Towards Using Gaze Properties to Detect Language Proficiency

Jakob Karolus
University of Stuttgart
Pfaffenwaldring 5a
70569 Stuttgart, Germany
jakob.karolus@vis.uni-stuttgart.de

Paweł W. Woźniak
University of Stuttgart
Pfaffenwaldring 5a
70569 Stuttgart, Germany
pawel.wozniak@vis.uni-stuttgart.de

Lewis L. Chuang
Max Planck Institute for
Biological Cybernetics
Spemannstr. 38
72076 Tübingen, Germany
lewis.chuang@tuebingen.mpg.de

NordiCHI '16, October 23 - 27, 2016, Gothenburg, Sweden

ACM 978-1-4503-4763-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2971485.2996753

Abstract
Humans are inherently skilled at using subtle physiological cues from other persons, for example gaze direction in a conversation. Personal computers have yet to explore this implicit input modality. In a study with 14 participants, we investigate how a user's gaze can be leveraged in adaptive computer systems. In particular, we examine the impact of different languages on eye movements by presenting simple questions in multiple languages to our participants. We found that fixation duration is sufficient to ascertain if a user is highly proficient in a given language. We propose how these findings could be used to implement adaptive visualizations that react implicitly on the user's gaze.

Author Keywords
Eye tracking; pattern recognition; adaptive visualization.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.; H.1.2 [Model and Principles]: User/Machine Systems—Human information processing

Introduction
During conversations, we usually pay attention to the view direction of our conversation partner [4, 5]. This enables us to detect if the other person is bored or not engaged in the

Language	ISO 639-1 [6]
English	en
German	de
Danish	da
Dutch	nl
Finnish	fi
French	fr
Greek	el
Romanian	ro
Spanish	es
Turkish	tr
Slovenian	sl
Arabic	ar
Hungarian	hu

Table 1: Languages in this study.

conversation. Changing the topic or requesting attention from the partner is a common countermeasure.

When we interact with personal computer systems communicating on this level is not yet possible. In virtual environments, gaze directional cues have been researched as a predictor of conversational attention [13, 14]. It has been shown that a system that reacts to the user's gaze, e.g. a virtual agent [8], increases emotional response and attentional allocation by the user. In our work, we examine how gaze characteristics can explain the user's behavior and their context. We then suggest how these characteristics can be utilized to build adaptive user interfaces.

We propose an approach that leverages the distinctiveness of gaze characteristics as a means to infer whether the user needs assistance or not. In particular, we looked at how the eye movements of users changed when confronted with different languages. In our study, we presented the participants with simple questions in varying languages while recording their gaze with an eye tracking device. While studies regarding people's language proficiencies are plentiful, they often rely on an extensive analysis given eye movement data from multiple read documents [9, 17]. For a real-time adaptive system, the prospect of having the users read a whole document first to recognize that they did not understand the content is not feasible. Thus, we specifically aim to recognize a user's understanding as accurately as possible given only a short sentence in the respective language.

Our preliminary results show that a model relying solely on a person's mean fixation duration is sufficient to ascertain whether the user is highly proficient in the respective language. We contribute results from an eye tracking study with 14 participants that show the feasibility of such an approach.

Study Design

Nineteen participants volunteered for this study, of which the data of 14 participants (7 female, age: 19-36 years) were used for further analysis, based on data quality[1]. Two participants already had prior experience with eye tracking studies. All of them were native German speakers.

We collected simple questions in 13 different languages (15 questions each). The respective translations were provided by either native or highly proficient users of that language. Table 2 shows a few example questions. Most are part of the Indo-European language family [2], yet we included some outliers such as Finnish or Hungarian to analyze their effect. See Table 1 for a complete overview.

Our setup consisted of a 22 inch LCD display and a remote eye tracker (SMI RED 250) that was attached at the bottom of the screen. Our participants were seated 0.5 to 1 meters away from the display in a small cubicle. Figure 2 shows a picture of the apparatus.

After introducing the participants to our study, they were asked to sign the provided consent form and supply demographic information as well as rate their reading level for specific languages based on the *Common European Framework of Reference for Languages* [1]. Additionally the participants were asked to provide their proficiency level for languages not listed.

During the experiment, we displayed a total of 150 questions (randomized) one by one to the participants. Each question was visible for exactly ten seconds and could be answered by a single key press, such as one letter or one number (see Table 2). During this time the user was able to provide an answer by pressing the respective key. Eye

[1] For the other participants the data was not reliable due to glasses or make-up interfering with the recording.

Language	Example questions	
English	How many days are within a week?	What is the first letter of your first name?
French	Combien de jours y a-t-il dans une semaine?	Quelle est la première lettre de votre prénom?
Danish	Hvor mange dage er der i en uge?	Hvad er det første bogstav i dit fornavn?
Finnish	Kuinka monta päivää on viikossa?	Mikä on etunimesi ensimmäinen kirjain?

Table 2: Two example questions in four different languages.

Figure 2: Apparatus showing LCD monitor with attached eye tracking device.

movements were recorded at a rate of $250\,Hz$. The experiment was conducted in two sessions with an intermediate break in between to allow for relaxation. Figure 1 illustrates an example scanpath of one participant.

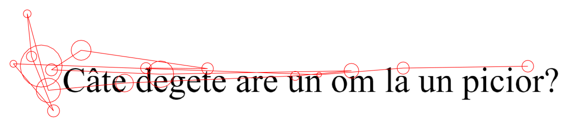

Figure 1: Fixations (circles) and saccades (lines) of one participant plotted on top of the respective question.

Results and Discussion

We applied several post processing steps to the obtained eye tracking data including event detection using a velocity based fixation algorithm [11] and blink removal. We discarded any eye movement data after the point in time when the participants provided an answer for the respective question. For the other questions, we decided to limit the examination time to five seconds after the stimulus was presented. Additionally we discarded outliers that were not sensible (lower than 50ms and higher than 600ms) given reported values for fixation duration during reading tasks in literature [10, 12].

In a preliminary analysis, we are interested in how different languages affect the user's fixations. A simple descriptive statistic is the average fixation duration. Thus, we discarded all saccades and computed the mean fixation duration for each participant for a given language. To compare these results within our participants we normalized the mean fixation duration over all languages per participant. As a normalization factor we used the mean fixation duration of each participant given all fixation data we had collected from the respective participant. Figure 3 shows a violin plot of the distribution of the mean fixation duration (normalized) of each participant per language.

To evaluate whether there is a significant effect of the language proficiency on the average fixation duration, we ran a one-way repeated measures ANOVA. The grand mean of the average fixation duration (normalized) was 1.0. The lowest was 0.80 for proficiency level C2[2], while the non-proficient level, i.e. no knowledge at all[3], showed the highest with 1.04. The main effect of proficiency level on the average fixation duration was statistically significant ($F_{6,175} = 6.583, p < .001$). A post-hoc Tukey HSD test revealed that the mean fixation durations of non-proficient readers were significantly longer than C1 ($p < 0.05$) and C2 ($p < 0.001$) readers.

[2] the highest proficiency level according to the CEFR [1]
[3] less than A1, with regard to the CEFR [1]

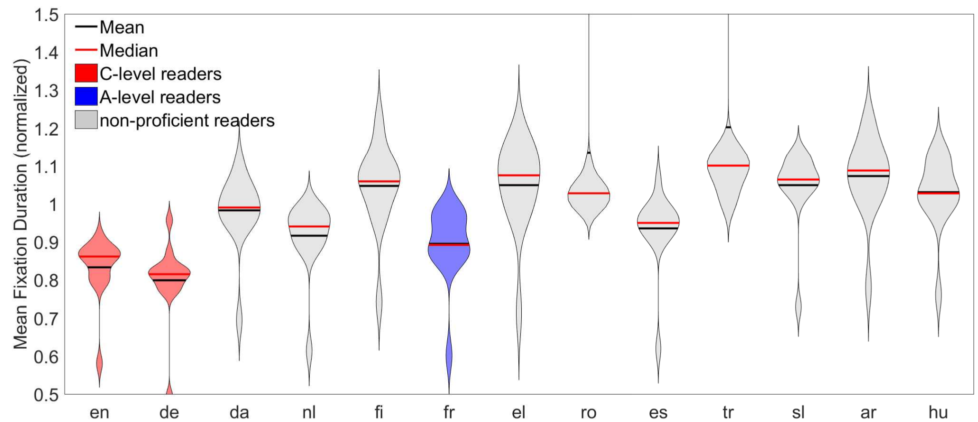

Figure 3: Violin plot showing the distribution of the average fixation duration (normalized) of each participants per language. Language code is given by ISO 639-1 [6] (see Table 1). Color coding refers to the most prominent proficiency level within the language group, such as C-level readers for German.

Conclusion and Future Work

Our results show that even short text snippets are sufficient to infer if a user is highly proficient given the respective language. The questions are not part of a single, long body of text and are randomized in the order that they were displayed. Thus, the independence of each question still holds.

We have shown that by comparing the average fixation duration over an interval of a maximum of five seconds, we can make an adequate statement on whether the user understands the shown content or not. Furthermore, this method does not require a disruptive calibration of the eye tracking device as relative gaze position is sufficient to calculate the fixation duration. Especially low-cost eye tracking devices struggle with accurate calibration and drifts over time, making them a prime target for our approach. In the future, an incorporation of eye tracking devices into already existing wearable computing interfaces, such as the Google Glasses might very well be possible.

Our vision is to utilize such distinct gaze characteristics in real-world scenarios where people interact with public displays. Gaze-based language switching enables foreigners to switch to a suitable language where explicit interaction might be troublesome or even impossible, e.g. at an airport or other commuting places (see Figure 4). Research already reports that off-the-shelf cameras are sufficient for simple gaze-based interaction in a tablet scenario [15] and in larger setups on wall-sized displays [16].

On a technical level, we will look into further gaze characteristics that might be key indicators of a user's understanding. Furthermore, we will examine the usefulness of low-level scanpath comparison [3] that utilizes geometrical representations to calculate similarities between different scanpaths [7].

In our future work, we will focus on building adaptive visualizations systems that leverage gaze properties demonstrated here and recognize when the users struggle with their task.

Figure 4: Public display in Japan. There is no way for non Japanese-speaking people to change the language. An adaptive gaze-based system could bridge this gap. © CEphoto, Uwe Aranas.

Acknowledgements

We thank all colleagues and friends that provided us with the necessary language translations. This research is financially supported by the German Research Foundation (DFG) within the project C02 of SFB/Transregio 161 and supported by the SimTech Cluster of Excellence. Ethical approval for this study was obtained from the Ethics Committee at the University of Konstanz.

References

[1] Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Applied Linguistics Non Series. Cambridge University Press, 2001.

[2] Deutscher, G. *The Unfolding of Language: An Evolutionary Tour of Mankind's Greatest Invention*. Henry Holt and Company, 2006.

[3] Feusner, M., and Lukoff, B. Testing for statistically significant differences between groups of scan patterns. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ETRA '08, ACM (2008), 43–46.

[4] Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., and Kingstone, A. Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition 117*, 3 (Dec. 2010), 319–331.

[5] Hirvenkari, L., Ruusuvuori, J., Saarinen, V.-M., Kivioja, M., Peräkylä, A., and Hari, R. Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer. *PLoS ONE 8*, 8 (Aug. 2013).

[6] International Organization for Standardization. Codes for the representation of names of languages – Part 1: Alpha-2 code. Standard, Geneva, CH, July 2002.

[7] Jarodzka, H., Holmqvist, K., and Nyström, M. A Vector-based, Multidimensional Scanpath Similarity Measure. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, ACM (New York, NY, USA, 2010), 211–218.

[8] Marschner, L., Pannasch, S., Schulz, J., and Graupner, S.-T. Social communication with virtual agents: The effects of body and gaze direction on attention and emotional responding in human observers. *International Journal of Psychophysiology 97*, 2 (Aug. 2015), 85–92.

[9] Martínez-Gómez, P., and Aizawa, A. Recognition of Understanding Level and Language Skill Using Measurements of Reading Behavior. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, ACM (New York, NY, USA, 2014), 95–104.

[10] Rayner, K., Slattery, T. J., and Bélanger, N. N. Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review 17*, 6 (2010), 834–839.

[11] Salvucci, D. D., and Goldberg, J. H. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ACM (2000), 71–78.

[12] Sereno, S. C., and Rayner, K. Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in cognitive sciences 7*, 11 (2003), 489–493.

[13] Steptoe, W., Wolff, R., Murgia, A., Guimaraes, E., Rae, J., Sharkey, P., Roberts, D., and Steed, A. Eye-tracking for Avatar Eye-gaze and Interactional Analysis in Immersive Collaborative Virtual Environments. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, ACM (New York, NY, USA, 2008), 197–200.

[14] Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, ACM (New York, NY, USA, 2001), 301–308.

[15] Wood, E., and Bulling, A. EyeTab: Model-based gaze estimation on unmodified tablet computers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, ACM (2014), 207–210.

[16] Yamazoe, H., Utsumi, A., Yonezawa, T., and Abe, S. Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ETRA '08, ACM (2008), 245–250.

[17] Yoshimura, K., Kise, K., and Kunze, K. The eye as the window of the language ability: Estimation of English skills by analyzing eye movement while reading documents. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, IEEE (2015), 251–255.