Eye typing in application: A comparison of two systems with ALS patients

Sebastian Pannasch
Applied Cognitive Research/Psychology III,
Technische Universitaet Dresden

Jens R. Helmert
Applied Cognitive
Research/Psychology III, Technische
Universitaet Dresden

Susann Malischke
Applied Cognitive Research/
Psychology III, Technische
Universitaet Dresden

Alexander Storch
Department of Neurology, Technische
Universitaet Dresden

Boris M. Velichkovsky
Applied Cognitive
Research/Psychology III,
Technische Universitaet Dresden

A variety of eye typing systems has been developed during the last decades. Such systems can provide support for people who lost the ability to communicate, e.g. patients suffering from motor neuron diseases such as amyotrophic lateral sclerosis (ALS). In the current retrospective analysis, two eye typing applications were tested (EyeGaze, GazeTalk) by ALS patients ($N = 4$) in order to analyze objective performance measures and subjective ratings. An advantage of the EyeGaze system was found for most of the evaluated criteria. The results are discussed in respect of the special target population and in relation to requirements of eye tracking devices.

Keywords: Eye-typing, Usability, Fixation Duration, Amyotrophic lateral sclerosis

Introduction

Being able to express oneself verbally is fundamental for quality of life. However, people suffering from high-level motor disabilities are often not able to carry out interpersonal communication fluently (Bates, Donegan, Istance, Hansen, & Räihä, 2007). In search for supporting communication abilities, progress has been made recently by implementations of brain computer interfaces. It was shown that EEG registration can allow for controlling a computer, e.g. typing (Nijboer et al., 2008). However, the same functionality, with less effort for the subjects and higher selection rates, can be achieved with gaze tracking systems. In fact, a broad range of gaze based computer interaction systems is already available, e.g. eye typing systems (GazeTalk [Hansen, 2006], Dasher [Ward & MacKay, 2002], EyeGaze [Cleveland, 1994]) and gaze based entertainment systems (adventure game Road to Santiago [Hernandez Sanchiz, 2007], puzzle [Vysniauskas, 2007], painting application EyeArt [Meyer & Dittmar, 2007]. For patients suffering from a lack of possibilities to communicate with their environment, especially eye typing systems represent a way to recover communicative abilities. Typed text spoken by a computer speech engine can allow for possibilities of verbal communication.

Several approaches for the design of such interfaces have been proposed. For instance, Ward and MacKay (2002) presented a system, where a dynamically modified display provide a highly efficient method of text entry. Huckauf and Urbina (this issue) could recently demonstrate the advances of two-dimensional, circular menus containing pie-shaped slices. However, most eye typing approaches are still based on on-screen keyboard-like interfaces that are driven by gaze movements (see for an overview Majaranta & Räihä, 2002). Disabled people can regain their ability to communicate by "typing" with their eyes. Some of these systems use hierarchical selection schemes with and without word prediction units (e.g.

Journal of Eye Movement Research
2(4):6, 1-8

Pannasch, S., Helmert, J.R., Malischke, S., Storch, A. & Velichkovsky, B.M. (2008)
Eye Typing in application: A comparison of two systems with ALS patients

Hansen, Hansen, & Johansen, 2001); others rely on a single level graphical layout with all possible input possibilities visible at the same time (e.g. Cleveland, 1994, see Figure 1, top panel). Using hierarchical selection schemes imply that less information is displayed at a time which allows larger buttons for the graphical interface (see Figure 1, bottom panel). Hence, gaze control becomes possible with less accurate (and therefore less expensive) eye tracking devices.

Besides merely technical requirements, usability becomes an important issue in evaluating these systems (see in particular ISO 9126-1, 2000; ISO 9241-1, 1997). A general problem of usability testing is however, that it is hardly possible to judge the usability of a system *per se* as the measures have no absolute scale. A reasonable approach is therefore to compare two or more systems using the same measures (Itoh, Aoki, & Hansen, 2006). This comparative approach can be applied to usability analysis of eye-typing systems. In the same vein, one of the most important aspects of usability is "learn ability". Taking into account a relatively small number of users in need of eye-typing interfaces, a small-sample within-subject (repeated measurement) procedure should be the method of choice.

Another issue that might be of importance for the usability evaluation of eye-typing software is the group of potential users. Since the speed of eye typing will always be below that of normal conversation (> 100 wpm) these systems are mainly developed for disabled people, especially because gaze typing represents their only means for communication. Although it is always emphasized in research investigating such systems, that eye-typing solutions are intended to provide support for people who cannot use standard keyboard or mouse (Majaranta & Räihä, 2002; Ward & MacKay, 2002) in most of the studies normal subjects participated (c.f. Hansen, Torning, Johansen, Itoh, & Aoki, 2004; Itoh et al., 2006; Majaranta, MacKenzie, Aula, & Räihä, 2006) with only few exceptions (e.g. Bates et al., 2007; Gips, DiMattia, Curran, & Olivieri, 1996). Apart from the fact that testing is easier and less time-consuming with normal subjects than with disabled people, e.g. ALS patients, this lack of empirical work is surprising. Inviting the intended users to participate in such an investigation should help to adopt the solutions to the users' needs. Moreover, it can be assumed, that the motivation of such users differs from that of normal subjects.

In the present study, two eye-typing systems, namely GazeTalk 5.0 (Hansen et al., 2001) and EyeGaze (Cleveland, 1994), were used and evaluated by patients suffering from amyotrophic lateral sclerosis (ALS). Although both systems have a keyboard like design, GazeTalk consists of a hierarchical menu structure including a word prediction (Hansen et al., 2001) whereas EyeGaze represents a rather simple on-screen keyboard. Moreover, we were interested if the participation in the experiment influenced the subjective experience of depression.

Methods

Participants

Three female and one male subject ranging in age from 47 to 79 years (mean age: 59 years) were included into retrospective data analysis. All were diagnosed to have a locked-in syndrome caused by ALS according to the El Escorial criteria with the subtype of the bulbar form. All subjects had normal vision, or by glasses corrected to normal vision and German as first language. None of the subjects were able to communicate by voice or by manual signs and none of them had previous experience with eye tracking systems.

Apparatus and questionnaire materials

A binocular Eyegaze Analysis System (LC Technologies, VA, USA) with remote binocular sampling rate of 120 Hz and an accuracy of about 0.45° was used in this investigation. Fixations and saccades were defined using the fixation detection algorithm supplied by LC Technologies: A fixation onset was identified if 6 successive samples were detected within a range of less than 25 pixels; accordingly the offset was detected if this criterion was not longer valid. Since the EyeGaze software uses an internal smoothing algorithm based on 10 data samples for cursor movements, we developed a similar eye-mouse program for the GazeTalk system using a moving average smoothing algorithm with a bin size of 10 samples. Therefore, the gaze prediction delays were comparable. The dwell selection time was set to 800 ms for both systems. The EyeGaze system was used in the standard QWERTY layout (see Figure 1, top panel). For GazeTalk the word prediction was limited to four words; in a pretest the word prediction was trained to the used stimuli and evaluated.

Journal of Eye Movement Research
2(4):6, 1-8

Pannasch, S., Helmert, J.R., Malischke, S., Storch, A. & Velichkovsky, B.M. (2008)
Eye Typing in application: A comparison of two systems with ALS patients

The questionnaire materials comprised two different inventories, which also were completed using the gaze to control the interface.

Information about the usability was gathered by the ISONORM 9241/10-Short questionnaire (Pataki, Sachse, Prümper, & Thüring, 2006). Therefore, seven items regarding the usability of both eye-typing systems had to be evaluated on a 4-point Liekert scale.

The rate of depression was explored using the ADI-12 (Kübler, Winter, Kaiser, Birbaumer, & Hautzinger, 2005), a short self-report screening questionnaire consisting of 12 items. None of the items refer to somatic or motor-related symptoms taking into account the progressive physical impairment which may culminate in severe motor paralysis and life sustaining treatment. The ADI-12 assesses a homogeneous, one-dimensional construct that is described as "mood, anhedonia and energy" (Kübler et al., 2005). Scores range from '0' (best possible) to '48' (worst possible) with scores between 22 and 28 indicating mild depression and those above 28 clinically relevant symptoms (Kübler et al., 2005). The validity of the ADI-12 was recently investigated and reported (Hammer, Hacker, Hautzinger, Meyer, & Kübler, 2008).

Procedure

The main task for the subjects was to type ten blocks of sentences with their gaze over a period of five recording sessions at five successive days. Each day two blocks were completed. Each block consisted of five sentences with 131 characters per block. Sentences were German translations of the "Phrase Set" by MacKenzie and Soukoreff (2003), which was specifically designed for experiments with eye typing. Subjects' task was to type the sentences shown on a laptop computer as fast and accurate as possible. For typing only small letters were used. Two of the subjects started with GazeTalk and other two with EyeGaze (see Figure 1).

Subsequently, the eye-typing systems were alternated between the recording sessions. Moreover, ten minutes breaks were given between the blocks. Each block started with a 9-point calibration procedure. Before and after each recording session the gaze-aware questionnaire was completed and every other day the ADI-12 was administered. At the end of the final session the usability of both eye-typing systems had to be evaluated with the ISONORM questionnaire.
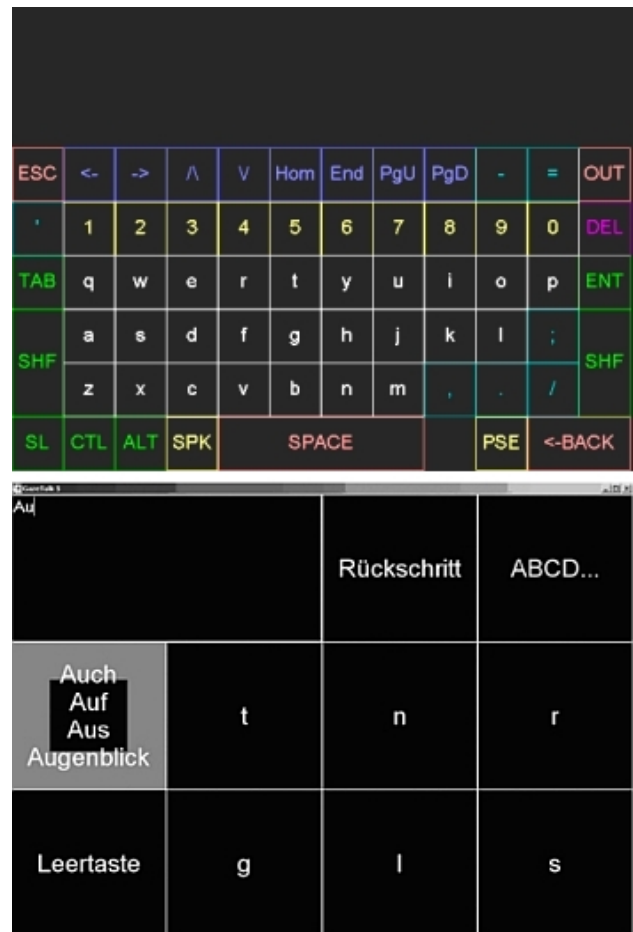

Figure 1. Screenshots of the tested eye typing systems, EyeGaze (top) and GazeTalk (bottom).GazeTalk used a word prediction restricted to four words (see mid left button). The grey rectangle indicates the currently selected button and shrinks with increasing dwell time. Note: "Rückschritt" means backspace and "Leertaste" designates space bar.

Data Analysis

Performance data were averaged per subject per session and applied to repeated measures ANOVAs. Typing speed (characters/min), task efficiency, total and corrected error rates were investigated. To estimate the error rates we used the taxonomy suggested by Soukoreff and MacKenzie (2003): number of correct inputs (C), number of keystrokes invested in error correction (F), number of errors made and corrected (IF) and number of errors but not corrected (INF). Following this categorisation, the total error rate would be $(INF + IF) / (C + INF + IF)$ x 100% and the corrected error rate would be $IF / (C + INF + IF)$ x 100% (see Soukoreff & MacKenzie, 2003 for further details).

Journal of Eye Movement Research
2(4):6, 1-8

Pannasch, S., Helmert, J.R., Malischke, S., Storch, A. & Velichkovsky, B.M. (2008)
Eye Typing in application: A comparison of two systems with ALS patients

Task efficiency was estimated considering quantity (*QN*, was set to 100% because our subjects always completed all sentences), quality (*QL*, 100% - total error rate) and time (*T*, total time of a session). Hence, task efficiency would be (*QN* + *QL*) / *T*. Questionnaire data was pooled and analysed according to the manual instructions (Kübler et al., 2005; Pataki et al., 2006).

One subject did withdraw from participation after the third session. Due to the rare availability of data from this group of patients, the data recorded so far was included in the further data processing.

Results

The data was analysed in order to compare typing speed, error rate and task efficiency for both eye-typing systems. Moreover, differences in the usability evaluation were investigated. Finally we compared the subjective experience of depression across the recording sessions.

Typing speed was analysed by a 2 (software) x 5 (session) factors repeated measures ANOVA. Significance was obtained for the main effects of software, $F(1,2) = 56.26$, $p = .017$, and session, $F(4,8) = 10.25$, $p = .003$. Testing also revealed a significant interaction between software and session, $F(4,8) = 6.84$, $p = .011$. Typing speed was always higher for EyeGaze than for GazeTalk (*M*s = 17 vs. 6.8 cpm). Moreover, the difference in typing speed between the first and last session was higher for EyeGaze than for GazeTalk (*M*s = 9.5 vs. 2.1 cpm) as qualified by the significant interaction (see Figure 2).
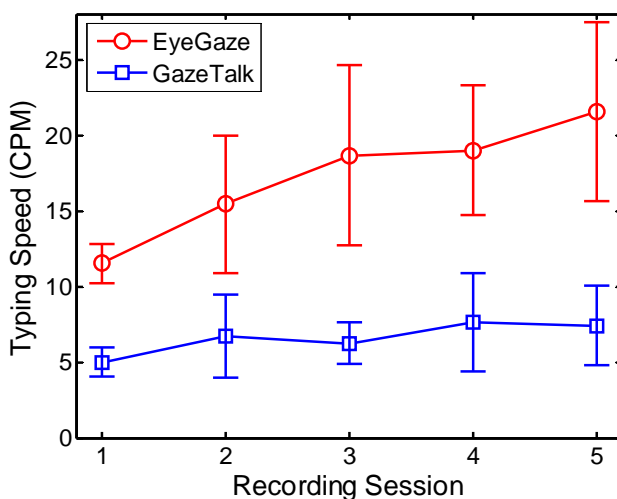
Figure 2: Mean typing speed across the recording sessions for both systems.

Percentage values of total and corrected error rates were entered into two 2 (software) x 5 (session) factors repeated measures ANOVAs. Testing for total error rates revealed no significance for the main effect software, $F(1,2) = 8.49$, $p = .100$, but for session, $F(4,8) = 5.28$, $p = .022$. No further interaction was obtained, $F < 1$. As depicted in Figure 3A, there is a general decrease of the total error rate across the sessions for both system (difference values first – last session: EyeGaze = 14.33; GazeTalk = 6.15). However, for the corrected error rate testing yielded a reliable effect for software, $F(1,2) = 50.26$, $p = .019$, but not for session, $F(4,8) = 3.25$, $p = .073$. No further interaction was found, $F < 1$. Thus, considering the corrected error rate reveals less errors for the use of the EyeGaze software in comparison to the GazeTalk system (*M*s = 10.94 vs. 22.42; see Figure 3B).
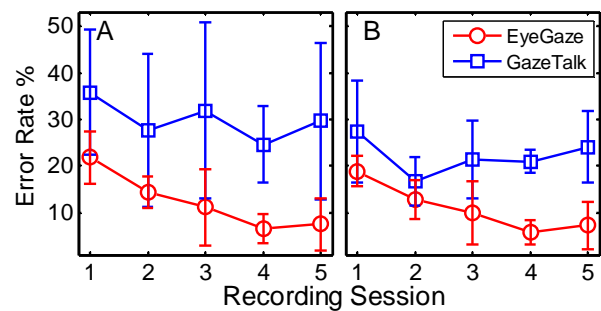
Figure 3. Mean percentage of error rates for total error (A) and corrected error (B) for both systems across the recording sessions.

Task efficiency was analysed by a 2 (software) x 5 (session) repeated measures ANOVA. Analysis yielded significance for software, $F(1,2) = 48.76$, $p = .020$, and for session, $F(4,8) = 11.02$, $p = .002$. Moreover a significant interaction was found, $F(4,8) = 6.74$, $p = .011$. The difference between both software systems suggests a higher task efficiency for the EyeGaze than the GazeTalk system (*M*s = 11.56 vs. 4.19; see Figure 4). In addition the task efficiency improves from the first to the last session for the EyeGaze but not for the GazeTalk system as qualified by the significant interaction (*M*s = 8.20 vs. 1.34).

Journal of Eye Movement Research
2(4):6, 1-8

Pannasch, S., Helmert, J.R., Malischke, S., Storch, A. & Velichkovsky, B.M. (2008)
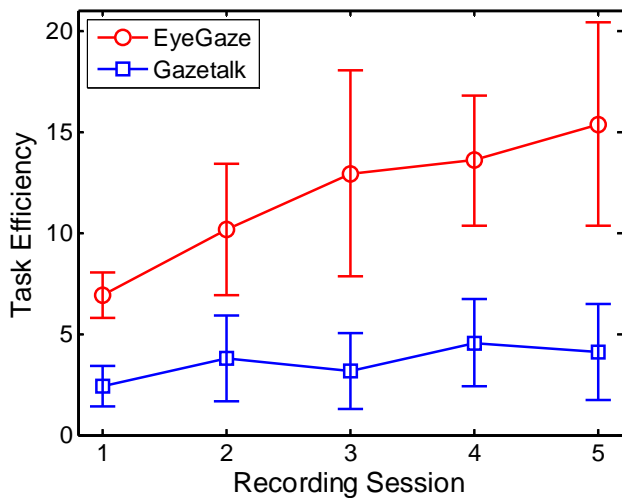Eye Typing in application: A comparison of two systems with ALS patients

Figure 4. Mean task efficiency for both systems across the recording sessions.

Moreover, the results of the evaluation of criteria obtained with the ISONORM usability questionnaire (Pataki et al., 2006) were analyzed. In general the EyeGaze software was rated more positive than GazeTalk on all scales (see Figure 5). However, testing for significance with single paired Wilcoxon tests for each criterion could not confirm this trend, $z \leq 1.84$, $ps \geq .05$.
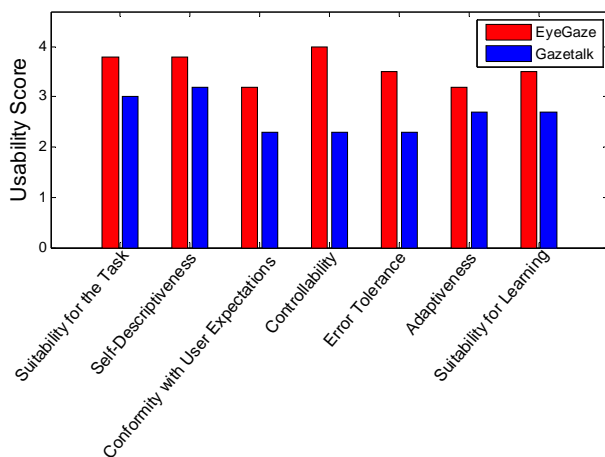

Figure 5. Mean usability scores for both software systems.

Depression rate values, collected after the first, third and fifth session were applied to a three factorial repeated measures ANOVA. We found a slight but not significant decrease in depression assessment, $F(2,4) = 4.92$. $p = .083$ (see Figure 6). Further data need to be collected in order to get more precise data about this relationship.
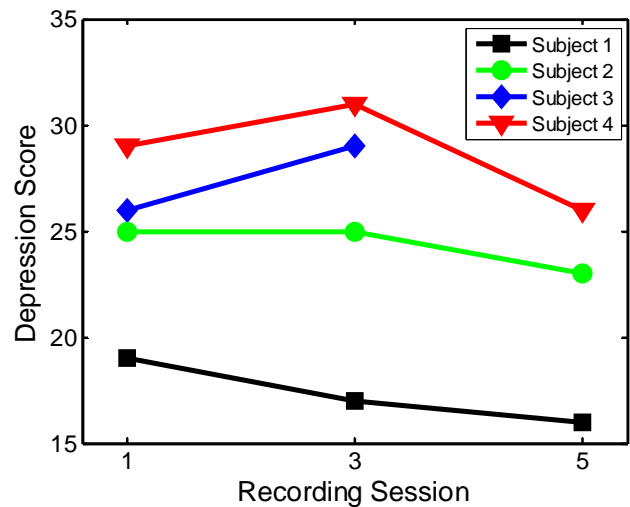

Figure 6. Depression scores at the three recording sessions.

Discussion

In this retrospective analysis, two eye typing software systems, namely GazeTalk and EyeGaze, were compared in terms of objective performance measures and subjective usability criteria. In contrast to most previous work (e.g. Itoh et al., 2006; Majaranta et al., 2006), in the current study patients suffering from ALS participated. Since it is documented in the literature that depression is associated with survival in ALS patients (McDonald, Wiedenfeld, Hillel, Carpenter, & Walter, 1994), we additionally used the ADI-12 depression inventory (Kübler et al., 2005).

Due to the very small sample size of $N = 4$ (and the additional withdrawal of one subject after completing half of the study) the results of the present research need a careful interpretation. Notwithstanding the development of gaze based interaction systems mainly aims to support communication abilities for people who are suffering from the so called locked-in state. The current study therefore included a sample from the target population into a systematic investigation.

The results from the objective measurements were rather straightforward. First of all, typing speed was higher for EyeGaze than for the GazeTalk system. This difference was obtained already within the first session; during the subsequent recordings a stronger increase in typing speed was found for the EyeGaze system. It is known that the speed of typing will always be below that

Journal of Eye Movement Research
2(4):6, 1-8

Pannasch, S., Helmert, J.R., Malischke, S., Storch, A. & Velichkovsky, B.M. (2008)
Eye Typing in application: A comparison of two systems with ALS patients

of verbal communication. However, using eye typing as a means for communication, speed is of great importance and will be one of the major criteria of successful applications. As our results suggest, best performance was achieved with a rather simple keyboard design in contrast to a system featuring word prediction.

Although not statistically significant, the total error rate was lower for EyeGaze than for GazeTalk. Even more the decrease of error rates from the first to the last session was 66% for EyeGaze in contrast to only 17% for GazeTalk. In terms of corrected errors EyeGaze was also superior, although this effect might be due to the lower overall error rate for EyeGaze. One reason for the different error rates could be the word prediction that was available in GazeTalk. In cases where the wrong word was selected it had to be deleted characterwise using the backspace key. With the EyeGaze words had to be typed letter by letter, thus such errors were not possible. Concerning the total error rate a general decrease over the recording sessions was found emphasizing that the quality of eye typing improved relatively fast for our test subjects. Nevertheless, during the five sessions fewer errors were made using simple keyboard layout without additional functions (e.g. word prediction).

The most important performance criterion is task efficiency, which includes error rate as well as time for task completion. Analysis for this criterion revealed that the EyeGaze was significantly better than GazeTalk with an efficiency value at the last session being three times higher. Taken together, all objective performance measures from the current investigation indicate that eye typing is faster and more efficient using the EyeGaze system. However, even if our results would underline this interpretation, at least two points need to be considered in relation to this outcome. First, since GazeTalk is based on a hierarchical menu system (Hansen et al., 2001) it might require more cognitive efforts to get familiar with the design and the use of the system. Once the user has become fully acquainted to the functionality of the system, the results of such a comparison could be completely different. Additionally, the GazeTalk interface supports also to control other computer programs such as multimedia and internet applications. Once the user knows how to use the GazeTalk system a huge range of possibilities is available. The second point is closely related to the first, the hierarchical design of GazeTalk has the advantage that this eye typing system is appropriate for

low spatial resolution eye-tracking devices (Itoh et al., 2006). GazeTalks design of the graphical user interface with the rather large buttons (see Figure 1) can compensate for spatial inaccuracy as well as for calibration errors. In contrast, for the EyeGaze system with the small keys a high spatial resolution and accuracy (<0.5°) is essential, which of course increases the cost for such a system. A comparison of both systems on a longer timescale could provide clarification about these issues.

However, the outcome of the objective measures is also reflected in the subjective assessments. Concerning the usability criteria both systems were evaluated rather positive, but a nonsignificant superiority of the EyeGaze was found. In addition, it should be pointed out that our subjects mostly had only little experience with computers and typing, both are considerable factors for such an investigation (see Helmert et al., this issue). Due to the fact that the mean age of our sample was 59 years, it can be assumed that at least the experience with computers will increase in the population for the next generations.

For the depression inventory no significant effects were obtained. However, it should be mentioned that none of our subjects did show clinical relevant symptoms of depression.

Indeed, the biggest critique on our study is the small sample size of only four persons, which make the application of elaborated statistical tools somewhat problematic. Nevertheless, we think that comparative usability studies should be preferably approached in small-size within-subject or even longitudinal designs in order to improve the eye-typing systems for the sake of those who need them most: the disabled users. This can be achieved comparing certain features more extensively as done in this study, which might result in an integration of those features that are the most appropriate.

Acknowledgements

Journal of Eye Movement Research
2(4):6, 1-8

Pannasch, S., Helmert, J.R., Malischke, S., Storch, A. & Velichkovsky, B.M. (2008)
Eye Typing in application: A comparison of two systems with ALS patients

References

Bates, R., Donegan, M., Istance, H. O., Hansen, J. P., & Räihä, K.-J. (2007). Introducing COGAIN: communication by gaze interaction. *Universal Access in the Information Society, 6*(2), 159-166.

Cleveland, N. R. (1994). Eyegaze Human-Computer Interface for People with Disabilities. In *Proceedings of the First Automation Technology and Human Performance Conference*. Washington, DC.

Gips, J., DiMattia, P., Curran, F. X., & Olivieri, P. (1996). Using EagleEyes - an electrodes based device for controlling the computer with your eyes - to help people with special needs. In *Fifth International Conference on Computers Helping People with Special Needs (ICCHP'96)*. Linz, Austria.

Hammer, E. M., Hacker, S., Hautzinger, M., Meyer, T. D., & Kübler, A. (2008). Validity of the ALS-Depression-Inventory (ADI-12) - A new screening instrument for depressive disorders in patients with amyotrophic lateral sclerosis. *Journal of Affective Disorders, 109*(1-2), 213-219.

Hansen, J. P. (2006). GazeTalk [Computer software and manual]. *Journal*. Retrieved from http://www.cogain.org/results/applications/gazetalk/about

Hansen, J. P., Hansen, D. W., & Johansen, A. S. (2001). Bringing Gaze-based Interaction Back to Basics. In *Proceedings of Universal Access in Human-Computer Interaction (UAHCI 2001)* (pp. 325–328). New Orleans, Louisiana.

Hansen, J. P., Torning, K., Johansen, A. S., Itoh, K., & Aoki, H. (2004). Gaze typing compared with input by head and hand. In *Eye Tracking Research & Applications Symposium, ETRA 2004* (pp. 131-138). New York: ACM Press.

Hernandez Sanchiz, J. (2007). Road to Santiago [Computer software and manual]. *Journal*. Retrieved from http://www.cogain.org/downloads/leisure-applications/road-to-santiago

ISO 9126-1. (2000). *Software Engineering - Product quality - Part 1: Quality Model*. Geneva, Switzerland: International Organization for Standardization.

ISO 9241-1. (1997). *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTS) - Part 1: General Introduction*. Geneva, Switzerland: International Organization for Standardization.

Itoh, K., Aoki, H., & Hansen, J. P. (2006). A comparative usability study of two Japanese gaze typing systems. In *Eye Tracking Research & Applications Symposium, ETRA 2006* (pp. 59-66). New York: ACM Press.

Kübler, A., Winter, S., Kaiser, J., Birbaumer, N., & Hautzinger, M. (2005). Das ALS-Depressionsinventar (ADI): Ein Fragebogen zur Messung von Depression bei degenerativen neurologischen Erkrankungen (amyotrophe Lateralsklerose)— A questionnaire to measure depression in degenerative neurological diseases. *Zeitschrift für Klinische Psychologie und Psychotherapie, 34*(1), 19-26.

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI*, 754-755.

Majaranta, P., MacKenzie, I. S., Aula, A., & Räihä, K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society, 5*(2), 199-208.

Majaranta, P., & Räihä, K.-J. (2002). Twenty years of eye typing: Systems and design issues. In *Eye Tracking Research & Applications Symposium, ETRA 2002* (pp. 15-22). New York: ACM Press.

McDonald, E. R., Wiedenfeld, S. A., Hillel, A., Carpenter, C. L., & Walter, R. A. (1994). Survival in amyotrophic lateral sclerosis. The role of psychological factors. *Archives of Neurology, 51*, 17-23.

Meyer, A., & Dittmar, M. (2007). EyeArt [Computer software and manual]. *Journal*. Retrieved from http://www.cogain.org/downloads/leisure-applications/eyeart/

Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., et al. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology, 119*(8), 1909-1916.

Pataki, K., Sachse, K., Prümper, J., & Thüring, M. (2006). ISONORM 9241/10-S: Kurzfragebogen zur Software-Evaluation. In F. Lösel & D. Bender (Eds.), *Berichte über den 45. Kongress der Deutschen Gesellschaft für Psychologie* (pp. 258-259). Lengerich: Pabst Science Publishers.

Soukoreff, R. W., & MacKenzie, I. S. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI*, 113-120.

Vysniauskas, V. (2007). Puzzle [Computer software and manual]. *Journal*. Retrieved from http://www.cogain.org/downloads/leisure-applications/puzzle

Ward, D. J., & MacKay, D. J. C. (2002). Artificial intelligence: Fast hands-free writing by gaze direction. *Nature, 418*(6900), 838-838.