Univ Access Inf Soc (2009) 8:297–310
DOI 10.1007/s10209-009-0152-5

LONG PAPER

Learning gaze typing: what are the obstacles and what progress to expect?

Hirotaka Aoki · John Paulin Hansen · Kenji Itoh

Published online: 19 March 2009

Abstract Gaze interaction is a promising input modality for people who are unable to control their fingers and arms. This paper suggests a number of new metrics that can be applied to the analysis of gaze typing interfaces and to the evaluation of user performance. These metrics are derived from a close examination of eight subjects typing text by gazing on a dwell-time activated onscreen keyboard during a seven-day experiment. One of the metrics, termed "Attended keys per character", measures the number of keys that are attended for each typed character. This metric turned out to be particularly well correlated to the actual numbers of errors committed ($r = 0.915$). In addition to introducing metrics specific for gaze typing, the paper discusses how the metrics could make remote progress monitoring possible and provides some general advice on how to introduce gaze typing for novice users.

Keywords Gaze typing · Learning process · Dwell time gaze typing system · Gaze metrics

H. Aoki (✉) · K. Itoh
Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2-12-1, Oh-okayama, Meguro-ku, Tokyo 152-8552, Japan
e-mail: aoki.h.ad@m.titech.ac.jp

J. P. Hansen
Innovative Communication Group, IT University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen S, Denmark
e-mail: paulin@itu.dk

1 Introduction

Gaze interaction has been recognized as a promising augmented and alternative communication (AAC) technology for severely disabled people who are unable to control their hands [26]. For example, it supports a number of people with late-stage ALS/MND (Amyotrophic lateral sclerosis), who have completely lost their ability to control muscle movement. By the use of eye-tracking technology, these people can communicate with friends and family just by gazing at a specific area of a computer screen to select the characters or commands they need. A typical gaze interaction system comprises an eye-tracking device and on-screen virtual keys for input. The eye tracker basically controls the pointer on the computer screen. Gaze interaction systems adopt one of the following three input principles:

1. *Point-and-click*. The position of a screen-pointer follows the users gaze. Selections are made by clicking the mouse button or activating an alternative switch. This requires that the user has the physical capacity to control a switch for selection.
2. *Dwell time*. An on-screen key gets activated when gazed at. The selection is completed by keeping the gaze fixed on the key for a certain time (i.e., dwell time). The purpose of this "clutch" function is to avoid unintended selections. Some severely disabled people who have no other stable muscle control other than eye movements use this input principle. GazeTalk [11] (Fig. 1), iKey [28], and ERICA [15] are examples of these kinds of systems.
3. *Continuous navigation*. Text typing is driven by continuous search and navigation controlled by an eye tracker. Dasher is an example of a gaze typing

Fig. 1 Screenshot of GazeTalk-J used in the present experiment

interface without dwell time activation [38]. Typing is controlled by continuous two-dimensional search and navigation in a column of characters that are ordered alphabetically and scaled according to their probabilities. The user makes the selections by searching for target letters without unnatural dwells breaking the flow. Stargazer [10] is another gaze typing system with continuous navigation. The user zooms in on characters displayed in a ring and a character is selected by flying into it.

This paper only examines a system with dwell-time selections. The difference between click and dwell selections was investigated by Hansen et al. [11]. Itoh et al. [17] analyze the difference between dwell time and continuous navigation input principles.

1.1 Why should we measure gaze typing?

Eye typing takes time to learn [6]. By experience, a few users can do it within minutes, while most people needs hours of practice. Novice users should be aware in advance that it is very common to type slowly in the beginning, and they should be presented with convincing documentation of the progress they make.

For more than a century, traditional touch typing skills have been measured by metrics. Words per minute (WPM) and error-rate are commonly used and frequently included as feedback in e-learning software that teaches students to type. Within ergonomics, new keyboard design or alternative input modalities are often evaluated by similar measures of the typing speed and accuracy [25]. In this paper, two traditional typing metrics are used to describe how users learn to master gaze interaction. But since gaze tracking provides unique information on the users' eye movements, it is also possible to include gaze data in the analysis. For instance, do the users look at the right places or do they look at irrelevant elements? Analysis of gaze data estimating users' behavior is the primal contribution of this paper. The investigation is motivated by the potentials for gaze performance measurements to be collected automatically at a user's home and sent off to, e.g., a communication specialist when users are on-line. Such remote monitoring of user progress could help communication specialists when timing their advice or visits. For instance, it may be recognized when a user has reached a stable, but sub-optimal level of performance, and better strategies could then be suggested to him/her. A disease like ALS/MND may suddenly affect the ability to gaze type. Eye trackers may fail partly without the user discovering it. Early detection of performance deterioration could warn the communication specialist that help was required.

Remote data collection [8] seems particular relevant for the design of AAC-systems for individuals with a mobility problem, and for user groups that are relatively few in numbers and thus spread widely in the community. Remote data collection can provide a large amount of data without intruding on the users. Several researchers and usability practitioners have emphasized the advantages of remote data collection [7, 14, 36, 37, 40, 41]. Ivory and Hearst [18] conducted an extensive survey of existing techniques, and provided useful reference material that systematically summarizes the range of applicability of each possible technique.

When evaluating gaze interaction systems, user data from large populations would enable researchers to benchmark gaze typing interfaces or eye-tracking equipment 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, the possibility to identify those few users with exceptionally fast typing could provide very useful inspiration from them on how to help others improve.

Text data from people's personal communication is to be treated with confidentiality. Therefore, the performance metrics should be calculated locally on the users PC and only the metrics, not the text itself, should be submitted, in order to protect the privacy of the user.

1.2 Pros and cons of gaze interaction

Previous research has indicated that gaze interaction may be faster than hand-mouse for pointing [39]. People can look at what they want more quickly than pointing with a finger, and usually they have gazed at the target already when they start to move their hands.

The very idea of gaze control itself seems intuitively understandable to users [34]. For instance, people frequently gaze at others to indicate whom they are addressing while talking. When users experience gaze interaction for the first time, most of them immediately become aware of the potentials of this technology and they readily understand the impact it may have [9].

Traditional AAC input devices like switches or joysticks may cause repetitive strain injury since they are operated all day long. The eyes have a very low mass compared to, e.g., an arm, and they have a very high amount of muscles to control them, so it is possible to move the eyes intensively for several hours without getting tired [33].

Despite these advantages of gaze input, there are also some serious challenges. One of the critical disadvantages is that it is difficult to consciously control the gaze all the time [21]. Gazing is performed more or less automatically when acquiring information from the environment. Gazing at a specific object for some time is difficult [21]. Moreover, pointing by gaze is less accurate than pointing by hand. This is partly because of anatomical features of the eyes (such as the size of fovea on retina [29]), and partly because of eye movement disturbances such as blinks, tremor, etc.

Consequently, the error-rate associated with gaze interaction is known to be much higher than for conventional input modalities [11, 30]. Bates [6] reported that novice users with less than 2 h of experience made 50% unnecessary key activations to complete a typing task. Even relatively experienced users who had used gaze typing for more than 6 h still made unnecessary key activations at a rate of approximately 28% of the minimum activations required. Istance et al. [16] observed that users of gaze typing needed to spend extra time to correct errors during gaze typing, resulting in very slow typing speeds of around one WPM. Aoki et al. [2] examined the usability of a large-button gaze typing system in comparison with other input modalities (mouse and head-tracking). This study revealed that although the error-rate for gaze typing was worse than that for mouse-based interaction, the improvements were dramatically better than improvements in mouse-pointer typing. In fact, it was predicted from the data that the error-rate for gaze would drop to the same level as mouse interaction after some days of practice.

1.3 Learnability of gaze interaction

Why can it be so difficult to control eye movements during gaze interaction? How long does it take to become acquainted with input by gaze? These two questions are of great importance to the AAC specialists. Once they understand the basic challenges of gaze interaction, they can better aid new users and they can introduce gaze interaction with clear expectations of their progress.

Jacob and Karn [22] suggested a taxonomy of eye movement-based human–computer interaction characterized by two distinctions: the nature of eye movements and the system's response. They argued that each of these distinctions can be considered as "natural" (meaning that there are similar types to be found in the real-world) or "unnatural" (meaning that there are no counterparts in the real-world). According to this taxonomy, most of the conventional gaze interaction systems fall into the category of unnatural eye movement/unnatural system's response. This indicates that people will have to learn how to use eye movements for input. Controlled eye movements are quite different from eye movements in the real world where people automatically acquire visual information without being aware of the scanning process.

The issue of learnability has been addressed by a few other studies. The majority of results from experiments with gaze interaction systems has been obtained from analysis with traditional measures of typing speed [38] and error-rate [13], but relatively little research has focused on measures including gaze behavior itself [4]. One of the few studies adopting eye movement patterns for evaluation of gaze interaction is the work by Majaranta et al. [27]. They used a metric termed "Number of Read Text Events per character", which measures how frequently a user directed his/her gazes toward the text field. This metric has no direct relation to the accuracy of gaze interaction as such, but may be regarded as an indirect measure, since people are more likely to look at the text they have produced when they are uncertain whether it has been typed correctly.

Aoki et al. [1] presented a learnability analysis by use of metrics termed "Rate of Midas Touch errors" and "Rate of premature movement errors". They showed that skills needed for gaze input were acquired within approximately 6 h of practice. However, they only included erroneous selections in their analysis. Additional analysis of gazes involved in correct selections will be needed to complete the full description of the learnability of gaze typing. The metrics by Aoki et al. [1] were derived in an ad hoc fashion without a unified understanding of the types of gaze actions possible. Consequently, metrics for all possible types of gaze actions on a dwell typing keyboard were not fully evolved. This paper therefore suggests a taxonomy of gaze actions involved in typing on a dwell-time keyboard and develops gaze metrics associated with each type of gaze action. The underlying objective is to perform a detailed analysis of how the metric values change with practice to highlight the learnability of gaze interaction.

2 Experiment

2.1 Subjects

Eight able-bodied Japanese university students—four female and four male, aged 18–22 years—participated in the experiment. All subjects had normal or corrected-to-normal vision, and none had any previous experience with gaze-typing systems. Each was paid approximately 5,000 Japanese yen (43 US dollars) for participating. In addition, a bonus of 5,000 yen was promised to the top three performing subjects (measured by their typing speed and accuracy). This was done to ensure that the subjects would maintain a high motivation throughout the experiment.

2.2 Gaze-typing interface

The Japanese version of GazeTalk, named GazeTalk-J, was selected for the experiment. GazeTalk [11] is a gaze interaction interface developed for severely disabled people such as people with ALS/MND. As shown in Fig. 1, it consists of large on-screen keys in a 3 × 4 keyboard layout. The upper left corner, a space the size of two adjacent keys, is used as a text-input field. The keys themselves are used for typing characters and commands such as delete, change mode, perform Kana-Kanji conversion (GazeTalk-J only), and so forth. The original version of GazeTalk is designed for European language users, and equipped with word-prediction functionality in order to ameliorate typing speed, the consequence of which is a keyboard with labels that change dynamically (Fig. 2). GazeTalk-J, on the other hand, does not provide word prediction because the structure of the Japanese language itself would render such functionality impractical. Instead, GazeTalk-J utilizes static key allocations within a hierarchical menu structure. Each character or command may be found in exactly the same position in the hierarchical menu, which allows to reproduce users' typing processes from text-input transition records.

Fig. 2 Person sitting in front of Japanese version of GazeTalk with a QuickGlance eye-tracking system mounted on the screen

Gazetalk-J was selected for three reasons. First, the system adopts the dwell time principle for key selections. Secondly, the onscreen keys are static in this system. This means that all the letters/commands are allocated to exactly identical positions. Finally, Gazetalk adopts large onscreen keys, which allows to ignore noises caused by eye-tracking inaccuracy, since this noise would have no impact on key selections. On a 17-inch monitor, the side of each onscreen key is around 5 cm. Most eye-tracking devices give a gaze position accuracy of 0.5–1.0°, which equals to approximately 0.5–1.0 cm on a monitor when the distance between the monitor and the subject is 60 cm. From the above-mentioned facts, it can be expected that the gaze behavior data obtained does not include any tracking inaccuracy-related noises.

2.3 Task

The experimental task was to transcribe Japanese text as quickly and accurately as possible using GazeTalk-J. Because the used eye-tracking device permits only minimal head movement, and also because members of the target end-user group have limited (if indeed any) muscle control themselves, the subjects were reminded to remain as still as possible during the experiment.

The experiment then ran as follows. The experimenter would read various Japanese sentences aloud. Simple and short sentences were chosen from a Japanese translation of Hans Christian Andersen's fairy tales, children's edition, selected because they could be easily understood.

Immediately after hearing each sentence, subjects were asked to type what they had heard and then convert the text into the correct combination of Kanji and Hiragana characters, including any alpha-numeric characters, using the built-in Kana-Kanji conversion program. They were told to correct immediately any typing errors that they happened to notice.

2.4 Apparatus

The GazeTalk-J was run on a personal computer (933-MHz CPU) operating with Windows 2000, which included the IME Kana-Kanji conversion program, as well as a 17-inch color monitor (1,024 × 768 pixels). The viewing distance between the subject and the screen was set at approximately 60 cm. As feedback to subjects, the key would be highlighted while the subject gazed at it, and a white area behind the key's label would shrink in size, indicating the dwell time remaining before activation. A QuickGlance system (EyeTech Digital System) was used as the eye-tracking device, with a tuning of 15 for the update rate and a setting of 7 for the smoothing factor.

2.5 Procedure

Each subject performed 22 experimental blocks in total over a period of seven working "days", as consecutively as possible, preferably day after day and with no more than a four-day interval. Each day comprised a certain number of "blocks", each of which was made up of five sentences, or a total of approximately 90 characters. On day 1, subjects performed one block of typing; on day 2, two blocks; on day 3, three blocks. For each of the following days, the subject performed four blocks. Between any two successive experimental blocks, the subject was permitted a short break of approximately 5 min.

Prior to the first block on day 1, subjects were instructed on how to perform the task; the typing system was also explained, along with the experimental procedure and the calibration of the eye-tracking device. Subsequently, subjects were given a 5-min demonstration of the gaze-typing system, which they observed, but did not experience yet. Afterward, the subject participated in a training exercise, involving the typing of 10–20 sentences using a mouse. The purpose of this exercise was to acquaint the subjects with the onscreen keyboard's hierarchical menu system, as well as all typing procedures. When subjects confirmed that they felt completely acquainted with the typing interface, they were permitted to proceed to the first experimental block of gaze typing. In this way, subjects acquired familiarity with the onscreen keyboard's hierarchical menu system but no physical training in gaze interaction. The eye-tracking system was calibrated before the start of each block.

3 Data analysis

The Gazetalk typing system had a logging facility that recorded all keys activated for more than 125 ms. This threshold is often used as the minimum time for defining a fixation among eye-tracking researchers [31], because it will exclude most random data recorded during a saccade. Since the threshold works on a key, the activations could possibly be made of several short fixations that felt within the same key area. The dwell time for the final completion of a key was set at 500 ms. If the eyes left a key during the dwell time, it would become deactivated, and the user would have to look at it again for 500 ms in order to complete the selection.

Key activations were classified into four categories on the basis of their duration and their deviation from the target key. A target key refers to the key that is the most correct next key to complete the task at any given moment. The target keys can be identified retrospectively by a close examination of all key activations, identifying every single keystroke made and deciding what key would then be the most correct next. For example, after activating an obviously incorrect key, one would decide that the subsequent target key would be "delete". At this moment, this identification process cannot be fully automated because there are many ways to produce a single sentence input in Japanese, including also the possibility of an error correction processes. They depend on users' chosen strategies, and identifying strategies automatically seems to be quite a challenge; more so, because the user may suddenly change strategies. Considering this, it was decided to perform the identification by hand even though this analysis is time consuming and requires a complete logging of all activations.

Figure 3 shows each activation category. The vertical axis indicates whether or not a specific activation matches completely with the target key. The horizontal axis makes a distinction between activations with longer or shorter durations than the predetermined dwell time.

Category 1 in Fig. 2 refers to correct activations of target keys for which the durations of those fixations was less than the pre-determined dwell time (i.e., <500 ms) but longer than the minimum activation threshold (i.e., >125 ms). This category consists of "Premature exits" initiated with the intention to activate a target key but interrupted before the end of the dwell time. As noted by Jacob [19], this category is a common cause for gaze-input errors.

Category 2 includes activations resulting in key selections that are correct.

Category 3, called "erroneous selections", consists of selections different from the target key. Erroneous selections can be related to the so-called "Midas Touch problem" termed by Jacob [20]: Everywhere the user looks

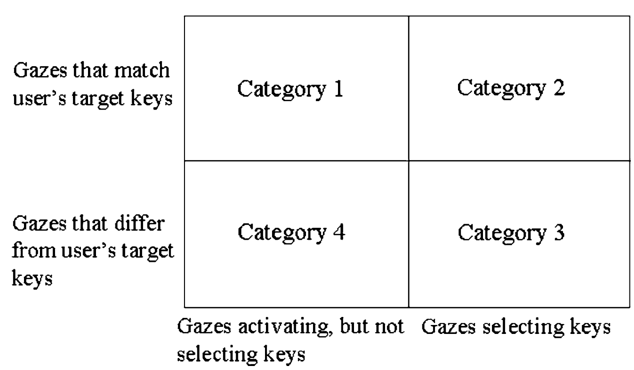

Fig. 3 Taxonomy of key activations in dwell time gaze typing systems

something may get selected, even without intention to do so. This kind of "slip" [32] happens despite the user's cognisance of the correct target key. The erroneous selections could also have been mistakes (c.f. Reason's distinction [32]) caused by a users inability to spell correctly. In the present experiment, however, the text was so easy that spelling mistakes were considered as negligible.

Category 4 is composed of activations on keys that are not targets, with durations that are less than the dwell time. It can be "scanning fixations" in a search for the target key, or it can be fixations without purpose (e.g., from erratic gaze behavior), that are too short to have any consequences in terms of selections.

3.1 Gaze interaction metrics

To analyze gaze interaction during the typing process, five metrics were identified corresponding to one of the activation categories in Fig. 2 and listed in Table 1.

3.1.1 Premature exits rate

The Premature Exits rate is calculated by dividing the number of premature exits by the number of characters typed. This metric is related to "Target re-entry" proposed by MacKenzie et al. [24]. The target re-entry metric is used by MacKenzie et al. to measure accuracy of pointing with computer pointing devices and obtained by counting the numbers of events in which the target is re-pointed within a pointer path. The premature exits rate represents a similar aspect as that of target re-entry.

3.1.2 Correct selections

This is calculated as the number of correct selections divided by the number of characters typed, representing the frequency of correct key activations, or matches with the target keys. The empirical measure of Correct Selections is slightly different from the Key Stroke Per Character (KSPC) index suggested by MacKenzie [23] since it only counts the correct selections, not the wrong selections. But the value of the minimum KSPC and the minimum Correct Selections will be identical. Both of the indexes are basically a characteristic of the keyboard design and the entry techniques that the keyboard affords.

Table 1 Relations of metrics with the taxonomy of key activations

Category	Metrics measuring frequency of the corresponding selections
1	Premature exits
2	Correct selections
3	Erroneous key selections
4	Activations during scanning
1 + 4	Attended keys per character

3.1.3 Erroneous selections

This covers Category 3 selections and is a measure of the frequency of incorrect but completed key selections. It is calculated as the number of erroneous selections divided by the number of characters typed.

3.1.4 Activations during scanning

Activations during scanning is a measure of the frequency of activations of non-target keys that are halted before completion. This metric is calculated as the number of activations during scanning divided by the number of characters typed.

3.1.5 Attended keys per character

The sum of Premature Exits and Activations during scanning compose another metric, "Attended keys per character", proposed by Aoki et al. [3], and defined as the number of keys attended divided by the number of characters typed.

3.2 Typing-speed metric

In addition to the above metrics, the traditional "Characters typed per minute" (CPM) was also calculated to evaluate typing speed. CPM is calculated as the number of characters typed divided by the time spent on typing it. It takes two Japanese characters to produce the text equivalent of an English word, so two CPM approximately equals one WPM.

4 Results

4.1 Characters typed per minute

The grand mean for the performed trials was 15.77 CPM ($\sigma = 4.66$), equivalent to 7.89 WPM. CPM were analyzed using a two-way analysis of variance (ANOVA) of the learning effect (Blocks 1–22) and individual variation (Subjects 1–8). The sentences for each block (5) were treated as repetitions. The same ANOVA procedure applies to the other metrics as well. The ANOVA result on CPM is shown in Table 2. Significant differences were found between the blocks and the subjects; no significant interaction effect was found from these two factors. To depict

Table 2 ANOVA result for CPM

Factor	*s.s.*	*df*	*V*	F_0
Block (A)	5,526.0	21	263.1	22.2**
Subject (B)	3,697.6	7	528.2	44.6**
A × B	1,484.5	147	10.1	0.852
Error	8,321.8	702	11.9	
Total	19,029.8	877		

** $p < 0.01$

Table 3 Result of ANOVA on premature exits

Factor	*s.s.*	*df*	*V*	F_0
Block (A)	10.00	21	7.137	14.86**
Subject (B)	5.28	7	10.09	21.01**
A × B	4.99	147	0.410	0.881
Error	25.06	703	0.480	
Total	45.34	878		

** $p < 0.01$

the effect of the block (or learning) and the subject, Fig. 4 plots mean CPM per block for each subject.

As Fig. 4 shows, all subjects seem to rapidly improve their typing speed until around Block 10, and their rate of improvement becomes moderate after around Block 13. Multiple-comparison (Friedman) tests between blocks revealed no significant differences between any two blocks after Block 9. (The range of $\chi_0^2(21) = 0.00–19.58$, $p > 0.50$.)

This indicates that the blocks can be divided into two stages: Blocks 1–8 and Blocks 9–22, in terms of typing speed. Mean CPM for Blocks 1–8 was 13.398 ($\sigma = 4.3046$), and 17.8349 ($\sigma = 4.1524$) for Blocks 9–22. Blocks 1–8 may be considered as a learning phase, in which subjects seem to learn how to master eye typing, whereas for Blocks 9–22, subjects demonstrate relatively stable typing performance.

4.2 Premature exits

The grand mean for premature exits was 0.110 ($\sigma = 0.227$), with the ANOVA result given in Table 3. For this metric, significant effects were noted between blocks and the subjects, but no significant interaction effect. Figure 5 depicts these effects, with mean premature exits per subject per block. As can be seen in this figure, premature exits drops suddenly for all subjects. The premature exits transitions flatten, and their values become very low (approximately 0.1). A multiple-comparison test was also performed between blocks. The results showed no significant differences between any two blocks after Block 2. (The range of $\chi_0^2(21) = 0.00-27.81$, $p > 0.10$.)

4.3 Correct selections

The mean minimum number of key activations per character necessary to produce the sentences used in the experiment was 2.79 ($\sigma = 0.268$). It can therefore be assumed that correct selections in the experiment can only be reduced to around 2.8.

The grand mean Correct Selections was 3.22 ($\sigma = 0.833$). Table 4 shows the ANOVA result for correct selections. We did find significant effects per block and subject, but no significant interaction effects between these two factors. Figure 6 shows the correct selections transitions per block per subject, illustrating the learning effect. It can be observed that the average of correct fixations per character drops suddenly after a few blocks for all subjects,

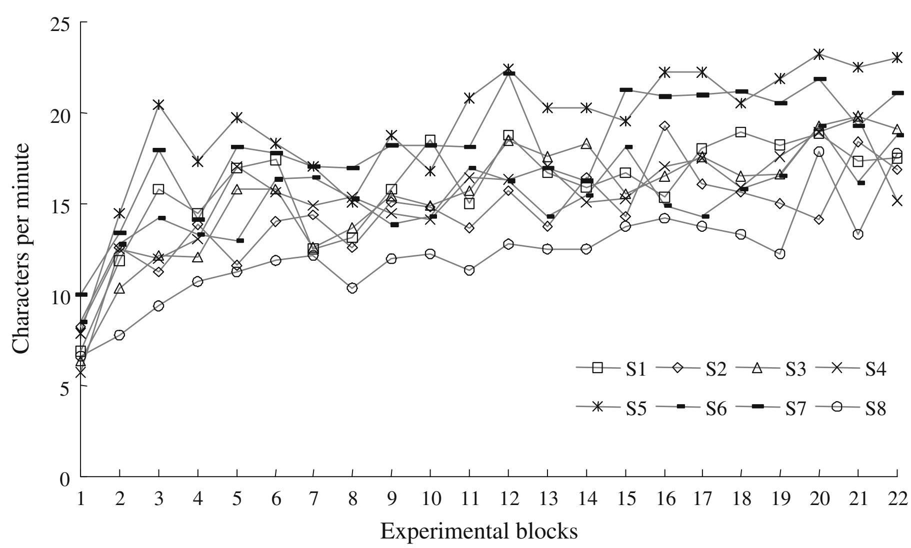

Fig. 4 CPM transitions per block per subject

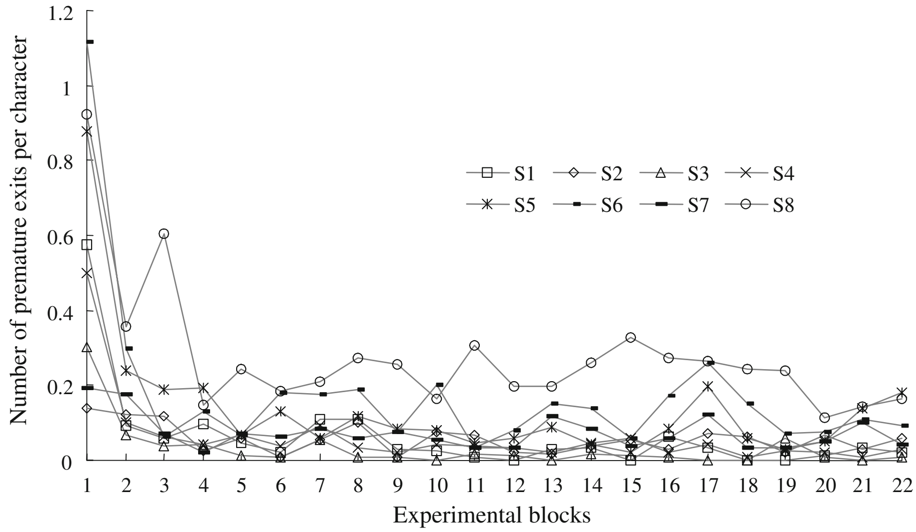

Fig. 5 Transitions of premature exits per block per subject

Table 4 ANOVA result for Correct Selections

Factor	*s.s.*	*df*	*V*	F_0
Block (A)	147.73	21	7.035	14.86**
Subject (B)	69.35	7	9.906	20.93**
A × B	58.81	147	0.400	0.869
Error	332.20	702	0.473	
Total	1,506.08	877		

** $p < 0.01$

and hovers constant at approximately 3.0 for the remaining blocks. A multiple-comparison test between blocks revealed no significant differences between any two after Block 2. (The range of $\chi_0^2(21) = 0.00–24.95$, $p > 0.20$.) The mean correct selections for Blocks 2–22 was 3.140 ($\sigma = 0.609$), while that for Block 1 was 4.94 ($\sigma = 2.110$). Most subjects were apparently able to reduce their Correct Selections to almost the minimal value (i.e., approx. 2.8) with very little training, namely, Block 1.

4.4 Erroneous selections

The grand mean for erroneous selections was 0.281 ($\sigma = 0.531$), with ANOVA results presented in Table 5. The tendency for this metric was identical to other metrics, namely, that there are significant effects per block, but no interaction effect. The effect per block per subject is depicted in Fig. 7. Great improvements can be observed in erroneous selections within just a few blocks, which the figure also illustrates. No significant differences between any two blocks after Block 2 were found with multiple-comparison testing. (The range of $\chi_0^2(21) = 0.00–27.81$, $p > 0.10$.) The mean erroneous selections for Blocks 2–22 was 0.222 ($\sigma = 0.302$), while the mean for Block 1 was 1.521 ($\sigma = 1.656$).

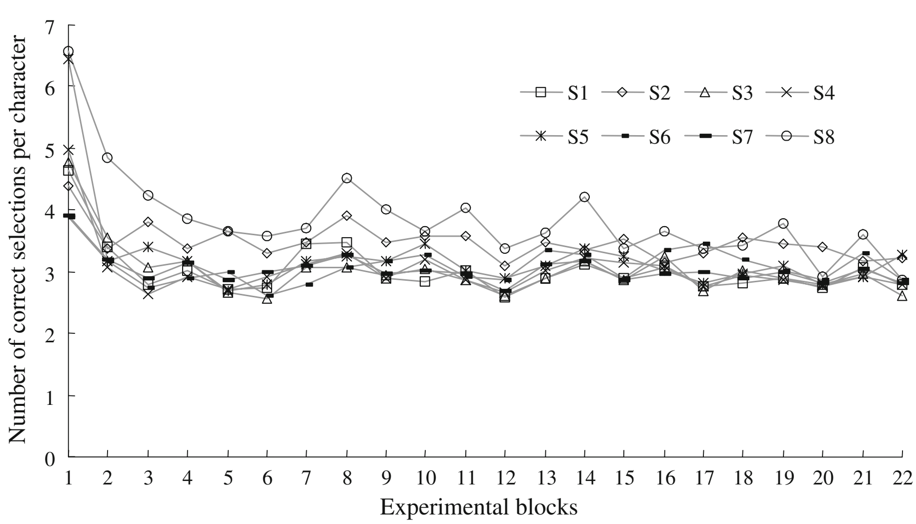

Fig. 6 Transitions of correct selections per block per subject

4.5 Activations during scanning

The grand mean Activations during scanning was 1.13 ($\sigma = 1.46$), and ANOVA results are reported in Table 6. Significant effects of block and subject, but no interaction effect, were found. Figure 8 illustrates the activation during scanning transitions for each subject. From this figure, a drop in activations during scanning is evident in the early blocks, but the progress of improvement seems to be relatively slower compared with Premature Exits, Correct Selections, and Erroneous Selections. Multiple-comparison testing indicated that there were no significant differences between any two of blocks after Block 8. (The range of $\chi_0^2(21) = 0.00$–28.23, $p > 0.10$.) The mean Scanning for Blocks 8–22 was 0.754 ($\sigma = 0.600$), while that for Blocks 1–7 was 1.94 ($\sigma = 2.24$).

4.6 Attended keys per character

The grand mean of attended keys per character was 1.24 ($\sigma = 1.63$), and ANOVA results are presented in Table 7. Significant effects of block and subject, and no interaction effect, were found. Figure 9 illustrates the attended keys per character transitions for each subject. From this figure, a drop in this index in the early blocks is observed. Multiple-comparison testing indicated that there were no significant differences between any two of blocks after Block 8. (The range of $\chi_0^2(21) = 0.00$–26.34, $p > 0.10$.)

4.7 Interrelations among metrics

The interrelations among the metrics are examined by calculating the correlation coefficients. Since no significant interaction effects between block and subject were observed for any the metrics, it can be assumed that the blocks affected each metric equally, regardless of subject. Therefore, the correlation coefficients were calculated using data obtained from all subjects instead of data from each subject individually. Table 8 shows the correlation coefficients calculated between CPM and the gaze typing metrics, as well as among the various gaze typing metrics. The gaze typing metrics correlated negatively with typing speed, but the correlations were not strong.

The gaze typing metrics, on the other hand, correlate positively with each other. Erroneous selections correlates especially strongly with Activations during Scanning and

Table 5 Results of ANOVA on erroneous selections

Factor	*s.s.*	*df*	*V*	F_0
Block (A)	70.15	21	3.34	18.51**
Subject (B)	18.96	7	2.71	15.01**
A × B	31.57	147	0.21	1.19
Error	127.07	704	0.18	
Total	247.76	879		

** $p < 0.01$

Table 6 ANOVA results for activations during scanning

Factor	*s.s.*	*df*	*V*	F_0
Block (A)	807.66	21	38.46	40.36**
Subject (B)	234.82	7	33.55	35.20**
A × B	164.00	147	1.12	1.17
Error	670.86	704	0.953	
Total	1877.35	879		

** $p < 0.01$

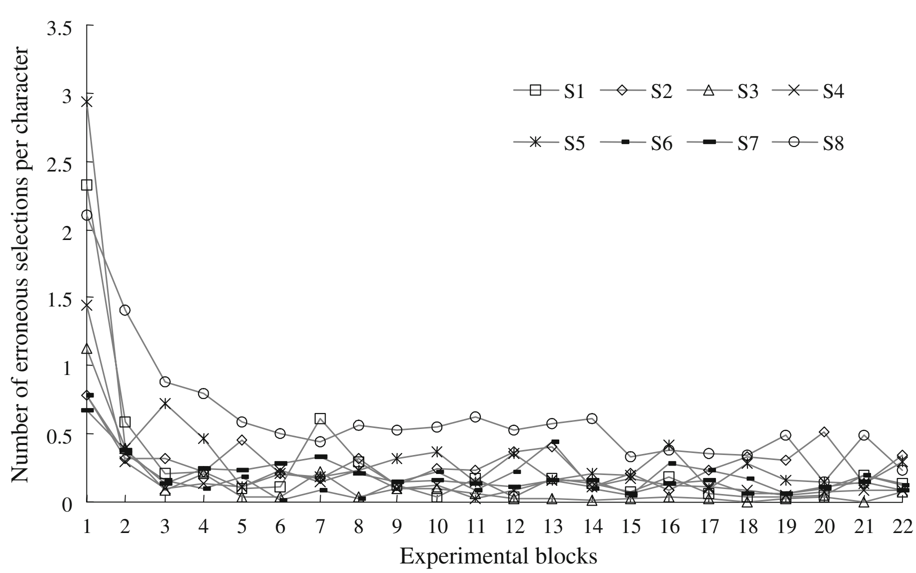

Fig. 7 Transitions of erroneous selections per block per subject

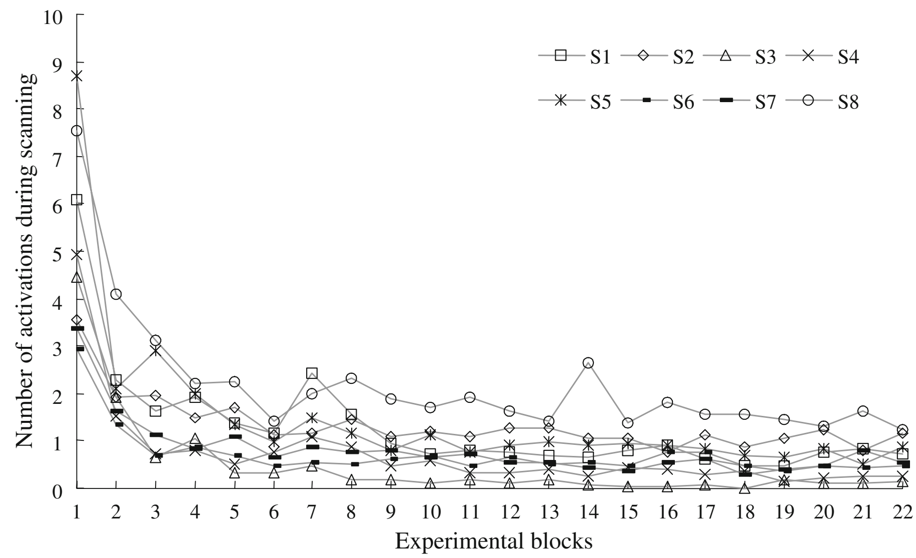

Fig. 8 Transitions for activations during scanning per block per subject

Table 7 ANOVA results for attended keys per character

Factor	*s.s.*	*df*	*V*	F_0
Block (A)	992.64	21	42.27	39.20**
Subject (B)	287.08	7	41.01	34.01**
A × B	192.27	147	1.31	1.08
Error	847.79	703	1.21	
Total	2,319.78	878		

** $p < 0.01$

attended keys per character ($r = 0.910$ and $r = 0.915$, respectively). Obviously, activation during scanning correlates with attended keys per character ($r = 0.995$), since the only difference is that premature exits are not included in activations during scanning.

Figure 10 shows the transition of the ratios of the four types of gaze typing actions in order to depict the general tendency. As can be seen from this figure, the ratios for erroneous and premature exits are almost constant except those pertaining to the early blocks. From this figure, it emerges that the frequencies for the four categories are almost constant, especially after Block 8.

5 Discussion

A general learning pattern for gaze typing (Fig. 9) can be extracted from the reported data. During the very first sentences, subjects may make many wrong selections and they move their eyes away from the correct buttons too soon. After just five sentences (equals to 1 block), they have learned to keep their eyes at the button they want, erroneous selections become fewer, but they continue to look at many buttons during scanning. When they have typed around 40 sentences (equals to 8 blocks), performance becomes stable and small improvements happen incrementally.

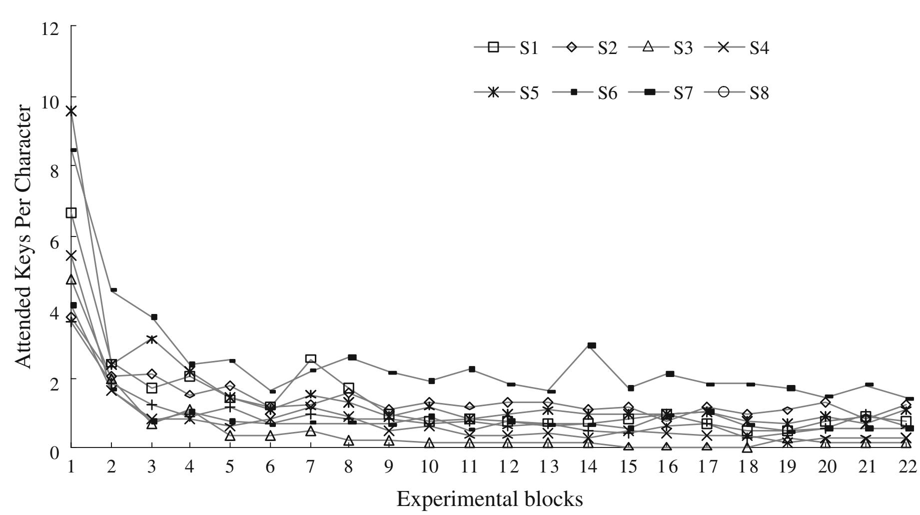

Fig. 9 Transitions for attended keys per character per block per subject

Table 8 Correlation coefficients among metrics

Combination of metrics	Correlation coefficients
CPM-premature exits	−0.426**
CPM-correct selections	−0.645**
CPM-erroneous selections	−0.525**
CPM-act. d. scanning	−0.526**
Premature exits-correct selections	0.636**
Premature exits-erroneous selections	0.698**
Premature exits-act. d. scanning	0.683**
Correct selections-erroneous selections	0.833**
Correct selections-act. d. scanning	0.830**
Erroneous selections-act. d. scanning	0.910**
CPM-attended keys per character	−0.533**
Premature exits-attended keys per character	0.754**
Correct selections-attended keys per character	0.838**
Erroneous selections-attended keys per character	0.915**
Act. d. scanning-attended keys per character	0.995**

** $p < 0.01$

The learning pattern reported in the present paper was observed among young and motivated student subjects. The students had been acquainted with the keyboard layout before they tried gaze interaction for the first time. Spelling the sentences was easy for them. The laboratory had good control over lightning and there were no external disturbances during the experiment. Learning to type by gaze at home or at hospital, with a severe disability like ALS/MND or other physical complications, may be quite different. Progress is most likely to be slower than what has been observed in the reported experiment. Even so, general experience can be extracted from this experiment that could potentially make a real-life learning situation more similar to the "ideal" experimental situation:

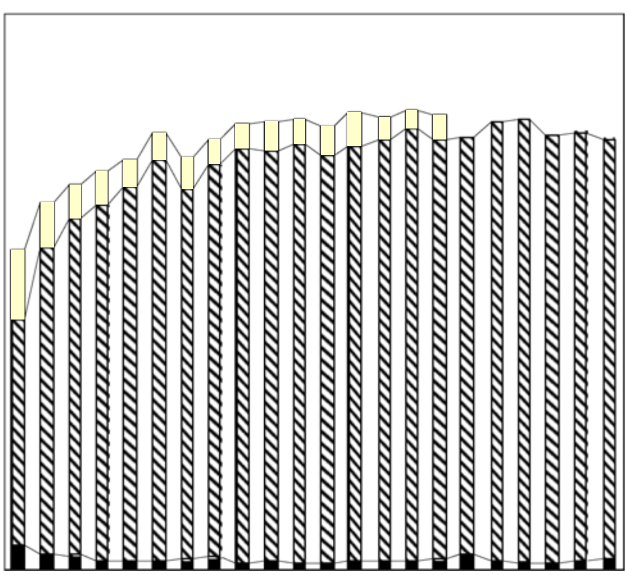

Fig. 10 Transitions of activation ratios for the four activation categories per block

1. Introduce the keyboard interface well in advance, and let the user try it with other input devices, for instance a mouse, trackball, joystick or head-tracker, if this is at all possible. Gazetalk and some other gaze typing systems may also be operated in a step-scan mode by single switches, and this is another possible way to familiarize the user with the layout, before gaze tracking is introduced.
2. Ask the user to type very easy and short sentences. No mental effort should be required to spell or remember the sentences, since learning gaze control will require all of the user's concentration.
3. Premature exits were found to be most frequent in the very first sentences. They may possibly be reduced if novice users are told to keep gazing at the target character once it is found. Most dwell time systems prevent double-clicking on the same button, so keeping the eye at the same place for longer than the dwell period should not do any harm. As indicated in the experimental result, the mean premature exits per character were around 0.110. Even if this could be regarded as a relatively seldom event, it means that approximately 3% of all correct selections gets effected by premature exits. Moreover, premature exits could be a constant source of annoyance since it seems to be persistent, at least to the end of this experiment.
4. After approximately five sentences have been typed, the user could be reminded not to dwell too long on a button while scanning. If the user often makes erroneous selections, dwell time may then be adjusted to a higher level. Some users need dwell times at more than 1 s, while other users can do well with dwell times as low as 400 ms. If the dwell time is increased, the communication specialist should encourage the user to decrease it again after a while, since long dwell times have a negative impact on the productivity of skilled users. The difference in productivity between a dwell time setting of, e.g., 900 ms and 500 ms will be significant by the end of the day. However, if the user can control a switch, this is most likely to be more productive than dwell time selections. Recent research indicates that facial muscles can control EMG-switches particularly fast and may work in conjunction with gaze tracking even for people with late-stage ALS/MND [12, 35].
5. Let the user start typing five sentences and then try to look in the log file for performance measures

(for instance CPM/WPM or error indications like "use of backspace"), attended keys per character, etc.). After a break the user should then type another five sentences. The typical progresses in performance are most likely to be encouraging.

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. In order to include time-based measures in the remote metrics, those breaks need to be filtered out. CPM (and WPM) obviously includes a time factor, but none of the eye movement related metrics suggested in this paper includes time. When collecting user data from larger populations outside experimental rooms, performance metrics are needed 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. In remote real-life conditions, users' behavior during typing cannot be controlled. This makes it impossible to obtain metrics such as "Overproduction Rate" [13] 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 this kind of personal text cannot be compared to any general dictionary standard, the quality of the productions cannot be judged by counting, for instance, the number of spelling errors. Even if this was possible, users differ enormously in their ability to spell. Whatever text the user produces, it has to be accepted as the target text.

Fortunately, gaze typing systems have unique advantages when collecting data "at home" since it is known where the user is looking. The attended keys per character metric can be measured without knowing what the user intended to type and independently from what was actually typed. It correlates very well with the actual error-rate according to the reported experimental results. The metric could be sent off by the end of the day, and would be a single, highly informative indication on the certainty the user had shown in his or her gaze operation of that particular system. Consequently, this metric is now calculated automatically in the log file that comes with the latest freeware version of Gazetalk, to be downloaded from http://www.cogain.org/downloads. It is hoped that the present work may help convincing providers of Augmented and AAC systems to include advanced logging facilities in future systems and designers of gaze typing systems in particular, to be aware of the potentials that gaze related metrics may provide.

The attended keys per character metric are inspired by the "Principles of Motion Economy" (e.g., Barnes [5]). These principles include guidelines for efficient manual work that conserve human energy. It is suggested here that this also applies to eye movements during skilled gaze typing. Although the "energy cost" of an eye movement is very small, the cumulated cognitive cost of all unnecessary fixations during routine operations can be very big. This cost is expected to be a strong predictor of long-term user satisfaction and perceived workload. The reported results confirm that attended keys per character are highly correlated to erroneous selections (Table 8), since every key that is unnecessarily fixated, also risks becoming a wrong selection, c.f. the Midas touch problem.

A metric like CPM (or WPM) can be measured on any text entry system and it is device independent, while the eye movement related 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 Attended keys per character seems to be, then gaze recording could become a standard procedure when designing typing systems for all kinds of input devices. The relation between attended elements per selection and error-rate may even hold true for other interfaces than just on-screen keyboards: The more (selectable) places people look at an interface before they make a selection, the higher the risk of the selection being wrong. It would be interesting to analyze if this was also the case for, e.g., web pages.

A 500 ms dwell time setting was used in this experiment. In general, increasing dwell time can be expected to increase the frequency of gazes shorter than dwell time, and to cause a decrease in gazes longer than dwell time. Therefore, it is expected that Premature Exits and Activations (but not selections) during scanning will become bigger with an increase of dwell time, and that Erroneous Selections become smaller. Correct Selections is not expected to be influenced by dwell time, since it mainly depends on the most efficient key activation stream required to complete a given task. Productivity will of course become lower, since longer dwell times decrease CPM/WPM. This is, however, just expectations and unforeseeable interactions between dwell time and, e.g., training could have a significant impact on performance. Therefore, more experiments are needed in order to analyze the relationship between dwell time and effectiveness of learning in gaze behavior. Ideally, some of these experiments could be carried out from the users home with remote data collection.

6 Conclusion

This study has analyzed the learning processes behind successful gaze interaction and has provided a taxonomy for classifying gaze activations. The involved subjects could reduce their rates of premature exits and erroneous selections within a mere five sentences of practice. The error-rate was found to be highly correlated to number of attended keys per character typed. This paper has argued that this new attended keys per character metric holds potentials for remote evaluation of gaze typing interfaces and monitoring of users performance progress. Finally, some advice has been provided on how to best facilitate the learning of gaze typing.

Acknowledgments This research was partly supported by the COGAIN Network of Excellence, IST IU 6, Contract Number 511598, by a scientist exchange program based on agreements/memoranda of understanding between the Danish Rectors Conference and the Japan Society for the Promotion of Science, and by the Grant-in-Aid for Encouragement of Young Scientist (A) (No. 19681017), The Japan Society for the Promotion of Science. We would like to acknowledge Hiroyuki Sakai, Tokyo Institute of Technology, for his assistance with data collection in the experiment.

References

1. Aoki, H., Itoh, K., Hansen, J.P.: Learning to type Japanese text by gaze interaction in six hours. In: Proceedings of the 1st Conference on Communication by Gaze Interaction, Copenhagen, Denmark, pp. 22–28 (2005)
2. Aoki, H., Itoh, K., Sumitomo, N., Hansen, J.P.: Usability of gaze interaction compared to mouse and head-tracking in typing Japanese texts on a restricted onscreen keyboard for disabled people. In: Proceedings of the 15th Triennial Congress of the International Ergonomics Association 1, Seoul, Korea, pp. 267–270 (2003)
3. Aoki, H., Hansen, J.P., Itoh, K.: Towards remote evaluation of gaze typing systems. In: Proceedings of the 2nd Conference on Communication by Gaze Interaction, Turin, Italy, pp. 94–101 (2006a)
4. Aoki, H., Itoh, K., Hansen, J.P.: Usability evaluation of gaze interface based on scan-path analysis. In: Proceedings of the 16th World Congress on Ergonomics, Maastricht, Netherlands (CD-ROM) (2006b)
5. Barnes, R.M.: Motion and Time Study. Wiley, New York (1949)
6. Bates, R.: Have patience with your eye mouse: eye-gaze interaction with computers can work. In: Proceedings of 1st Cambridge Workshop on Universal Access and Assistive Technology, pp. 33–38. http://www.rehab-www.eng.cam.ac.uk/cwuaat/02/7.pdf (2002)
7. Bernheim Brush, A.J., Ames, M. Davis, J.: A comparison of synchronous remote and local usability studies for an expert interface. Extended Abstracts of CHI 2004, pp. 1179–1182 (2004)
8. Drey, S., Siegel, D.: Remote possibilities? International usability testing at a distance. Interactions **11**(2), 10–17 (2004)
9. Engell-Nielsen, T., Glenstrup, A.J., Hansen, J.P.: Eye gaze interaction: a new media—not just a fast mouse. In: Itoh, K., Kuwano, S., Komatsubara, A. (eds.) Handbook of Human Factors/Ergonomics (in Japanese), pp. 445–455. Asakura Publishing, Tokyo (2003)
10. Hansen, D.W., Skovsgaard, H.H., Hansen, J.P., Møllenbach, E.: Noise tolerant selection by gaze-controlled pan and zoom in 3D. In: Proceedings of the 2008 Symposium on Eye Tracking Research and Applications (Savannah, Georgia, 26–28 March 2008), ETRA ‘08. ACM Press, New York, pp 205–212 (2008)
11. Hansen, J.P., Johansen, A.S., Hansen, D.W., Itoh, K., Mashino, S.: Command without a click: dwell time typing by mouse and gaze selections. In: Rauterberg, M., et al. (eds.) Human-Computer Interaction—INTERACT ‘03, pp. 121–128. IOS Press, Zürich (2003)
12. Hansen, J.P., Junker, A.M.: Gaze pointing and facial EMG clicking. Paper presented at COGAIN 2006: Gazing into the future (2006)
13. Hansen, J.P., Tørning, K., Johansen, A.S., Itoh, K., Aoki, H.: Gaze typing compared with input by head and hand. In: Proceedings of the 2004 Symposium on Eye Tracking Research and Applications, San Antonio, Texas, 22–24 March 2004, ETRA ‘04, pp. 131–138. ACM Press, New York (2004)
14. Hartson, H.R., Castillo, J.C., Kelso, J., Neale, W.C.: Remote evaluation: the network as an extension of the usability laboratory. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground, Vancouver, Canada, pp. 228–235 (1996)
15. Hutchinson, T.E., White, K.P., Martin, W.N., Reichert, K.C., Frey, L.A.: Human-computer interaction using eye-gaze input. IEEE Trans. Syst. Man. Cybern. **19**(6), 1527–1534 (1998)
16. Istance, H.O., Spinner, C., Howarth, P.A.: Providing motor impaired users with access to standard graphical user interface (GUI) software via eye-based interaction. In: Proceedings of 1st European Conference on Disability, pp 109–116. Virtual Reality and Associated Technology, Maidenhead (1996)
17. Itoh, K., Aoki, H., Hansen, J.P.: Usability of gaze-based system for typing Japanese text having hierarchical and static menus: a comparative study with interface by continuous navigation. In: Proceedings of the 2006 Symposium on Eye Tracking Research and Applications, pp. 59–66, San Diego (2006a)
18. Ivory, M.Y., Hearst, M.A.: The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. **33**(4), 470–516 (2001)
19. Jacob, R.K.: The use of eye movements in human-computer interaction techniques: what you look at is what you get. ACM Trans. Info. Syst. **9**(3), 152–169 (1991)
20. Jacob, R.K.: Eye-movement-based human-computer interaction techniques: toward non-command interfaces. In: Hartson, H.R., Hix, D. (eds.) Advances in Human Computer Interaction, vol. 4, pp. 151–190 (1993)
21. Jacob, R.K.: Eye tracking in advanced interface design. In: Barfield, W., Furness, T.A. (eds.) Virtual Environments and Advanced Interface Design, pp. 258–288. Oxford University Press, New York (1995)
22. Jacob, R.J.K., Karn, K.S.: Commentary on Sect. 4. Eye tracking in human-computer interaction and usability research: ready to deliver the promises. In: Hyönä, J., Radach, R., Deubel, H. (eds.) The Minds Eye, pp. 573–605. Elsevier, Amsterdam (2003)
23. MacKenzie, I.S.: KSPC (keystrokes per character) as a characteristic of text entry techniques. In: Proceedings of the Fourth International Symposium on Human Computer Interaction with Mobile Devices, pp. 195–210. Heidelberg, Germany (2002)
24. MacKenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing devices. In: Proceedings of the ACM CHI 2001 Conference on Human Factors in Computing Systems, Seattle, pp. 9–16 (2001)
25. MacKenzie, I.S., Soukoreff, R.W.: Text entry for mobile computing: models and methods, theory and practice. Hum. Comput. Interact. **17**, 147–198 (2002)
26. Majaranta, P., Räihä, K.-J.: Twenty years of eye typing: systems and design issues. In: Proceedings of ETRA, New Orleans, pp. 15–22 (2002)

27. Majaranta, P., Aula, A., Räihä, K.-J.: Effects of feedback on eye typing with a short dwell time. In: Proceedings of ETRA, San Antonio, pp. 139–146 (2004)
28. Majaranta, P., MacKenzie, I.S., Aula, A., Räihä, K.-J. Auditory and visual feedback during eye typing (short paper). In: Extended Abstracts of CHI '03 Conference on Human Factors in Computing Systems, Lauderdale, pp. 766–767 (2003)
29. Majaranta, P., MacKenzie, I.S., Aula, A., Räihä, K.-J.: Effects of feedback and dwell time on eye typing and accuracy. Univ. Access. Inf. Soc. **5**, 199–208 (2006)
30. Ohno, T.: Features of eye gaze interface for selections tasks. In: Proceedings of the 3rd Asia-Pacific Computer Human Interaction, Yokohama, pp 1–6 (1998)
31. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychol. Bull. **124**, 372–422 (1998)
32. Reason, J.: Human Error. Cambridge University Press, New York (1990)
33. Saito, S.: Does fatigue exist in a quantitative measurement of eye movements? Ergonomics **35**(5/6), 607–615 (1992)
34. Stampe, D.M., Reingold, E.M.: Selection by looking: a novel computer interface and its application to psychological research. In: Finddlay, J.M., Walker, R., Kentridge, R.W. (eds.) Eye Movement Research: Mechanism, Processes and Applications, pp. 467–478. Elsevier, Amsterdam (1995)
35. Surakka, V., Illi, M., Isokoski, P.: Gazing and frowning as a new technique for human-computer interaction. ACM Trans. Appl. Percept. **1**, 40–56 (2004)
36. Thompson, K.E., Rozanski, E.P., Haake, A.R.: Here, there, anywhere: remote usability that works, Proceedings of SIGITE, pp. 132–137 (2004)
37. Tullis, T., Fleschman, S., McNulty, M., Cianchette, C., Bergel, M.: An empirical comparison of lab and remote usability testing of web sites, Proceedings of Usability Professional's Association 2002. http://home.comcast.net/~tomtullis/publications/RemoteVsLab.pdf (2002). Accessed 27 July 2007
38. Ward, D.J., MacKay, D.J.C.: Fast hands-free writing by gaze direction. Nature **418**(6900), 838 (2002)
39. Ware C., Mikaelian, H.H.: An evaluation of an eye tracker as a device for computer input. In: Proceedings of CHI/GI'87, pp. 183–188. ACM Press, Cambridge (1987)
40. Waterson, S., Landay, J., Matthews, T.: In the lab and out in the wild: remote usability testing for mobile devices. Extended Abstracts of CHI, pp. 796–797 (2002)
41. West, R., Lehman, K.: Automated summative usability studies: an empirical evaluation. In: Proceedings of the ACM CHI 2006 Conference on Human Factors in Computing Systems CHI'06, Montréal, pp. 631–639 (2006)