Extending the Mind-Writing Pupil: Interpreting Blinks and Autocomplete

Manvi Agarwal
s2817047
University of Groningen, The Netherlands

ABSTRACT

Pupillometry-based BCIs use information from pupil size changes to interpret covert attention. Such systems can be set up inexpensively and need very little preparation or calibration, unlike EEG-based methods, while producing similar high rates of accuracy. However, they require the user to maintain sustained levels of concentration and are quite slow to use. To ameliorate these issues, we propose an extension to these spellers by incorporating blink detection and autocomplete. With these features, users can select suggested words with a voluntary blink. To assess the effect of this on the system's efficiency and usability, we design and conduct a study. We find that our extended speller is faster and more user-friendly than the standard speller. Potential applications include enabling faster communication for MND patients.

Author Keywords

human-computer interface; speller; attention; autocomplete; pupillometry; locked-in syndrome.

INTRODUCTION

Brain-computer interfaces have helped individuals with motor neuron diseases (MNDs), such as Locked-in Syndrome (LIS), Amyotrophic Lateral Sclerosis (ALS), and severe brain or spinal cord injuries, to communicate their thoughts to the outside world. For instance, Stephen Hawking was able to 'type' using a cheek muscle. Many BCI spellers analyse EEG activity, such as Event-Related Potential and motor imagery, to translate thoughts into words [10]. However, these paradigms have significant drawbacks, such as the time-consuming process of setting up the system and calibrating it to classify a user's brain activity, and the technical challenges inherent in using EEG systems, which limit their usability [7].

A different methodology using pupillometry was proposed to improve the ease of use. The Mind-Writing Pupil uses covert attention and pupillary light response to guide stimulus selection [9]. Characters are presented in circles with alternating patterns of brightness. Participants select a letter by covertly attending to it, without any overt movement. Changes in pupil size indicate which stimulus the participant intended to select. While the rates of selection accuracy and information transfer reported are at par with state-of-the-art non-invasive covert-attention BCIs, no results on user experience are provided. During the course of our own study, it was observed that participants found the task of spelling a pre-defined phrase demanding, because the time taken to select each letter tended to be close to one minute, and it was difficult to sustain attention over long periods of time. This means that the user

indicates

experienced frustration and high task load, despite the accuracy of the system. Moreover, an application domain proposed for this system is serving as a communication channel for patients with Locked-in Syndrome (LIS). Patients with incomplete LIS and classic LIS retain use of their eyelids [5]. However, this speller takes into account only pupil dilation, not blinks. Furthermore, akin to the autocomplete feature on smartphones, we think it is possible to reduce the number of character selections a user has to perform by introducing a way for the system to make word suggestions and for the user to select suggested words.

Our team extended the Mind-Writing Pupil by adding blink detection and autocomplete functionality. Stimuli selection occurs by not only detecting covert attention, but also classifying and processing blinks. In addition to selecting a letter by changes in pupil size, the system can select a word via autocomplete by tracking voluntary blinks by the user. We hypothesize that these additions to the speller will make it more efficient and user friendly.

DESIGN A

To test this hypothesis, we design and conduct a behaviour

Word Prediction with Autocomplete

Motivated by the ease of typing with autocomplete, we added an analogous feature to the speller. When a user selects a letter, the autocomplete function suggests a word, which appears onscreen below the fixation point, much like word suggestions that appear above the virtual keyboard on phones. The user can then use a long blink to indicate that they want to choose this word.

Word suggestions are generated by using a Python library [1], suitably modified for our purposes. This library uses Hidden Markov Models to generate probability distributions of possible words. Trained on a corpus, such a model takes into account the current incomplete word being typed and the previous complete word before it. The conditional probability of each candidate word is calculated, given that it begins with the letters of the word being typed, and is preceded by the previously typed word in the corpus. Finally, the word with the highest conditional probability is suggested.

The autocomplete model for our speller was trained on the Brown Corpus, compiled in 1967 [6]. While this corpus is quite outdated, it was chosen because it is sizeable, comprehensive, and freely available. Nevertheless, since we are not assessing corpus suitability, but the benefits of adding autocomplete, the Brown Corpus is adequate.

Word Selection with Blink Detection

In the Mind-Writing Pupil speller, one cycle in the letter selection loop consists of a 500-ms transition time to change the brightness level of the stimulus, a 500-ms adaptation time to allow the pupil to adjust in response to the changed brightness level of the attended stimulus, and a 250-ms collection time to collect pupil data and compare it with the inversion pattern. The median pupil size is obtained from the data collected during the last phase. This median is calculated for consecutive cycles, and the change in pupil size is determined using the Proportional Pupil Size Difference (PPSD). This is used to alter the probabilities of which stimulus group the user is attending to. When the proportional difference between the probabilities exceeds a certain threshold, the system takes this as a reliable measure of stimulus selection. The group with the highest selection probability is chosen, and the others are discarded. The loop continues on to the next cycle, with the winning group subdivided into smaller groups, until a single letter is chosen.

This loop was adapted to include detection and interpretation of blinks. Involuntary blinks are usually between 100 ms and 150 ms long [4]. Pupil size data is collected both for the adaptation time and the collection time. If no pupil is detected - for instance, when the eyes are closed - the pupil size is set to zero. The median pupil size is calculated, and if this value is zero, this means that the pupil was absent was for at least half the time in one of the two data collection windows. This means that this pupil absence time has to be a minimum of 125 ms to be classified as an intentional blink. For a duration as short as 125 ms to be misclassified, it would have to occur in such a way that it only partially overlaps with the collection time, which is itself also a very short time span. Since such an occurrence is not very likely, we think this is a reliable way to distinguish between involuntary and voluntary blinks.

USER STUDY

The purpose of this study is to understand the effect of adding autocomplete functionality to the Mind-Writing Pupil speller system. Our hypothesis is that the Blinking Pupil speller is faster and easier to use than the Mind-Writing Pupil speller.

To test this claim, we collect data on the speed of spelling, both for the entire test phrase and individual characters, in both systems. This is used to calculate the Information Transfer Rate (ITR) for the Blink Speller.

And the mind-writing pupil too btw

After testing both conditions, the NASA Task Load Index tool is administered to analyse and compare the experiences of using the two systems. The NASA-TLX assesses perceived workload, and is divided into six subscales. A participant assigns weights to each subscale by making 15 pairwise comparisons. This weightage is used for both the conditions. For each condition, a participant assigns a score of 1 to 21 on each subscale. Individual subscale scores are modified based on the corresponding weights, and the total weighted score, between 0 and 100, indicates the overall task load experienced by the participant. The unweighted subscale scores and overall task load score are used as quantitative indicators of a participant's subjective workload.

Users and Equipment

Five users (4 male, 1 female) participated in the study. Users ranged from 23 to 26 years of age, with distribution (M=24, SD=1.22). All had normal or corrected-to-normal vision. Pupillary response and blinks were measured using a desktop-mounted SR Inc. Eyelink 1000 eye-tracking system.

Experimental Design

Firs, participants were given a short explanation of how the Mind-Writing Pupil speller works. Then, they were given a period of training, with no fixed time limit, where they could practice free-writing and become familiar with the system.

In the testing period, the phrase that the participants had to spell was 'The city in'. This phrase was chosen to minimise the number of letter selections required before autocorrect suggests the next word in the Blinking Pupil speller. The two conditions being tested were:

- *Standard Condition*: Using the Mind-Writing Pupil speller system *without* autocomplete
- *Autocorrect Condition*: Using the extended Min-Writing Pupil speller system *with* autocomplete

In Phase 1, participants had to select one character at a time. They had to select a minimum of 12 characters, including spaces and a terminating character, provided they made no mistakes and did not require to use the backspace. In Phase 2, in addition to character selection, they had the option of selecting a word by blinking. The word suggested by autocorrect, with a space added at the end, appeared below the fixation point. Participants were told that they could blink if they wanted to select that word. This means, at the minimum, they had to select four characters ('T', 'C', 'I' and the terminating character). They could use up to three blinks to spell the phrase. Participants were allowed to restart the spelling process as many times as they wanted during the two phases.

The time difference between the start and end of the spelling period was calculated for each subject and phase. The selection times for each character was also recorded. After the testing was completed, the NASA-TLX assessment tool was administered and responses were collected for each participant. Inbuilt functions in R were used for testing normality with the Shapiro-Wilk normality test [2], and for testing statistical significance with the Repeated Measures t-test [3].

RESULTS

Below, we present a comparison between the two spellers on the metrics identified in our hypothesis - efficiency and usability. All significance levels were calculated with Paired two-sided t-tests.

Efficiency

The mean response time for the speller with autocomplete (3.37 ± 0.95 mins) was significantly lower ($t = 8.35$, $p = 0.001$) than that for the standard speller without autocomplete (8.85 ± 1.87 mins), as shown in Figure 1.

The Information Transfer Rate (ITR) for the speller with autocomplete (18.57 ± 6.64 bits per min) was significantly higher ($t = -5.66$, $p = 0.005$) than that for the standard speller without

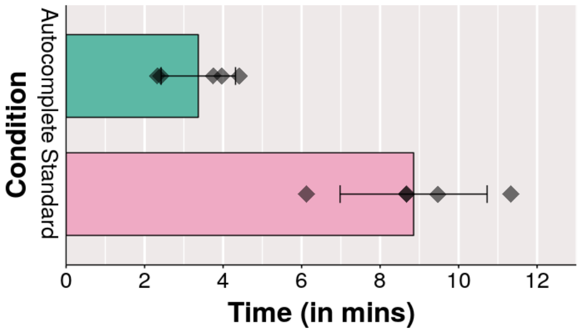

Figure 1. Time taken (in minutes): Bars indicate mean response time, and error bars indicate one standard deviation. Dots indicate individual participants.

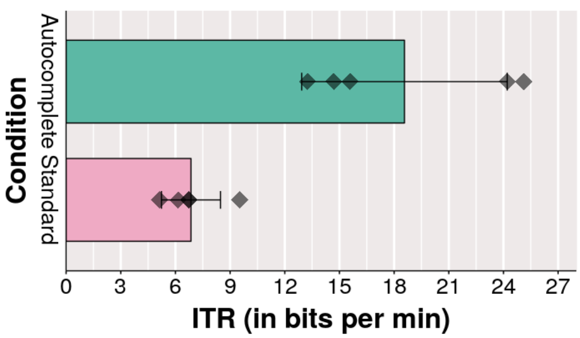

Figure 2. Information Transfer Rate (in bits per minute): Bars indicate ITR, and error bars indicate one standard deviation. Dots indicate individual participants.

autocomplete (6.85 ± 1.62 bits per min), as shown in Figure 2.

No blinks were misclassified. This means that word selection did not work by accident, and all voluntary blinks were treated as intended.

Usability

The mean Task Load Index for the speller with autocomplete (53.6 ± 4.61) was significantly lower (t = 4.87, p = 0.008) than that for the standard speller without autocomplete (74.6 ± 7.28), as shown in Figure 4.

Participants assigned the highest weights to Frustration. They assigned the lowest weights to Performance and Physical Demand, which in some cases were also zero, as is shown in Fig. 3. Of the six subscales, two demonstrated significant effect of condition. The mean weighted score for Effort for the speller with autocomplete (156 ± 83.85) was significantly lower (t = 4.09, p = 0.015) than that for the standard speller without autocomplete (241 ± 104.67). The mean weighted score for Frustration for the speller with autocomplete (244 ± 112.88) was significantly lower (t = 3.86, p = 0.018) than that for the standard speller without autocomplete (378 ± 127.54).

In both conditions, participants had the option of restarting the experiment at any time. We noticed that participants requested restarts multiple times while using the standard speller, whereas this did not happen with the autocomplete

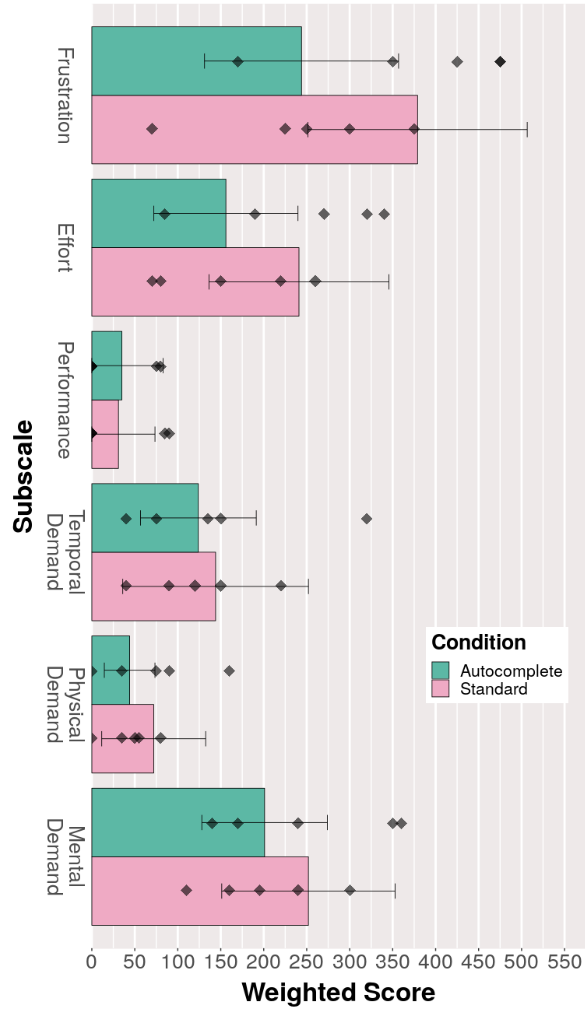

Figure 3. Information Transfer Rate (in bits per minute): Bars indicate ITR, and error bars indicate one standard deviation. Dots indicate individual participants.

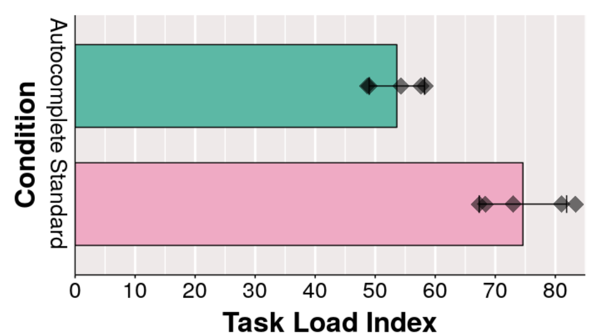

Figure 4. Information Transfer Rate (in bits per minute): Bars indicate ITR, and error bars indicate one standard deviation. Dots indicate individual participants.

version of the speller. This means that in the standard condition, when participants made spelling mistakes, they preferred discarding all their work over using covert attention to first select the backspace character several times and then select the correct letters. Moreover, they used the autocomplete option as many times as was possible and immediately after a suggested word appeared.

DISCUSSION

Though this study indicates that extending the Mind-Writing Pupil speller system with blink detection and autocomplete functionalities does improve its performance in terms of speed and user-friendliness, there are some finer points regarding our experimental setup and the system itself to consider.

The behavioural experiments designed by Mathôt check for gaze-stabilization, to ensure that character selection is driven by covert attention alone. Thus, when the pupil is not fixated on the central dot, the system pauses until the pupil is detected again. This feature had to be altered, such that when no pupil is detected, the system does not pause, but records the pupil size as being zero. This implies that the findings from our experiment hold under the assumption that participants maintained fixation throughout the experiment. Further, there was no way to distinguish between the absence of the eye altogether and a closed eye during a blink, as both were recorded as being values of zero.

The detection of blinks in our system was done by checking the median of pupil sizes during two collection windows, which are of different durations. This means that what can qualify for a voluntary blink in the short window will not be similarly interpreted in the long window. Moreover, the median itself may not be a suitable measure, since this does not take into account the distribution of pupil sizes. Thus, it might be a good design decision to reorganize the collection of pupil size data and employ more robust methods of detecting voluntary blinks while controlling for fixation.

Once a reliable system for detecting and interpreting blinks is deployed, it is possible to experiment with other types of blinks. Different combinations of blinks can be used to control the system in different ways. For instance, a double blink can indicate a backspace. Besides brightness in direct and peripheral vision, the pupil also responds to other types of stimuli, such as color, and cognitive factors [8]. This suggests that combinations of various stimuli can be tested to asses whether they can lead to reliably detectable changes in pupil size.

The word suggestion model for the autocomplete feature was trained on an outdated corpus, and cannot give personalized recommendations to individual users. A possible area for improvement would be to provide user-specific corpora, such as samples of writing from their social media presence. Moreover, the trained model does not have to be static, but can also evolve with time to adapt to a user better. This way, the word suggestions can be made highly personalized, and speed up the spelling procedure even more.

Our experiment always presented the participants with the standard speller first, and then the speller with autocomplete. However, due to lack of resources, we were unable to perform the same experiment with the order of task presentation reversed. Although our intuition is that there would be no notable effects of this, it would be good to confirm it with empirical evidence. Also, our sample size of five participants is quite small, which does not allow us to utilize the full statistical power of the methods that require us to properly estimate the population distribution.

While it was demonstrated in this study that participants experienced lower workload, frustration, and effort while using the speller with autocomplete, as compared to the standard speller, it would have been even more interesting to detect and study the precise sources of such effects. For instance, the trials that participants chose to interrupt with a restart were not recorded. These might have been an important source of data to pinpoint the sources of frustration. Additionally, we do not record other modalities of information such as fixation times or saccadic traces. These might have been used to also find out the extent to which the participants forgo the use of covert attention and instead resort to overt attention. This would provide a measure of the participant's willingness to change the experimental conditions (technically 'cheat') and hence will also serve as a measure of the level of perceived workload that the subject can endure before willing to employ overt attention for the task.

For the deployment of such a system at scale, where it can be used on a regular basis by LIS patients, several improvements have to be made. First, the system must be built outside the experimental confines of OpenSesame, so that lags in loading the program are minimized. Second, the current design of the speller is heavily dependent on the ability of users to sustain high levels of concentration over a long period of time. This can be mitigated by, for instance, giving users the ability to pause the spelling cycle whenever they are tired. Third, there is no provision to cut short a character selection loop where the wrong subset of characters have been shortlisted and carried forward to the next cycle within the loop.

CONCLUSION

The Mind-Writing Pupil speller system uses pupillometry and the decoding of covert attention to guide stimulus selection. While this selection procedure is highly accurate, it is slow and requires sustained attention over long period of time. This impacts the effectiveness of the system. To improve the efficiency and usability of the speller, we added two features: blink detection and autocomplete. With these, in addition to selecting individual letters, a user can select words suggested by autocomplete with a voluntary blink.

To compare the standard speller with our extended version, we designed a behavioural study, where participants were asked to spell a pre-defined phrase with both spellers. An analysis of the data obtained indicates that the extended speller not only enables the user to spell faster, but also reduces the workload, effort, and frustration experienced by the user. This confirms our expectations.

ACKNOWLEDGEMENTS

We would like to thank all the participants. We extend our gratitude to Prof. Jelmer Borst, Oscar Portoles Marin and Stefan Huijser for their helpful comments.

REFERENCES

1. Palacios, r. - autocomplete pypi. `https://pypi.org/project/autocomplete/`. Last accessed: 04-11-2018.
2. R: Shapiro-wilk normality test. `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/shapiro.test.html`. Last accessed: 04-11-2018.
3. R: Student's t-test. `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/t.test.html`. Last accessed: 04-11-2018.
4. Feb 2006.
5. Klein, E., Peters, B., and Higger, M. Ethical considerations in ending exploratory braincomputer interface research studies in locked-in syndrome. *Cambridge Quarterly of Healthcare Ethics 27*, 4 (2018), 660674.
6. Kučera, H., and Francis, W. N. *Computational analysis of present-day American English*. Dartmouth Publishing Group, 1967.
7. Lazarou, I., Nikolopoulos, S., Petrantonakis, P. C., Kompatsiaris, I., and Tsolaki, M. Eeg-based braincomputer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21st century. *Frontiers in Human Neuroscience 12* (2018), 14.
8. Mathôt, S. Pupillometry: Psychology, physiology, and function. *Journal of Cognition 1*, 1 (2018).
9. Mathôt, S., Melmi, J.-B., Van Der Linden, L., and Van der Stigchel, S. The mind-writing pupil: A human-computer interface based on decoding of covert attention through pupillometry. *PloS one 11*, 2 (2016), e0148805.
10. Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. Braincomputer interface spellers: A review. *Brain Sciences 8*, 4 (Mar 2018), 57.