Text entry on smart glasses

A comparison of two input techniques

Roderick McCall
SnT, University of Luxembourg
Luxembourg City, Luxembourg
roderick.mccall@uni.lu

Andrei Popleteev
SnT, University of Luxembourg
Luxembourg City, Luxembourg
andrei.popleteev@uni.lu

Benoît Martin
LCOMS, Université de Lorraine
Metz, France
benoit.martin@univ-lorraine.fr

Nicolas Louveton
SnT, University of Luxembourg
Luxembourg City, Luxembourg
nicolas.louveton@uni.lu

Thomas Engel
SnT, University of Luxembourg
Luxembourg City, Luxembourg
thomas.engel@uni.lu

Abstract— **This paper presents a comparative pilot usability study of Dasher and an on-screen keyboard on a head-mounted display. Interaction logging data was captured along with subjective responses (via the SUS questionnaire). The results indicate that there is a strong need to develop text entry systems for smart glasses rather to simply adopt those that are already available. However, both approaches are useful when there is a need to enter private or sensitive data.**

Keywords— smart glasses; text input; keyboard

I. INTRODUCTION

Smart glasses that include sensors are capable of detecting gestures. For example, small head movements may be a good way of text entry for people with impaired mobility. Also for other users the ability to enter text via a personal screen may provide a way to enter sensitive or private data that may normally be visible on mobile phone screens or overheard if speech interaction is used. This paper explores these challenges by conducting a comparative pilot study of Dasher [1] and an on-screen keyboard both of which are implemented on an Epson Moverio BT-200 augmented reality headset.

The paper starts by reviewing existing approaches to text entry, in particular the range of benefits and challenges each approach introduces. It then introduces the on-screen keyboard and Dasher approaches via smart glasses and provides results from a pilot user study comparing both methods.

II. BACKGROUND

The number of mobile devices is growing each year with a seemingly endless set of new phones, tablets, watches and glasses being released. Furthermore, the range of new interaction techniques brings more opportunities and challenges. Many solutions exist and can be used but they have to be adapted to the new contexts. In this section, we list different alternatives that appear suitable for electronic glasses even if they were not originally proposed for this context.

Text entry is a task where the user must enter and edit text. We will only focus on entering text. The most common way is to use a standard physical QWERTY or AZERTY keyboard. For the electronic glasses this keyboard must be considered as an external device and can be difficult to handle. Also, it is non-ambiguous, since each key produces only one letter and therefore the overall keyboard requires significant space.

To reduce the overall size of the keyboard, the size of the keys can be reduced but this also restricts the level of user comfort and possibly also the overall usability. Ambiguous keyboards can also be used instead, where one key gives access to more than one letter. The disambiguation can be done by the user or by the computer. An example of such an approach is Multitap where the user can tap the same key in order to access the desired letter. However, this can slow down the text entry process. This is why LetterWise introduced a letter-level optimization: it proposes the letter on the basis of probability depending on the context [2]. The user must still browse the letters but the order is optimised. T9 introduced a word-level optimization [3]: the user taps only one time each key that contains the desired letter and at the end of the word, T9 proposes the most probable word from the list of possible options. It uses a dictionary and if the proposed word is not the correct one, the user must browse the list. In addition, it is impossible to write a word that doesn't belong to the dictionary.

In the chord keyboard, for example Twiddler [4], the letter is obtained by pressing several keys at the same time. Chord keyboards are mainly eyes-free which can be interesting in the context of electronic glasses: the user can write without looking at the keys. The keyboard can remain in the pocket. The disadvantage is the learning curve as the layout is totally new, which might strain the user's patience.

The need to use an external device is not limited to the cases when input is required via a keyboard. It is possible to use several different sensors to capture the gestures of the user, for example camera, accelerometer, gyroscope, sensitive surface, etc. There is no need to have a keyboard. The system must interpret gestures as letters. For example, a sensor attached to the wrist sends movements to a computer system that translates relevant signals into text. This allows users to write letters in the air, as if they are using an invisible board or pad [5].

It is also possible to use a virtual keyboard. A virtual keyboard is a logical keyboard that can have a different layout from standard mechanical keyboards. In the context of electronic glasses, the display can be done on an external surface (mobile phone, table, wall, etc.) or directly on the glasses. New projected keyboards use an external surface [6]. The image of the keyboard is projected onto a surface, generally the table in front of the user. When the user touches the surface covered by an image of a key, the corresponding keystroke is detected. Such a keyboard only requires a project and a suitable surface for the image to be displayed on. While this solution is tempting it is not suitable for all usage contexts.

It is also possible to provide the virtual keyboard on the screen of a secondary device or directly on the electronic glasses. Depending of the screen size, however, the keyboard size might become a problem. However, many new options benefit from the dynamic aspect of the display. Completely new layouts have been explored [7] as well as dynamic keyboard that reorganizes the layout after each key pressed [8]. Highly dynamic solutions that are based on navigating through the space of letters also exists [9][1]. For example, Dasher [1] is a dynamic system: the user steers through an animated space populated by characters. More probable characters have a larger area making them easier to select. When the user steers through the area of a symbol, this symbol is entered. Any input device with either continuous position input or one or more buttons can be used with Dasher. In this particular example visual feedback is a key aspect.

When the virtual keyboard is directly displayed on the glasses, the main challenge is to propose suitable interaction techniques to replace the direct contact between the fingers and the physical keyboard or the surface of projection. Many alternatives can be used:

- External device
- Scanning
- Gestures
- Eye tracking
- Head tracking
- Etc.

External devices can also act as pointing devices. For example, in TypingRing, a standard QWERTY keyboard is divided into multiple zones of three keys [10]. This keyboard is underneath the hand of the user but is invisible to them. The user wears a ring on their middle finger. They move their hand and the ring indicates the active zone and the platform provides visual feedback to the user about that zone. Then the user carries out a typing gesture using the forefinger, the middle finger or the ring finger depending on the desired key.

Gestures can also be undertaken on top of the virtual keyboard. They can be done directly on the external device that displays the virtual keyboard [12]. Furthermore, mid-air gestures can also be used [13][14]. In Vulture, the gestures are considered at word level as in ShapeWriter [12]. By moving their hand, the user places the cursor over the first letter of the word and after pinching they trace the word in the air. After releasing the pinch, the five words that best match the gesture are proposed; the top match is pre-selected. The main drawback is perhaps the huge gestures that must be done by the user. The users can thus disturb their environment and people around them. This remark is relevant for all mid-air gesture systems.

Scanning is also widely used for disabled people. The system in [15] scans the different keys and the user has to confirm when the desired key is highlighted. The drawback is the very slow writing speed, although the system does support some data entry privacy as it needs no large gestures, just a simple interaction for validation.

In the domain of systems for disabled users, it is common to use eye-trackers to offer an alternative way to select keys on a virtual keyboard. The user looks directly at the target to select it. For example, Dasher is well-suited to be used with an eye – tracker [16].

Finally, the increasing use of gyroscopes in most of the new devices provides the possibility to write via tilting [17]. A typical solution is provided by Hex [9]. The user navigates through an infinite hexagonal grid. The first hexagon contains 36 letters. The 36 letters are split into the 6 neighboring hexagons (6 letters per hexagon). By tilting their device, the user enters into a neighboring hexagon. The 6 letters are then split into the 6 neighboring hexagons (1 letter per hexagon). The user continues tilting to enter in the neighboring hexagon that contains the desired letter. The entry is then validated and the interaction restarts. Hex uses a linguistic model to ease tilting toward more probable letters. Visual feedback also provides an important cue to the user.

Overall, a clear challenge of text entry on restricted devices (small form factor and absence or limited support for external devices) is to reduce the operational distance between each data entry. Both layout [9][1] and semantic [6][7] optimisations have been proposed. In this paper we compare a QWERTY on-screen visual keyboard, which is a non-optimal technique for head-gesture (yet a familiar metaphor), to Dasher which is a technique optimised for reducing user movements (such as in motor impairment situations). It should be noted that within the on-screen keyboard version participants were asked not to use the predictive text features.

TABLE I. TRAINING PHRASES

Hello world
The cat sat on the mat

TABLE II. STUDY PHRASES

A	The University of Luxembourg
B	Paris is the capital of France
C	Interdisciplinary Centre of Security, Reliability and Trust

III. PILOT EVALUATION

A within-subjects randomized order experiment was conducted that contained two test conditions. The first condition was the Dasher interface [1] which was implemented on the Epson Moverio BT-200 headset (Fig. 1). The second interface was the headset's standard on-screen keyboard version (Fig. 2). Both interfaces were described earlier. The study design was consistent across both interfaces and consisted of the following overall steps:

- Demographic Questionnaire
- Training stage with both interface styles
- Tests of each interface with different pieces of text provided both written and dictated. Error rates and task completion time were recorded.
- Completion of SUS Questionnaire
- Debriefing session.

Basic demographic information was collected, such as age, gender, familiarity with the technologies in question. This included an examination of experience of any similar interfaces.

A. *Training Phase*

During the training phase the evaluator would explain how each interface work. In the case of the dasher interface this was also demonstrated using a mobile phone version. After the explanation the participants were asked to enter the training phrases (see Table 1: Training Phrases). We used short simple phrases during this stage. Initially this was provided in text form so that those who were non-native English speakers could familiarise themselves with any spelling issues. Each phrase was also read out to the participant. After entering each phrase they were asked if they had any problems and the evaluator would provide some information if required.

B. *Study Phase*

The study phase followed a similar pattern to the training session. Each participant was asked to enter text using both interaction techniques. The order in which the devices were used and which piece of text was provided were randomized. In order to overcome problems with spelling etc, we chose phrases which were very familiar to people (see Table 2: Study Phrases). They were given the same excerpt on both devices. The participants were told to take down the dictated text (without looking at the written document). They were informed that they did not have to capitalize letters or use punctuation. They were also informed that they should only make corrections to the current word (as in [1][18]).

C. *Participants*

In total 7 participants took part in the study. Of these six completed the study and five provided completed SUS questionnaires. One participant was unable to use the visor due to vision problems and their data has been removed from the analysis. The participants were aged 23-34, with an average age of 29.7. Of these 4 were male and 2 female. All were working in computer science research or project management at the university. Some users had to wear their glasses as well as the visor in order to see the content. They were all regular computer users and none (although they were not asked) appeared to have any problems with interaction via touchpads or other traditional input devices.

In total only two people reported to having used an on-screen keyboard in augmented reality before while only one reported to having used the Dasher Interface under augmented reality before. It should be noted however that the user who had used both interface styles did not complete the SUS questionnaires, therefore this will have had no impact on the rated usability level. Furthermore, both users rated their experience of the two interfaces as low with their last experiences ranging from 2 days to one month prior to the study.

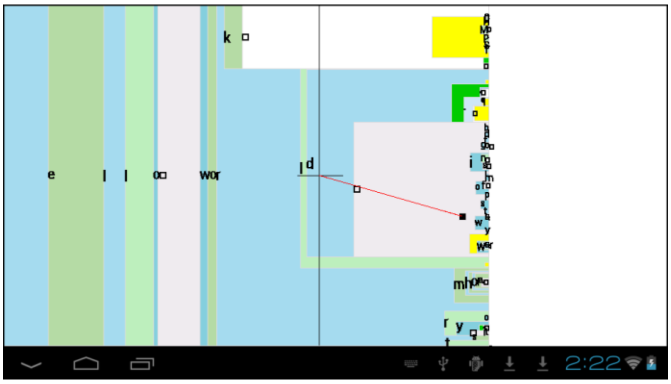
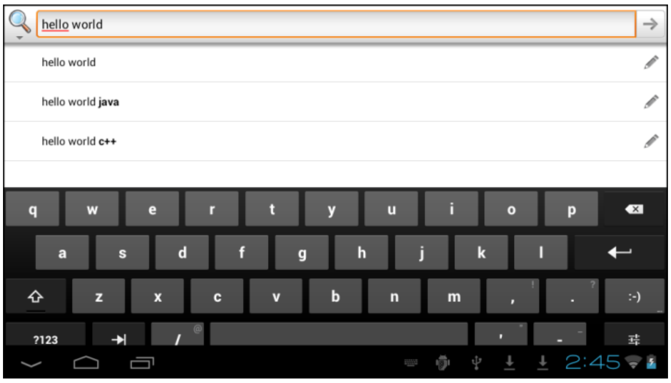


Fig. 1. Dasher interface on Epson Moverio glasses.



hello world

Fig. 2. On-screen keyboard of Epson Moverio BT-200 glasses. The letters are selected with a cursor controlled by an external touchpad.

TABLE III. QUESTIONNAIRE ITEMS

	Item
1	I think that I would like to use (device) frequently
2	I found the (device) unnecessarily complex
3	I thought the (device) was easy to use
4	I think that I would need the support of a technical person to be able to use this (device)
5	I found the various functions in this (device) were well integrated
6	I thought there was too much inconsistency in this Device
7	I would imagine that most people would learn to use this (device) very quickly
8	I found the (device) very difficult to use
9	I felt very confident using the (device)
10	I needed to learn a lot of things before I could get going with this (device)

D. *Data Collected*

During the experiments, we performed automated screen recording from the headset, taking a screenshot at approximately 1 Hz rate. Performance evaluation was based on the analysis of these timestamped screenshots.

For each textual phrase the start and end time was logged along with all entries made (both correct and incorrect). For example, which letter was selected or when the delete key was used to make a correction. This data was then compared to the planned text and the number of errors was logged. In order to ensure that there was no learning effect between users, the prediction caches were re-initialized so that the system had to relearn for each user.

E. *System Usability Questionnaire*

The SUS [10] was applied after each device had been used. The SUS is a standard questionnaire which has been used across a range of IT products and is used to assess the users subjective feelings of usability about the system. Although it consists of a number of questions the author indicates that only the overall score should be considered and not the individual components. In general if a system receives a score of above 68 (out of 100) it is deemed to be average. There are in total ten questions put to the user (see Table 3) and they are asked to rate this on a 5 point scale. For the purposes of scoring, a strong disagreement with the proposition counts as a 0, while a strong agreement counts as a 4.

F. *Debriefing*

At the end of the study the participant was asked if they had any further questions to ask and were provided with more information if needed. As the true purpose of the study was stated to participants at the outset there was no need to divulge the true nature of the experiment beyond reminding them of the purpose.

IV. RESULTS

A. *Task Completion Time*

The task completion time for each text fragment and each interface type are presented in Table 4 and Fig. 3. Task completion time is calculated from the entry of the first character in each phrase until the last character has been completed. We would expect that some phrases take longer to enter than others. Also phrase C proved problematic as the Dasher's predictive engine was unaware of the word “interdisciplinary”; therefore, entering this word proved more laborious than the other words, for which probabilities were properly estimated and suggested letters were conveniently arranged into most probable sequences.

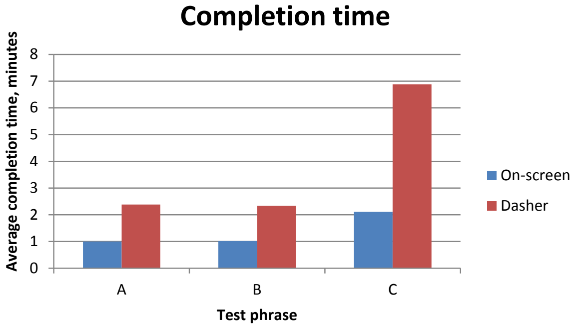


Fig. 3. Average task completion time.

TABLE IV. AVERAGE TASK COMPLETION TIME PER INTERFACE AND PER PHRASE (IN MINUTES).

	Phrase	On-Screen Keyboard	Dasher
A	The University of Luxembourg	1.00	2.38
B	Paris is the capital of France	1.01	2.34
C	Interdisciplinary Centre of Security, Reliability and Trust	2.11	6.88

B. *Error and Correction Rates*

The error and correction rates were defined as the absolute number of errors and corrections divided by the length of the corresponding test phrase. As shown in Fig. 4 and Table 5, the on-screen keyboard was more accurate, with error rates varying between 2-3%, whereas Dasher's error rate was considerably higher, reaching up to 13%.

Error rates for both interfaces were calculated on the following basis, (1) total failure and (2) number of errors. A total failure was taken to be when a participant was unable to complete the task (2 out of 7 participants); these data were excluded from analysis. One participant stopped the experiment after long struggle with the last test phrase, this samples was also excluded. For other errors these were simply taken to be the total number of incorrect characters relative to the original text phrase. Account was taken of when English vs US English spelling errors occurred for example “Center” instead of “Centre”. Other common errors were “Luxembourg” or “Luxemburg”. Although the participants were given the correct spelling and asked to be aware of this at the outset some made these errors. In general such errors were ignored as they are not indicative of a mistake when using the device.

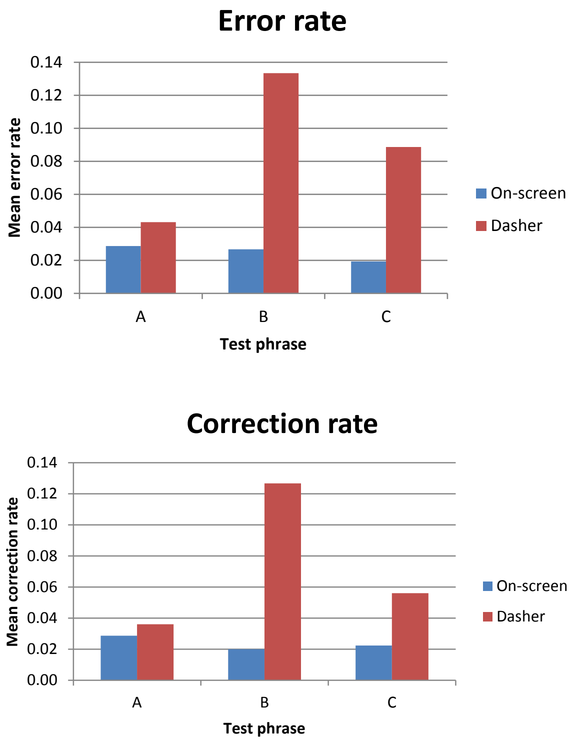


Fig. 4. Error and correction rates per interface.

TABLE V. THE ERROR (CORRECTION) RATES PER INTERFACE.

	Phrase	On-Screen Keyboard	Dasher
A	The University of Luxembourg	2.9% (2.9%)	4.3% (3.6%)
B	Paris is the capital of France	2.7% (2.0%)	13.3% (12.7%)
C	Interdisciplinary Centre of Security, Reliability and Trust	1.9% (2.2%)	8.9% (5.6%)

C. *System Usability Score for Dasher and On-Screen Keyboard*

The onscreen keyboard received an SUS score of 73.5 on average, while the Dasher interface received a score of 44. This is not entirely unexpected as the Dasher interface is targeted at users with limited mobility. Therefore it may not be viewed as user-friendly as devices with which the users are more familiar.

It should be noted that the sample size is very small as this is only a pilot study. However, the range of results also varies. For example, with Dasher the SUS scores range from 12.5 to 72.5. While for the on-screen keyboard they range from 55 to 90.

V. DISCUSSION

It is not unexpected that touch pad interaction via the onscreen keyboard would prove faster, less error prone and more user friendly. This can be explained from a number of perspectives, namely that the Dasher system is intended for those who cannot normally use a keyboard or similar set up and would therefore naturally be slower. Also that it probably requires more training than was provided in this study. However, what is more interesting is that the longest and most complex phrase yielded the lower error rate for on-screen keyboard entry and was the second lowest when used in Dasher. This can however be explained as one user had a much higher error rate than the others. Also interesting to note is that the average completion time for the on-screen keyboard does not rise as quickly as the phrases become longer when compared to the Dasher interface. This is due to the Dasher interface not being able to semi-predict “Interdisciplinary” which meant that users had to manually find the letters. This also increased the completion time and increased the error rate.

Regardless of the comparative nature of the results, both the Dasher and On-screen keyboard interfaces bring benefits to end-users. Firstly they allow for the entry of private data as passersby cannot see what is being entered; unlike a traditional keyboard or when speech interaction is used. Also in the case of Dasher no external device is used.

VI. CONCLUSION

In this paper we compared two text entry methods for smart glasses: a predictive gesture-based Dasher method and a standard on-screen keyboard controlled by an external touchpad. The biggest drawback of the onscreen keyboard is that it requires an external device in its present form in order to allow interaction. In contrast, the Dasher approach when accompanied with gesture recognition requires no additional device but does suffer from a number of limitations. As can be seen in both cases, however, the amount of time it takes to enter text is more than would be the case when using a normal keyboard. As a result there is clearly a need to develop interfaces that are much more suited to entering text on smart glasses rather than simply adapting those that already exist. However, despite these drawbacks, both approaches allow relatively discrete and private text entry. This may be beneficial when entering private information.

This paper presented only two of the potential input methods with results from a small pilot study. On-going work is exploring how to improve not only the interaction within these approaches but also other text entry methods, with the aim to publish a larger scale study with a range of different augmented reality text entry interfaces.

ACKNOWLEDGEMENTS

This work has been performed as part of the eGlasses project which is partially funded by NCBiR, FWF, SNSF, ANR and FNR under the framework of the ERA-NET CHIST-ERA II, project eGLASSES.

REFERENCES

[1] Ward, D. J., Blackwell, A. F. and MacKay, D. J. C. 2000. Dasher - a data entry interface using continuous gestures and language models. In Proceedings of UIST 2000, ACM, 129-137.

[2] MacKenzie, I. S., Kober, H., Smith, D., Jones, T., Skepner, E. (2001). LetterWise: Prefix-based disambiguation for mobile text input. Proceedings of the ACM Symposium on User Interface Software and Technology - UIST 2001, pp. 111-120. New York: ACM.

[3] Grover, D.L., King, M.T. and Kushler, C.A. Reduced keyboard disambiguating computer. US Patent 5818437 (1998).

[4] K. Lyons, T. Starner, D. Plaisted, J. Fusia, A. Lyons, A. Drew, and E. Looney. Twiddler typing: Onehanded chording text entry for mobile phones. In Proceedings CHI. ACM Press, 2004.

[5] Christoph Amma and Tanja Schultz. 2013. Airwriting: bringing text entry to wearable computers. XRDS 20, 2 (December 2013), 50-55.

[6] Helena Roeber, John Bacus, and Carlo Tomasi. 2003. Typing in thin air: the canesta projection keyboard - a new method of interaction with electronic devices. In CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03). ACM, New York, NY, USA, 712-713.

[7] Frank Chun Yat Li, Richard T. Guy, Koji Yatani, and Khai N. Truong. 2011. The 1line keyboard: a QWERTY layout in a single line. In Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11). ACM, New York, NY, USA, 461-470.

[8] Xiang 'Anthony' Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions. In Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14). ACM, New York, NY, USA, 615-620.

[9] Williamson J. and Murray-Smith R. Hex: Dynamics and Probabilistic Text Entry. In Proc. of the Hamilton Summer School on Switching and Learning in Feedback systems, LNCS, vol. 3355, 333-342, 2005.

[10] Brooke, J. SUS-A quick and dirty usability scale. Usability evaluation in industry, vol. 189(194), 4-7. 1996.

[11] Shahriar Nirjon, Jeremy Gummeson, Dan Gelb, and Kyu-Han Kim, "TypingRing: A Wearable Ring Platform for Text Input", Accepted at the 13th International Conference on Mobile Systems, Applications, and Services (MobiSys '15),14 pages, Florence, Italy, 2015.

[12] Zhai, S., Kristensson, P.O., Shorthand Writing on Stylus Keyboard, in Proceedings of CHI 2003, ACM Conference on Human Factors in Computing Systems, Fort Lauderdale, Florida, April 5-10, 2003. pp 97-104.

[13] Barbariyskiy Nikolay, Mid-Air Text Entry Methods, Master's Thesis. https://dip.felk.cvut.cz/browse/pdfcache/barbanik_2014dipl.pdf

[14] Anders Markussen, Mikkel Rønne Jakobsen, and Kasper Hornbæk. 2014. Vulture: a mid-air word-gesture keyboard. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI '14). ACM, New York, NY, USA, 1073-1082.

[15] Octavio Rivera Romero, Alberto Jesus Molina Cantero, Isabel María Gómez González, A Study of Two-Inputs Scanning Methods to Enhance the Communication Rate, Assistive Technology from Adapted Equiment to Inclusive Environments, At Florence (Italy), Volume: 25, 132-137.

[16] Daniel Rough, Keith Vertanen, and Per Ola Kristensson. 2014. 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). ACM, New York, NY, USA, 169-176.

[17] Eleanor Jones, Jason Alexander, Andreas Andreou, Pourang Irani, and Sriram Subramanian. 2010. GesText: accelerometer-based gestural text-entry systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10). ACM, New York, NY, USA, 2173-2182.

[18] Tuisku, O., Majaranta, P., Isokoski, P., & Räihä, K. J. 2008. Now Dasher! Dash away!: longitudinal study of fast text entry by Eye Gaze. In Proceedings of the 2008 symposium on Eye tracking research & applications. ACM, New York, NY, USA. 19-26.