Gaze Typing Compared with Input by Head and Hand

John Paulin Hansen*
Kristian Tørning*
Anders Sewerin Johansen*
IT University of Copenhagen

Kenji Itoh†
Hirotaka Aoki†
Tokyo Institute of Technology

Abstract

This paper investigates the usability of gaze-typing systems for disabled people in a broad perspective that takes into account the usage scenarios and the particular users that these systems benefit. Design goals for a gaze-typing system are identified: productivity above 25 words per minute, robust tracking, high availability, and support of multimodal input. A detailed investigation of the efficiency and user satisfaction with a Danish and a Japanese gaze-typing system compares it to head- and mouse (hand) - typing. We found gaze typing to be more erroneous than the other two modalities. Gaze typing was just as fast as head typing, and both were slower than mouse (hand-) typing. Possibilities for design improvements are discussed.

CR Categories: B.4.2 [Input/Output Devices]: Channels and controllers. C.4 [Performance of Systems]: Design studies. H.5.2. [User Interfaces]: Input devices and strategies

Keywords: Eye typing, eye tracking, eye mouse, head mouse, assistive technology, alternative communication, computer input devices.

1. Introduction

Usability is generally defined as: ”the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments” [ISO DIS 9241-11]. The effectiveness of gaze interaction has been demonstrated through twenty years of eye typing. Most gaze typing systems consist of an eye tracking system in combination with one of the standard on-screen keyboards (e.g., “Point for Windows”, “Wivik” and others), but new on-screen keyboards that are designed specifically for eye typing have recently been introduced, e.g. Ward and MacKay [2002]. Typing speed is often just a few words per minute [Majaranta and Räihä, 2002].

*e-mail:{paulin, toerning, dduck}@itu.dk
†e-mail: {ken, aoki}@ie.me.titech.ac.jp

Designers of gaze typing systems should seek to improve the efficiency of their systems, especially because disabled people use gaze typing as their only means of communication. Whereas it may never be possible to reach the speed of a normal conversation (>100 wpm), Pierpont [1997] observed - from many years of experience - that typing in the 25 – 30 wpm range is enough for extensive personal interchange within the domain of radiotelegraphy, “to keep the thought moving.” Also, it is a common experience that people typing at around 40 wpm are able to have enjoyable conversations in chat rooms. Could gaze typing advance into this range? And can it be done without sacrificing efficiency and user satisfaction? These are the challenges addressed in this paper.

2. Previous work

Experimental investigations of gaze-based selections have found them to be faster than mouse selections (e.g., Sibert and Jakob [2000]). Once the target is located, the pointer is already there. However when gaze-based selections are used for more involved tasks such as typing or switch selections, this superiority has not been manifest. The speed of gaze selections has often been measured to be very similar to that of hand (mouse) selections (e.g., Calhoun [1986], Miyoshi and Murata [2001]), but exhibiting a higher error rate (e.g., Ohno [1998], Hansen et al [2003]). The productivity of gaze typing using on-screen keyboards has been relatively low, compared with other input modalities. For example Instance, Spinner and Howarth [1996] reported subjects to produce as little as one word per minute (wpm = 5 characters, including space) when entering their name and address; mainly because they spent much time correcting entry errors (e.g., typing the same character twice). In contrast, Sears [1991] found that people could type 25 words per minute with a touch screen keyboard, 17 wpm using the mouse, and 58 wpm when using the keyboard. Spaepen et al. [1996] found performance to be approximately 7 wpm, and Stampe & Reingold [1995] obtained similar results on their system. Hansen et al. [2003] reported typing speed to be approximately 16 Japanese characters per minute when using only hiragana and katakana, which is equivalent to a typing speed of approximately 6 wpm for English or Danish. Mouse dwell selections were found to be 33 % more efficient than eye dwell selections in this experiment.

Jakob [1991] identified “The Midas Touch” usability problem for gaze based interactivity, namely that selections can happen unintentionally, simply because the user is studying the interface. He also noticed that it could be difficult for some people to stare at will in order to do a dwell-time selection. Naturally, the eyes are moved whenever a piece of information has been noticed and a decision to act has been taken. But if this is done before the end of the dwell time, the selection is cancelled. These two problems may explain why gaze selection falls short on usability in more demanding tasks. The hypothesis would be that the speed of gaze typing is comparable to that of hand (mouse), but the accuracy

should be expected to be lower because of the inherent usability problems of gaze input.

3. Improving efficiency and user satisfaction

There are several ways to accelerate gaze selection and/or to reduce time-consuming error correction:

1) Using word or character predictions to minimize search time for target locations (e.g., Ward and MacKay [2002])
2) Reducing or eliminating the dwell time for each selection (e.g., Salvucci [1999])
3) Using task models to interpret inaccurate input (e.g., Salvucci and Anderson [2000]).
4) Designing keys especially for gaze operation (e.g., Majaranta et al. [2003a and 2003b]).
5) Extensive use of trivial undo functions (e.g., Jacob [1991]).
6) Increasing tolerance of noise on the gaze input by using large and/or well-separated selection areas (e.g., Hansen et al. [2003] and Bates & Istance [2003]).

Ward and MacKay [2002] successfully demonstrated the use of 1) and 2) in a novel data-entry interface named "Dasher," in which selections have become a fully integrated part of a continuous search and navigation process. Users are reported to type by gaze at more than 25 wpm after one hour of practice and experts could type at 34 wpm. Error rates were found to be less than 5% compared with error rates of approximately 20% for an on-screen keyboard. By our experience, the novelty of the Dasher interfaces necessitates a certain amount of training, and the high spatial compression of the character display requires an accurate and well-calibrated eye tracking system.

Salvucci [1999] demonstrated the potential of fixation tracing by use of hidden Markov models to predict selections on an on-screen keyboard. Inferring the most likely intended word from a gaze path on a sequence of key candidates compensated for the lack of precision in eye-tracking systems and eliminated the need for a dwell time delay. He found that eye typing averaged 822 ms per character, which equals almost 15 wpm. However, systems like this will be indistinct when confronted with misspellings and unknown words, which include family names or local places. Therefore, they are of limited practical value in tasks that require a free communication, but they may be of great value for constrained input.

Majaranta et al [2003a and 2003b] showed that basic design issues are important for the efficiency of gaze typing. On a standard qwerty on-screen keyboard they increased the average typing speed by 0.5 wpm (up from approximately 7 wpm to 7.5 wpm), simply by adding a click sound to each character selection, compared with selections with no audio feedback or selections with a spoken character feedback [2003a]. They also investigated the use of motion, specifically animations of the character shrinking, in on-screen keys as a feedback on remaining dwell time [2003b]. This had a significant impact on typing speed (mean = 7.02 wpm) compared with keys without the shrinking motion (mean = 6.65 wpm). The shrinking effect also had a significant bearing on how many times the user (unnecessarily) refocused on the same key. Shrinking helped keep focus on the centre of a key and therefore the user on average only focused approximately 1.2 times per key stroke compared with 1.3 times when not using the animated buttons [2003b].

Hansen et al. [2001 and 2003] are currently developing a gaze-based communication tool, "GazeTalk," designed for people with amyotrophic lateral sclerosis (ALS) who have lost their voice and mobility and may only be able to move their eyes. In order to make the tool widely available, they intend to use standard consumer camera technology to determine gaze positions (Hansen et al. [2002]). The relative low resolution of this camera technology requires large on-screen buttons, and therefore only 12 keys can be reliably selected on a standard 15 inch monitor. The design of the system has been tested in a set of usability experiments described later in this paper.

Efficiency is not the only objective to consider when designing a user-friendly gaze communication system. Hansen et al. [2001] referred to additional user requirements for a system to be satisfying. The system should be easy to install, maintain and update. It should consist of standard consumer hardware components that can be replaced immediately when something breaks down. Calibrations should be performed easily and quickly. Tracking should be sufficiently robust to allow for mobile use with occasional changes in light conditions, use of glasses, and minor changes in head position. Prolonged use should not cause fatigue or cause the eyes to dry out. The price of the system should not be prohibitively high. Finally, the system should not make the disabled person look awkward. For instance, members of an ALS user group have told us that they would prefer not to wear any kind of peculiar equipment on their head, and that the tracking hardware should either look familiar or be invisible.

4. Improving effectiveness

Some people have to give up their preferred communication tool because of repetitive strain injury or because they eventually lose the motor control necessary for a certain input mode. Johansen and Hansen [2002] listed how the progression of ALS may influence the effectiveness of augmented and alternative communication (AAC) systems:

	Symptoms	Input devices
1st stage	Fatigue is noticeable. Reduced mobility and strength in arms and hands. Often slurred speech.	Keyboard with hand/arm rests and modified operation (sticky shift, no repeat).
2nd stage	Fatigue is a factor. Unable to move arms due to lack of strength, but mobility is usually retained in one or both hands. Severely slurred speech, largely unintelligible to outsiders.	Mouse, joystick, reduced keyboard (5-10 keys).
3rd stage	Almost full lock-in. No speech function. Severely reduced mobility of all extremities.	One or two switches, eye or head tracking.
4th stage	Full lock-in.	Eye tracking.

Table 1: Typical progression of ALS (from Johansen and Hansen [2002])

For several reasons, AAC systems for ALS users must be designed with multimodal input in mind:

- The user should be able to use the same system through all stages of the disease.
- Many ALS patients have little or no previous experience with computers and are quite busy adapting to the severity of their situation; keyboards may be the only input form that they are familiar with.
- Several caretakers must be able to help the user complete letters, edit text and use other functions in the program without having to learn an unfamiliar interaction method.
- Limited time resources among the specialist responsible for introducing and configuring the system means that the duration and quality of user training often are severely limited.
- The progression through the stages of ALS is gradual, and the fatigue factor often makes it necessary for the user to switch to a less efficient input method during the day.

To summarize, major usability requirements for an AAC gaze-typing system are: productivity above 25 wpm, robust tracking, high availability, and support of multimodal input.

We have designed a typing interface "GazeTalk" with the ambition to reach these design goals. In the next section we report on a usability experiment conducted to test the likely efficiency with which users react to this interface, particularly in terms of efficiency of text production. We use three different modalities: hand (mouse), head and gaze. The gaze-tracking components of the system are currently being modified to improve their robustness (Hansen et. al. [2002]). Therefore we have tested the interface separately with standard head- and eye-tracking equipment. Our focus on initial performance stems from a concern that first impressions of system may have a major influence on the users' determination to master it. User motivation is of highest importance to AAC systems as they are often introduced in periods of a life crisis, e.g., during recovery from an accident or during a serious progressing disease such as ALS.

5. Experiment with a dynamic Danish keyboard

Twelve non-disabled students with normal or corrected-to-normal vision (5 female and 7 male) were paid approximately 120 US dollars to participate in the experiment. Mean subject age was 32 years. The typing system was run on a PC (1 GHz) with a 17-inch colour monitor (1024x768 pixels). Dwell time for a key input was set at 500 ms for all the input devices. A "QuickGlance" system (from "EyeTech Digital System") was used for eye tracking with an update rate of 15 frames per second and a smoothing factor of 7 samples. A "Smart-Nav™" hands free mouse (from "Natural Point") was used for head tracking with the speed factor set to maximum, and the smoothness factor set to "motion fast". This head tracker uses four infrared LEDs located behind an opaque front lens to illuminate reflective material (3M safety material) such as a dot placed in the forehead of the user. A camera inside the unit picks up this reflection and transmits the image data to a chip where it is processed (Fig.1).

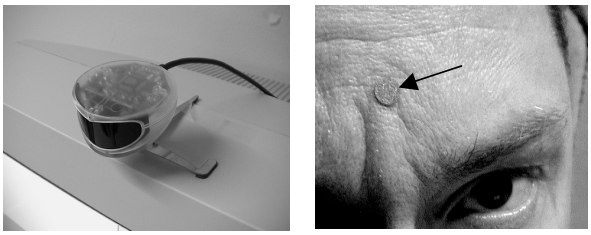

Fig. 1: *The "Smart-Nav™" head tracker used for the Danish and Japanese experiments. The unit - placed on top of a monitor - tracks IR reflections from a dot in the users forehead.*

Each subject entered 12 Danish sentences from the tales of Hans Christian Andersen for each of the three input modes (hand, head and mouse) in two blocks, one each day. For each day the order of the three input modes was randomised across subjects, and there was a break of at least ten minutes between each of the input conditions. All 12 sentences were shown on a clipboard, mounted by the top left corner of the monitor. Subjects were instructed to type as quickly and accurately as possible. The total number of sentences typed in Danish was 864 (12 subjects x 12 sentences x 3 input modes x 2 blocks). The average length of the 72 different sentences used for the experiment was 7.1 words (SD = 1.7 words).

The sentences were typed in on the system shown in figure 2. The size of each button was approximately 8 cm by 8 cm, and the text field (top left corner) was 16 cm by 8 cm. Text entered into this field was displayed in a 12-point boldface Helvetica font. The system was configured for dwell time activation (with the progress bar shown). Buttons were highlighted when pointed to with the cursor.

The primary letter-entry mode featured a dynamic keyboard with six buttons arranged in a three by two matrix, which allowed the user to type the currently most likely six letters directly. The built-in context-sensitive letter prediction algorithm was used to supply the six most likely letters for the dynamic keyboard. Furthermore, this mode featured buttons for backspace and space as well as buttons for access to word prediction/completion mode and an alphabetical letter-entry mode ("A to Z"). The word prediction/completion mode presented the current eight most likely words (the actual words are shown on the "eight most likely words" button) in a four by two matrix, and featured buttons for access to alphabetical letter-entry mode and the primary letter-entry mode. Once a full word had been chosen from the suggestions, the system remained in a full-word suggestion –mode. The alphabetical letter-entry mode enabled the user to select the desired letter in a two-stage process first by selecting a group of letters (e.g., "ABCDEFG") containing the desired letter, and then by selecting the letter itself. Learning features of the word prediction system were not enabled. The prediction algorithm had been trained on the collected work of Hans Christian Andersen, minus the 72 sentences used in the experiment. The keystroke per character (KSPC) was 0.9 for error-free performance, when the model had been trained on this corpus.

This is the text f_		A to Z	Backspace
[8 most likely words]	A	I	O
Space	R	L	U

Fig. 2*: Layout of the Danish on-screen keyboard. The subject is typing, "This is the text field." Letter and word predictions are refined continuously as the user types. The progress bar behind the character indicates the remaining time before the "I" button is activated by the dwell time selection system.*

After completion of all the sessions, each subject evaluated the usability attributes using a five-point scale. Data analysis and results of this experiment will be reported in section 7 along with data analysis of a similar experiment with a hierarchical interface for typing in Japanese.

6. Experiment with a hierarchical Japanese interface

Fifteen non-disabled Japanese students with normal or corrected-to-normal vision (3 female and 12 male) participated in this experiment. Mean age was 21 years. The participants were paid approximately 25 US dollars for participating. As in the Danish experiment, each subject performed two replications of all three input devices over two days. In each experimental session, subjects typed 12 Japanese sentences composed in Hiragana, Katakana, Kanji, and some alphanumeric letters. Subjects were instructed to type a sentence as quickly and accurately as possible. The sentences were Japanese translations of the same sentences from H. C. Andersen's tales that were used for the Danish experiment, and they were also shown to the subjects on a clipboard placed beside the monitor. In total the Japanese subjects produced 1080 sentences (15 subjects x 12 sentences x 3 input modes x 2 blocks). The average length of the 72 sentences was 18.7 characters (SD = 3.99 characters).

The equipment and basic layout of the Japanese interface were identical to the Danish. Unlike the Danish version, which had character prediction, the Japanese system had static typing menus. The menu structure was hierarchical according to the order of the Japanese character system. At the top menu (cf. Figure 3) one syllabary group was allocated to each key. When a user activates one of the keys in the top menu, eight keys corresponding to this syllabary group appear at the subsequent menu. In this way, a syllabary (e.g., "Hiragana" or "Katakana") can be typed by two or three key activations. A Japanese text comprises Hiragana, Katakana – a syllabary mostly used for words imported from other languages – and Kanji (Chinese characters), as well as alphanumeric letters and symbols. When a string of Hiragana characters has been typed via the menus, it can be converted into a text with the right combination of Kanji and Hiragana characters. The Japanese version of GazeTalk used for the experiment had a so-called Kana-Kanji conversion function and extra keys for executing this function was included in the menu, but besides the suggestions made by this conversion function, there were no predictive or adaptive features in the Japanese system. The average keystroke per character (KSPC) for the interface was 3.0 (error-free performance).

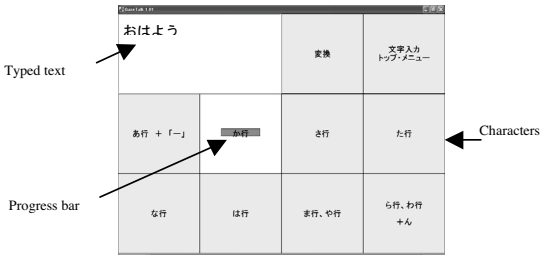

Figure 3*: The static Japanese interface (top level) with a progress bar in the activated button and 8 keys to type characters.*

7. Results

Data analysis was performed as a 2-factor ANOVA for each interface, with input mode (hand, head or gaze) and block (day 1 or day 2) as the independent variables (subjects were treated as repetitions). For the Danish version, word per minute (wpm) was analysed as the dependent variable, and for the Japanese version characters (Hiragana, Katakana and Kanji) per minute (cpm) was analysed. All sentences, including erroneous and corrected sentences, were included.

The grand mean of wpm was 6.22 and the grand mean of cpm was 11.71. The Danish result is within the same speed range as similar eye-typing systems, e.g., Stampe and Reingold [1995] and Majaranta and Räihä [2002].

There was a significant learning effect in both experiments, $F(1,66) = 4.22$, $p < 0.05$ and $F(1,84) = 84.94$, $p < 0.0001$. Danish subjects improved their wpm, from 5.82 on day one to 6.61 on day two, and Japanese subjects also improved, from 10.16 cpm to 13.24 cpm on day two. There was a significant main effect from input mode in both experiments, $F(2,66) = 13.05$, $p < 0.01$ and $F(2,84) = 87.96$, $p < 0.0001$. Mouse/hand was the fastest input on day two for both interfaces, 7.45 wpm and 16.08 cpm respectively, whereas the head input yielded 6.10 wpm and 12.29 cpm respectively on day two. Gaze input was found to be 6.26 wpm and 11.37 cpm respectively on the second day. The difference between head and gaze input was not significant in any of the experiments. Figure 4a and 4b summarise the results of both experiments.

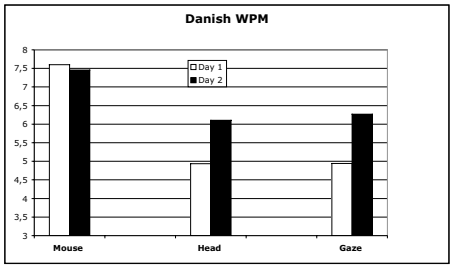

***Figure 4a**: Typing speed for mouse, head and gaze input on a dynamic Danish on-screen keyboard with dwell time (=500 ms) selections, N = 12.*

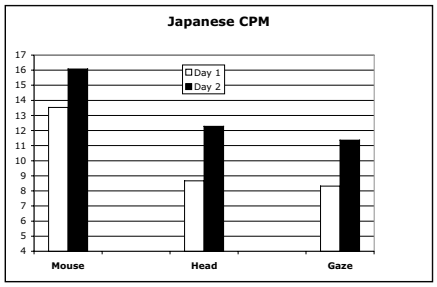

***Figure 4b**: Typing speed for mouse, head and gaze input on a static Japanese on-screen keyboard with dwell time (=500 ms) selections, N = 15.*

Accuracy was measured by the number of sentences that either were erroneous or had been corrected by the subjects in percentages of all sentences typed. Errors were very rare in the Japanese experiment for all three input devices. Only 3 % of the sentences typed by hand (mouse) were erroneous or had been corrected, whereas 5 % of the head-typed sentences and 6 % of the gaze-typed sentences had errors or corrections. The Danish subjects were much less accurate. They committed errors in 14 % of the sentences for hand (mouse), 14 % for head and 28 % for gaze.

A more detailed analysis of the high number of Danish errors was conducted. Errors that had been corrected by the subject were measured in terms of keystrokes used for the corrections in percentages of all keystrokes used to type the sentence. A main learning effect and a main effect of input modality were found; $F(1,66) = 7.77$, $p<0.05$ and $F(2,66) = 16.20$, $p<0.0001$. Corrective keystrokes decreased from 3.5 % on day one to 2.4 % on day two. The corrective keystrokes for hand (mouse) were only 0.65 %, which was significantly different from the corrective keystrokes for head (3.47 %, $p<0.005$) and for gaze (4.29 %, $p<0.0001$). Head and gaze were not significantly different. Remaining errors that had not been corrected by the subject were analysed by the "minimum string distance" (MSD) method, suggested by Soukoreff and MacKenzie [2001]. This method calculates accurately how many basic actions (insertions, deletions and substitutions) it would have required to correct the remaining errors for each sentence. Again, there were significant main effects for input modalities $F(2,66) =6.14$, $p<0.01$. MSD for gaze (MSD = 1.09) was significantly different from hand (mouse) interaction (MSD = 0.46) and from head (MSD = 0.49). In summary, gaze typing did cause the most errors, and because the subjects did not correct them with greater effort, the final text was more erroneous than text produced by head or mouse input.

The Japanese error analysis was performed in a different way because the MSD-method does not (yet) apply to the Japanese character system. Instead, the overproduction rate (OR) was calculated. This index refers to additional – over-produced – inputs due to corrections and to less-than-optimal selection strategies. The OR was calculated as the total number of activated keys minus the number of keys required for producing a sentence, divided by the number of required key activations. Fig. 5 illustrates the change in the mean ORs from day one to day two and the improvement ratios. ORs for gaze and head decrease with the two experimental days. In contrast, this decrease could not be found in OR for hand (mouse) interaction, but the mean OR for hand (mouse) was much lower than that for other devices, and the ORs for gaze were much higher compared to the other input modalities. Moreover, the mean improvement ratio for gaze interaction is lower than for head interaction. In summary, the Japanese results confirm the Danish findings that gaze interaction is more error prone than head and - especially – hand (mouse). This inherent disadvantage is most likely to be explained by "The Midas Touch Problem" and that staring is unnatural (Jacob [1991]).

We made a rough estimate of the performance to be expected from a well-trained subject. This was done by using the "*power law of practice*": $T_n=T_1n^{-a}$, where n=number of sentence, T_n=times required to type the nth sentence, and a=learning coefficient. We obtained learning curves for all the combinations of subjects and devices. Based on the curves, we estimated the variables (T_1 and a) derived from the regression results. In addition, we calculated wpm and cpm for the 1000th sentence (n=1000) for each subject based on the obtained variables. Estimated means and ranges of wpm and cpm are shown in Table 2. The results indicate that no differences should be expected between gaze and hand (mouse) after 1000 repetitions. However, the wide range of estimated wpm and cpm are caused by poor regressions. Therefore, longitudinal studies are required to determine expert performance levels more accurately.

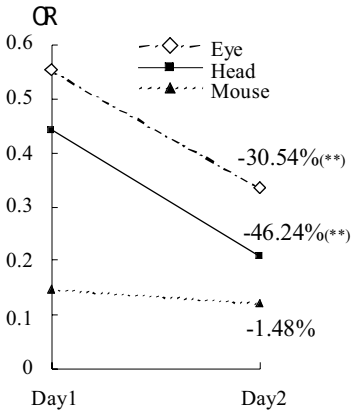

***Figure** 5: The change in overproduction rate (OR) from day 1 to day 2 and the improvement rations, Japanese interface, N=15, (** significant at p<0.01)*

Estimated indices	Eye		Head		Mouse	
	Mean	Range	Mean	Range	Mean	Range
wpm	9.36	2.04-26.4	12.1	5.52-35.0	10.1	2.15-14.9
cpm	29.9	13.1-50.9	33.1	22.1-64.6	23.5	14.7-38.4

***Table 2**: Estimated means and ranges of wpm and cpm after 1000 sentences*

After the last session on the second day the subjects were asked to rate their subjective impression of the efficiency and satisfaction with the system on a six-point scale, in which 1-2 was scored as "negative", 3-4 as "neutral" whereas 5-6 was scored as "positive." There were no significant differences in rating between the 12 Danish and the 15 Japanese subjects, so they are all included in figure 6. There was a positive correlation (Pearsons) between the subjects' rating of efficiency and their mean wpm and cpm, $r= 0.43$ $p < 0.01$ and $r = 0.42$, $p < 0.01$, respectively. There was also a negative correlation between the Japanese subjects' rating of satisfaction and their overproduction rate, $r = - 0.33$, $p < 0.05$. As can be seen from this figure, the subjects seemed most pleased with the hand (mouse) input. The satisfaction with gaze input was rather mixed, while the efficiency ratings of head and gaze were almost identical. Bates & Instance [2003] also found users to be more satisfied with head mouse interaction compared to eye mouse interaction.

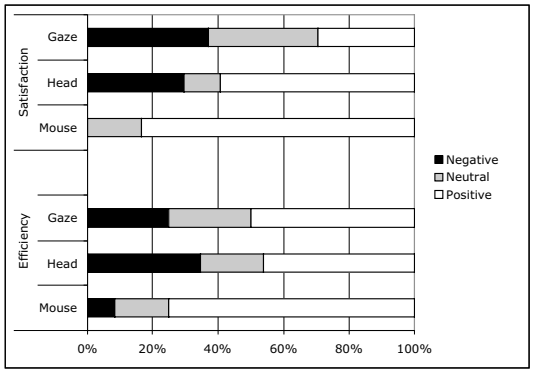

***Figure 6**: Subjective ratings of efficiency and satisfaction of 3 input devices, N = 27.*

8. Discussion

Dwell-time based gaze interaction seems to be the least efficient and satisfying input modality of the three tested. The advantage of this modality is though that it may be the only possible access tool for a highly disabled person or it may supplement head tracking or mouse input, if fatigue or RSI forces the user to switch modality. The main reason for gaze input to fall behind is the higher number of errors it provokes, especially for the novice. Most of the errors committed by the subjects in our experiment were unintended activation of a button, while they were searching for the next button to hit – a clear case of the "Midas Touch" problem. They often reported that they found the dwell time of 500 ms too fast, especially for the button on which the eight most likely words were shown in the Danish interface. A typical comment was: "It's a bit difficult to get used to not dwelling at unintended buttons". Another user explained that is was "difficult to orient oneself without activating something". Obviously, the dwell time setting could have been increased, but that would have had an impact on the typing speed. In the Danish interface, for instance, it took 45 keystrokes per minute to type at 10 words per minute (KSPC = 0.9, error free) but if another 250 ms were added to the dwell time setting, the effective words per minute would go down to 8.4 wpm. At the Japanese interface (KSPC = 3.0, error free) the similar consequences would be a drop in productivity from 15 cpm to 12.6 cpm.

Another problem with the gaze-activated interface, recognized by Jacob [1991], is that staring is unnatural. One of our subjects addressed this thus: "It can be hard to keep staring at the buttons you want". In order to get an impression of just how unnatural it is for the user to keep the eyes on the button that is to be selected, we recorded the spontaneous eye movements of one subject (not included in the experiment) while he was dwell typing by mouse on the Japanese interface. The recording was done by use of an ASL4000 head-mounted eye tracking system. Of 35 selections analyzed, in 40% of the cases the gaze remained at the dwelling key; in 40% of the cases it moved to another key during the dwell period; in 12% of the cases it moved to the text field during the dwell period and in 8% of the cases the eye was focused on another key during the whole dwell period. This indicates that in 60% of all selections the user has to keep the eyes immobile against his normal (mouse-) control patterns. This can of course be learned, but will be a source of errors and frustrations for the novice user.

Dwell times could be set individually for each button, so for instance the button displaying the eight word suggestions might be given a dwell time of 1000 ms, while single character buttons in a static menu structure may be configured for a dwell time of 500 ms and characters in a dynamic menu structure could be set at 600 ms. Also, dwell times for each button type could be adaptive. Lesher et. al. [2000] developed a method for automatic, real-time adjustments of delays in a so-called row/column scanning interface. Using typing errors as feedback to the system, the delay dropped off quickly from a starting point at 2000 ms stabilizing at 270 milliseconds after approximately 1200 selections. If this adaptive method were applied to the GazeTalk interface, and the 270-ms optimum could be reached for all gaze dwells, it might eventually improve the efficiency of a Danish user from 10 wpm (dwell time set at 500 ms) to 12.1 wpm and a Japanese user from 15 cpm to 18.2 cpm. Most importantly, it would also remove the discomfort associated with the unnatural staring, as each button would be activated with "just enough" fixation time for that particular type of button.

Trivial undo functions are especially important for dwell-typing systems, as it is not possible with these to delete full words or longer pieces of text simply by clicking backspace rapidly and repeatedly. An activation of the backspace function takes at least one dwell time to perform. This is particularly annoying to the user when the user inadvertently selects an unintended full word in a single activation and must then correct the error by deleting the word one character at a time. We have conducted simple GOMS-estimations, which indicate that it may take a user 27 seconds to delete a 9-letter word, but only 9 seconds to activate an undo-word selection function (by four strokes). Therefore we have decided to give undo functions a prominent position in the next version of the system.

Unfortunately, even though these design improvements may increase user satisfaction, none of them will improve the efficiency to a level anywhere nears the +25 wpm goal range (or +40 cpm on the Japanese version). The solutions suggested by Ward and MacKay [2002] eliminate the dwell time and reduce the search operation by extensive use of character predictions in Dasher. Salvucci [1999] also eliminate the dwell time and let people type on the familiar QWERTY-format, on which the search time could be expected to be just the eye movement time. However as previously mentioned, the Dasher design trade-offs on immediate ("walk up and use"-) usability and the design by Salvucci trade-offs on its ability to handle unknown words. None of these designs would work with the low resolution of our gaze tracker. Thus, the most obvious way to increase typing speed would be to increase the keystroke per character-factor. The highest efficiency possible in the Danish version is 0.2 KSPC, if only full-word suggestions were used all the time. The system could then peak at its theoretical maximum performance of around 40 wpm, assuming a search time of 1200 ms and a dwell time of 300 ms. The 1200 ms search time is estimated on the basis of results obtained from the present experiment, in which the fastest subjects produced around 35 keystrokes per minute.

Full-word typing may only occur when the user stays within a vocabulary of phrases and expressions, which have either been typed previously, or were included in the training material used for building the language model. It is most likely to happen when common requests are typed, for instance: "I would like some [tea, coffee, water, etc.]". When engaged in a composition task, it's a reasonable assumption that the user will often accept a synonym for the desired word, rather than spending time explicitly requesting the desired word. This feature may lead to so-called "parrot speech", i.e., the text composed with the system lacks the individual style that text produced by other means would exhibit. However, as the primary concern is to aid a disabled user, who struggles to keep up with normal spoken conversation, we consider this an acceptable design trade-off. A Danish ALS-patient, Arne Lykke Larsen [2003], had a take on this possibility. In an ALS-newsletter posting entitled "No, but I have read the book" he argues, without being serious, for the development of a communication system with only 25 to 30 sentences built in: "The major part of all conversations among common people is very predictable and by using just 25-30 sentences <...> you should be able to keep the conversation going for a couple of hours without anybody noticing it. If, for instance, one of your friends asked you if you had seen that movie, you would just reply: "No, but I have read the book" by using a few keystrokes." - and then he adds: "Computer nerds may actually think of a method by which it could be done by using just one keystroke". This is not yet the case with our system. It would take him at least 7 keystrokes, and communication speed would be 32 wpm, supposing that he were as fast as the best subjects in our experiment.

When Japanese people communicate via e-mail or text, the Kanji characters are essential, but in daily Japanese communication it is not always necessary to use Kanji characters. Full sentences can be expressed – in the most direct manner – by using only nouns. Although this is generally considered impolite, it will increase the efficiency. For instance, the sentence "Please give me a cup of coffee" would require at least 15 characters, including Kanji characters, equal to 45 keystrokes. But if it is said in the direct way – "Coffee" – it only consists of 4 characters, which would take only 12 keystrokes. It is socially acceptable in Japan for disabled people to communicate in this direct manner. Conversations among friends typically would be a mixture of short, direct commands and more polite, advanced expressions. We estimate the total daily conversation to involve approximately 50 % fewer characters than the sophisticated sentences of H. C. Andersen.

Our estimates of what wpm and cpm rates can be achieved in daily use are just speculations and extrapolations. Therefore we have decided to include a regular typing test in GazeTalk as part of the communication tool. The results of these regular tests will be sent to us automatically by the built-in e-mail system. This way we hope to be able to track the development of gaze-typing skills under real conditions. With easy access to the actual users in their particular environments, we believe to have a sound basis for incremental improvement of the system that eventually will make it possible to meet our design goals.

9. Conclusion

Gaze-typing systems should be fast, efficient, and reliable. Novice users tested on the GazeTalk system do not achieve the goal range of +25 wpm or +45 cpm. Gaze interaction was found to be slower than mouse/hand interaction and more erroneous than head interaction. Usability evaluations of gaze-typing systems should not focus only on efficiency, however. High availability of system components and multimodal input options are also important factors contributing to effectiveness use of the systems. Progress in performance is essential for motivating the user, and in this respect the present experiment has yielded promising learning effects for both head and gaze interaction. Some design improvements are still possible; therefore a measurable and noticeable improvement in typing speed can be expected in future versions, at least for standard phrases.

Acknowledgements

This research was partly supported by Grant-in-Aid for Scientific Research (A), No. 14208040, the Japan Society for the Promotion of Science, by the Danish Ministry of Science, Technology and Innovation and by the Nordic Academy for Advanced Study (NorFA).

Literature

Bates, R. & Istance, H.O. (2003): Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing device. *Univ Access Inf Soc*, 2: 280 – 290.

Calhoun, G. L. (1986): Use of eye control to select switches. *Proceedings of the Human Factors Society*. 30th Annual Meeting, 154 – 158.

Hansen, D. W., Hansen, J. P., Nielsen, M., Johansen, A. S. & Stegmann, M. B. (2002), Eye Typing using Markov and Active Appearance Models, *IEEE Workshop on Applications on Computer Vision*, 132-136.

Hansen, J. P., Hansen, D. W., Johansen, A. S. (2001) Bringing Gaze-based Interaction Back to Basics. *Proceedings of*

Universal Access in Human-Computer Interaction (UAHCI 2001), New Orleans, Louisiana

Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K. & Mashino, S. (2003). Command Without a Click: Dwell Time Typing by Mouse and Gaze Selections. *Human-Computer Interaction – INTERACT'03*. M. Rauterberg et al. (Eds.) IOS Press, 121 – 128.

Instance, H. O., Spinner, C. & Howarth, P. A. (1996). Providing motor impaired users with access to standard Graphical User Interface (GUI) software via eye-based interaction. *Procceedings of 1st European Conference on Disability, Virtual Reality ans Associated Technology*, ECDVRAT, UK.

ISO 9241-11: Guidance on Usability (1998)

Jacob, R. K. (1991). The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. ACM Transactions on Information Systems, Vol. 9, No. 3, April 1991, 152 – 169.

Johansen, A. S. & Hansen, J. P. (2002), Augmentative and alternative communication: The future of text on the move. *Proceeding at 7th ERCIM Workshop "User Interfaces for all"*, Paris (Chantilly), 367-386.

Larsen, A. L. (2003). No, But I Have Read The Book. Posting at the Danish ALS Newsletter, August 2003, Available: http://www33.brinkster.com/alsforum/artikel.asp?id=81

Lesher, G.W., Higginbotham, D.J., & Moulton, B.J. (2000), Techniques for automatically updating scanning delays. *Proceedings of the RESNA 2000 Annual Conference*, 85-87.

Majaranta, P. & Räihä, K. J. (2002), Twenty Years of Eye Typing: Systems and Design Issues, *Proceedings of the Symposium on ETRA 2002: Eye Tracking Research & Applications Symposium 2002,* New Orleans, 15– 22.

Majaranta, P. I. MacKenzie, S., Aula, A. and Räihä, K.(2003a). Auditory and Visual Feedback During Eye Typing *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems* CHI 2003. New York: ACM 2003.

Majaranta, P. I. MacKenzie, S. and Räihä, K.(2003b). Using motion to guide the focus of gaze during eye typing. *Abstract Proceedings of ECEM12* 20-24 August 2003, Dundee, Scotland

Miyoshi, T. & Murata, A. (2001), Input Device Using Eye Tracker in Human-Computer Interaction, *IEEE International Workshop on Robot and Human Interactive Communication*, 580-585.

Pierpont, William G. (1997): The Art and Skill of Radio-Telegraphy. NOHFF 1997. 3rd edition. Available: http://www.qsl.net/n9bor/n0hff.htm

Ohno, T. (1998). Features of Eye Gaze Interface for Selections Tasks. *Proceedings of The Third Asia Paicific Computer Human Interaction – APCHI'98*. IEEE Computer Society. 1 – 6

Salvucci, D. D. (1999), Inferring intent in eye-movement interfaces: Tracing user actions with process models, *Human Factors in Computing Systems: CH1 99 Conference Proceedings,* Pittsburgh, PA, ACM Press, 254-261.

Salvucci, D. D. (2000), Intelligent Gaze-Added Interfaces, *CH1 2000 Conference Proceedings,* The Hague, Amsterdam, ACM Press, 273 – 280.

Sears, Andrew. 1991. Improving Touchscreen Keyboards. Univ. of Maryland CS technical reports, CS-TR-2536, March 1991

Sibert, L. E. & Jacob, R. J. K. (2000), Evaluation of Eye Gaze Interaction, *Human Factors in Computing Systems: CH1 2000 Conference Proceedings*, The Hague, Amsterdam, 281 – 288.

Stampe, D. M. & Reingold, E. M. (1995). Selection by looking: A novel computer interface and its application to psychological research. Eye Movement Research. L. (Eds). Elsevier Science B. V. 467 - 478

Ward, David J. & MacKay, David J.C., (2002), Fast Hands-free Writing by Gaze Direction, Nature 418, p. 838 (22nd August 2002).