes of a communication and media systems for the target group will be developed. The data-flow in our system is depicted on Figure 2.

A mighty thin framework

Our open source and hardware-agnostic thin framework provides support for:

- connecting multiple geographically distributed multi-touch tables together
- connecting mobile/smart-phones and tablets locally with the tables (via WLAN)
- connecting mobile or desktop browsers with the tables (via internet)
- connecting local sensors (e.g., Kinect , Leap Motion, and eye-tracking systems) with the tables
- using fiducial markers on the tables, if supported by the table hardware
- sharing content from the internet, a centralised server, or from a local storage of each table
- using well-known audio-visual content-creation frameworks and programming environments for application development.

For details, please refer to [22].

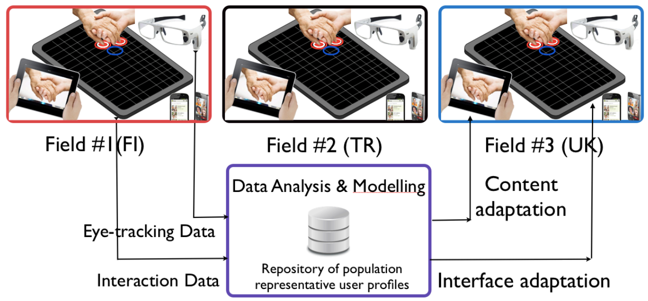

Figure 2. Data flow in our system

Previous research has found touch displays to be especially suitable for the target group, so the implementation of the system will be based on multi-touch tables. In addition to eye-tracking, peripheral tablets and smart-phones (see, e.g., [1,14,19]) and Natural User Interfaces (NUIs)[1] provide important affordances for our target group. Finally, 3 application domains stand out as very potent for our purposes: 1) Crowd-sourced and shared news content adapted and presented to an individual, based on her cognitive abilities, 2) Serious games for motivation and engagement [7,10], and 3) Creative interactive art applications to maintain and develop cognitive abilities, similar to those developed by Yamamoto and colleagues (see, e.g., [23]). This third domain is motivated by the statistical correlation between the involvement of elderly people in creative artistic activities, the scores of their cognitive ability and skills, and their self-accounts on creativity and late-life development [13].

Our environment will be based an an open source and hardware-agnostic thin framework. This framework provides support for various *connections*, i.e., sharing interaction data and content, as outlined in the sidebar. It has been developed within the *AaltoWindow* project of the last author, and has been previously used for co-learning and collaboration in higher education, allowing fast deployment even for inexperienced developers [22]. The learnability aspect of the framework is important, as we aim for participatory design that will empower our participants as co-designers, instead of being passive users of the technology [21]. Moreover, adhering to open standards will allow us focus more on the interesting and novel aspects of interaction.

Conceptual view of interaction

Figure 3 depicts the overall conceptual view of interactions with the system. The architectures and methods we will develop perform constant monitoring of interest and engagement. They model the everyday behavioral patterns of the users by gathering interaction data through eye-tracking and other sensors. Internally, the system maintains dynamic user models and performs collaborative filtering. The framework estimates the condition of the user to deal with the changing needs, such as decreasing cognitive function or a change from single user scenarios to collaborative environment.

The user models are enriched by three sources of information: collaborative modeling techniques, crowd-powered information, and by closest relatives' and caretakers' input. The prior knowledge is compared to a constantly updated knowledge base and adaptation is performed to actual needs and context, delivering the

---

[1] In the interview "CES 2010: NUI with Bill Buxton" released by Microsoft Research in 2010, Buxton defines the "Natural" in NUIs as "using the existing skill-set", which aligns well with our strength-based design approach,

best and personalized options. Appropriate interaction means are selected for all users. With other users, the functionality of the application is identical, however different adaptations offer different routes and form to reach the user experience goal.

Conclusions

In this position paper, we have provided a framework for developing applications that combine gaze and natural interaction on a multi-touch table network. Currently, by using off-the-shelf sensors such as MS Kinect, we are able to track a user's head orientation, skeleton, and positions of fingertips in 3D above the table. This enables, for instance, 3D rendering based on the viewpoint of the user, detecting left and right hand touch points separately, and gestures above the surface. We are working on adding eye tracking and multimodal interaction capabilities, and aim for meaningful applications and life-enhancing experiences for our co-developers with varying cognitive strength.

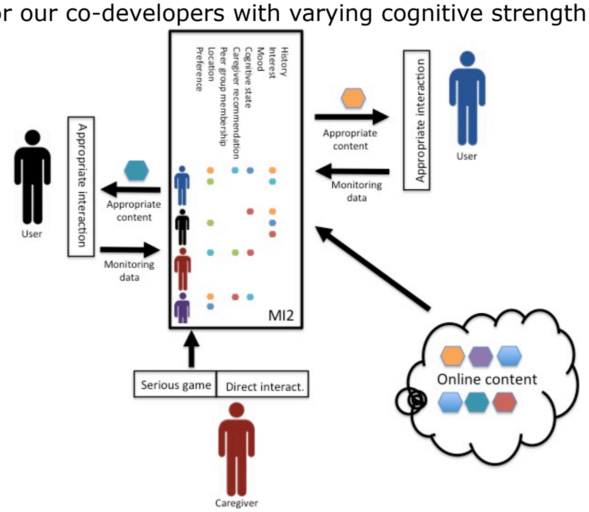

Figure 3. A conceptual overview of user interaction.

The integration of interaction modalities has been classically known to cause specific complexity in multimodal frameworks [24]. Fortunately, our framework is constructed with integration in mind and supports a frame-based solution similar to [24], in addition to events. Applications can override frame-based input methods, but if an application does not handle the frames, our platform iterates the frame content and converts them into individual events (e.g., all fingers in a frame are converted to multiple single touch events). The application can then respond to the input on the event level. Other examples how eye-tracking is integrated in multimodal environments can be found in [11].

Acknowledgements

We thank our colleagues at the University of Eastern Finland, Institute for Inclusive Science and Solutions.

References

[1] Al-Razgan, M.S., Al-Khalifa, H.S., et al. Touch-Based Mobile Phone Interface Guidelines and Design Recommendations for Elderly People: A Survey of the Literature. In *Proc. Intl. Conf. Neural Information Proc.*, Springer (7666), Berlin, 2012, 568–574.

[2] Alper, M., Hourcade, J.P., and Gilutz, S. Adding reinforced corners: designing interactive technologies for children with disabilities. *interactions* 19, 6 (2012).

[3] Bednarik, R., Gowases, T., and Tukiainen, M. Gaze interaction enhances problem solving. *J. Eye Movement Research 3*, 1 (2009), 1–10.

[4] Bednarik, R., Gowases, T., and Tukiainen, M. Gaze-augmented interaction improves problem-solving: new evidence from verbal protocols. *Behaviour & Information Technology*, (2012), 1–9.

[5] Bulling, A., Dachselt, R., Duchowski, A., Jacob, R., Stellmach, S., and Sundstedt, V. Gaze interaction in the

post-WIMP world. In Proc. CHI EA ’12, ACM (New York, NY, USA, 2012). 1221–1224.

[6] Buxton, B. Sketching user experience. Morgan Kaufmann, San Francisco, CA, USA, 2007.

[7] Carvalho, D., Bessa, M., Peres, E., Magalhaes, L., Guedes, C., and Oliveira, L. Developing a multi-touch serious game to fight the digital divide. In *Proc. CISTI’12* (Madrid, Spain, 2012), 1–6.

[8] Czaja, S.J. and Lee, C.C. Information Technology and Older Adults. In *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates, 2008, 777–792.

[9] Duchowski, A. Eye Tracking Methodology: Theory and Practice. Springer, London, UK, 2007.

[10] Hoffman, B. and Nadelson, L. Motivational engagement and video gaming. *Educational Technology Research and Development* 58, 3 (2010), 245–270.

[11] Jacob, R.J.K., Girouard, A., Hirshfield, L.M., et al. Reality-based interaction: a framework for post-WIMP interfaces. In Proc. CHI’08 (Florence, Italy, 2008), 201-210.

[12] Jylhä, A. Sonic Gestures and Rhythmic Interaction between the Human and the Computer. 2012. PhD Thesis, Aalto University, Espoo, Finland

[13] Lindauer, M.S. Aging, Creativity, and Art. Kluwer, New York, NY, USA, 2003.

[14] Loureiro, B. and Rodrigues, R. Multi-touch as a Natural User Interface for elders: A survey. In *CISTI’11* (Chaves, Portugal, 2011), 1–6.

[15] Majaranta, P. and Räihä, K.-J. Twenty years of eye typing: systems and design issues. ACM (2002), 15–22.

[16] Newell, A.F., Carmichael, A., Gregor, P., Alm, N., and Waller, A. Information Technology for Cognitive Support. In *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates, 2008.

[17] Obrenović, Ž., Abascal, J., and Starcevic, D. Universal accessibility as a multimodal design issue. *Communications of the ACM* 50, 5 (2007), 83–88.

[18] Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372–422.

[19] Rocha, R. and Paschoarelli, L.C. Digital Tablets as a Tool for Digital Inclusion of Elderly People. In *Advances in Usability*, CRC Press (2012).

[20] Vertegaal, R., Shell, J.S., Chen, D., and Mamuji, A. Designing for augmented attention: Towards a framework for attentive user interfaces. In *Computers in Human Behavio*r 22, 4 (2006), 771–789.

[21] Voutilainen, M., Vellonen, V., and Kärnä, E. Establishing a Strength-based Technology-enhanced Learning Environment with and for Children with Autism. In *AACE* (2011), 601–606.

[22] Vuorimaa, P. and Kleimola, J. AaltoWindow: Open Multi-Touch Framework for Campus Learning Centers. In *EUNIS Congress* (2012), 1–10.

[23] Yamamoto, M., Komeda, M., Nagamatsu, T., and Watanabe, T. Development of eye-tracking tabletop interface for media art works. In *Proc. ITS’10*, ACM (New York, NY, USA, 2010), 295–296.

[24] Blattner, M. and Glinert, E., Multimodal integration, IEEE Multimedia, vol. 3, no. 4, pp. 14–24, 1996.

[25] Bednarik, R., Eivazi, S. and Hradis, M. Gaze and conversational engagement in multiparty video conversation: an annotation scheme and classification of high and low levels of engagement. In *Proc. Gaze-In '12*. ACM (Santa Monica, CA, USA, 2012), Article 10.