Adam NOWOSIELSKI

West Pomeranian University of Technology, Szczecin

doi:10.15199/48.2016.12.62

Acceleration of touchless typing with head movements by limiting the set of potential characters

***Abstract**. In this paper, a proposition of touchless typing interface operated by head movements is proposed. It is designed for users with motor impairments unable to operate the standard computer input devices but able to move their heads. The interface has a form of single row alphabetical keyboard operated through left and right directional head movements and up and down head rotary movements. To accelerate the typing the interface is adapting by limiting the set of next characters. The initially typed phrase is dictionary checked for possible complements.*

***Streszczenie.** W artykule zaproponowano interfejs bezdotykowego pisania za pomocą ruchów głowy. Jest on przeznaczony dla użytkowników z motorycznymi ograniczeniami, którzy nie są w stanie obsługiwać standardowych urządzeń obsługi komputera, ale mogą poruszać głową. Proponowany interfejs ma formę jednowierszowej klawiatury uporządkowanej alfabetycznie obsługiwanej poprzez kierunkowe ruchy głowy w lewo i prawo oraz ruchy obrotowe w górę i w dół. Celem przyspieszenia pisania interfejs adaptuje się ograniczając zbiór potencjalnych znaków. Wstępnie wpisana fraza jest sprawdzana słownikowo, aby uzyskać możliwe uzupełnienia. (**Przyspieszanie bezdotykowego pisania za pomocą ruchów głowy poprzez ograniczanie zbioru potencjalnych znaków**).*

Keywords: human-computer interaction, touchless typing, gesture typing, gesture interaction, virtual keyboard.
Słowa kluczowe: komunikacja człowiek-komputer, bezdotykowe pisanie, pisanie gestem, interakcja gestem, wirtualna klawiatura.

Introduction

Head operated interfaces can be defined as touchless interaction interfaces where manipulation is achieved by a form of head movements. These interfaces may also focus directly on user's face or facial features. Head operated interfaces are of particular importance for physically challenged people who are unable to operate the standard computer input devices. Touchless interaction enable users to operate without additional medium and intermediate equipment. There exist solutions for conventional mouse replacement in non-contact environment. The touchless typing, however, still remains a challenge.

Humans have relatively high degree of freedom in moving the head. Unfortunately, not all movements are suitable for touchless human-computer interaction. Head movements in horizontal directions are easy to make for the user and are very well detectable. Rotary left and right movements are also easy to make but since the viewing direction changes it is uncomfortable to look at the monitor at an angle. In the vertical direction the head movements are easy to detect but are difficult or inconvenient for the user to perform when sitting. Contrary, the rotary up and down movements (tilt of the head up and down) are natural and easy to perform. They can be also detected with computer vision techniques.

Some approaches to touchless typing with head movements have already been reported in the literature [1, 2, 3]. Touchless text entry is also possible with other parts of the body used for steering purposes. Non-contact finger operated text entry interface is proposed in [4] and hand operated in [5]. Other alternatives include speech recognition or eyetracking. Nevertheless, the problem of touchless typing is neglected in the contemporary research and vision-based text entry interfaces are still rare, considered as a novel issue [4] and insufficiently studied [2]. Moreover, text entry is the tasks in human-computer interaction equally important as the mouse input.

Related works

Most head operated interfaces focus on conventional mouse replacement and the pointer manipulation in the graphical user interface. Touchless interaction with head movements have already been reported in the literature and some examples are referred below.

In [6] the cursor is controlled by relative position of the nostrils to the face region. The skin colour method is used first for the face detection. The nostrils are then found using heuristic rules.

In [7] the mouse control is achieved with eye and nose tracking. The eye regions are processed for recognition of winks (used as the mouse clicks). The user's face is detected initially by means of the well known Viola and Jones algorithm [8]. Then, a Gaussian model in RGB is used to represent the skin colour probability density function. Another solution that uses the image plane position of the eyes is presented in [9]. The changes in the eyes localization are transferred to a control system. Eye blink is used for confirmations. The involuntary blink and steering blink are distinguished on the basis of the closure duration.

[1] proposes a touchless interface using multiple facial feature detection and tracking. Mouse manipulation is implemented based on the user's eye movements. Clicking events, on the other hand, are based on the user's mouth shapes (opening/closing). Face detection is achieved here using a skin-color model and connected-component analysis. The eye regions are localized using a neural network texture classifier. The mouth region is localized using an edge detector. After the detection, the eye regions are tracked using a mean-shift algorithm. The mouth regions are tracked using template matching. Authors of the solution also provided a "spelling board" which substitute for a standard keyboard (operated with eye movements and mouth shapes).

Touchless typing with head movements is also proposed in [2]. Here, the head movements are supported with three face gestures chosen for a key selection: mouth open, brows up and brows down.

The above referred solutions are examples found in the scientific literature. There are also examples of finished products available for final users. One example of mouse replacement with the head movements is the SmartNav from NaturalPoint (http://www.naturalpoint.com/smartnav/). It is a commercial product that offers hands-free cursor control also with movements of other part of the body. This is a typical marker-based solution. It works in infrared light and uses: infrared emitters, reflector and an infrared camera. Another proposition for hands-free mouse alternative with head movements is the Enable Viacam (eViacam, http://eviacam.sourceforge.net). It is an open source solution that works in visible spectrum without the need of additional markers. It works on standard PC equipped with a webcam. Face and mouth detection and subsequent tracking are of most importance here.

Proposition of the interface

Head movements in head operated interface should be simple and straightforward, easy to understand, learn and operate by new users. For the interface the directional horizontal movements and vertical rotary movements have been selected. Touchless typing can be achieved using only these four movements without any dwell, eye blink or other techniques. The details are provided below.

To minimize the number of different interactions the single row alphabetical keyboard is used. With a single row layout the selection of a letter requires only movements to the left and right. The keypress can be simulated with the movement in the downward direction. The up direction can be arranged for the backspace (letter deletion). Such proposition is very straightforward: 4 directional head movements are required for the text entry. For the comparison, with the QWERTY keyboard four direction movements are required to the letter selection and the problem with the pressing occurs (another interaction is required).

The change of the head position in the horizontal direction should be mapped to the interface (i.e. one key should be active and highlighted). There are two possible solutions for the mapping from the head position. The first one is the absolute measure of displacement: the distance of the head shift corresponds directly to a particular letter. Such approach is fast but prone to inaccuracies. We are investigating this solution with the assistance of swipe-like typing technique and hope to report the results in the near future.

The second approach to mapping procedure is the relative displacement where the head shift moves the active letter to the neighbor no matter what is the range of the shift (provided it is sufficient enough and does not correspond to small involuntary movements). When the user moves the head left or right the active letter changes accordingly. While remaining in the shifted position consecutive letters are periodically activated (repetition of the action). On the return to the central position the last activated letter remains active and with the nod gesture can be typed.

The main problem with the relative mapping is the necessity of the transition through many letters. A good solution is the usage of dictionary suggested words. However, the number of suggestions can be large and impossible to display in the readable and usable form. Here, the new concept is proposed. Instead of providing the user with suggestions of the whole words the interface is modified limiting the set of next characters. The initially typed phrase is dictionary checked for possible complements. Letters not allowing the creation of proper dictionary words are deactivated. The user is able to display all the letters with the right utmost key (a function key). The whole concept of the interface is presented in the graphical form in Fig. 1. The letters in the interface are deactivated with the shadowing. In the presentation the 'X' mark is used.

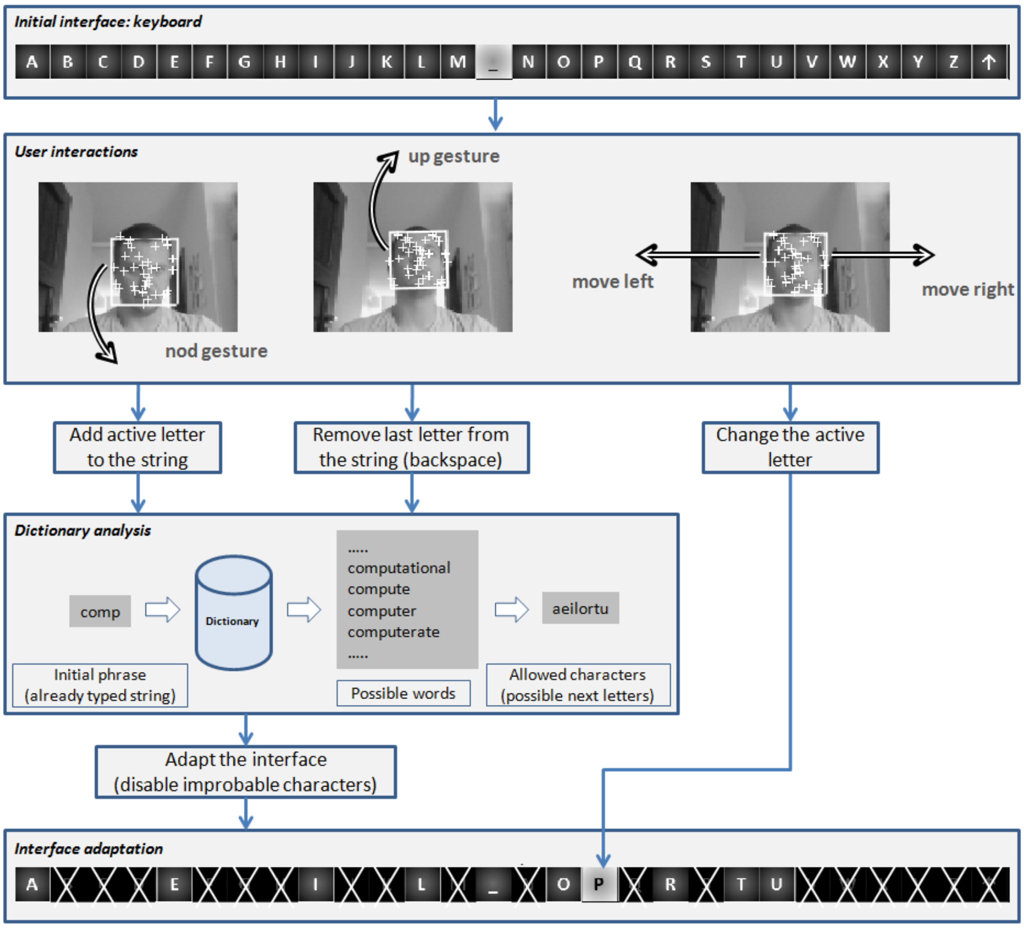

Fig.1. The interface concept

For the task of dictionary analysis the dictionaries provided for the OpenOffice suite are utilized. Experiments have been performed using the United States (US) English dictionary. The used version of en_US.dic consisted of 52890 entries.

There are many data structures and algorithms for dictionary representation and text retrieval. A good example is the trie data structure. In our solution the dictionary have been divided to subdictionaries containing words beginning with the same characters. Based on the first letter typed by the user the dictionary is analyzed for possible next characters with 18 ms average (standard PC with 1 GHz processor). For longer strings the retrieval is faster (for two characters string it is 11 ms and decreases with successive letters).

An example of the interaction

The proposed solution of deactivating the improbable characters can accelerate the process of typing. Figure 2 presents an example interaction during the head typing of the "computer" word. The figure is divided horizontally presenting all the steps of the word entry.

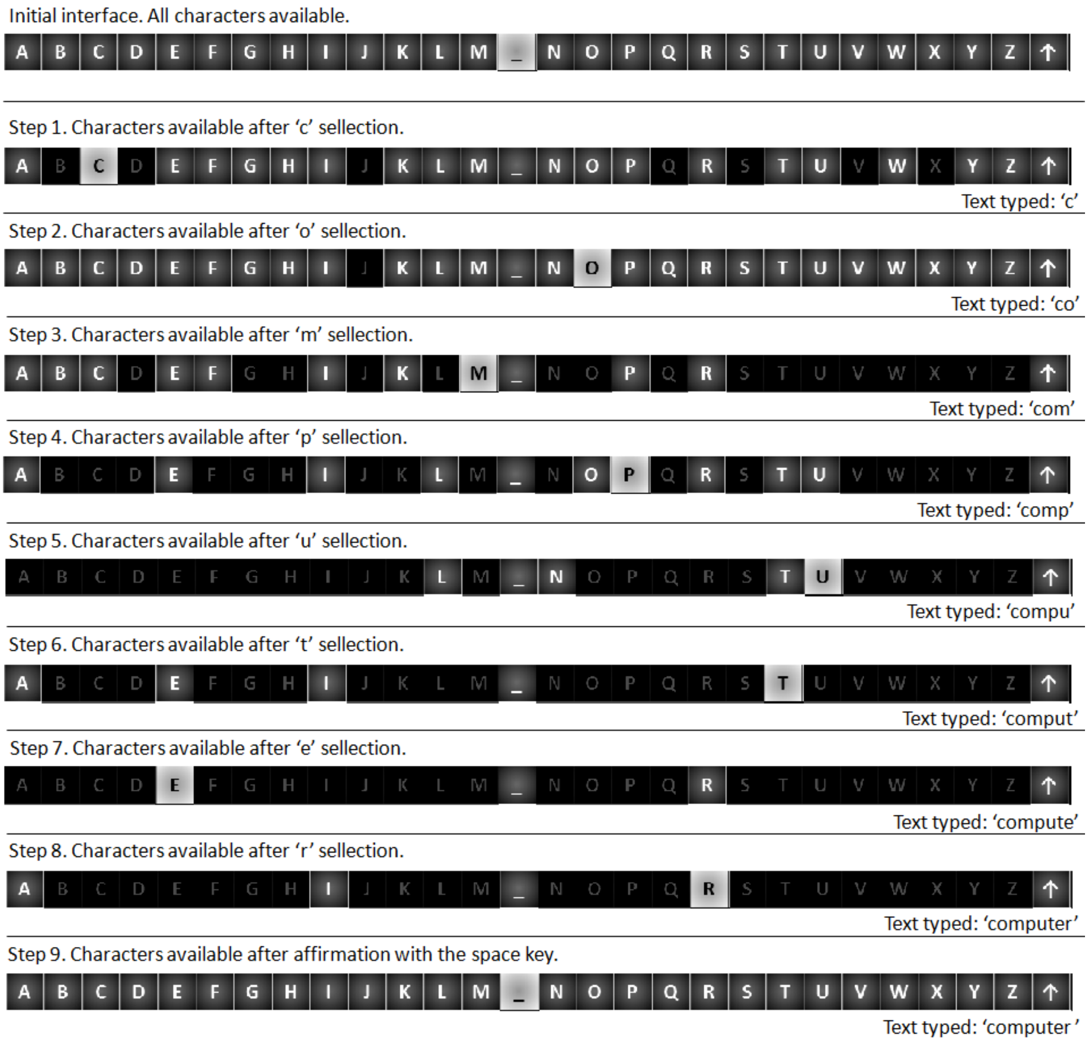

Fig.2. The exemplary interaction during "computer" word entry

With the new letter typed, set of potential characters is limited thus shortening the path to the next letter. The shortening is as follows (the path to the first letter is omitted):

- from 13 to 12 for 'c' to 'o' transition,
- from 3 to 2 for 'o' to 'm' transition,
- from 4 to 3 for 'm' to 'p' transition,
- from 5 to 3 for 'p' to 'u' transition,
- 1 (no change) for 'u' to 't' transition,
- from 17 to 3 for 't' to 'e' transition,
- from 14 to 2 for 'e' to 'r' transition,
- from 5 to 1 for 'r' to space transition.

The total number of horizontal direction movements have been reduced from 62 to 27 (i.e. 56.45%.) for this example. For other words the reduction is the obvious issue. The calculated performance rates for typing are provided in the Summary section.

Detection and tracking of the user's head

The main computer vision problems in head operated interfaces are the face detection and tracking. Some solutions found in the literature (already referred) use markers attached to distinctive parts of the head (e.g. middle of the forehead). Their aim is to simplify the processes of detection and tracking. Considering the visible spectrum without the use of additional markers, the user's face in front of the screen can be detected and tracked using the following computer vision techniques:

- continuous face detection (i.e. tracking by detection),
- face detection using individual frame and tracking in subsequent frames,
- stereo vision,
- background modelling.

The continuous face detection is reserved for fast methods. Those methods frequently offer approximate positions and are prone to errors. More accurate approaches are slower and not sufficient for the real-time interaction. They are, however, willingly combined with faster tracking algorithms. In stereo vision the scene is observed using a pair of calibrated cameras or a single stereoscopic camera. Individual images are combined to form a depth map which can be easily analyzed for the presence of the user silhouette, upper body or a head. Background modelling approaches also offer a robust method for user detection. They are, in turn, sensitive for changes in the scene background and not suited for dynamic surroundings.

The Viola and Jones approach [8] is well known and widely used for the face detection task. It offers good results even for complex scenes. In the human-computer interaction context, when the user is present in front of the screen, few frames are available, and the lighting conditions are at least moderate, the results are perfect. The algorithm has been used in the proposed interface for the initial face detection task. After the detection the face is tracked. Such approach offers most robust operations even on slower hardware configurations.

In the proposed interface the Viola and Jones approach for face detection is assisted with the Kanade-Lucas-Tomasi (KLT) feature-tracking procedure. The algorithm has the following steps. First, initial points for tracking are selected. In the interface these are corners calculated using the minimum eigenvalue algorithm [10]. Other possibilities include [11, 12]: corners calculated using Harris-Stephens algorithm, corners calculated using FAST algorithm, SURF features and others. In the second stage, for each point the tracker attempts to find the corresponding point in the new frame. Then, the geometric transformation based on matched pairs is calculated. Outliers are excluded with the MSAC algorithm [13] (a variant of the RANSAC algorithm). Finally, the calculated transform is applied to the bounding box of the previous face localization.

The tracking algorithm is stable for rigid objects. For a face some points can be lost. However, the procedure of new points selection is fast and imperceptible by the user. It is launched when too many points are lost or periodically when the head is over the neutral area (between consecutive words typing).

The referred approach offers stable and reliable head position for the real-time interaction. The centre of the bounding box around the face is used for the steering purposes - detection of the horizontal and vertical head movements. Movements are detected in relation to the initial head centre stored as the reference point. Coordinates of the reference point are updated with involuntary natural user movements when head position corresponds to the centre area (adaptation). The motion detection in a given direction automatically locks the perpendicular direction. This procedure prevents the accidental transitions between letters or the accidental pressing. During horizontal head movements users tend to tilt their heads sideways causing unwanted accidental press. Similarly, without the locking procedure, during vertical head movements, shifts to an adjacent characters could also occur.

Summary

In the paper, the problem of acceleration of touchless text entry using head movements have been addressed. Head operated interfaces are limited to finite number of possible interaction. Such interfaces are of particular importance for physically challenged people who are unable to operate the standard computer input devices. With most non-contact interfaces, reported in the literature, focusing on conventional mouse replacement the touchless text entry still remains a challenge.

The head typing requires many individual movements. The number of movements may be decreased using delayed repetition during the freeze of the motion. Further acceleration can be achieved with the proposed mechanism of deactivation of improbable next letters. The solutions have been subjected to examination and achieved 15.07% increase in typing speed. The experiments have been conducted among group of 8 participants (computer science students) familiar with touchless interfaces. They achieved on standard interface with the relative mapping of head position to active letter the performance rate of 15.93 cpm (chars per minute). With the proposed mechanism of limitation the potential characters the average typing speed of 18.33 cpm have been achieved (the best trail was almost 21 cpm).

In the future the proposed approach will be subjected to more extensive studies involving the comparisons with other techniques (especially the prototyped swipe-based head typing solution).

***Authors**: dr inż. Adam Nowosielski, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Wydział Informatyki, ul. Żołnierska 52, 71-210 Szczecin, E-mail: anowosielski@wi.zut.edu.pl.*

REFERENCES

[1] Shin, Y., Ju, J.S., Kim, E.Y., Welfare interface implementation using multiple facial features tracking for the disabled people, *Pattern Recognition Letters*, 29 (2008), 1784-1796

[2] Gizatdinova, Y., Spakov, O., Surakka, V., Face Typing: Vision-Based Perceptual Interface for Hands-Free Text Entry with a Scrollable Virtual Keyboard, *IEEE Workshop on Applications of Computer Vision 2012*, (2012), Breckenridge, CO, USA, 81-87

[3] Nowosielski, A., Chodyła, Ł., Touchless input interface for disabled, *Advances in Intelligent Systems and Computing*, 226 (2013), 701-709

[4] Markussen, A., Jakobsen, M.R., Hornbæk, K., Vulture: a mid-air word-gesture keyboard, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'14*, (2014), 1073-1082

[5] Nowosielski, A., QWERTY- and 8pen- based Touchless Text Input with Hand Movement, *Lecture Notes in Computer Science*, 8671 (2014), 470-477

[6] Morris, T., Chauhan, V., Facial feature tracking for cursor control, *Journal of Network and Computer Applications*, 29 (2006), 62-80

[7] Varona, J., Manresa-Yee, C., Perales, F.J., Hands-free vision-based interface for computer accessibility, *Journal of Network and Computer Applications*, 31 (2008), No. 4, 357-374

[8] Viola, P., Jones, M., Robust real-time face detection, *International Journal of Computer Vision*, 57 (2004), No. 2, 137-154

[9] Santis, A., Iacoviello, D., Robust real time eye tracking for computer interface for disabled people, *Computer Methods and Programs in Biomedicine*, 96 (2009), No. 1, 1-11

[10] Shi, J., Tomasi, C., Good Features to Track, *Proceedings CVPR'94*, (1994), 593-600

[11] Rosten, E., Drummond, T., Fusing Points and Lines for High Performance Tracking, *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, (2005), Vol. 2, 1508-1511

[12] Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding*, 110 (2008), No. 3, 346-359

[13] Torr, P.H.S., Zisserman, A., MLESAC: A New Robust Estimator with Application to Estimating Image Geometry, *Journal of Computer Vision and Image Understanding*, 78 (2000), No. 1, 138-156