Life in the Fast Lane: Effect of Language and Calibration Accuracy on the Speed of Text Entry by Gaze

Kari-Jouko Räihä

Life in the Fast Lane: Effect of Language and Calibration Accuracy on the Speed of Text Entry by Gaze

Kari-Jouko Räihä

School of Information Sciences, University of Tampere, Finland

kari-jouko.raiha@uta.fi

Abstract. Numerous techniques have been developed for text entry by gaze, and similarly, a number of evaluations have been carried out to determine the efficiency of the solutions. However, the results of the published experiments are inconclusive, and it is unclear what causes the difference in their findings. Here we look particularly at the effect of the language used in the experiment. A study where participants entered text both in English and in Finnish does not show an effect of language structure: the entry rates were reasonably close to each other. The role of other explaining factors, such as calibration accuracy and experimental procedure, are discussed.

Keywords: text entry · gaze input · performance · entry speed · error rate · comparative evaluation · longitudinal study

1 Introduction

Augmentative and alternative communication techniques are crucial for a considerable share of the world's population. Without such techniques they would, for instance, not be able to use many of the services that are increasingly being offered through the internet. Techniques that have been developed include speech synthesis and Braille printers for the vision impaired, and speech-to-text solutions for the hearing impaired.

A particularly challenging user group is those with severe motor-neuron diseases, such as the locked-in syndrome or ALS. In such situations the cognitive abilities are intact, but the ability to control the muscles deteriorates and eventually disappears. The eyes are often the last muscle that works. Techniques that allow communication using just the eyes have been developed and studied extensively [6].

Here we focus on using the eyes for text entry. This, too, is an actively studied field; see [4, 8] for reviews. However, as will be discussed in the next section, the results concerning the text entry speed that can be reached by using the eyes vary a lot. We will tackle one possible explanation: the difference in the language used in the experiments, and study it systematically. In addition, we will look at a number of other factors that might explain the results.

Why is it important to know the limits of text entry by gaze? By understanding the subtle differences in the implementation details of a technique we can produce better

tools for the people that need the software. Sufficiently fast text entry can, for instance, in the future enable the development of applications where one partner enters text on a mobile phone with the eyes, the phone uses text-to-speech to transmit it to a hearing partner at the other end, who can then respond. To carry out a real conversation the process cannot take too long. Therefore "How fast can one type with the eyes?" is an equally interesting and important research question as "How fast can one type with the typewriter" was for decades.

2 Previous work

Results from past experiments with text entry by gaze have been collected in Tables 1 (experiments with a soft keyboard) and 2 (experiments with other techniques). We have included only longitudinal experiments where the participants came back to the lab on several days and thus had a chance to improve their performance through experience. The data in the tables is teased out from the publications. For [17] the exact numbers were not reported and are therefore estimated from the graphs in the paper. The same holds for the MSD rate in [9].

Several papers reported on more than one studies. For [10], there were results for the learning phase (denoted by lp in Table 1) and the advanced phase (denoted by ap). In [11] there were two experiments, denoted by e1 and e2. Pedrosa et al. [9] carried out a number of experiments; two longitudinal studies with able-bodied participants, followed by individual trials with users with motor disabilities. Here we report the numbers from their second longitudinal study that produced the fastest text entry rates.

Table 1. Data from experiments on text entry by gaze with soft keyboards

Reference	Tracker	Participants	Session length	Sessions	Dwell threshold	WPM	MSD
[5]	Tobii 1750	11	15 min	10	adjustable	19.89	0.36%
[17]	Tobii 1750	8	8 phrases	14	330 ms	> 8	< 0.01%
[10] lp	Tobii T60	10	15 min	10	adjustable	18.45	0.58%
[10] ap	Tobii T60	10	15 min	5	20-250 ms	18.98	0.46%
[11] e1	Tobii P10	9	10 min	9	1000 ms	6.0	2.20%
[11] e2	Tobii P10	12	15 min	9	adjustable	7.0	4.00%

The WPM column shows the standard metric for measuring text entry speed: words per minute (WPM), where "word" means five letters (including spaces and punctuation). In other words, WPM is the number of characters entered in a minute divided by 5. MSD indicates uncorrected errors, i.e. errors that remain in the text that was entered. It is computed as the mean string distance of the entered phrase and the model phrase [16].

Table 2. Data from experiments on text entry with novel techniques

Reference	Tracker	Participants	Session length	Sessions	Technique	WPM	MSD
[17]	Tobii 1750	8	8 phrases	14	EyeWrite	> 6	< 0.01%
[15]	SR-Research Eyelink2	9	3 sayings	20	pEYEwrite-WoDyn	13.47	0.01%
[14]	Tobii 1750	12	15 min	10	Dasher	17.26	0.57%
[11] e1	Tobii P10	9	10 min	9	Dasher	12.4	1.70%
[11] e2	Tobii P10	12	15 min	9	Dasher	14.2	3.30%
[9]	Tobii REX	6	20 min	6	Filter-yedping	15.95	< 0.3%

Since all the experiments included in Tables 1 and 2 consisted of several sessions, the numbers reported indicate the averages of all participants for the best session in the series (typically the last one). Individual participants could reach better performance in some sessions. The best reported single session text entry rate was 23.10 WPM for soft keyboards [10] and 23.11 WPM for Dasher [14].

It is understandable that for different interaction techniques there is high variation in the performance, but it is surprising that for the same basic technique (soft keyboards in Table 1 and Dasher in Table 2) the results vary so much. There are a number of factors that are not shown in the tables and that do affect the efficiency; we will discuss them extensively later. However, one aspect that stands out is that all the fastest results ([5, 10, 14]) were achieved with Finnish as the language of the phrases that were entered. In all the other studies the language was probably English (the language is not reported in [15] and it could also be German).

Could it be that language has such a drastic effect on the text entry speed? We cannot really say based on the results quoted above, because, e.g., different participants took part in each experiment. To shed light on this issue, we carried out an experiment where the same participants entered text using gaze both in English and in Finnish. Longitudinal studies are laborious and we were only able to do the experiment for soft keyboards, but there the difference in past results is anyhow the largest.

3 Method

We followed a similar experimental design as has become the norm (with small variations) in studies of text entry by gaze. The details are presented next.

3.1 Participants

Eight participants (5 male, 3 female) were recruited for the experiment. Seven of them had participated in a previous experiment where eye tracking was used passively just to follow participants' eye movements while viewing a set of images. None had previously used gaze as an input technique. Two participants wore eyeglasses. Selection

of participants was based on their demonstrated ability to be tracked by the eye tracker, and on their self-reported knowledge of written English. Six participants reported their skill level as excellent (“mistakes in grammar and vocabulary are rare”) and two reported it as good (“I can communicate but I make mistakes”). All had Finnish as their mother tongue.

Each participant was paid €15 for each one hour test session. In addition, a bonus of €70 was paid both to the participant that reached the highest text entry rate, and to the participant that showed the biggest improvement in text entry rate over the sessions.

3.2 Apparatus

A Tobii T60 eye tracker with a 17-inch TFT color monitor with 1280 × 1024 resolution was used to track the gaze. A PC running Windows XP was used for the experiment. The software for text entry was an in-house application based on a soft keyboard. The application window filled the screen, as shown in Fig. 1.

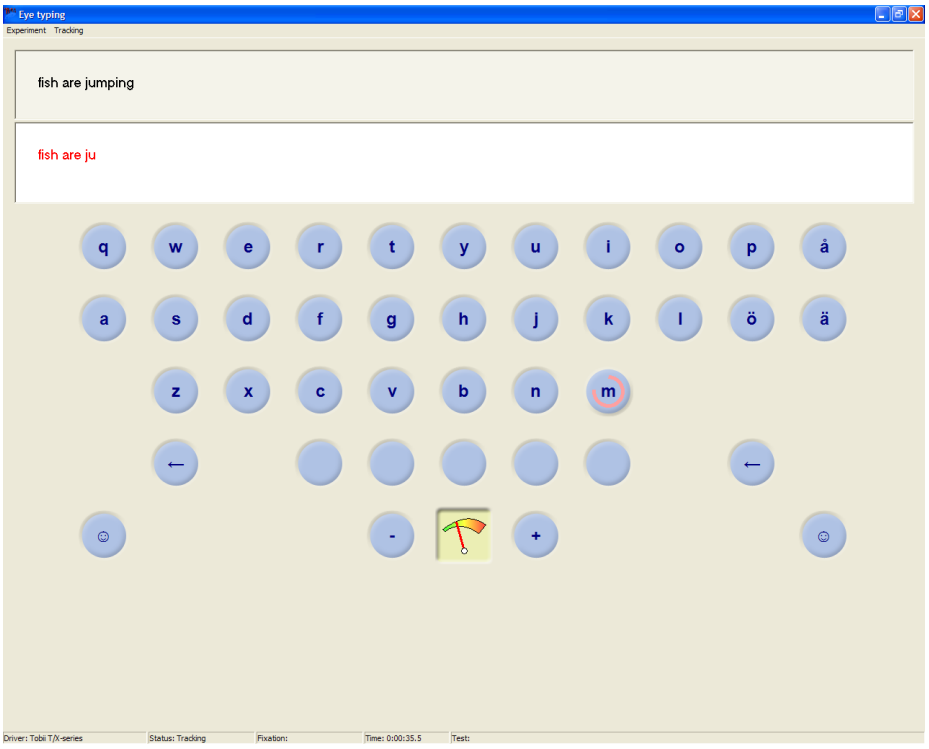

Fig. 1. A screenshot of the application used in the experiment. The QWERTY keyboard shows the Scandinavian layout used

The top frame of the window is used to display the model phrase that the participant must enter. The text entered is shown in the next frame in red. These are later referred to as the source and target frames, respectively.

The layout of the keys follows the layout on a standard Scandinavian qwerty keyboard. The letters on the keys are displayed in 18 pt Arial bold. There is no shift key, since the phrases that had to be entered did not have upper case letters. Similarly, the punctuation marks are missing for the same reason. On the row below there are five space keys and two backspace keys. The motivation for multiple keys with the same function was to provide a short path to these frequently used keys, as suggested in [10].

When a key is gazed at, a red curve starts to grow on the key, as shown for the letter "m" in Fig.1. Once the curve has made a full circle, the letter gets selected: a sharp clicking sound is played and the letter appears in the target frame.

The bottom left and right corners hold "ready" keys (smileys) to be selected when the participant has finished entering the model phrase. This causes the next phrase to appear in the source frame and the target frame to be cleared.

The time needed for a key to be selected can be adjusted by the participant at any point during the experiment. This is done using the two gaze-controlled keys on the bottom row; the speedometer between them is an indicator of the selected threshold. Clicking on the "+" key increases speed, i.e., shortens the dwell time. The decrement depends on the current level of the dwell threshold. At 1000 ms (the initial setting), clicking on the "+" lowers the dwell to 900 ms. Moving to faster settings the decrement decreases, so that at 260 ms and below, clicking on "+" shortens the required dwell by 20 ms. The "–" key works with similar increments in the opposite direction.

In addition to the experimental text entry software, the TraQuMe application [1] was used to measure calibration accuracy in the beginning and end of each test session. In our experiment TraQuMe was used in a mode where it displayed five targets (much like in a typical calibration procedure), each for 1.5 s, and collected the gaze coordinates produced by the eye tracker during that time. The average x and y coordinates of the data points and their distance to the center of the target were computed to produce the *offset*, a measure for the accuracy of the calibration. Similarly, precision was measured by computing the *dispersion* (the standard deviations) of the gaze points during the 1.5 s span.

Of the two measures, small offset is more important for our eye typing software than small dispersion, because of the way that gaze is mapped to the keys on the keyboard. Since the algorithm is of crucial importance for obtaining good entry rates, we describe it here briefly. First, the expansion algorithm makes objects larger than their visual appearance. Therefore there are actually no gaps between the keys. The mapping algorithm further works with this expanded layout, where the effective size of each key is 100×100 pixels. From a distance of 60 cm the side of each key was thus about 2.5 degrees.

Each key has a time counter that is increased (by the amount of time between two successive gaze points) when a new gaze point hits the key, and decreased (by the same amount of time) otherwise. With each sample only one key increases its counter, all others decrease their counters or keep them at 0. Once a counter exceeds the dwell

time, the corresponding key becomes selected, and all counters are reset. The dwell time progress circle is shown on the key that has the greatest counter value at the moment.

3.3 Procedure

Each participant first came to the lab for a briefing of the software and the experiment. The functionality of the software was explained and the procedure to be followed in the test sessions was described. The participant then had a chance to try out the software. Each participant first entered in Finnish two phrases with a 1000 ms dwell threshold. They were then instructed to try decrementing the threshold with one or two notches and to enter two phrases in English. They were further offered the chance to experiment with two more phrases, but only two participants made use of that opportunity; in general everyone got the hang of it during the first four phrases. Finally the experimenter stepped in the user's role and showed with two phrases how text entry looks like when the dwell threshold is in the order of 300 ms.

The participants were instructed that the goal of the experiment was to find out how fast they can enter correct text. They were advised to correct errors that they spotted immediately, but not to back up long stretches of already entered text if they noticed an error in the beginning.

The actual sessions took place on consecutive working days. Each participant was requested to carry out the test on at least five and at most eight consecutive days. All eight participants came back for at least six days, meaning that everybody had a weekend in between the sessions at some point. One participant did the experiment for seven days and four participants for eight days.

Each day of the experiment consisted of two 15-minute sessions with a 10-minute break in between. One session was in English and the other in Finnish. The order of the languages was swapped each day. Half of the participants started with Finnish, the other with English in the first day.

In the first session the participants were asked to use the initial dwell threshold setting of 1000 ms for the first phrase, and after that to adjust the threshold at will. At the end of the first day (Sessions 1 and 2), the rules of how to get the bonus were revealed. In particular, the metric to be used (entered words per minute minus error percentage) was explained.

The participants were seated at a distance of about 60 cm from the screen. In each of the two sessions the eye tracker was first calibrated for the participant using a 9-point calibration. The calibration procedure was repeated until the result looked adequate in the calibration window of the ETU driver [13] that was used by the software. Then TraQuMe [1] was used to quantitatively measure the quality of the calibration. The measurement was repeated at the end of the 15-minute session.

3.4 Task

The participant's task was to transcribe the model phrases as quickly and accurately as possible. The English phrases came from the 500-phrase collection [3] that is cus-

tomarily used in such experiments. The upper case letters in the phrases were replaced by lower case letters for this experiment. The Finnish phrases mostly came from the Finnish translation [2] of the English corpus. However, some English phrases that were lines in song lyrics (e.g., "starlight and dewdrops" from *Beautiful Dreamer*) did not make sense as direct translations, since the lyrics were not translated literally for the Finnish version of the song. In such cases popular Finnish songs were used instead as a source of phrases.

The text entry application chose the model phrases at random. It logged the text entered and full gaze data. It also kept time, so that when 15 minutes had passed, it allowed the participant to finish entering the current phrase but at the click of the ready key a save dialog would pop up. The experimenter then stepped up to an adjacent display, saved the log and carried out the calibration quality measurement.

4 Results

Altogether 114 text entry sessions took place. In them a total of 5 338 phrases and 146 607 letters were entered. In addition, in 38 cases the participant unintentionally clicked the ready key while entering the phrase, and these (0.7% of all cases) were removed from the data.

Fig. 2 plots the average text entry speed over the sessions. The graph shows the data from all participants over six days (12 sessions) and separately the data from those four participants that continued for the full eight days.

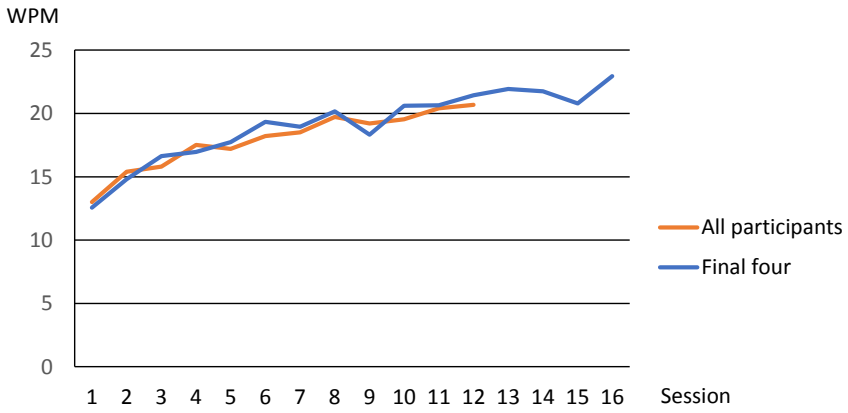

Fig. 2. Text entry rate as a function of session number

For easier comparison with previous research, the entry rates in Fig. 2 and elsewhere in this paper are expressed in WPM without taking into account the errors in the remaining text. The participants were extremely conscientious in correcting the errors they spotted. On average, the error rate (measured as string edit distance, MSD) was 0.39% – that is, approximately one incorrect character for every 260 correct charac-

ters. The errors decreased from 0.58% in the beginning to 0.28% in the last session. Their effect is so negligible that the speed metric that takes errors into account follows very closely the same pattern as the WPM metric in Fig. 2.

The best average session rate, 25.6 WPM, was achieved by participant P5 in Session 12. He stopped then and did not continue for the final two days. In that session the resulting text had 4 errors, meaning an error rate of 0.06%. His data is shown in Fig. 3 as an example of how experience affected the measures.

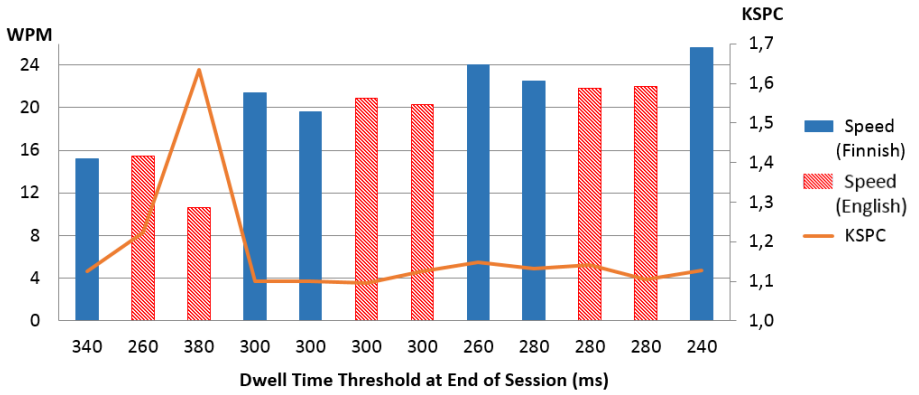

Fig. 3. Development of text entry rate (as bars) and keystrokes per character (as line graph) for participant P5. Blue bars denote sessions with Finnish, striped red bars with English.

Fig. 3 also plots the keystrokes per character (KSPC), a common metric to measure how many errors were made and corrected during a session [16]. P5 was one of the participants who moved quickly to very short dwell times. That backfired in Session 3, where poor calibration created a need to backspace often, as Fig. 3 shows. High KSPC correlates strongly with lower WPM rates.

Fig. 3 indicates that for P5, results for Finnish were somewhat better than for English. Fig. 4 shows for the first six days, and for all participants, the text entry speed separately for Finnish and English. Each day each participant carried out one session in each language.

A two-way ANOVA with language and day as within-subject factors revealed a significant effect of day on WPM ($F_{7,5} = 17.48$, $p < .0001$). Although the graph for Finnish stays above the one for English for every day, the difference is not big enough to be statistically significant ($F_{7,1} = 5.25$, ns).

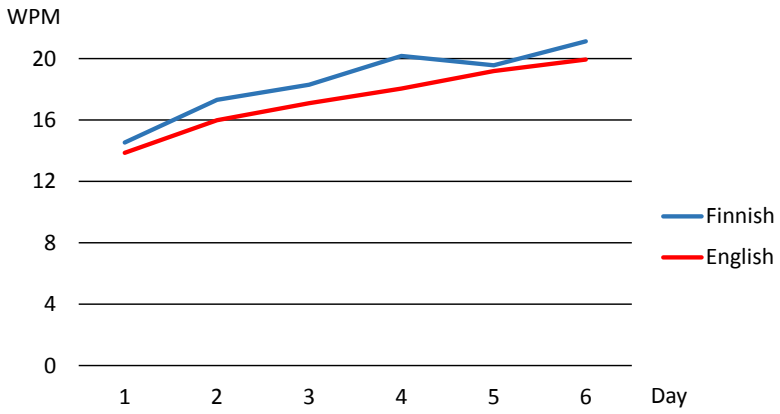

Fig. 4. Text entry rates for Finnish and English as a function of day of experiment

On average, calibration accuracy remained on the same level throughout the sessions. The average offset at the beginning of sessions was 0.6 degrees (with SD of 0.06 degrees), and the average dispersion was 0.3 degrees (SD 0.05 degrees). The figures are low and as good as can be expected from the eye tracker used in the experiment. However, although the variation between sessions, taken over all participants, was thus low, there was higher variation for individual participants. The WPM rate for each session is plotted in Fig. 5 for each participant against the calibration quality at the beginning of the session. For simplicity, the sum of offset and dispersion was used as the combined quality measure.

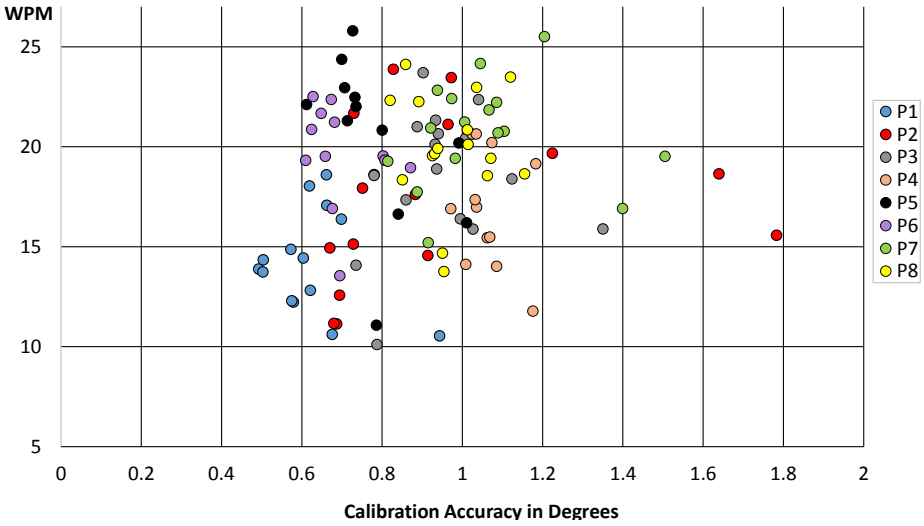

Fig. 5. WPM vs. quality of calibration at start of session for each participant

We also measured the calibration accuracy at the end of each session. However, those measurements are less reliable, since some participants had the tendency to relax at the end of the session and change their posture considerably. Therefore the measurements are not reported here.

Another factor that directly affects text entry rate is the dwell time threshold. We recorded it at the start and end of each session. Fig. 6 plots text entry rate against the dwell time threshold at the end of each session; note that a faster or slower pace can have been used at some point during the session. However, with the exception of the first day, the within session adjustments of the threshold were moderate.

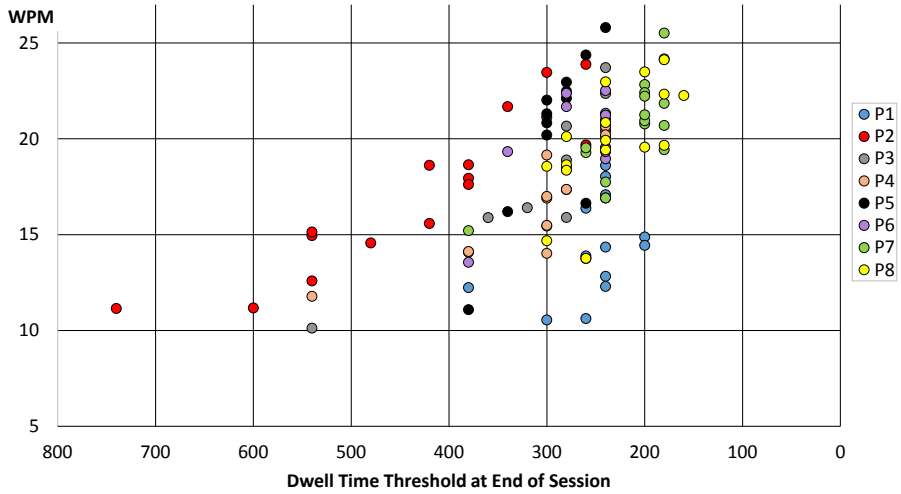

Fig. 6. WPM vs. dwell time threshold at end of session for each participant

The time taken to transcribe the model phrases is affected by how easy it is to recall them and their spelling. The number of glances back to the source and target frames at the top of the window are plotted in Fig. 7 for each participant. The first graph shows the overall number for the two frames, whereas the second graph focuses on the glances to the source frame only and shows the numbers for English and Finnish phrases separately.

A two-way ANOVA with language and day as within-subject factors revealed a significant effect of language on glances both to the source ($F_{7,1} = 15.24$, $p < .01$) and target ($F_{7,1} = 7.62$, $p < .05$) frames, with English giving rise to more glances than Finnish per phrase (4.0 vs. 2.7 for source, 2.7 vs. 2.2 for target). Day did not have a significant effect, nor was there interaction between language and day.

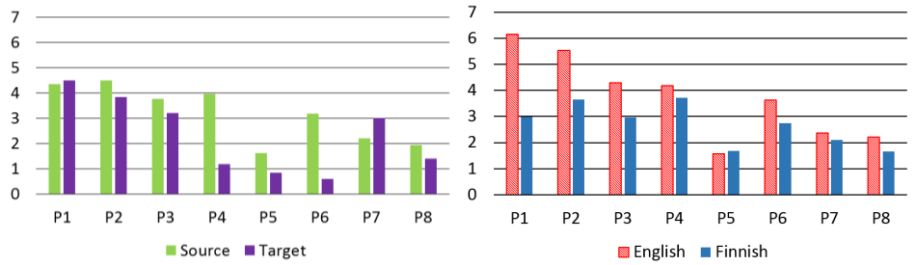

Fig. 7. On the left the average number of glances to the source and target frames per phrase and participant. On the right the average number of glances to the source frame for the English and Finnish phrases.

5 Discussion

5.1 Speed of Text Entry

This study did not provide with convincing evidence of the effect of language on the speed of text entry by gaze one way or the other. The trends in Fig. 4 appear clear on the surface, but the ANOVA analysis did not show a statistically significant difference (although it was fairly close with $p = .056$).

How much difference can one expect between the participants' mother tongue and another language that they know well? In a similar study that used a regular keyboard instead of eye gaze as the input technique, Isokoski and Linden [2] found a bigger difference: 49.7 WPM for Finnish vs. 41.8 WPM for English, a difference of about 16%. In this light the difference of the average rates in the participants' best sessions, 22.8 WPM for Finnish and 20.9 WPM for English, i.e. a difference of about 8%, is considerably smaller than expected.

Text entry by gaze is in many ways different from using a keyboard, even if the layout of the keys is similar. With a hard keyboard the typist can use at least two hands, if not all ten fingers. In text entry by gaze the process is strictly sequential, with the same "device" (eyes) used to select in succession all the keys one by one. Then the distance to be covered when moving from one key to the next can come into play. Could it be different for the two languages?

To analyze this, we computed the bigram frequency matrices for both corpora. A bigram is a pair of letters that appear in succession in the text, and the frequency matrix gives the number of occurrences of each bigram in the corpus. A similar matrix was created for the distance between the letters in each bigram on the soft keyboard. Multiplying these matrices gives us a bigram distance matrix weighted with the frequencies, and taking the average over all bigrams then tells us whether we can expect an effect that is caused by the bigram frequencies being different.

Table 3 summarizes the key characteristics of the corpora.

Table 3. Characteristics of the corpora.

	Finnish	English
phrases	500	500
words	1834	2713
characters	13878	14310
distinct bigrams	352	389
double letters	1027	324
average weighted bigram distance	324.9	304.4

There is a difference of 304.4 pixels vs. 324.9 pixels (about 6.5%) in average weighted bigram distance in favor of English, which can be expected to affect the text entry speed. Adding this to the observed speed difference of 8% brings us closer to the findings in [2].

Another factor that, however, works in the opposite direction, is the proportion of double letters (the same letter appearing twice in succession) in the two languages. Double letters are more frequent in Finnish than in English. To enter the double letters with our text entry software the user can simply keep looking at the letter. After the first of the pair of letters is entered, there is a fixed threshold of 150 ms before the dwell time counter is restarted. This is both quick and convenient, making the entry of double letter bigrams faster than entering the other bigrams: the average time to move the gaze from one character to the next was typically more than 200 ms. The share of double letter bigrams in the corpora is 7.4% for Finnish and 2.3% for English.

Finally, it should be noted that Fig. 2 indicates that the entry speeds were still improving after the 6th day (12th session). Our analysis was based on the first six days only to have a sufficient number of participants.

In addition to learning from day to day, one may ask if there was difference between the two daily sessions. Indeed, the overall grand mean of text entry speed was slightly higher for the second daily session (18.5 WPM vs. 17.4 WPM). However, the difference was not statistically significant ($F_{7,1} = 5.39$, ns).

In summary, there are factors that have an adverse influence on the expected text entry speed: the mother tongue (other than English) puts English to a disadvantaged position, but the characteristics of the language can have features that affect the entry speed in both directions.

5.2 Errors

The extremely low level of errors in the text entered, 0.28% in the last sessions, may come as a surprise. It is largely explained by the feel of text entry with a low dwell threshold. Typing then becomes essentially an activity of moving the gaze from one letter to the next, with a very short time for waiting for the letter to get selected. The typist hears the click and realizes if the gaze was on the intended letter or not. If not, it feels natural to click on the backspace and correct the error immediately. In fact, a

large portion of the remaining errors are not caused by entering a letter that the typist did not intend, but rather by recalling the model phrase incorrectly.

With a low dwell threshold it becomes easier to get into a regular typing rhythm. The importance of rhythm has been noted before [4]. In our study one participant (P8) voluntarily commented on this; even if correcting the errors is an almost automatic activity, it still interferes with the typing rhythm, which she considered crucial for fast text entry.

Compared to typing with a conventional keyboard, the sequential nature of text entry by gaze has a positive effect on the level of errors. The two hands can be in action at the same time and may arrive at their target keys in the wrong order without the typist noticing this before looking at the text entered. With gaze such a competition situation is not possible and the typist can trust the order of letters to be correct.

This becomes evident also in analyzing the glances to the source and target frames. Fig. 7 (left) shows that most participants looked more at the source frame than the target frame. The two participants for which this does not hold (P1 and P7) had to struggle in many sessions with calibration that deteriorated during some sessions. The number of glances is particularly low for the overall fastest typist, P5. The formula for fast text entry speed seems to be a combination of a suitably small dwell threshold combined with trust in the correct letters getting entered, and thus no need to check the text frames frequently.

Trust is, of course, affected by calibration accuracy and thereby the software behaving "as expected". Nevertheless, Fig. 5 indicates that people can learn to adjust their gaze position to compensate for poorer calibrations as well. Participant P7, in particular, who was often struggling with a lower quality calibration, still put up respectable entry speed numbers, and reached even the second best result of all in spite of the calibration accuracy in that session being as poor as 1.2 degrees.

Getting back to the glances, Fig. 7 (right) shows that the source phrase is glanced at more often for English than for Finnish. This is understandable for people who don't have English as first language. In addition to some rare words with difficult spelling that might give problems even to native speakers (e.g., "conscience", "exasperation", "dewdrop"), the problem is compounded by the different spellings in dialects of English (e.g., "favourite" vs. "favorite"). No such problems exist with Finnish, which is written as pronounced. It is understandable that participants rechecked the source phrase in such cases to make sure that they got the spelling right.

In summary, the participants took many means to ensure the correctness of their entered text: correcting errors immediately and checking the spelling of words when in doubt. Excluding from the analysis errors that were the result of memory lapses would have further improved the correctness rate substantially.

5.3 Why Are the Findings Across Studies So Different?

If the language of the phrases does not explain the difference in the results obtained in different experiments, what does? Perhaps most importantly, even if the basic techniques are the same, they may differ significantly in important implementation details. This holds for the experiments with Dasher [11, 14], and the same is true for soft

keyboards. For instance, feedback modes have a significant effect on text entry speed [7]. The dwell time threshold similarly affects both the efficiency and the experienced workload [10].

The fact that the user is in control of adjusting the dwell time threshold during a session, as pioneered in [5], is of crucial importance. In [11], the dwell time was also adjustable, but the paper does not give details on how this was done. Judging from the screenshot of the interface in [11], this was only possible between the sessions. It is understandable that users are not eager to jump to big decrements in the dwell threshold if they are unsure of how it affects them. When they have the tool to increase or decrease the threshold at will, they can experiment with shorter dwell times to see if it works for them.

The briefing given at the beginning of the experiment is also important. We showed ourselves what text entry with a short dwell time threshold looks like before the participants were engaged in the first session. This presumably caused them to move to short thresholds much faster than in any previous experiments. As Fig. 6 shows, a clear majority (71%) of the sessions ended with the threshold being in the 200–300 ms range. In 8% of the sessions it was below 200 ms and in 21% of the sessions above 300 ms. Only one participant used higher than 300 ms thresholds after the third day of the experiment.

It should be emphasized again that a short threshold not only enables fast text entry, but it gives a very different feeling to the user. Text entry changes from interacting with icons that represent keys to selecting letters that one wants to enter. In other words, interacting with the interface as an artefact fades to the background and the user can focus on the task at hand.

One obvious factor affecting the results is the corpus used [12]. Especially in our case, when we decided to use only lower case letters to limit any confounding factors, it can be expected that the results are better than in previous studies. Comparing our best single session average (25.61 WPM) and overall last session average (20.36 for the 12^{th} session with all participants, 22.66 WPM for the 16^{th} session with final four participants) to the results quoted in Section 2 (23.10 WPM and 19.89 WPM, respectively), we see that this is indeed the case.

The apparatus used in the experiments is another significant factor. Most experiments reported in the literature have been carried out with Tobii eye trackers. Trackers with a higher sampling rate would provide an interesting comparison. Importantly, the algorithm for mapping gaze points to keys of the keyboard plays a central role. We suspect that the algorithm described in Section 3.2 contributes greatly to achieving high text entry rates, and it may explain much of the differences in the results of other soft keyboard experiments. This, of course, requires further experimentation.

Finally, the choice of participants and their motivation are also important factors. In most studies cited here the participants were compensated for their participation and often rewarded for best performance, so this should not be a major cause of variation. Our participants were native speakers of Finnish and fluent with English. The participants in [11] were also fluent in English, but the paper does not tell if this means that they were natives. Participants having English as their mother tongue can

be expected to get even better results than those in our study, even if the spelling controversies discussed in Section 5.2 still remain to some extent.

6 Conclusion

We have carried out a controlled experiment where the participants entered text using a soft keyboard with their gaze from three to four hours (12 to 16 sessions of 15 minutes each), half of it in English and the other half in Finnish. No statistically significant difference was found between the languages. For reasons discussed in Section 5.3, our participants reached somewhat higher entry rates than has been reported in the literature before. In particular, the speed of text entry in English (best single session rate of 22.31 WPM, and overall average rate of 19.95 WPM for the sixth day and 21.07 WPM for the eighth day) is much higher than has been reported before.

We don't intend to claim the superiority of one technique over the other. Instead, we want to underline the importance of the many factors that affect the results in such experiments. To compare two designs with each other it is important to control as many of the experimental variables as possible: same participants, same briefing, same goal setting, same compensation, comparable calibration quality and same tracker. Without control over such factors the results are not fully commensurate.

Finally, it should be noted that our experiment (as most similar experiments) was carried out with able bodied users. The users that really need the technology of text entry by gaze can develop into very fast eye typists, as evidenced e.g. by some videos on the Internet for English[1] and Finnish[2]. To trace the world record speed in text entry by gaze, the experiment should involve such users who use the technique not for hours, but for days, weeks and months.

Acknowledgments

I am grateful to Per Ola Kristensson for suggesting this experiment and to Päivi Majaranta, Poika Isokoski, Saila Ovaska and Oleg Špakov for discussions and comments on the manuscript. In particular, Oleg Špakov implemented the soft keyboard and the gaze tracking algorithm used in the experiment. The statistical analysis was done using I. Scott MacKenzie's Java tools[3]. Finally, the participants in the experiment were committed and responsible – thank you very much!

This work was supported by the Academy of Finland (project MIPI).

[1] http://www.eyegaze.com/eye-tracking-assistive-technology-device/

[2] http://www.iltalehti.fi/iltv-doc/201411070115600_dc.shtml (text entry e.g. at 01:00 and 03:48)

[3] http://www.yorku.ca/mack/RN-Anova.html

References

1. Akkil, D., Isokoski, P., Kangas, J., Rantala, J., Raisamo, R.: TraQuMe: A Tool for Measuring the Gaze Tracking Quality. In: Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’14), pp. 11–18. ACM, New York (2014)
2. Isokoski, P., Linden, T.: Effect of Foreign Language on Text Transcription Performance: Finns Writing English. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction (NordiCHI ’04), pp. 109–112. ACM, New York (2004)
3. MacKenzie, I.S., Soukoreff, R.W.: Phrase Sets for Evaluating Text Entry Techniques. In: CHI ’03 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’03), pp. 754–755. ACM, New York (2003)
4. Majaranta, P.: Text Entry by Eye Gaze. Dissertations in Interactive Technology 11, University of Tampere (2009)
5. Majaranta, P., Ahola, U.-K., Špakov, O.: Fast Gaze Typing with an Adjustable Dwell Time. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’09), pp. 357–360. ACM, New York (2009)
6. Majaranta, P., Aoki, H., et al. (eds): Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies. IGI Global, Hershey (2012)
7. Majaranta, P., MacKenzie, I.S., Aula, A., Räihä, K.-J.: Effects of Feedback and Dwell Time on Eye Typing Speed and Accuracy. Univ Access Inf Soc 5, 199–208 (2006)
8. Majaranta, P., Räihä, K.-J.: Text Entry by Gaze: Utilizing Eye-tracking. In: MacKenzie, I.S., Tanaka-Ishii,K. (eds.) Text Entry Systems: Mobility, Accessibility, Universality, pp. 175–187. Morgan Kaufmann, San Francisco (2007)
9. Pedrosa, D., da Graça Pimentel, M., Wright, A., Truong, K.N.: Filteryedping: Design Challenges and User Performance of Dwell-Free Eye Typing. ACM Transactions on Accessible Computing 6, Issue 1, Article 3 (2015)
10. Räihä, K.-J., Ovaska, S.: An Exploratory Study of Eye Typing Fundamentals: Dwell Time, Text Entry Rate, Errors, and Workload. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12), pp. 3001–3010. ACM, New York (2012)
11. Rough, D., Vertanen, K., Kristensson, P.O.: An Evaluation of Dasher with a High-Performance Language Model as a Gaze Communication Method. In: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI ’14), pp. 169–176. ACM, New York (2014)
12. Sanchis-Trilles, G., Leiva, L.A.: A Systematic Comparison of 3 Phrase Sampling Methods for Text Entry Experiments in 10 Languages. In: Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI 2014), pp. 537–542. ACM, New York (2014)
13. Špakov, O.: iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications. Dissertations in Interactive Technology 9, University of Tampere (2008). See also http://www.sis.uta.fi/~csolsp/downloads.php
14. Tuisku, O., Majaranta, P., Isokoski, P., Räihä, K.-J.: Now Dasher! Dash Away! Longitudinal Study of Fast Text Entry by Eye Gaze. In: Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’08), pp. 19–26. ACM, New York (2008)
15. Urbina, M.H., Huckauf, A.: Alternatives to Single Character Entry and Dwell Time Selection on Eye Typing. In: Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’10), pp. 315–322. ACM, New York (2010)

16. Wobbrock, J.O.: Measures of Text Entry Performance. In: MacKenzie, I.S., Tanaka-Ishii,K. (eds.) Text Entry Systems: Mobility, Accessibility, Universality, pp. 47–74. Morgan Kaufmann, San Francisco (2007)
17. Wobbrock, J.O., Rubinstein, J., Sawyer, M.W., Duchowski, A.T.: Longitudinal Evaluation of Discrete Consecutive Gaze Gestures for Text Entry. In: Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '08), pp. 11–18. ACM, New York (2008)