A multiscript gaze-based assistive virtual keyboard

H. Cecotti[1] *Senior Member, IEEE*, Y. K. Meena[3], B. Bhushan[3], A. Dutta[3], G. Prasad[4] *Senior Member, IEEE*

Abstract— **The recent development of inexpensive and accurate eye-trackers allows the creation of gazed based virtual keyboards that can be used by a large population of disabled people in developing countries. Thanks to eye-tracking technology, gaze-based virtual keyboards can be designed in relation to constraints related to the gaze detection accuracy and the considered display device. In this paper, we propose a new multimodal multiscript gaze-based virtual keyboard where it is possible to change the layout of the graphical user interface in relation to the script. Traditionally, virtual keyboards are assessed for a single language (e.g. English). We propose a multiscript gaze based virtual keyboard that can be accessed for people who communicate with the Latin, Bangla, and/or Devanagari scripts. We evaluate the performance of the virtual keyboard with two main groups of participants: 28 people who can communicate with both Bangla and English, and 24 people who can communicate with both Devanagari and English. The performance is assessed in relation to the information transfer rate when participants had to spell a sentence using their gaze for pointing to the command, and a dedicated mouth switch for commands selection. The results support the conclusion that the system is efficient, with no difference in terms of information transfer rate between Bangla and Devanagari. However, the performance is higher with English, despite the fact it was the secondary language of the participants.**

I. INTRODUCTION

The emergence of novel inexpensive sensors and input devices provide new tools for the creation of assistive technologies, which can increase the independence and improve the quality of life of a large population of severely disabled people. To ensure that assistive technologies improve users' quality of life, the emphases should focus on consumer involvement in the selection and evaluation of appropriate assistive technology, and ways to make technologies more widely available and affordable [1], such as in developing countries. The availability, the cost, and the user experience, are key criteria that go beyond the raw performance in terms of user acceptance of a new technology. There are disabilities, e.g. neuro-locomotor disabilities or amyotrophic lateral sclerosis, that represent research challenges for caregivers and assistive technology [2]. A virtual keyboard based on gaze detection is primarily aimed at people with severe speech and motor impairment who are unable to speak nor use sign language to communicate. Even for patients

who are able to talk, speech recognition for typing is not private, people around can hear what the user wants to type, it is not efficient in a noisy environment, and it can be cumbersome to edit the text. Therefore, they need adapted human-computer interfaces to communicate [3], [4]. In addition, assistive technology devices have to be adapted in relation to the constraints imposed by the user, such as his language. These constraints can be avoided to some extent by modifying existing commercial devices, or the creation of new communication means, e.g. brain-machine interface (BCI) for locked-in patients. While there exists an appeal for BCI, it represents the only means of communication for only a small number of people, because a large number of severely disabled people are able to control their gaze. Severely disabled people may also be able to do some gestures, and the detection of a gesture can be used as a signal to validate an item pointed by the user with his gaze, e.g., people with quadriplegia. The ability of gaze control is actually least affected by disabilities: eye movement is not affected by severe disabilities such as spinal cord injuries. Virtual keyboards using eye-tracking can therefore serve a substantial number of patients and disabled people. Paraplegia has a high impact on individuals and their ability to be employed. A study conducted in India reported the employment rate of 41%, and people who were employed were living either in centers run by armed force or in specialized centers [5]. In addition, none of the people with tetraplegia living in the community was employed. Furthermore, it is critical to both evaluate and improve their quality of life [6]. Multiple virtual keyboard based on gaze detection have been proposed recently, they vary in the type of layout, the strategy for detecting the zones of interest on the screen, the mode of operation synchronous vs. asynchronous [7], the use of a dwell time or type of action (e.g. hand gesture [8]) that is used for determining the selection of an item. In [9], a gaze based virtual keyboard was proposed with a mouth switch for the selection of the commands and demonstrated the usefulness of such an approach for typing text using an English keyboard arranged in alphabetical order.

In this paper, we present a new multimodal multiscript virtual keyboard that can be used with the mouse, the touch screen, and through gaze detection and a mouth switch for the selection of the commands. The key novelty of the approach lies in the combination of 3 scripts within the layout of the same graphical interface. Indeed, in order to provide a generic and usable interface for typing, it is necessary to consider different scripts and assess the performance in these different scripts. The proposed solution includes

(1) Department of Computer Science, College of Science and Mathematics, Fresno State University, Fresno, Ca, USA. `hcecotti@csufresno.edu`

(2) Department of Computer Science, Swansea University, Swansea, UK.

(3) Centre of Mechatronics, Department of Humanities and Social Sciences, Indian Institute of Technology (IIT), Kanpur, India.

(4) Intelligent Systems Research Centre, Ulster University, Londonderry, UK.

the Latin (i.e. English characters), Bangla, and Devanagari scripts. The contributions of this paper are related to the effective presentation of a multiscript gaze-based virtual keyboard and its evaluation with people speaking different languages. In particular, this paper assess the differences of performance between people who speak Hindi (Devanagari script) or Bangla as a first language, and English (Latin script), as a secondary language. The remainder of the paper is organized as follows: first, we describe the gaze based virtual keyboard and the different scripts in Section II. The experimental protocol is then detailed in Section III. The system performance is presented in Section IV and discussed in Section V.

II. METHODS

Given the ratio between the number of symbols and the possible number of buttons that can be placed simultaneously on the screen, it is often necessary to consider a menu-tree interface for symbol selection. It is the case for virtual keyboards using BCI [10] and eye-tracking [11]. We consider such an approach in this system because both Bangla and Devanagari have a large number of symbols that cannot be placed directly on the screen without jeopardizing the detection accuracy of the command due to the reduction of the visual angle delta between two buttons. The graphical user interface of the virtual keyboard is composed of two main parts. The first part represents the output, which is presented in the center of the screen, and therefore directly accessible to the user without any gaze shifts. In the output section, the text that is typed by the user, the current speed, and the estimated gaze position are displayed. The second part of the layout corresponds to the edge of the screen that contains the different buttons. The graphical user interface for the three scripts is depicted in Fig. 1. The top panel represents the layout in the Latin script, the panel in the middle corresponds to the layout of the Bangla script, while the bottom panel corresponds to the Devanagari script. In relation to the aspect ratio of the common display devices, the buttons are placed in 3 rows of 4 columns, and the space for 2 buttons in the second row is used to display the output text. Hence, 10 buttons are present to access the different letters and editing commands. This choice was driven by the size of the screen and prior evaluation of such a system [12]. While it is possible to have a reliable performance with a button per letter with the Latin script [9], such an approach cannot be considered for the Devanagari or Bangla scripts that have a larger number of letters. Here, we assume that the 10 buttons will be reliably detected with the eyetracker at the expense of using a tree selection approach with two levels for the selection of all the different letters. The first level of the selection tree structure contains 9 letters and the "go back" command in each button. The layout of the second level is displayed inside each button at the first level. For instance, the letter 'D', which is in the 1st row, 4th column is displayed as such in the image assigned to the button in the first level. The "go back" command is employed for going back to the first level in case of a wrong selection at the first level. It

is worth noting that contrary to the layout proposed in [8] that included both upper and lower case letters, the present layout in the Latin script includes characters with diacritics to match the necessary diacritics that are present in Bangla and Devanagari. The same way that an experienced user will not look at the keyboard while typing, an experienced user with a gaze based virtual keyboard will not look at the typed text in the output box, and the user's gaze will shift directly from one command to the next. To keep track of the typed letters using covert attention, i.e, paying attention without moving the eyes, the last five spelled-out letters are displayed under each button. In addition, an auditory stimulus (a simple beep sound) is played after command selection. In addition to the changes related to the layout of the GUI after the selection of a command, another visual feedback is provided by changing the color of the frame around the selected button. The last visual feedback corresponds to the estimated gaze position of the user to help users to adjust their posture. For each script, there are 88 symbols and 12 commands that are used to navigate or edit the text, i.e. go back, delete. For the Latin script, the layout includes the 26 characters, 10 digits, 9 diacritics, 15 symbols, including punctuation marks, delete, delete all, go back, and space. For the Bangla script and Devanagari script, the digits, delete, delete all, go back, and space commands are conserved. For Devanagari, it includes 45 letters, 17 different matras (diacritics) and halants (killer strokes). For Bangla, it includes 47 letters, 15 different diacritics.

Command selection is achieved through two steps. In the first step, the user is pointing his/her gaze towards the desired button, the estimated gaze position is then compared to the center position of each command, and the position of the frame representing the typed text. If the gaze position is outside the output part, then the button that is the closest (using the Euclidean distance) to the gaze position is selected as the potential candidate for the selection. It is therefore possible to select a command while the gaze position is not inside the bounding box of a button. In the second step, the selection of the current candidate is made through a switch. In this present case, we consider a custom made mouth switch, which aims at providing a communication device for severely disabled people who are not able to talk. The software was written in C# with Microsoft Visual Studio 2017, and the gaze data was acquired with functions from the SDK provided with the Tobii 4C [13]. The momentary push button switch is presented in Fig. reffig:f3. The switch is directly connected to the right click button of a USB computer mouse that is connected to the computer.

III. EXPERIMENTAL PROTOCOL

A. Procedure and design

The system was evaluated at the Indian Institute of Technology (IIT) Kanpur, India. The participants were all undergraduate and graduate students from India. Some students are Hindi (Devanagari script) native speakers while some others are Bangla native speakers. All the students had some reasonable knowledge of English, allowing them to
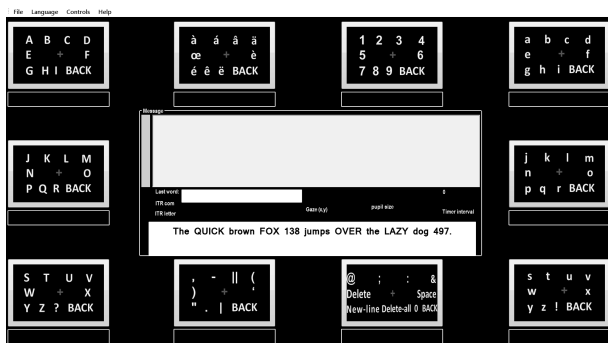
(a) Latin
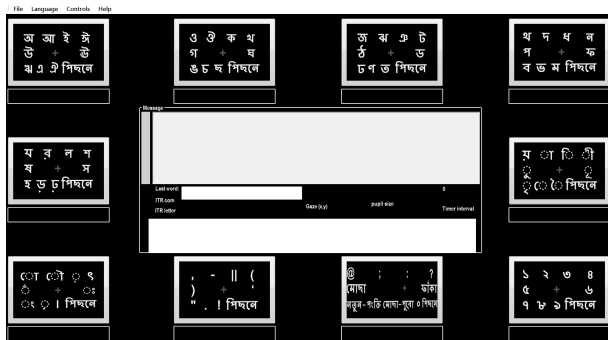
(b) Bangla

(c) Devanagari

Fig. 1. GUI of the application for **(a)** Latin script, **(b)** Bangla script, and **(c)** Devanagari script.

type sentences in English using the virtual keyboard. There was no financial reward provided to the participants. The Helsinki Declaration of 2000 was followed while conducting the experiments. 56 healthy adult participants took part to the study. Among these participants, 28 completed a session with English and Bangla, while 24 completed a session for English and Hindi. These two groups are analyzed separately to consider pairs of values. Each participant had to type a sentence in two scripts (Bangla and English, or Devanagari and English). For instance, the sentence in English was: "The QUICK brown FOX 138 jumps OVER the LAZY dog 497." The order of the scripts was randomized across participants. Each task had to be fully completed, without any mistake; errors would be corrected using the appropriate editing commands.
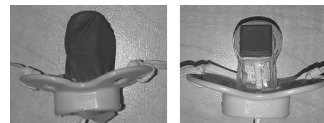

Fig. 2. Mouth switch (push-button switch) with protection.

B. Performance evaluation

The performance was assessed in relation to the information transfer rate (ITR) (in bits/min) at the command level (n=10) ITR_{com} and at the symbol level ITR_{symb}. The ITR at the command level represents the number of commands that can be produced per minute ($log_2(10)$), while the ITR at the symbol level represents the amount of produced symbols ($log_2(88)$). Because the Bangla and Devanagari scripts contain a large number of diacritics that must specified in relation to a letter, a large number of letters represent the composition of several elements. Hence, the typing speed in terms of number of letters spelled-out per minute is not a reliable measurement for comparing the performance across scripts. For computing statistical significance, the Wilcoxon signed-rank non-parametric test is used to evaluate the paired performances for participants testing Bangla vs. English, and Devanagari vs. English. To evaluate the differences between Bangla and Devanagari, we consider the Wilcoxon rank sum non-parametric test.

IV. RESULTS

The performance of the virtual keyboard is presented in Table I. The table presents the mean and standard deviation (SD) of performance across participants in terms of information transfer rate at both the command level (ITR_{com}) and the symbol level (ITR_{symb}). The remaining columns of the table display the average time (in second) for each of the 10 buttons. Given the number of diacritics and composition of letters in Devanagari and Bangla, it is difficult to properly compare the performance in terms of number of letters typed per minute across different scripts, including the Latin script for the English language. When we analyze the ITR at the command level across the different conditions, a pairwise comparison indicates that there exists a substantial difference between English and Bangla ($p < 10e-3$, $z = 2.82$), with an ITR of 82.22 ± 23.11 and 72.57 ± 19.78 bits/min for English and Bangla, respectively. Similarly, there exists a difference between English and Devanagari ($p < 10e - 3$, $z = 3.46$), with an ITR of 88.63 ± 19.31 and 75.53 ± 17.58 bits/min for English and Devanagari, respectively. However, we observe no difference of performance related to the ITR between Bangla and Devanagari, showing that the performance level was similar, independently of the chosen script. The same pattern of performance is observed for the ITR at the letter level, with a decrease of the values due to commands dedicated to the correction of errors. For instance, the ITR at the letter level is 57.61 ± 14.14 and 50.45 ± 13.62 bits/min for English and Bangla (Bangla speaker subjects), while it is at 57.06 ± 22.06 and 45.19 ± 15.34 bits/min for English and Devanagari (Hindi speaker subjects).

TABLE I

PERFORMANCE OF THE VIRTUAL KEYBOARD FOR THE DIFFERENT SCRIPTS: INFORMATION TRANSFER RATE AND COMMAND SELECTION DURATIONS.

Script		ITR_{com}	ITR_{symb}	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B_{mean}
Latin	Mean	85.28	57.86	2.51	1.92	2.33	2.88	2.43	2.18	2.49	2.06	2.60	3.23	2.46
	SD	21.36	18.24	0.71	0.81	1.23	0.90	0.74	0.73	0.76	0.51	0.70	1.22	0.83
Bangla	Mean	72.35	50.63	2.00	2.17	3.68	4.24	3.16	2.59	3.58	2.01	2.98	2.76	2.92
	SD	19.15	13.16	0.63	0.94	1.35	1.09	1.04	0.96	1.60	0.57	0.96	0.93	1.01
Devanagari	Mean	75.53	45.19	1.96	2.74	3.79	5.69	2.8	2.41	2.91	1.96	2.63	2.45	2.93
	SD	17.58	15.34	0.67	1.15	1.31	2.36	0.95	0.68	0.98	0.57	0.69	0.65	1.00

V. DISCUSSION AND CONCLUSION

The abilities to communicate verbally and in a written form are critical skills that are needed in our daily life. A large number of alternative solutions as a means of communication have been proposed in the last decades, from eye-tracking solutions to brain-machine interfaces. These progresses have been made possible through both the improvement of the hardware, e.g. the sensors, and the software, e.g. signal processing techniques, that don't progress at the same speed. This discrepancy in the progress implies changes in terms of the choice of the best solution at a particular time. Thanks to the recent advances of portable consumer grade eye-trackers, it is possible to create reliable gaze based virtual keyboards. Yet, the design of the graphical user interface should take into account the type of input as a constraint. In addition, the input script has an influence on the graphical user interface design, as the number of letters and the way letters can be combined have a significant impact on the layout arrangement. A new multiscript (Latin, Bangla, and Devanagari) gaze based virtual keyboard has been proposed. The system can be used by a large number of people as 2.6 billion people (36% of the world population) use the Latin alphabet, and about 1 billion people (14%) use the Devanagari script. There is a difference between the script and the language as there are 510, 490, and 215 million speakers of English, Hindi, and Bangla, respectively [14]. The system can be directly used for the Marathi/Konkani speakers (70 million) by adding a single letter (i.e., ऋ). While it can be used for other languages using the Latin script, the proposed system has some limitations. It includes only a subset of the Latin script symbols. Scandinavian and Baltic languages require more diacritics (e.g. å, ä, ö), so a different layout would be needed for these languages. This issue raises the difference between a script and a language, in terms of prior probability to use a letter and/or diacritic to type a word. Hence, further adaptations may need to decouple the script from the language to optimize the arrangement of the letters on the screen. This project represents one step towards increasing cultural and linguistic diversity in the use of augmentative and alternative communication devices [15], reaching developing countries (e.g., India, Bangladesh). When researching a generic solution, one has to find an approach that can be adapted for different scripts. A key question to address in such a system is the stability of the performance across scripts in relation to the choices made for the design of the graphical user interface. The evaluation of the proposed system has shown that the performances for Bangla and Devanagari were similar, and the performance with the Latin script provided better performance despite the fact that participants were not native English speakers. In a large number of developing countries, there exists the need of portable and robust solutions for patients and/or disabled people who are unable to communicate or write with a regular keyboard. Further work will deal with the addition of other languages in order to provide a universal virtual keyboard.

REFERENCES

[1] M. J. Scherer, "Outcomes of assistive technology use on quality of life," *Disability and Rehabilitation*, vol. 18, 1996.

[2] R. Lupu, R. Bozomitu, F. Ungureanu, and V. Cehan, "Eye tracking based communication system for patient with major neuro-locomotor disabilities," in *Proc. IEEE 15th ICSTCC*, Oct. 2011, pp. 1–5.

[3] V. Raudonis, R. Simutis, and G. Narvydas, "Discrete eye tracking for medical applications," in *Proc. 2nd ISABEL*, 2009, pp. 1–6.

[4] H. A. Caltenco, L. N. Andreasen Struijk, and B. Breidegard, "Tonguewise: Tongue-computer interface software for people with tetraplegia," in *Proc IEEE Eng. Med. Biol. Soc.*, 2010, pp. 4534–7.

[5] N. Gupta, J. Solomon, and K. Raja, "Employment after paraplegia in india: a postal survey," *Spinal Cord.*, vol. 49, no. 7, pp. 806–811, Jul. 2011.

[6] R. Singh, S. S. Dhankar, and R. Rohilla, "Quality of life of people with spinal cord injury in northern india," *Int. J. of Rehabilitation Research*, vol. 31, pp. 247–251, Sept. 2008.

[7] Y. K. Meena, H. Cecotti, K. Wong-Lin, and G. Prasad, "Design and evaluation of a time adaptive multimodal virtual keyboard," *Journal on Multimodal User Interfaces*, pp. 1–19, 2019.

[8] H. Cecotti, Y. K. Meena, and G. Prasad, "A multimodal virtual keyboard using eye-tracking and hand gesture detection," in *Proc. 40nd Int. IEEE Conf. of the Engineering in Medicine and Biology Society*, 2018, pp. 3330–3333.

[9] S. Sanjay and H. Cecotti, "A gaze-based virtual keyboard using a mouth switch for command selection," in *Proc. 40nd Int. IEEE Conf. of the Engineering in Medicine and Biology Society*, 2018, pp. 3334–3337.

[10] H. Cecotti, "A self-paced and calibration-less SSVEP based brain-computer interface speller," *IEEE Trans. on Neural Systems and Rehab. Eng.*, vol. 18, pp. 127–133, 2010.

[11] ——, "A multimodal gaze-controlled virtual keyboard," *IEEE Trans. Human-Machine Syst.*, vol. 46, no. 4, pp. 601–606, 2016.

[12] Y. K. Meena, H. Cecotti, K. Wong-Lin, and G. Prasad, "A novel multimodal gaze-controlled hindi virtual keyboard for disabled users," in *Proc. Annual Conf of IEEE Systems, Men, and Cyberntics (SMC)*, 2016, pp. 1–6.

[13] "Tobii technology, danderyd, sweden," http://www.tobii.com/, 2016, accessed: 2015-06-01.

[14] R. G. J. Gordon, Ed., *Ethnologue: Languages of the World, 15th Edition*. Dallas, Texas: SIL International, 2005.

[15] J. Light and D. McNaughton, "The changing face of augmentative and alternative communication: Past, present, and future challenges," *Augmentative and Alternative Communication*, vol. 28, no. 4, pp. 197–204, 2012.